Wilfried Plaßmann
Detlef Schulz *Hrsg.*

Handbuch Elektrotechnik

Grundlagen und Anwendungen für Elektrotechniker

7. Auflage

Springer Vieweg

Wilfried Plaßmann · Detlef Schulz
Herausgeber

Handbuch Elektrotechnik

7., neu bearbeitete Auflage

Springer Vieweg

Herausgeber

Wilfried Plaßmann
Bad Nenndorf, Deutschland

Detlef Schulz
Hamburg, Deutschland

ISBN 978-3-658-07048-9
ISBN 978-3-658-07049-6 (eBook)
DOI 10.1007/978-3-658-07049-6

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer Vieweg
© Springer Fachmedien Wiesbaden GmbH 1999, 2002, 2004, 2007, diese Auflagen erschienen unter dem Titel „Böge/-Plaßmann, Vieweg Handbuch Elektrotechnik", 2009, 2013, 2016
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.
Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.

Springer Vieweg ist Teil von Springer Nature
Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Strasse 46, 65189 Wiesbaden, Germany

Vorwort

Das *Handbuch Elektrotechnik* erscheint nunmehr in der 7. Auflage. Seit seiner Einführung ist es gedacht als eine Kombination aus studienbegleitendem Werk und einem Nachschlagewerk. Dieses Konzept hat sich bewährt und wurde auch in dieser Auflage beibehalten.

Das Werk wurde gegenüber der 6. Auflage neu bearbeitet. Bewährte Kapitel, die die Grundlagen behandeln, wurden belassen; andere Kapitel wurden durch inhaltliche Änderungen dem Stand der Technik angepasst. Das Kapitel *Elektrische Energieversorgung* wurde völlig neu gestaltet und trägt damit den aktuellen Entwicklungen dieses Zweiges der Energietechnik Rechnung. Dazu gehört auch das neu verfasste und aktualisierte Kapitel *Leistungselektronik*. Damit erfährt die Energietechnik, ihrer Bedeutung entsprechend, eine angemessenere Beachtung.

Das Handbuch ist als Informationsbasis für Studierende und in der Praxis tätige Ingenieure und Techniker ausgelegt. In diesem Werk finden die Anwender alle notwendigen Formeln, Hinweise, Tabellen, Schaltpläne und Normen. Zur Sicherung sachkundiger Anwendungen werden wichtige Berechnungsgleichungen ausführlich hergeleitet. Zahlreiche anwendungsbezogene Beispiele in jedem Kapitel erhöhen das Verständnis für die oft komplexen Zusammenhänge und geben die zur Problemlösung unerlässliche Sicherheit.

Die Herausgeber danken den Lesern für kritische Anmerkungen zum Buch und sind auch weiterhin für Anregungen und Verbesserungsvorschläge dankbar. Weiterhin danken sie Frau Broßler und Frau Himmel sowie Herrn Dapper vom Verlag Springer-Vieweg für die gute Zusammenarbeit bei der Realisierung des umfangreichen Werkes.

Die E-Mail-Adressen der Herausgeber lauten:

wilfried.plassmann@hs-hannover.de

detlef.schulz@hsu-hh.de

Hannover/Hamburg, November 2016

Wilfried Plaßmann/Detlef Schulz
Herausgeber

Inhaltsverzeichnis

Teil I Mathematik

1 Arithmetik . . . 3

1.1 Mengen . . . 4

1.2 Aussageformen und logische Zeichen . . . 4
1.2.1 Aussageformen . . . 4
1.2.2 Logische Zeichen . . . 5
1.2.3 Vollständige Induktion . . . 5

1.3 Einteilung der Zahlen . . . 6

1.4 Grundrechenarten . . . 7

1.5 Grundlegende Rechenregeln . . . 8
1.5.1 Buchstabenrechnen . . . 8
1.5.2 Kehrwert, Quersumme . . . 8
1.5.3 Teilbarkeitsregeln . . . 8
1.5.4 Punktrechnung vor Strichrechnung . . . 8
1.5.5 Potenzrechnung vor Punktrechnung . . . 8
1.5.6 Grundgesetze der Addition und Multiplikation . . . 9
1.5.7 Indizes, Summenzeichen, Produktzeichen . . . 9
1.5.8 Binomische Formeln . . . 10
1.5.9 Fakultäten, Binomialkoeffizienten und Pascal'sches Dreieck . . . 10
1.5.10 Binomischer Lehrsatz . . . 11

1.6 Potenz- und Wurzelrechnung . . . 12
1.6.1 Definition der Potenz . . . 12
1.6.2 Regeln der Potenzrechnung . . . 12
1.6.3 Definition der Wurzel . . . 14
1.6.4 Regeln der Wurzelrechnung . . . 14

1.7 Dezimalzahlen und Dualzahlen . . . 17
1.7.1 Dezimalsystem . . . 17
1.7.2 Dualsystem . . . 17
1.7.3 Runden . . . 18

1.8 Logarithmen . . . 18
1.8.1 Definition des Logarithmus . . . 18
1.8.2 Spezielle Basen . . . 19
1.8.3 Regeln der Logarithmenrechnung . . . 19
1.8.4 Zusammenhang von Logarithmen mit verschiedenen Basen . . . 20
1.8.5 Dekadische Logarithmen . . . 21

1.9 Ungleichungen . . . 21
1.9.1 Definitionen und Rechenregeln . . . 21
1.9.2 Absolutbetrag . . . 22
1.9.3 Intervalle . . . 22

1.10 Komplexe Zahlen . . . 23
1.10.1 Algebraische Form . . . 23
1.10.2 Trigonometrische Form . . . 24

1.10.3 Addieren und Subtrahieren komplexer Zahlen 25
1.10.4 Multiplizieren komplexer Zahlen 25
1.10.5 Dividieren komplexer Zahlen 26
1.10.6 Potenzieren komplexer Zahlen 27
1.10.7 Radizieren komplexer Zahlen 27
1.10.8 Euler'sche Formel 28

2 Gleichungen 29

2.1 Gleichungsarten 30

2.2 Äquivalente Umformungen 31

2.3 Lineare Gleichungen 32

2.4 Proportionen 33

2.5 Quadratische Gleichungen 34
2.5.1 Definitionen 34
2.5.2 Lösungsverfahren 34
2.5.3 Satz von Viëta für quadratische Gleichungen 37

2.6 Algebraische Gleichungen höheren Grades 37
2.6.1 Kubische Gleichungen 37
2.6.2 Polynomdivision 39
2.6.3 Gleichungen vierten Grades 40
2.6.4 Gleichungen n-ten Grades 42
2.6.5 Satz von Viëta für Gleichungen n-ten Grades 43

2.7 Auf algebraische Gleichungen zurückführbare Gleichungen 43
2.7.1 Bruchgleichungen 43
2.7.2 Wurzelgleichungen 44

2.8 Transzendente Gleichungen 45
2.8.1 Exponentialgleichungen 45
2.8.2 Logarithmische Gleichungen 46
2.8.3 Trigonometrische Gleichungen 46

2.9 Lineare Gleichungssysteme 47
2.9.1 Definitionen 47
2.9.2 Zwei lineare Gleichungen mit zwei Variablen 47
2.9.3 Drei lineare Gleichungen mit drei Variablen 49
2.9.4 Matrizen und Determinanten 50

3 Planimetrie 57

3.1 Geraden und Strecken 58

3.2 Winkel 58

3.3 Projektion 60

3.4 Geometrische Örter 60

3.5 Dreiecke 60
3.5.1 Allgemeine Dreiecke 60
3.5.2 Gleichschenklige Dreiecke 61
3.5.3 Gleichseitige Dreiecke 62
3.5.4 Rechtwinklige Dreiecke 62
3.5.5 Besondere Geraden, Strecken und Kreise 62
3.5.6 Flächensätze im rechtwinkligen Dreieck 64

3.6 Vierecke 66
3.6.1 Allgemeine Vierecke 66
3.6.2 Trapeze 66
3.6.3 Parallelogramme 67

3.6.4 Rhomben . . . 67
3.6.5 Rechtecke . . . 67
3.6.6 Quadrate . . . 68
3.6.7 Drachen . . . 68
3.6.8 Sehnenvierecke . . . 68
3.6.9 Tangentenvierecke . . . 69

3.7 Reguläre n-Ecke . . . 69

3.8 Polygone . . . 70

3.9 Kreise . . . 71
3.9.1 Definitionen . . . 71
3.9.2 Kreissektoren . . . 71
3.9.3 Kreissegmente . . . 72
3.9.4 Kreise und Geraden . . . 72
3.9.5 Bogenmaß . . . 72

3.10 Symmetrie . . . 73
3.10.1 Punktsymmetrie . . . 73
3.10.2 Achsensymmetrie . . . 73

3.11 Ähnlichkeit . . . 73
3.11.1 Zentrische Streckung . . . 73
3.11.2 Strahlensätze . . . 74
3.11.3 Ähnliche Figuren . . . 74

4 Stereometrie . . . 77

4.1 Prismen . . . 78
4.1.1 Allgemeine Prismen . . . 78
4.1.2 Parallelepiped und Würfel . . . 78

4.2 Zylinder . . . 79
4.2.1 Allgemeine Zylinder . . . 79
4.2.2 Gerade Kreiszylinder . . . 79
4.2.3 Hohlzylinder . . . 80

4.3 Pyramiden . . . 80
4.3.1 Allgemeine Pyramiden . . . 80
4.3.2 Gerade quadratische Pyramiden . . . 80

4.4 Kegel . . . 81
4.4.1 Allgemeine Kegel . . . 81
4.4.2 Gerade Kreiskegel . . . 82

4.5 Cavalieri'sches Prinzip . . . 82

4.6 Kugeln . . . 82
4.6.1 Definitionen . . . 82
4.6.2 Kugelsegmente . . . 83
4.6.3 Kugelsektoren . . . 83
4.6.4 Kugelschichten . . . 83

5 Funktionen . . . 85

5.1 Definition und Darstellungen von Funktionen . . . 86
5.1.1 Definitionen . . . 86
5.1.2 Funktionsgleichung . . . 86
5.1.3 Graph einer Funktion . . . 87
5.1.4 Wertetabelle einer Funktion . . . 87

5.2 Verhalten von Funktionen . . . 88
5.2.1 Monotone Funktionen . . . 88
5.2.2 Symmetrische Funktionen . . . 88

5.2.3 Beschränkte Funktionen . . . 89
5.2.4 Injektive Funktionen . . . 89
5.2.5 Surjektive Funktionen . . . 89
5.2.6 Bijektive Funktionen . . . 90
5.2.7 Periodische Funktionen . . . 90
5.2.8 Umkehrfunktionen . . . 90
5.2.9 Reelle und komplexe Funktionen . . . 91

5.3 Einteilung der elementaren Funktionen . . . 91

5.4 Ganze rationale Funktionen . . . 93
5.4.1 Konstante Funktionen . . . 93
5.4.2 Lineare Funktionen . . . 93
5.4.3 Quadratische Funktionen . . . 95
5.4.4 Kubische Funktionen . . . 99
5.4.5 Ganze rationale Funktionen n-ten Grades . . . 100
5.4.6 Horner-Schema . . . 101

5.5 Gebrochene rationale Funktionen . . . 102
5.5.1 Nullstellen, Pole, Asymptoten . . . 102
5.5.2 Partialbruchzerlegung . . . 105

5.6 Irrationale Funktionen . . . 106

5.7 Transzendente Funktionen . . . 108
5.7.1 Exponentialfunktionen . . . 108
5.7.2 Logarithmusfunktionen . . . 109

6 Trigonometrie . . . 111

6.1 Definition der trigonometrischen Funktionen . . . 112

6.2 Trigonometrische Funktionen für beliebige Winkel . . . 113

6.3 Beziehungen für den gleichen Winkel . . . 114

6.4 Graphen der trigonometrischen Funktionen . . . 114

6.5 Reduktionsformeln . . . 116

6.6 Additionstheoreme . . . 116

6.7 Sinussatz und Kosinussatz . . . 117

6.8 Arkusfunktionen . . . 118

7 Analytische Geometrie . . . 121

7.1 Koordinatensysteme . . . 122
7.1.1 Kartesisches Koordinatensystem der Ebene . . . 122
7.1.2 Polarkoordinatensystem der Ebene . . . 122
7.1.3 Zusammenhang zwischen kartesischen und Polarkoordinaten 123
7.1.4 Kartesisches Koordinatensystem des Raums . . . 124

7.2 Geraden . . . 124
7.2.1 Geradengleichungen . . . 124
7.2.2 Abstände . . . 127

7.3 Kreise . . . 128
7.3.1 Kreisgleichungen . . . 128
7.3.2 Berechnung von Kreisen . . . 129

7.4 Kugeln . . . 130

7.5 Kegelschnitte . . . 130
7.5.1 Ellipsen . . . 132
7.5.2 Hyperbeln . . . 133

7.5.3 Parabeln . . . 135
7.5.4 Anwendungen . . . 136

7.6 Vektoren . . . 137
7.6.1 Definitionen . . . 137
7.6.2 Multiplikation eines Vektors mit einem Skalar . . . 138
7.6.3 Addition und Subtraktion zweier Vektoren . . . 138
7.6.4 Komponentendarstellung von Vektoren in der Ebene . . . 139
7.6.5 Komponentendarstellung von Vektoren im Raum . . . 139
7.6.6 Skalarprodukt . . . 140
7.6.7 Vektorprodukt . . . 141
7.6.8 Spatprodukt . . . 141

8 Differential- und Integralrechnung . . . 143

8.1 Folgen . . . 144
8.1.1 Grundbegriffe . . . 144
8.1.2 Arithmetische Folgen . . . 144
8.1.3 Geometrische Folgen . . . 145
8.1.4 Grenzwert einer Folge . . . 145
8.1.5 Einige Grenzwerte . . . 146
8.1.6 Divergente Folgen . . . 146

8.2 Reihen . . . 146
8.2.1 Definitionen . . . 146
8.2.2 Arithmetische Reihen . . . 148
8.2.3 Geometrische Reihen . . . 148
8.2.4 Harmonische Reihen . . . 149
8.2.5 Alternierende Reihen . . . 149

8.3 Grenzwerte von Funktionen . . . 150
8.3.1 Grenzwert an einer endlichen Stelle . . . 150
8.3.2 Einseitige Grenzwerte . . . 150
8.3.3 Grenzwert im Unendlichen . . . 151
8.3.4 Rechenregeln für Grenzwerte . . . 151
8.3.5 Unbestimmte Ausdrücke . . . 152
8.3.6 Stetigkeit einer Funktion . . . 152
8.3.7 Unstetigkeitsstellen . . . 153

8.4 Ableitung einer Funktion . . . 153
8.4.1 Definitionen . . . 153
8.4.2 Differentiationsregeln . . . 154
8.4.3 Höhere Ableitungen . . . 157
8.4.4 Ableitungen einiger algebraischer Funktionen . . . 157
8.4.5 Ableitungen einiger transzendenter Funktionen . . . 157
8.4.6 Sekanten und Tangenten . . . 159
8.4.7 Extremwerte von Funktionen . . . 160
8.4.8 Krümmungsverhalten von Funktionen . . . 161
8.4.9 Wendepunkte von Funktionen . . . 161
8.4.10 Kurvendiskussion . . . 162
8.4.11 Anwendungsbeispiele . . . 163
8.4.12 Näherungsverfahren zur Nullstellenbestimmung . . . 163

8.5 Integralrechnung . . . 165
8.5.1 Unbestimmtes Integral . . . 165
8.5.2 Integrationsregeln . . . 165
8.5.3 Unbestimmte Integrale einiger algebraischer Funktionen . . . 168
8.5.4 Unbestimmte Integrale einiger transzendenter Funktionen . . . 169
8.5.5 Bestimmtes Integral . . . 169
8.5.6 Hauptsatz der Differential- und Integralrechnung . . . 170
8.5.7 Eigenschaften des bestimmten Integrals . . . 170
8.5.8 Einige Anwendungen der Integralrechnung . . . 172

8.6 Funktionenreihen . . . 174
8.6.1 Definitionen . . . 174
8.6.2 Potenzreihen . . . 175
8.6.3 Fourier-Reihen . . . 177

9 Anhang . . . 183

9.1 Symbole und Bezeichnungsweisen . . . 184

9.2 Das griechische Alphabet . . . 185

Teil II Physik

10 Einführung . . . 189

10.1 Physikalische Größen . . . 190
10.1.1 Skalare . . . 190
10.1.2 Vektoren . . . 190

10.2 SI-System . . . 190

11 Mechanik . . . 193

11.1 Kinematik des Massenpunktes . . . 194
11.1.1 Eindimensionale Bewegungen . . . 194
11.1.2 Zusammengesetzte Bewegungen . . . 196
11.1.3 Kreisbewegung . . . 197

11.2 Dynamik . . . 199
11.2.1 Newton'sche Axiome . . . 199
11.2.2 Kraft . . . 199
11.2.3 Impuls . . . 201
11.2.4 Arbeit, Leistung, Wirkungsgrad und Energie . . . 203
11.2.5 Stoßprozesse . . . 205
11.2.6 Rotation . . . 206
11.2.7 Gravitation . . . 209

11.3 Elastische Verformung fester Körper . . . 210

11.4 Mechanik der ruhenden Flüssigkeiten und Gase . . . 210
11.4.1 Druck . . . 210
11.4.2 Kompressibilität . . . 210
11.4.3 Volumenausdehnung . . . 211
11.4.4 Hydrostatischer Druck in Flüssigkeiten . . . 211
11.4.5 Schweredruck in Gasen . . . 211
11.4.6 Auftrieb . . . 211

11.5 Hydrodynamik . . . 212
11.5.1 Kontinuitätsgleichung . . . 212
11.5.2 Bernoulli-Gleichung . . . 212
11.5.3 Innere Reibung . . . 212

12 Thermodynamik . . . 213

12.1 Grundbegriffe . . . 214

12.2 Temperatur . . . 214
12.2.1 Einheiten . . . 214
12.2.2 Temperaturmessung . . . 214

12.3 Thermische Ausdehnung . . . 214
12.3.1 Feste Stoffe . . . 215
12.3.2 Flüssigkeiten . . . 215
12.3.3 Gase . . . 215

12.4 Ideale Gase . . . 216
12.4.1 Allgemeine Zustandsgleichung idealer Gase . . . 216
12.4.2 Kinetische Gastheorie . . . 216
12.4.3 Wärmeenergie . . . 217
12.4.4 Zustandsänderungen idealer Gase . . . 217
12.4.5 Kreisprozesse . . . 218

12.5 Wärmeübertragung . . . 219
12.5.1 Wärmeleitung . . . 219
12.5.2 Wärmeströmung . . . 219
12.5.3 Wärmestrahlung . . . 219

13 Schwingungen . . . 221

13.1 Freie ungedämpfte harmonische Schwingungen . . . 222

13.2 Gedämpfte Schwingungen . . . 223

13.3 Erzwungene Schwingungen . . . 223

13.4 Überlagerung harmonischer Schwingungen . . . 224
13.4.1 Schwingungsrichtung parallel zueinander . . . 224
13.4.2 Schwingungsrichtung senkrecht zueinander . . . 225

14 Wellen . . . 227

14.1 Harmonische Wellen . . . 228
14.1.1 Ausbreitung . . . 228
14.1.2 Interferenz . . . 229

14.2 Huygens'sches Prinzip . . . 230
14.2.1 Reflexion . . . 230
14.2.2 Brechung . . . 230
14.2.3 Beugung . . . 231

14.3 Dopplereffekt . . . 231

15 Akustik . . . 233

15.1 Schallausbreitung . . . 234

15.2 Reflexion, Transmission, Absorption . . . 234

15.3 Ultraschall . . . 235

16 Optik . . . 237

16.1 Eigenschaften des Lichtes . . . 238

16.2 Geometrische Optik . . . 238
16.2.1 Reflexion des Lichtes . . . 238
16.2.2 Brechungsgesetz . . . 240
16.2.3 Optische Geräte . . . 243

16.3 Wellenoptik . . . 246
16.3.1 Interferenz . . . 246
16.3.2 Beugung . . . 246

16.4 Photometrie . . . 248
16.4.1 Strahlungsphysikalische Größen . . . 248
16.4.2 Lichttechnische Größen . . . 250

16.5 Licht als Korpuskel . . . 251

Literatur . . . 252

17 Anhang 253
17.1 Physikalische Größen und Einheiten 254
17.2 Zahlenwerte physikalischer Größen 254

Teil III Werkstoffkunde

18 Stoffe 257
18.1 Eigenschaften der Stoffe 258
18.2 Atombau und Periodensystem 258
18.3 Aufbau der festen Körper 260
18.4 Chemische Grundzusammenhänge 262
18.5 Elektrochemie 264
19 Elektrische Leitfähigkeit 267
19.1 Leitungsmechanismus 268
19.2 Isolator 268
19.3 Halbleiter 269
19.4 Normalleiter 269
19.5 Supraleiter 270
19.6 Halleffekt 270
20 Elektrische Leiter 273
20.1 Normalleiter 274
20.2 Halbleiter 274
20.3 Supraleiter 276
21 Magnetische Leitfähigkeit 279
21.1 Modellvorstellung 280
21.2 Verhalten von Materie im Magnetfeld 280
21.3 Magnetisierung 281
21.4 Magnetisierungskurve 282
21.5 Permeabilität 283
22 Magnetika 285
22.1 Metalloxide (Ferrite) 286
22.2 Weichmagnetika 286
22.3 Hartmagnetika (Dauermagnete) 290
23 Dielektrische Eigenschaften 293
23.1 Modellvorstellungen zur dielektrischen Polarisation 294
23.2 Dielektrische Materialeinteilung 295
23.3 Elektrische Materialeinteilung 295

24 Dielektrika . . . 297
24.1 Natürliche anorganische Dielektrika . . . 298
24.2 Natürliche organische Dielektrika . . . 299
24.3 Künstliche anorganische Dielektrika . . . 299
24.4 Künstliche organische Dielektrika . . . 300
24.5 Silikone . . . 301
Literatur . . . 301

Teil IV Grundlagen der Elektrotechnik

25 Grundbegriffe . . . 305
25.1 Aufbau der Atome . . . 306
25.2 Ladungsträger . . . 306
25.3 Spannung . . . 306
25.4 Strom . . . 306
25.4.1 Bewegung von Ladungsträgern . . . 306
25.4.2 Stromstärke . . . 307
25.4.3 Stromdichte . . . 307
25.5 Das Ohm'sche Gesetz . . . 307
25.6 Spezifischer Widerstand, Leitfähigkeit . . . 307
25.7 Temperaturabhängigkeit des elektrischen Widerstandes von Metallen 308

26 Der Gleichstromkreis . . . 311
26.1 Zählpfeilsysteme . . . 312
26.2 Kirchhoff'sche Gesetze . . . 312
26.2.1 Knotenregel . . . 312
26.2.2 Maschenregel . . . 312
26.3 Schaltung von Widerständen . . . 313
26.3.1 Reihenschaltung . . . 313
26.3.2 Parallelschaltung . . . 313
26.3.3 Stern-Dreieck Umwandlung . . . 314
26.3.4 Messbereichserweiterung . . . 315
26.4 Reale Spannungsquelle . . . 315
26.4.1 Kombination von Spannungsquellen . . . 316
26.5 Ersatzstromquelle . . . 316
26.6 Netzwerkberechnung . . . 317
26.6.1 Gemischte Schaltungen . . . 317
26.6.2 Überlagerungsverfahren . . . 318
26.6.3 Ersatzspannungsquelle . . . 318
26.6.4 Nichtlineare Gleichstromkreise . . . 319
26.7 Energie, Leistung, Wirkungsgrad . . . 320
26.7.1 Leistungsanpassung . . . 320
26.7.2 Leistungsverlust auf Leitungen . . . 321
26.7.3 Wirkungsgrad . . . 321
26.7.4 Umwandlung elektrischer Energie . . . 321

27 Das Elektrische Feld . . . 323

27.1 Grundgrößen . . . 324
27.1.1 Kräfte zwischen Ladungen . . . 324
27.1.2 Feldstärke . . . 324
27.1.3 Feldlinien . . . 325
27.1.4 Potential, Spannung . . . 325
27.1.5 Äquipotentiallinien . . . 326
27.1.6 Elektrischer Fluss . . . 326
27.1.7 Energie geladener Teilchen im elektrischen Feld . . . 326

27.2 Materie im elektrischen Feld . . . 326
27.2.1 Leiter . . . 327
27.2.2 Nichtleiter . . . 327

27.3 Kondensatoren . . . 327
27.3.1 Kapazität . . . 327
27.3.2 Schaltungen mit Kondensatoren . . . 329
27.3.3 Energie des elektrostatischen Feldes . . . 331
27.3.4 Laden und Entladen eines Kondensators . . . 331
27.3.5 RC-Reihenschaltung . . . 333

28 Das Magnetische Feld . . . 335

28.1 Feldlinien . . . 336
28.1.1 Dauermagnet . . . 336
28.1.2 Stromdurchflossene Leiter . . . 336
28.1.3 Stromdurchflossene Spule . . . 336
28.1.4 Magnetfeld der Erde . . . 337

28.2 Magnetische Grundgrößen . . . 337
28.2.1 Feldstärke . . . 337
28.2.2 Fluss, Flussdichte . . . 340

28.3 Kräfte im Magnetfeld . . . 341
28.3.1 Kräfte auf bewegliche Ladungsträger . . . 341
28.3.2 Stromdurchflossener Leiter . . . 341
28.3.3 Magnetisches Moment . . . 342
28.3.4 Kräfte zwischen zwei parallelen Leitern . . . 342
28.3.5 Hall-Effekt . . . 343

28.4 Energie des Magnetfeldes . . . 344

28.5 Materie im Magnetfeld . . . 344
28.5.1 Diamagnetismus . . . 345
28.5.2 Paramagnetismus . . . 345
28.5.3 Ferromagnetismus . . . 346
28.5.4 Antiferromagnetismus . . . 348
28.5.5 Ferrimagnetismus . . . 348

28.6 Magnetische Kreise . . . 348
28.6.1 Magnetische Spannung . . . 348
28.6.2 Magnetischer Widerstand . . . 349
28.6.3 Unverzweigte Kreise . . . 349
28.6.4 Verzweigte Kreise . . . 351

29 Induktion . . . 353

29.1 Induktion bei Änderung der Fläche . . . 354

29.2 Induktion bei Änderung des Magnetfeldes . . . 355

29.3 Die Induktivität einer Spule . . . 356
29.3.1 Selbstinduktion . . . 356
29.3.2 Gegeninduktion . . . 357

29.3.3 Energie im Magnetfeld einer Spule . . . 358
29.3.4 Ein- und Ausschaltvorgänge . . . 358
29.3.5 Zusammenschalten von Induktivitäten . . . 359

30 Wechselstrom . . . 361

30.1 Grundbegriffe des Wechselstroms . . . 362
30.1.1 Erzeugung einer sinusförmigen Wechselspannung . . . 362
30.1.2 Phasenverschiebung . . . 363
30.1.3 Effektivwert . . . 363
30.1.4 Darstellungsarten . . . 363

30.2 Grundschaltelemente im Wechselstromkreis . . . 364
30.2.1 Ohm'scher Widerstand . . . 364
30.2.2 Kapazität . . . 365
30.2.3 Spule . . . 365

30.3 Schaltungen von Wechselstromwiderständen . . . 367
30.3.1 Reihenschaltung von Wechselstromwiderständen . . . 367
30.3.2 Parallelschaltung von Wechselstromwiderständen . . . 371
30.3.3 Gemischte Schaltungen . . . 374

30.4 Passive Filter . . . 376
30.4.1 Hochpassschaltung mit RC- und RL-Glied . . . 376
30.4.2 Tiefpassschaltung mit RC- und RL-Glied . . . 377
30.4.3 Bandpassschaltung . . . 377

30.5 Schwingkreise . . . 378
30.5.1 Reihenresonanz . . . 378
30.5.2 Parallelresonanz . . . 379

30.6 Leistung und Arbeit im Wechselstromkreis . . . 379
30.6.1 Leistung und Arbeit bei Phasengleichheit von Spannung und Strom . . . 379
30.6.2 Leistung und Arbeit bei Phasenverschiebung von Spannung und Strom . . . 379
30.6.3 Leistung in komplexer Schreibweise . . . 381
30.6.4 Leistungsfaktor . . . 381

31 Drehstrom . . . 383

31.1 Erzeugung von mehrphasigem Wechselstrom . . . 384

31.2 Phasenverkettung . . . 384
31.2.1 Sternschaltung . . . 384
31.2.2 Dreieckschaltung . . . 385

31.3 Leistung des Dreiphasenstroms . . . 386

31.4 Das unsymmetrische Dreiphasensystem . . . 387
31.4.1 Das unsymmetrische Dreileiternetz . . . 387
31.4.2 Das unsymmetrische Vierleiternetz . . . 388

Literatur . . . 389

Teil V Elektronik

32 Leitungsmechanismen bei Halbleitern, pn-Übergang . . . 393

32.1 Einführung in die Halbleiterphysik . . . 394

32.2 Der pn-Übergang . . . 396

33 Dioden . . . 397

33.1 Kennlinien . . . 398

33.2 Kenndaten und Grenzwerte . . . 399

33.3 Kennzeichnung von Halbleiter-Bauelementen . . . 401

33.4 Diodenarten . . . 402
33.4.1 Kapazitätsdioden . . . 403
33.4.2 Schalterdioden . . . 404
33.4.3 Schottky-Dioden . . . 404
33.4.4 Gleichrichter-Dioden . . . 405
33.4.5 Z-Dioden . . . 405

33.5 Anwendungsschaltungen . . . 406
33.5.1 Begrenzerschaltungen . . . 406
33.5.2 Gleichrichter . . . 408
33.5.3 Spannungsvervielfacher . . . 412
33.5.4 Diode als Konstantspannungsquelle (Z-Diode) . . . 413

34 Mehrschichtdioden und -trioden . . . 415

34.1 Vierschichtdioden . . . 416

34.2 Thyristoren . . . 417

34.3 Diac . . . 420

34.4 Triac . . . 421

34.5 Schutz der Dioden und Trioden . . . 421

34.6 Zündmethoden . . . 422

35 Transistoren . . . 425

35.1 Bipolare Transistoren . . . 426
35.1.1 Transistoreffekt . . . 426
35.1.2 Transistorkennlinien . . . 427
35.1.3 Kenn- und Grenzwerte des Transistors . . . 429

35.2 Feldeffekttransistoren (FET) . . . 432
35.2.1 Aufbau und Wirkungsweise des Sperrschicht-FET . . . 432
35.2.2 Aufbau und Wirkungsweise des MOSFET . . . 433
35.2.3 Kennlinien von FET . . . 434
35.2.4 Kennwerte von FET . . . 436

36 Besondere Halbleiter-Bauelemente . . . 441

36.1 Unijunction-Transistor (Doppelbasisdiode) . . . 442

36.2 Darlington-Transistor . . . 443

36.3 VMOS-Transistoren . . . 444

36.4 SIPMOS-Transistoren . . . 444

36.5 IGBT . . . 446

37 Analoge Verstärker . . . 449

37.1 Bipolarer Transistor als Verstärker . . . 450
37.1.1 Grundschaltungen . . . 451
37.1.2 Arbeitspunktstabilisierung . . . 454
37.1.3 Emitterschaltungen . . . 457
37.1.4 Kollektorschaltungen . . . 460
37.1.5 Basisschaltung . . . 461

37.2 Feldeffekt-Transistor als Verstärker . . . 461
37.2.1 Arbeitspunkteinstellung und -stabilisierung . . . 462
37.2.2 Grundschaltungen von FET . . . 464
37.2.3 Weitere Anwendungen . . . 469
37.3 Mehrstufige Verstärker . . . 470
38 Endstufen . . . 477
38.1 Betriebsarten . . . 478
38.2 Schaltungen . . . 478
39 Operationsverstärker . . . 485
39.1 Einführung . . . 486
39.2 Differenzverstärker . . . 486
39.3 Grundlagen des OP . . . 487
39.4 Operationsverstärker als Verstärker . . . 489
39.4.1 Verstärker mit frequenzunabhängiger Gegenkopplung . . . 489
39.4.2 Verstärker mit frequenzabhängiger Gegenkopplung . . . 492
39.4.3 OP als Leistungsverstärker . . . 494
39.4.4 Aktive Filterschaltungen . . . 495
40 Elektronische Schalter, Kippstufen . . . 499
40.1 Transistor als Schalter . . . 500
40.2 Kippschaltungen mit Transistoren . . . 503
40.2.1 Bistabile Kippstufe . . . 503
40.2.2 Monostabile Kippstufe . . . 504
40.2.3 Astabile Kippstufe . . . 505
40.2.4 Triggerschaltungen . . . 505
40.3 Operationsverstärker als Schalter . . . 507
40.4 Kippschaltungen mit Operationsverstärker . . . 507
40.4.1 Triggerschaltungen mit Operationsverstärker . . . 507
40.4.2 Astabile Kippstufe mit Operationsverstärker . . . 509
40.4.3 Monostabile Kippstufe mit Operationsverstärker . . . 509
40.4.4 Bistabile Kippstufe mit Operationsverstärker . . . 510
40.5 Zeitgeber 555 . . . 510
40.6 Trigger TCA 345 A . . . 511
41 Oszillatoren . . . 513
41.1 Allgemeines . . . 514
41.2 Sinusgeneratoren (RC-Oszillatoren) . . . 514
41.3 Funktionsgeneratoren . . . 516
42 Schaltungstechniken . . . 519
42.1 Integrierte Schaltungen . . . 520
42.2 SMD-Technik . . . 522
43 Optoelektronik . . . 525
43.1 Grundsätzliche Überlegungen . . . 526
43.2 Optoelektronische Bauelemente . . . 526
43.2.1 Fotowiderstand (LDR – light dependent resistor) . . . 526

43.2.2 Fotodiode und Fotoelement . . . 529
43.2.3 Fototransistoren . . . 531
43.2.4 Lumineszenzdioden und Flüssigkristalle . . . 532
43.3 Anzeigeeinheiten . . . 533
43.4 Signalübertragung mit Optokoppler . . . 535
43.5 Faseroptische Übertragungsmittel . . . 535
44 Analog-Digital-Wandler . . . 537
44.1 Grundlagen . . . 538
44.2 Spannungs-Frequenz-Wandler . . . 538
44.3 Sägezahnverfahren . . . 538
44.4 Dual-Slope-Verfahren . . . 539
44.5 Flash-Wandler . . . 539
44.6 Wandler nach dem Wägeverfahren . . . 540
44.7 Integrierte Wandler . . . 540
45 Digital-Analog-Wandler . . . 541
45.1 Grundlagen . . . 542
45.2 D/A-Wandler-Varianten . . . 542
45.3 Integrierte Wandler . . . 543

Teil VI Technische Kommunikation/Technisches Zeichnen

46 Grundlagen der zeichnerischen Darstellung . . . 547
46.1 Zeichengeräte . . . 548
46.2 Normen für Technische Zeichnungen . . . 548
46.3 Normteile und Konstruktionselemente . . . 549
46.4 Nutzen der Normung . . . 559
47 Schaltungsunterlagen . . . 561
47.1 Schaltzeichen nach DIN . . . 564
47.2 Elektrische Betriebsmittel . . . 566
47.3 Schaltungsunterlagen der Energietechnik . . . 566
47.4 Schaltungsunterlagen der Elektronik . . . 571
47.4.1 Allgemeines . . . 571
47.4.2 Schaltzeichen nach DIN 40900 Teil 12 Binäre Elemente . . . 571
47.4.3 Entwurf von Schaltungen . . . 574
47.5 Projektierung . . . 575
48 Schaltungssynthese und -analyse . . . 577
48.1 Beispiele aus der Elektrotechnik . . . 578
48.2 Beispiele aus der Elektronik . . . 578

49 CAD-Technik . . . 581
49.1 Allgemeines . . . 582
49.2 Hardware und Software . . . 582
49.3 Erstellen von Schaltplänen . . . 582
49.4 Erstellen von Layouts . . . 582
49.5 Anwendungen in der Elektronik . . . 586
49.6 Auswahl von CAD-Systemen . . . 586

Teil VII Datentechnik

50 Digitaltechnik . . . 589
50.1 Grundbegiffe der Digitaltechnik . . . 590
50.2 Logische Grundschaltungen . . . 590
50.2.1 Grundverknüpfungen . . . 590
50.2.2 Realisierungsmöglichkeiten logischer Verknüpfungen . . . 591
50.3 Schaltalgebra . . . 593
50.3.1 Allgemeines . . . 593
50.3.2 Normalform einer binären Funktion . . . 594
50.4 Zahlensysteme in der Digital- und Datenverarbeitung . . . 602
50.4.1 Dualsystem Synthese . . . 602
50.4.2 Hexadezimalsystem . . . 603
50.4.3 Rechnen mit Dualzahlen . . . 603
50.4.4 Zahlen in Rechenanlagen . . . 603
50.5 Codes . . . 606
50.5.1 Allgemeines . . . 606
50.5.2 Binär-Code . . . 606
50.5.3 BCD-Code . . . 606
50.5.4 Gray-Code . . . 607
50.5.5 Codierung alphanumerischer Zeichen . . . 608
50.5.6 Fehlererkennung und Redundanz . . . 608
50.6 Digitale Grundschaltungen . . . 611
50.6.1 Allgemeines . . . 611
50.6.2 Schaltnetze . . . 611
50.6.3 Schaltwerke . . . 617
50.6.4 Sonderschaltungen . . . 629

51 Integrierte Schaltkreise der Digitaltechnik . . . 633
51.1 Allgemeines . . . 634
51.2 Umgang mit integrierten Schaltungen . . . 634
51.3 Daten und Begriffe der Logikschaltungen . . . 634
51.3.1 Grenz- und Kenndaten . . . 634
51.3.2 Pegel . . . 635
51.3.3 Störsicherheit . . . 635
51.3.4 Lasteinheit . . . 635
51.3.5 Temperaturbereich . . . 636
51.3.6 Gatterlaufzeit . . . 636
51.3.7 Verlustleistung . . . 637
51.4 TTL-Familie . . . 637
51.4.1 Eigenschaften und Kenndaten . . . 637

51.4.2 Standard-TTL . . . 638
51.4.3 Schaltungen mit 3-state . . . 640
51.4.4 Schottky-TTL und Low-Power-Schottky . . . 641

51.5 Emittergekoppelte Logik . . . 641

51.6 Integrierte MOS-Schaltungen . . . 642
51.6.1 NMOS- und PMOS-Technik . . . 642
51.6.2 CMOS-Inverter . . . 642

51.7 Interfaceschaltungen . . . 646

51.8 Anwendungsspezifische integrierte Schaltungen . . . 647
51.8.1 Allgemeines . . . 647
51.8.2 Kundenspezifische IC's . . . 647
51.8.3 Programmierbare Logikbausteine . . . 648

51.9 Gehäuse . . . 659

52 Mikrocomputertechnik . . . 661

52.1 Komponenten eines Mikrocomputers . . . 662

52.2 Mikroprozessoren . . . 662
52.2.1 Allgemeines . . . 662
52.2.2 Architektur . . . 662
52.2.3 Übersicht gängiger Mikroprozessoren . . . 664
52.2.4 8-Bit-Mikroprozessoren . . . 664
52.2.5 16-Bit-Prozessoren . . . 672

52.3 Halbleiterspeicher . . . 676
52.3.1 Allgemeines . . . 676
52.3.2 Kenndaten und Technologie . . . 676
52.3.3 Bedeutung der Anschlüsse . . . 677
52.3.4 Organisation und Aufbau . . . 677
52.3.5 Zeitverhalten . . . 679
52.3.6 Speichertypen . . . 680
52.3.7 Speichererweiterung . . . 687
52.3.8 Zentralspeicher . . . 689

52.4 Peripheriebausteine . . . 689
52.4.1 Allgemeines . . . 689
52.4.2 BUS-Treiber . . . 690
52.4.3 Einfache E-/A-Bausteine für den parallelen Betrieb . . . 690
52.4.4 Programmierbare Schnittstellenbausteine . . . 692
52.4.5 Zeitgeberbausteine . . . 696
52.4.6 Programmierbarer E/A-Baustein mit Speicher und Zeitgeber . . . 699
52.4.7 Eingabe-Ausgabe-Bausteine für den seriellen Betrieb . . . 702
52.4.8 Bausteine mit Sonderfunktionen . . . 705

52.5 Mikrocontroller . . . 708
52.5.1 Allgemeines . . . 708
52.5.2 8-Bit-Mikrocontroller . . . 709
52.5.3 16-Bit-Mikrocontroller . . . 715

52.6 Maschinensprache . . . 718
52.6.1 Allgemeines . . . 718
52.6.2 Maschinencode . . . 718
52.6.3 Befehlsaufbau . . . 719
52.6.4 Befehlsdarstellung . . . 719
52.6.5 Befehle . . . 720
52.6.6 Befehlszyklus und Befehlszeiten . . . 723

52.7 Befehlsvorrat . . . 726

52.8 Hinweise zur Programmierung und Progammbeispiele . . . 726

53 Computertechnik . . . 729

53.1 Komponenten eines Computers . . . 730

53.2 Massenspeicher . . . 730
53.2.1 Magnetplatten . . . 730
53.2.2 CD-ROM- und CD-Laufwerk . . . 735

53.3 Eingabegeräte . . . 736
53.3.1 Tastatur . . . 736
53.3.2 Maus . . . 736

53.4 Ausgabegeräte . . . 736
53.4.1 Datensichtgeräte . . . 736
53.4.2 Drucker . . . 740

54 Programmiertechnik . . . 741

54.1 Programmiersprachen . . . 742
54.1.1 Assembler . . . 742
54.1.2 ADA . . . 742
54.1.3 ALGOL . . . 743
54.1.4 BASIC . . . 743
54.1.5 C . . . 743
54.1.6 FORTRAN . . . 743
54.1.7 PASCAL . . . 743
54.1.8 PL/M . . . 744

54.2 Grundlagen der Programmierung . . . 744
54.2.1 Interpreter . . . 744
54.2.2 Compiler . . . 744
54.2.3 Editor . . . 744
54.2.4 Integrierte Entwicklungsumgebung . . . 744
54.2.5 Methoden der Programmentwicklung . . . 745
54.2.6 Problembeschreibung . . . 745
54.2.7 Top-Down-Methode . . . 745
54.2.8 Bottom-Up-Methode . . . 745
54.2.9 Bewertung der Methoden . . . 746
54.2.10 Programm-Test . . . 746

55 Datenkommunikation . . . 747

55.1 Einführung . . . 748
55.1.1 Warum Datenkommunikation? . . . 748
55.1.2 Geschichte der Datenkommunikation . . . 748

55.2 Computernetzwerke . . . 749
55.2.1 Peer-to-Peer-Netz . . . 749
55.2.2 Client/Server-Netz . . . 749

55.3 Einteilung von Netzwerken . . . 750
55.3.1 Ausdehnung . . . 750
55.3.2 Dienstarten und Verbindungen . . . 750
55.3.3 Basisband – Breitband . . . 751
55.3.4 FDM – DWDM – TDM . . . 751

55.4 Netzwerk-Topologien . . . 752
55.4.1 Sternnetz . . . 752
55.4.2 Busnetz . . . 752
55.4.3 Ringnetz . . . 753
55.4.4 Baumnetz . . . 753
55.4.5 Vollverbindung . . . 753
55.4.6 Mischformen . . . 753
55.4.7 Sonderformen . . . 754

55.5 Das OSI-Referenzmodell . . . 754
55.5.1 Aufteilen langer Nachrichten . . . 754
55.5.2 Schicht 0: Übertragungsmedien . . . 756
55.5.3 Schicht 1: Bitübertragung . . . 758
55.5.4 Schicht 2: Datensicherung . . . 760
55.5.5 Schicht 3: Vermittlung . . . 762
55.5.6 Schicht 4: Transport . . . 766
55.5.7 Aufgaben der Schichten 5–7 . . . 768
55.6 Industrial Ethernet . . . 769
Literatur . . . 769

Teil VIII Automatisierungstechnik

56 Einführung . . . 773

57 Automatisierungsgeräte . . . 775

58 Grundzüge der SPS-Norm IEC 61131-3 . . . 779
58.1 Programmorganisationskonzept . . . 780
58.2 Deklaration von FB- und FC-Bausteinen . . . 780
58.3 Variablen . . . 782
58.4 Programmiersprachen . . . 784

59 Programmstrukturen . . . 787
59.1 Lineares Programm . . . 788
59.2 Gegliedertes Programm . . . 788
59.3 Parametrierbares Programm . . . 788

60 Eingabe- und Ausgabesignale . . . 789
60.1 Binäre Signale . . . 790
60.2 Digitale Signale . . . 790
60.3 Analoge Signale . . . 790

61 Eingabe-/Ausgabebaugruppen . . . 791

62 Verknüpfungssteuerungen . . . 793
62.1 Logische Grundverknüpfungen in verschiedenen Darstellungen . . . 794
62.2 Zusammengesetzte logische Grundverknüpfungen . . . 794
62.3 Schließer- und Öffnerkontakte, Drahtbruchsicherheit, Erdschlussgefahr 794
62.4 Speicherfunktionen . . . 795
62.5 Flankenauswertung . . . 797
62.6 Darstellung und Eigenschaften elektropneumatischer Stellglieder . . . 797
62.7 Regeln für das Umsetzen von Schützschaltungen in SPS-Programme . . . 798
62.8 Zeitfunktionen . . . 799
62.9 Zählerfunktionen . . . 799
62.10 Vergleichsfunktionen . . . 800
62.11 MOVE-Funktion . . . 801
62.12 EN/ENO-Mechanismus . . . 801

63 Aufruf und Wertübergaben zwischen Bausteinen nach IEC 61131-3 803
63.1 Aufrufhierarchie der Bausteine P, FB und FC 804
63.2 Aufruf von Funktionsbausteinen in FBS 804
63.3 Aufruf von Funktionsbausteinen in AWL 804
63.4 Aufruf von Funktionen in AWL 805

64 Ablaufsteuerung 807
64.1 Ablauf-Funktionsplan 808
64.2 Grafische Darstellung von Ablaufsteuerungsfunktionen 808
64.3 Betriebsartenteil und Bedienfeld 811

65 Kommunikation in Automatisierungssystemen 813
65.1 Bussysteme 814
65.2 PROFINET – Offener Industrial Ethernet Standard 815
65.3 OPC-Technologie 816

66 Steuerungssicherheit 817
66.1 Europäische Richtlinien und Sicherheitsnormen 818
66.2 Sicherheitsbegriff 818

67 Regelungstechnische Grundbegriffe der Automatisierungstechnik 821
67.1 Unterschied zwischen Steuern und Regeln, regelungstechnische Größen 822
67.2 Regler-Technologien 823

68 Regelstrecken 825
68.1 Beispiele für Regelstrecken 826
68.2 Beschreibungsmittel zur Darstellung von Regelstreckeneigenschaften 826

69 Regler 831
69.1 P-Regler, P-Regelfunktion 832
69.2 I-Regler, I-Regelfunktion 834
69.3 PI-Regler, PI-Regelfunktion 834
69.4 PID-Regler, PID-Regelfunktion 834
69.5 Vergleich der verschiedenen Reglertypen 835
69.6 PID-Reglerbaustein für digitale Abtastregelung 835
69.7 SPS als kontinuierlicher PID-Abtastregler 836
69.8 Digitaler Schrittregler mit PI-Verhalten 837
69.9 Zweipunktregler, Zweipunkt-Regelfunktion 838
69.10 Regelgüte 838
Literatur 839

Teil IX Regelungstechnik

70 Einführung 843

71 Strukturbild und Übertragungsglieder . . . 847

71.1 Strukturbild . . . 848

71.2 Übertragungsglieder . . . 848

71.3 Klassifikation von Übertragungsgliedern . . . 849
71.3.1 Statische und dynamische Übertragungsglieder . . . 849
71.3.2 Lineare und nichtlineare Übertragungsglieder . . . 849
71.3.3 Zeitinvariante und zeitvariante Übertragungsglieder . . . 849
71.3.4 Kausale und akausale Übertragungsglieder . . . 849

71.4 Beschreibung von Übertragungsgliedern um einen Arbeitspunkt . . . 849
71.4.1 Arbeitspunkt eines Übertragungsgliedes . . . 849
71.4.2 Linearisierung um den Arbeitspunkt . . . 850

71.5 Impulsantwort und Sprungantwort einfacher Übertragungsglieder . . . 850
71.5.1 P-Glied . . . 850
71.5.2 I-Glied . . . 851
71.5.3 D-Glied . . . 851
71.5.4 TZ-Glied . . . 851
71.5.5 PT_1-Glied . . . 851
71.5.6 PT_2-Glied . . . 851

72 Übertragungsverhalten linearer, zeitinvarianter Übertragungsglieder . . . 853

72.1 Übertragungsfunktion . . . 854

72.2 Frequenzgangfunktion . . . 854

72.3 Frequenzkennlinien elementarer Übertragungsfunktionen . . . 855
72.3.1 Verstärkungsfaktor . . . 855
72.3.2 Integrierer . . . 855
72.3.3 Differenzierer . . . 855
72.3.4 Reeller Pol . . . 855
72.3.5 Reelle Nullstelle . . . 856
72.3.6 Konjugiert komplexes Polpaar . . . 856
72.3.7 Konjugiert komplexes Nullstellenpaar . . . 857
72.3.8 Totzeit . . . 857
72.3.9 Allpass erster Ordnung . . . 858
72.3.10 Allpass zweiter Ordnung . . . 858

73 Stabilität . . . 859

73.1 Definition der Stabilität . . . 860
73.1.1 Asymptotische Stabilität . . . 860
73.1.2 BIBO-Stabilität . . . 860

73.2 Stabilitätskriterien . . . 860
73.2.1 Grundlegendes Stabilitätskriterium . . . 860
73.2.2 Hurwitz-Kriterium . . . 861
73.2.3 Nyquist-Kriterium . . . 862
73.2.4 Phasenreserve und Amplitudenreserve . . . 862

74 Entwurf von Regelkreisen . . . 865

74.1 Forderungen an die Regelung . . . 866
74.1.1 Stabilität . . . 866
74.1.2 Stationäre Genauigkeit . . . 866
74.1.3 Hinreichende Dämpfung . . . 867
74.1.4 Hinreichende Schnelligkeit . . . 867

74.2 Reglertypen . . . 867
74.2.1 Standardregler . . . 867

74.2.2 Korrekturglieder . . . 868

74.3 Faustregeln für die Wahl der Reglerstruktur und der Reglerparameter . . . 869

74.4 Reglerentwurf mit dem Frequenzkennlinienverfahren . . . 870
74.4.1 Entwurf eines PI-Reglers für ein Verzögerungssystem . . . 870
74.4.2 Reglerentwurf für eine doppelt integrierende Strecke . . . 870

74.5 Einstellregeln bei speziellen Regelstrecken . . . 871
74.5.1 Betragsoptimum . . . 871
74.5.2 Symmetrisches Optimum . . . 872

Literatur . . . 872

Teil X Messtechnik

75 Grundlagen und Grundbegriffe der Messtechnik . . . 875

75.1 Begriffe . . . 876

75.2 Einheiten . . . 876

75.3 Messabweichung, Messfehler . . . 877
75.3.1 Systematische Abweichungen . . . 877
75.3.2 Zufällige Abweichungen . . . 877
75.3.3 Arithmetischer Mittelwert, Erwartungswert . . . 878
75.3.4 Standardabweichung . . . 878

75.4 Abweichungsfortpflanzung, Fehlerfortpflanzung . . . 879

75.5 Fehlerangaben von Messgeräten . . . 881
75.5.1 Analog anzeigende Messgeräte . . . 881
75.5.2 Digital anzeigende Messgeräte . . . 881

75.6 Arithmetischer Mittelwert und Effektivwert von Wechselgrößen . . . 882

75.7 Häufigkeitsverteilung, Vertrauensbereich . . . 882

75.8 Bearbeitung und Auswertung von Messwerten . . . 883

76 Analog anzeigende Messgeräte . . . 885

76.1 Grundlagen . . . 886

76.2 Drehspul-Messwerk . . . 886

76.3 Dreheisen-Messwerk . . . 887

76.4 Elektrodynamisches Messwerk . . . 887

76.5 Symbole und Instrumentenbeschriftungen . . . 888

77 Oszilloskop . . . 889

77.1 Übersicht . . . 890

77.2 Aufbau eines Oszilloskopes . . . 890

77.3 Zusatzeinrichtungen . . . 892

78 Schreibende Messgeräte . . . 895

78.1 Betriebsarten . . . 896

78.2 Schreibstiftauslenkung . . . 896

79 Digital anzeigende Messgeräte . . . 897

79.1 Digitalvoltmeter . . . 898

79.2 Digitalmultimeter . . . 898

79.3 Messung von Kapazitäten, Frequenzen und Stromverstärkungen . . . 899

79.4 Messung von Temperaturen . . . 899

80 Messverfahren zur Messung elektrischer Größen . . . 901

80.1 Messung von Gleichspannungen . . . 902
80.1.1 Analog anzeigende Spannungsmessgeräte . . . 902
80.1.2 Digital anzeigende Spannungsmessgeräte . . . 902
80.1.3 Messabweichung durch den Innenwiderstand des Spannungsmessers . . . 902
80.1.4 Spannungsmessung mit dem Kompensator . . . 903

80.2 Messung von Gleichströmen . . . 903
80.2.1 Analog anzeigende Strommessgeräte . . . 903
80.2.2 Digital anzeigende Strommessgeräte . . . 903
80.2.3 Messabweichung durch den Innenwiderstand des Strommessers . . . 904

80.3 Messbereichserweiterung . . . 904
80.3.1 Spannungsmessung . . . 904
80.3.2 Strommessung . . . 904

80.4 Messung von Wechselspannungen . . . 904
80.4.1 Analog anzeigende Wechselspannungsmessgeräte . . . 904
80.4.2 Digital anzeigende Wechselspannungsmessgeräte . . . 905

80.5 Messung von Wechselströmen . . . 906
80.5.1 Analog anzeigende Wechselstrommessgeräte . . . 906
80.5.2 Digital anzeigende Wechselstrommessgeräte . . . 906

80.6 Widerstands- und Impedanzmessung . . . 906
80.6.1 Gleichstrom-Messbrücken zur Widerstandsmessung . . . 906
80.6.2 Wechselstrom-Messbrücken zur Widerstands- und Impedanzmessung . . . 909
80.6.3 Vergleich mit bekanntem Widerstand – Spannungsvergleich . 910
80.6.4 Messung von Strom und Spannung . . . 910
80.6.5 Widerstandsmessung mit analogen Multimetern . . . 911

80.7 Leistungsmessung . . . 911
80.7.1 Wirkleistungsmessung . . . 912
80.7.2 Blindleistungsmessung . . . 914
80.7.3 Scheinleistungsmessung . . . 915
80.7.4 Messbereichserweiterung bei der Leistungsmessung . . . 915
80.7.5 Leistungsfaktormessung . . . 915

80.8 Messung der Arbeit . . . 916

80.9 Messung von L, C, Gütefaktor und Verlustfaktor . . . 917
80.9.1 Messung von $|\underline{Z}_L|$ oder $|\underline{Z}_C|$ . . . 917
80.9.2 Messung von Z_L, Z_C, Gütefaktor und Verlustfaktor . . . 917

80.10 Messung magnetischer Größen . . . 918
80.10.1 Magnetischer Fluss . . . 918
80.10.2 Magnetische Flussdichte . . . 919
80.10.3 Magnetische Feldstärke . . . 919
80.10.4 Permeabilität . . . 920

81 Messverfahren zur Messung nichtelektrischer Größen . . . 921

81.1 Messaufnehmer . . . 923
81.1.1 Ohm'sche Aufnehmer . . . 923
81.1.2 Kapazitive Aufnehmer . . . 924
81.1.3 Induktive Aufnehmer . . . 925
81.1.4 Optische Aufnehmer . . . 925
81.1.5 Ladungsliefernde Aufnehmer . . . 926
81.1.6 Thermische Aufnehmer . . . 927
81.1.7 Chemische Aufnehmer . . . 929
81.1.8 Aufnehmer zur Messung von Gaskonzentrationen allgemein . 929

81.2 Messverfahren . . . 930
81.2.1 Kraftmessung mit Dehnungsmessstreifen (DMS) . . . 930
81.2.2 Füllstandsmessung und Messung der Foliendicke . . . 933
81.2.3 Drehzahlmessung . . . 934
81.2.4 Durchflussmessung . . . 935
81.2.5 Zeit- und Frequenzmessung . . . 937
81.2.6 Weg- und Winkelmessung . . . 938
81.2.7 Beschleunigungsmessung . . . 939

82 Messdatenaufbereitung . . . 943

82.1 Verringerung der Störeinflüsse von außen . . . 944

82.2 Messverstärker . . . 945

83 Bussysteme für die Messtechnik . . . 947

83.1 Grundbegriffe . . . 948

83.2 IEC-Bus . . . 948

83.3 DIN-Messbus . . . 949

83.4 Aktuator-Sensor-Interface (ASI) . . . 949

83.5 CAN-BUS . . . 952

83.6 USB-Übertragung . . . 952

83.7 LAN (*L*ocal *A*rea *N*etwork) . . . 952

84 Probleme bei der Digitalisierung analoger Messwerte . . . 953

84.1 Fehler bei der Digitalisierung . . . 954

84.2 Signal-Quantisierungs-Geräuschabstand . . . 954

84.3 Verbesserung des Signal-Rausch-Verhältnisses . . . 955

84.4 Abtast-Halte-Glied . . . 955

84.5 Aliasing . . . 955

84.6 Erfassung von Momentanwerten . . . 956

85 PC-gestützte Messverfahren und Messsignalanalyse . . . 957

85.1 Statistische Verfahren zur Messsignalauswertung . . . 958

85.2 Graphische Darstellung . . . 958

85.3 Ermittlung von Kenngrößen, Klassierung . . . 959

85.4 Messsignalanalyse . . . 959
85.4.1 Verfahren . . . 959
85.4.2 Anwendungen . . . 960

85.5 Automatisierung von Messabläufen . . . 960

Literatur . . . 960

Teil XI Energietechnik

86 Elektrische Maschinen . . . 963

86.1 Transformatoren . . . 964
86.1.1 Aufgaben eines Transformators . . . 964
86.1.2 Bauteile eines Transformators . . . 964
86.1.3 Wirkungsweise eines Einphasen-Transformators . . . 965
86.1.4 Aufbau und Schaltung von Drehstrom-Transformatoren . . . 970
86.1.5 Parallelschalten von Transformatoren . . . 971
86.1.6 Transformatorschutz . . . 972
86.1.7 Überlastung von Transformatoren . . . 972
86.1.8 Aufstellen von Transformatoren . . . 973
86.1.9 Sondertransformatoren . . . 973
86.1.10 Messwandler . . . 975

86.2 Drehstrommaschinen . . . 976
86.2.1 Die Drehstromasynchronmaschine . . . 976
86.2.2 Linearmotor . . . 985
86.2.3 Drehstromsynchronmaschinen . . . 985

86.3 Einphasen-Asynchronmotoren . . . 988
86.3.1 Einsträngiger Motor . . . 988
86.3.2 Zweisträngiger Motor . . . 988
86.3.3 Kondensatormotor . . . 988
86.3.4 Spaltpolmotor . . . 989

86.4 Drehstrommotor im Einphasenbetrieb . . . 989

86.5 Sonderbauformen . . . 989
86.5.1 Schrittmotor . . . 989
86.5.2 Servomotor . . . 990

86.6 Gleichstrommaschinen . . . 991
86.6.1 Aufbau und Wirkungsweise . . . 991
86.6.2 Betriebsverhalten von Gleichstrommaschinen . . . 992
86.6.3 Betriebsverhalten von Gleichstromgeneratoren . . . 997
86.6.4 Gleichstrommaschine am Wechsel- oder Drehstromnetz . . . 998
86.6.5 Universalmotor . . . 1001

86.7 Auswahl von Motoren . . . 1001
86.7.1 Auswahl unter Berücksichtigung der Normen . . . 1001

Literatur . . . 1007

87 Elektrische Energieversorgung . . . 1009

87.1 Energiewirtschaftliche Rahmenbedingungen . . . 1010
87.1.1 Einführung . . . 1010
87.1.2 Gesetzliche Rahmenbedingungen . . . 1010
87.1.3 Struktur der Elektrizitätswirtschaft . . . 1010
87.1.4 Vertragsbeziehungen und Energiehandel . . . 1012

87.2 Elektrische Energieerzeugung . . . 1013
87.2.1 Elektrischer Energieverbrauch . . . 1013
87.2.2 Anteile an der Stromerzeugung und Kraftwerksleistung . . . 1014
87.2.3 Überblick und Klassifizierung von Kraftwerksarten . . . 1015
87.2.4 Fossil befeuerte Kraftwerke . . . 1015
87.2.5 Regenerative Kraftwerke . . . 1022

87.3 Elektrische Energieübertragung und Energieverteilung . . . 1034
87.3.1 Spannungssysteme . . . 1034
87.3.2 Aufbau und Betrieb von Energieversorgungsnetzen . . . 1035
87.3.3 Frequenzregelung . . . 1036

87.3.4 Spannungsregelung 1038
87.3.5 Herausforderungen bei einer Energieversorgung mit fluktuierender Erzeugung 1045

87.4 Elektrische Betriebsmittel 1053
87.4.1 Freileitungen 1053
87.4.2 Kabel 1063
87.4.3 Schaltanlagen 1068
87.4.4 Netzformen und Schutz auf der Niederspannungsebene 1071

Literatur 1074

88 Elektrische Energieanwendung 1075

88.1 Kompensationsanlagen 1076

88.2 Beleuchtungsanlagen 1077
88.2.1 Grundgrößen der Lichttechnik 1078
88.2.2 Lichtquellen 1078
88.2.3 Glühlampen 1078
88.2.4 Leuchtstofflampen 1079
88.2.5 Entladungslampen 1079
88.2.6 Leuchten 1079
88.2.7 Berechnung von Beleuchtungsanlagen 1080

Literatur 1081

Teil XII Leistungselektronik

89 Halbleiter in der Leistungselektronik 1085

90 Leistungshalbleiter 1087

90.1 Halbleiterdioden 1088
90.1.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte 1088
90.1.2 Dynamisches Verhalten, TSE 1088
90.1.3 Verlustleistung 1089
90.1.4 Bauarten 1089
90.1.5 Schutzschaltungen, TSE-Beschaltung 1090
90.1.6 Sicherungen 1090
90.1.7 Zusammenschaltung von Dioden 1090

90.2 Bipolartransistoren 1091
90.2.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte, SOAR 1092
90.2.2 Dynamisches Verhalten 1093
90.2.3 Verlustleistung 1095
90.2.4 Bauarten 1095
90.2.5 Schutzschaltungen 1096
90.2.6 Parallelschaltung von bipolaren Transistoren 1096
90.2.7 Anwendungen 1097

90.3 Feldeffekttransistoren 1097
90.3.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte 1097
90.3.2 Dynamisches Verhalten 1098
90.3.3 Verlustleistung 1100
90.3.4 Schutzschaltungen 1100
90.3.5 Zusammenschaltung von Feldeffekttransistoren 1100

90.4 IGBT-Transistoren 1100
90.4.1 Aufbau von IGBT-Transistoren 1101
90.4.2 BIMOS-Schaltstufe 1101

90.5 Thyristoren 1101

90.5.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte 1101
90.5.2 Dynamisches Verhalten 1103
90.5.3 Verlustleistung 1104
90.5.4 Schutzschaltungen 1104
90.5.5 Zusammenschaltung von Thyristoren 1105
90.5.6 Löschen von Thyristoren 1105
90.5.7 Einsatz in kritischen Anwendungsfällen 1106
90.5.8 Spezielle Bauarten von Thyristoren 1106

91 Einsatz von Leistungshalbleitern als Schalter 1109

91.1 Schalter allgemein 1110

91.2 Halbleiter-Relais 1111

91.3 Halbleiterkühlung 1112
91.3.1 Grundlagen, statisches Verhalten 1112
91.3.2 Dynamisches Verhalten 1114

91.4 Ansteuerschaltungen 1115
91.4.1 Ansteuerschaltungen allgemein 1115
91.4.2 Ansteuersignale für Bipolartransistoren 1116
91.4.3 Ansteuersignale für Feldeffekttransistoren und IGBT 1117
91.4.4 Ansteuersignale für Thyristoren 1118
91.4.5 Ansteuersignale für GTO-Thyristoren 1119

91.5 Schutzschaltungen 1120
91.5.1 Überspannungsschutz 1120
91.5.2 Überstromschutz 1122

Teil XIII Nachrichtentechnik

92 Grundlagen der Nachrichtenübertragung 1127

92.1 Prinzip der elektrischen Nachrichtenübertragung 1128

92.2 Aufgaben der Nachrichtentechnik 1129

92.3 Grundbegriffe 1129

92.4 Nachricht, Information und Signal 1130
92.4.1 Informationsgehalt 1130
92.4.2 Signale in der Nachrichtentechnik 1130
92.4.3 Entropie 1131
92.4.4 Redundanz 1131
92.4.5 Informationsfluss 1132
92.4.6 Kanalkapazität, Dynamik 1132
92.4.7 Nachrichtenquader 1132
92.4.8 Signale im Zeitbereich: Analog, digital, kontinuierlich, diskret 1133
92.4.9 Signale im Frequenzbereich 1133
92.4.10 Abtasttheorem von Shannon 1138
92.4.11 Zufällige (stochastische) Signale 1138
92.4.12 Verzerrungen 1141

92.5 Kenngrößen der Übertragungsstrecke 1143
92.5.1 Dämpfungsfaktor 1143
92.5.2 Übertragungsfaktor 1143
92.5.3 Dämpfungsmaß 1144
92.5.4 Übertragungsmaß, Verstärkungsmaß 1144
92.5.5 Pegel 1144

93 Zweitore, Vierpole 1147

93.1 Zweitor allgemein 1148
93.1.1 Grundlagen 1148
93.1.2 Zweitorgleichungen, Zusammenschaltung von Zweitoren . . . 1148
93.1.3 Bestimmung der Zweitorparameter 1149
93.1.4 Elementarzweitore 1150
93.1.5 Betriebskenngrößen 1151

93.2 Spezielle Zweitore 1155
93.2.1 Übertragungssymmetrische (reziproke) Zweitore 1155
93.2.2 Widerstandssymmetrische Zweitore 1156
93.2.3 Längssymmetrische Zweitore 1156
93.2.4 Rückwirkungsfreie Zweitore 1156

93.3 Wellenparameter passiver Zweitore 1157
93.3.1 Allgemeine passive Zweitore 1157
93.3.2 Längssymmetrische passive Zweitore 1157
93.3.3 Wellenwiderstand bei passiven längssymmetrischen Zweitoren 1158
93.3.4 Übertragungsmaß bei passiven längssymmetrischen Zweitoren 1158
93.3.5 Spezielle Zweitore 1159

94 Leitungen 1161

94.1 Leitungsbeläge und Leitungsgleichungen 1162

94.2 Leitung mit sinusförmigen Spannungen und Strömen 1163
94.2.1 Allgemeine Lösung 1163
94.2.2 Wellenwiderstand 1164
94.2.3 Ausbreitungskoeffizient 1164
94.2.4 Verlustlose Leitung 1164
94.2.5 Lösung mit Zeigerdarstellung 1164
94.2.6 Unendlich lange Leitung 1165
94.2.7 Anpassung 1165
94.2.8 Phasengeschwindigkeit, Gruppengeschwindigkeit 1165

94.3 Leitung mit sinusförmigen Spannungen und Strömen und beliebiger Abschlussimpedanz $\underline{Z}_e$ 1166
94.3.1 Reflexionsfaktor, Übertragungsfaktor 1166
94.3.2 Eingangsimpedanz 1167
94.3.3 Verzerrungsfreie Leitung 1167
94.3.4 Leitung als Vierpol 1168

94.4 Verlustlose Leitung 1169
94.4.1 Eigenschaften 1169
94.4.2 Wanderwellen bei Reflexion am Leitungsein- und -ausgang . 1169
94.4.3 Elektrisch lange Leitung 1170
94.4.4 Leitung als Transformator 1170
94.4.5 Stehende Wellen 1171
94.4.6 Kettenleiter 1171
94.4.7 Wellenfilter 1171
94.4.8 Filterentwurf 1172
94.4.9 Dielektrische Filter 1175

94.5 Daten von Leitungen 1177

94.6 Hochfrequenzleitungen 1177
94.6.1 Hochfrequenz-Koaxialkabel 1177
94.6.2 Hohlleiter 1179
94.6.3 Streifenleitungen 1182

94.7 $\underline{s}$-Parameter 1184
94.7.1 Signalflussdiagramm 1184
94.7.2 Leistungsverstärkung 1186

94.8 Kreisdiagramm . . . 1186
94.8.1 Doppel-Kreisdiagramm . . . 1189
94.8.2 s-Parameter im Kreisdiagramm . . . 1190

95 Antennen . . . 1193

95.1 Grundlagen . . . 1194

95.2 Kenngrößen . . . 1196

95.3 Ausführungsformen von Antennen . . . 1198
95.3.1 Vertikalantenne . . . 1198
95.3.2 Rahmenantenne . . . 1198
95.3.3 Ferritantenne . . . 1198
95.3.4 $\lambda/2$-Dipol, $\lambda/2$-Faltdipol . . . 1199
95.3.5 Breitbanddipol . . . 1199
95.3.6 Gruppenstrahler . . . 1199
95.3.7 Yagi-Antenne . . . 1199
95.3.8 Langdrahtantenne . . . 1200
95.3.9 Rohrschlitzstrahler . . . 1200
95.3.10 Parabolantenne . . . 1200

95.4 Wellenausbreitung . . . 1200
95.4.1 Boden- und Raumwelle . . . 1200
95.4.2 Erdatmosphäre . . . 1201
95.4.3 Wellenausbreitung im Plasma . . . 1201
95.4.4 Wellenausbreitung im Bereich 30 kHz bis 30 GHz . . . 1202

96 Modulation . . . 1203

96.1 Übersicht . . . 1204

96.2 Sinusträger – mit Analogsignal moduliert . . . 1204
96.2.1 Amplitudenmodulation (AM) . . . 1205
96.2.2 Sonderformen der Amplitudenmodulation . . . 1209
96.2.3 Technische Ausführung der Amplitudenmodulation . . . 1210
96.2.4 Winkelmodulation . . . 1211

96.3 Sinusträger – mit Digitalsignal moduliert . . . 1217
96.3.1 Amplitudenumtastung (ASK) . . . 1217
96.3.2 Frequenzumtastung (FSK) . . . 1218
96.3.3 Phasenumtastung (PSK) . . . 1218

96.4 Pulsträger uncodiert . . . 1221
96.4.1 Pulsamplitudenmodulation (PAM) . . . 1221
96.4.2 Pulsfrequenz- und Pulsphasenmodulation (PFM, PPM) . . . 1222
96.4.3 Pulsdauermodulation (PDM) . . . 1223
96.4.4 Pulscodemodulation (PCM) . . . 1224

97 Filter . . . 1229

97.1 Einfache passive R-C-Filter . . . 1230
97.1.1 Allgemeines . . . 1230
97.1.2 Passive R-C-Filter . . . 1231

97.2 Weitere Ausführungsformen . . . 1235
97.2.1 Quarzfilter, keramische Filter . . . 1235
97.2.2 Digitale Filter . . . 1236
97.2.3 Filter mit geschalteten Kondensatoren, SC-Filter . . . 1237

97.3 Allgemeiner Filterentwurf . . . 1238
97.3.1 Allgemeine Eigenschaften und Entwurfskriterien . . . 1238
97.3.2 Tiefpassfilter . . . 1240
97.3.3 Filtertransformationen . . . 1242
97.3.4 Besondere Filterschaltungen . . . 1244

98 Empfängerschaltungstechnik 1245
98.1 Geradeausempfänger 1246
98.2 Überlagerungsempfänger 1246
98.3 Automatische Verstärkungsregelung (AVR) 1246
98.4 Weitere Schaltungskonzepte 1247

99 Ton- und Bildübertragung 1249
99.1 Rundfunk-Stereoübertragung 1250
99.2 Digitale Farbfernsehtechnik 1251

100 Mehrfachübertragung – Multiplexverfahren 1257
100.1 Zeitmultiplexverfahren 1258
100.1.1 Analoge Signalübertragung 1258
100.1.2 Digital codierte Signalübertragung 1259
100.2 Frequenzmultiplexverfahren 1259
100.3 Raummultiplexverfahren 1260

101 Richtfunktechnik 1261

102 Nachrichtenübertragung über Satellit 1265

103 Nachrichtenübertragung über Lichtwellenleiter (LWL) 1269
103.1 Physikalische Grundlagen 1270
103.2 Grundmodelle von Lichtwellenleitern 1271
103.2.1 Mehrmoden-Stufenindex mit Totalreflexion 1271
103.2.2 Mehrmoden-Gradientenindex 1271
103.2.3 Einmoden-Stufenindex 1272
103.3 Technische Ausführung von Lichtwellenleitern 1273
103.4 Lichtsender 1274
103.5 Lichtempfänger 1274
103.6 Verbinden von Lichtleitern 1275
103.7 Optische Schalter 1275

104 Funkmesstechnik 1277
104.1 Radar 1278
104.1.1 Grundlagen, Kenngrößen 1278
104.1.2 Daten von Radaranlagen 1280
104.2 Funkortungssystem OMEGA 1280
104.3 GPS – Global Positioning System 1280

105 Elektroakustische Wandler 1283
105.1 Definitionen, Kenngrößen 1284
105.2 Schallempfänger 1285
105.2.1 Kenngrößen für Mikrofone 1285
105.2.2 Mikrofonsysteme 1286
105.2.3 Daten und Eigenschaften verschiedener Mikrofonsysteme . . . 1288
105.3 Schallsender 1289
105.3.1 Lautsprecher- und Hörer-Systeme 1289
105.3.2 Kenngrößen, Daten 1290

106 Vermittlungstechnik 1293
106.1 Grundbegriffe 1294
106.2 Vermittlung 1294
106.3 Verkehrstheorie 1295
106.4 Endgerät 1296

107 Kommunikations- und Datennetze 1299
107.1 Strukturen lokaler Kommunikations- und Datennetze 1300
107.2 Öffentliche Kommunikations- und Datensysteme 1300
107.2.1 Einführung 1300
107.2.2 ISDN 1301
107.3 Öffentliche Funknetze 1305
107.3.1 Mobilfunk mit GSM-Technik 1305
107.3.2 Mobilfunk mit UMTS-Technik 1306
107.4 Private Funknetze 1307
107.4.1 Schnurlostelefone 1307
107.4.2 Bluetooth 1307
107.4.3 WLAN 1308

108 Optimierte Nachrichten- und Datenübertragung 1309
108.1 Codierung 1310
108.1.1 Quellencodierung 1310
108.1.2 Kanalcodierung 1312
108.2 Optimalfilter 1316
108.3 Anwendung der Korrelation bei gestörten Signalen 1317
108.4 Verschlüsselung von Daten 1318
108.5 Fensterung 1321
Literatur 1322

Teil XIV Signal- und Systemtheorie

109 Einführung 1327
109.1 Darstellung in der Zeit- und in der Frequenzebene 1328
109.2 Hinweise zur Anwendung 1329

110 Grundbegriffe 1331

111 Periodische nichtsinusförmige zeitkontinuierliche Signale 1333
111.1 Reelle und komplexe Fourierreihe 1334
111.2 Beispiele und Anwendungen 1335

112 Nichtperiodische zeitkontinuierliche Signale 1339
112.1 Fouriertransformation 1340
112.1.1 Transformationsregeln 1340
112.1.2 Eigenschaften 1341
112.1.3 Korrespondenztabelle 1341
112.1.4 Beispiele 1342
112.2 Laplacetransformation 1343

112.2.1 Einführung . . . 1343
112.2.2 Transformation . . . 1343
112.2.3 Sätze zur Laplacetransformation . . . 1344
112.2.4 Anwendung der Laplacetransformation bei bekanntem $H(s)$ . . . 1345
112.2.5 Bestimmung von $H(s)$ . . . 1346
112.2.6 Beispiele . . . 1347

113 Spezielle Signale . . . 1355

113.1 Stoßfunktion, δ-Funktion, Dirac-Impuls . . . 1356

113.2 Sprungfunktion $\sigma(t)$ . . . 1356

113.3 Verknüpfung von σ- und δ-Funktion . . . 1357

113.4 Harmonische Schwingungen . . . 1357

114 Leistung . . . 1359

115 Faltungsintegral . . . 1363

116 Abtasttheorem . . . 1367

117 Nichtkontinuierliche (zeitdiskrete) Signale . . . 1369

117.1 Diskrete Fouriertransformation (DFT) . . . 1370

117.2 Schnelle Fouriertransformation (FFT) . . . 1371

117.3 z-Transformation . . . 1372

118 Zufällige Signale . . . 1375

118.1 Grundbegriffe und Kenngrößen . . . 1376

118.2 Verteilungen . . . 1379
118.2.1 Binomialverteilung . . . 1379
118.2.2 Poissonverteilung . . . 1379
118.2.3 Normalverteilung, Gaußverteilung . . . 1380

118.3 Rauschen . . . 1381

118.4 Signalerkennung bei gestörter Übertragung . . . 1382
118.4.1 Erkennen versteckter Periodizitäten . . . 1382
118.4.2 Signalerkennung allgemein . . . 1382
118.4.3 Signalangepasste Filter (matched filter) . . . 1382

Literatur . . . 1383

Sachverzeichnis . . . 1385

Mathematik

Teil I



1 Arithmetik 3
2 Gleichungen 29
3 Planimetrie 57
4 Stereometrie 77
5 Funktionen 85
6 Trigonometrie 111
7 Analytische Geometrie 121
8 Differential- und Integralrechnung 143
9 Anhang 183

1 Arithmetik

Arnfried Kemnitz



1.1 Mengen 4
1.2 Aussageformen und logische Zeichen 4
1.3 Einteilung der Zahlen 6
1.4 Grundrechenarten 7
1.5 Grundlegende Rechenregeln 8
1.6 Potenz- und Wurzelrechnung 12
1.7 Dezimalzahlen und Dualzahlen 17
1.8 Logarithmen 18
1.9 Ungleichungen 21
1.10 Komplexe Zahlen 23

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_1

1.1 Mengen

Die in der Mathematik betrachteten Gegenstände werden oftmals durch Symbole, meistens Buchstaben, bezeichnet. Dabei kennzeichnen manche Symbole feste Dinge, zum Beispiel π das Verhältnis zwischen Umfang und Durchmesser eines beliebigen Kreises. Andere Symbole sind Veränderliche (auch Variable oder Platzhalter genannt), das heißt, sie können jeden Gegenstand einer Klasse von Gegenständen bezeichnen.

In der Mathematik wird jede Zusammenfassung von bestimmten wohl unterscheidbaren Objekten zu einer Gesamtheit eine Menge genannt. Eine Menge ist definiert, wenn feststeht, welche Objekte zu dieser Menge gehören und welche nicht. Die zur Menge gehörenden Objekte heißen ihre Elemente. Mengen werden meistens mit großen lateinischen Buchstaben bezeichnet und die Elemente mit kleinen Buchstaben.

Es gibt zwei Möglichkeiten, Mengen zu definieren: Durch Aufzählen ihrer Elemente, die in beliebiger Reihenfolge zwischen geschweiften Klammern (Mengenklammern) gesetzt sind und durch Kommata getrennt werden (Schreibweise: {Element 1, Element 2, ...}).

Durch Angabe einer die Elemente charakterisierenden Eigenschaft (Schreibweise: $\{x \mid x$ erfüllt Eigenschaft$\}$).

Eine Menge von Punkten heißt Punktmenge.

Beispiele

1. $A = \{1, 2, 3\}$ (die Menge A besteht aus den Elementen 1, 2 und 3)
2. $B = \{x \mid x^2 - 1 = 0\}$ (die Menge B besteht aus den Elementen x, für die $x^2 - 1 = 0$ gilt)
3. $B = \{1, -1\}$ (da $x^2 - 1 = 0$ die Lösungen $x = 1$ und $x = -1$ besitzt, kann man die Menge B auch in dieser Form schreiben)
4. $C = \{-1, 0, 1, 2, 3, 4, 5\}$ (die Menge C besteht aus den Elementen -1, 0, 1, 2, 3, 4, 5) ◀

Gehört ein Objekt a einer Menge M an, so schreibt man $a \in M$ (gelesen: a ist Element von M). Gehört a nicht zu M, so schreibt man $a \notin M$.

Wenn jedes Element einer Menge M auch Element einer Menge N ist, so nennt man M Teilmenge von N und schreibt $M \subseteq N$. Nach dieser Definition ist offenbar jede Menge Teilmenge von sich selbst.

Die leere Menge $\varnothing = \{;\}$ enthält kein Element.

Beispiele

$2 \in A$; $2 \in C$; $4 \in C$; $4 \notin A$; $A \subseteq C$; $\varnothing = \{x \mid x \neq x\}$ ◀

Die Vereinigung $A \cup B$ zweier Mengen A und B besteht aus denjenigen Elementen, die in A oder in B, also in mindestens einer der beiden Mengen A, B enthalten sind:

$$A \cup B = \{x \mid x \in A \text{ oder } x \in B\}$$

Der Durchschnitt $A \cap B$ zweier Mengen A und B besteht aus denjenigen Elementen, die sowohl in A als auch in B, also gleichzeitig in beiden Mengen A, B enthalten sind:

$$A \cap B = \{x \mid x \in A \text{ und } x \in B\}$$

Beispiel

$A = \{1, 2, 3\}$, $B = \{1, -1\}$; $A \cup B = \{-1, 1, 2, 3\}$, $A \cap B = \{1\}$ ◀

Eine Menge heißt endlich, wenn sie nur endlich viele Elemente besitzt. Die Anzahl der Elemente einer endlichen Menge M heißt Mächtigkeit der Menge, bezeichnet mit $|M|$.

Beispiele

1. $M = \{2, 4, 6, 8, 10\} \Rightarrow |M| = 5$
2. $M = \{1, 2, 3, \ldots, 99, 100\} \Rightarrow |M| = 100$ ◀

1.2 Aussageformen und logische Zeichen

1.2.1 Aussageformen

Eine Aussageform ist ein mathematischer Ausdruck, in dem Variable vorkommen.

Aussageformen erhalten einen Wahrheitswert, wenn allen in ihnen vorkommenden Variablen ein Wert zugeordnet wird.

Beispiele

1. Die Aussageform „$x - 3 = 5$“ wird zu einer wahren Aussage, wenn man für x die Zahl 8 einsetzt ($x = 8$ ist die Lösung der Gleichung).
2. Die Aussageform „$x^2 = 1$“ wird zu einer wahren Aussage, wenn man für x die Zahl 1 oder -1 einsetzt ($x_{1,2} = \pm 1$ sind die Lösungen der quadratischen Gleichung).
3. Die Aussageform „$x + 1 = 3$“ wird zu einer falschen Aussage, wenn man für x die Zahl 1 einsetzt (denn die Lösung der Gleichung ist $x = 2$). ◀

1.2.2 Logische Zeichen

In der Mathematik ist es häufig sinnvoll, kompliziertere Aussagen mit Hilfe logischer Zeichen zu formalisieren.

Sind A und B Aussagen, dann bedeutet

$A \wedge B$,	dass A und B gelten,
$A \vee B$,	dass A oder B gilt,
$\neg A$ (nicht A),	dass das Gegenteil von A gilt,
$A \Rightarrow B$,	dass B aus A folgt,
$A \Leftrightarrow B$,	dass sowohl $A \Rightarrow B$ als auch $B \Rightarrow A$ gelten.

Die logischen Zeichen bezeichnet man auch als Junktoren. Das Symbol $\vee$ ist das nicht ausschließende Oder (also nicht entweder ... oder).

Eine Aussage $A \Rightarrow B$ heißt eine Implikation, man sagt: A impliziert B. Man nennt A die Prämisse, B die Konklusion. Die Prämisse enthält die Voraussetzungen, unter denen die Aussage B gilt.

Gilt $A \Leftrightarrow B$, so sagt man, die beiden Aussagen A und B sind äquivalent oder gleichwertig.

Beispiele

1. Für eine natürliche Zahl $n \geq 1$ ist die Implikation „6 teilt $n \Rightarrow 2$ teilt n“ wahr. Die umgekehrte Implikation gilt nicht.
2. „6 teilt n“ und „2 teilt n und 3 teilt n“ sind zwei äquivalente Aussagen. ◀

1.2.3 Vollständige Induktion

Mathematische Aussagen müssen in der Regel bewiesen werden. Neben direkten Beweisen und Widerspruchsbeweisen ist eine häufig verwendete Beweismethode die vollständige Induktion. Dieses Beweisverfahren lässt sich bei Aussagen über natürliche Zahlen (vgl. Abschn. 1.3) anwenden.

Ein Beweis mit vollständiger Induktion, dass eine Aussage $A(n)$ (eine Eigenschaft oder eine Formel) für alle natürlichen Zahlen $n \geq m$ (also von m an) richtig ist, besteht aus drei Schritten:

1. Induktionsanfang (Induktionsverankerung):
   $A(n)$ ist richtig für $n = m$. Dies muss meistens auf direktem Weg nachgewiesen werden.
2. Induktionsannahme (Induktionsvoraussetzung):
   Die Aussage $A(n)$ ist für eine beliebige natürliche Zahl n_0 ($n_0 \geq m$) richtig, es gilt also $A(n_0)$.
3. Induktionsschluss (Induktionsschritt):
   Unter Benutzung der Induktionsannahme wird gezeigt, dass die Aussage $A(n)$ dann auch für $n_0 + 1$ richtig ist, das heißt, aus $A(n_0)$ folgt $A(n_0 + 1)$. Man nennt diesen Schritt auch Schluss von n_0 auf $n_0 + 1$.

Man beachte, dass sowohl der Induktionsanfang als auch der Induktionsschluss durchgeführt werden müssen. Der Induktionsschluss von n_0 auf $n_0 + 1$ macht den Nachweis der Gültigkeit für $n = m$ (Induktionsanfang) nicht überflüssig. Der Induktionsschluss kann gelingen, auch wenn die Aussage $A(n)$ für alle natürlichen Zahlen falsch ist.

Beispiel

Behauptung:

Für alle natürlichen Zahlen $n \geq 0$ gilt

$$0 + 1 + 2 + 3 + \ldots + n = \frac{n \cdot (n+1)}{2}$$

Beweis der Behauptung mit vollständiger Induktion

Induktionsanfang:

Für $n = 0$ ist die Behauptung richtig, denn auf beiden Seiten des Gleichheitszeichens steht Null: $0 = \frac{0 \cdot 1}{2} = 0$

Induktionsannahme:

Für eine beliebige natürliche Zahl n_0 ist die Behauptung richtig:

$$0 + 1 + 2 + 3 + \ldots + n_0 = \frac{n_0 \cdot (n_0+1)}{2}$$

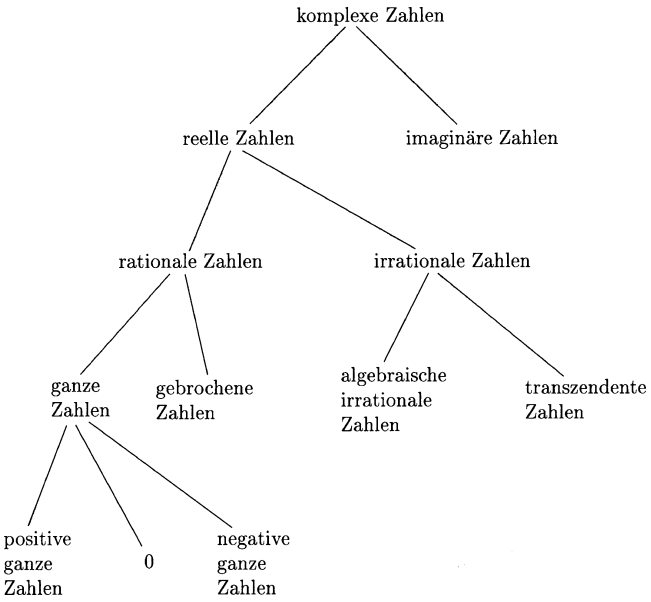
Induktionsschluss (von n_0 auf $n_0 + 1$):

Addiert man auf beiden Seiten der Induktionsannahme $n_0 + 1$, so folgt

$$\begin{aligned} &0 + 1 + 2 + 3 + \ldots + n_0 + (n_0 + 1) \\ &= \frac{n_0 \cdot (n_0+1)}{2} + (n_0 + 1) \\ &= (n_0 + 1) \cdot \left(\frac{n_0}{2} + 1 \right) \\ &= \frac{(n_0+1)(n_0+2)}{2} \\ &= \frac{(n_0+1)[(n_0+1)+1]}{2} \end{aligned}$$

Aus der Richtigkeit der Annahme für n_0 folgt somit auch die Richtigkeit für $n_0 + 1$. Damit gilt die Formel für alle $n \geq 0$. ◀

1.3 Einteilung der Zahlen



Einige der Zahlenbereiche werden häufig in Mengenschreibweise dargestellt:

$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
Menge der natürlichen Zahlen

$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
Menge der ganzen Zahlen

$\mathbb{Q} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0 \right\}$
Menge der rationalen Zahlen

$\mathbb{R}$ Menge der reellen Zahlen

$\mathbb{C} = \{z = a + bj \,|\, a, b \in \mathbb{R}, j = \sqrt{-1}\}$
Menge der komplexen Zahlen

Die natürlichen Zahlen sind die positiven ganzen Zahlen und die Null.

Eine Teilmenge der natürlichen Zahlen sind die Primzahlen. Eine Primzahl ist eine natürliche Zahl größer als 1, die nur durch 1 und durch sich selbst ohne Rest teilbar ist.

Die Primzahlen sind die Zahlen 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, ..., die Zahl 1 ist keine Primzahl. Es gibt unendlich viele Primzahlen, das heißt, es gibt keine größte Primzahl, zu jeder Primzahl gibt es noch größere. 2 ist die einzige gerade Primzahl. Alle Primzahlen zusammen bilden die Menge $\mathbb{P}$ der Primzahlen, die eine Teilmenge der Menge $\mathbb{N}$ der natürlichen Zahlen ist.

Jede natürliche Zahl $n \geq 2$ lässt sich in ein Produkt von Primzahlen zerlegen, die Zerlegung ist eindeutig bis auf die Reihenfolge der Faktoren (sogenannte Primfaktorzerlegung).

Beispiele zur Primfaktorzerlegung

$100 = 2 \cdot 2 \cdot 5 \cdot 5 = 2^2 \cdot 5^2$; $546 = 2 \cdot 3 \cdot 7 \cdot 13$. ◀

Die ganzen Zahlen setzen sich zusammen aus den natürlichen Zahlen und den negativen ganzen Zahlen.

Die rationalen Zahlen sind alle ganzen und gebrochenen Zahlen. Rationale Zahlen lassen sich als Brüche aus ganzen Zahlen darstellen. Jede rationale Zahl kann als endlicher oder unendlicher periodischer Dezimalbruch dargestellt werden.

Der Dezimalbruch einer rationalen Zahl ist die Darstellung der rationalen Zahl als Dezimalzahl, also als Zahl „mit Stellen hinter dem Komma“ (siehe auch Abschn. 1.7). Bei einem endlichen Dezimalbruch ist die Anzahl der Stellen nach dem Komma endlich, bei einem periodischen Dezimalbruch wiederholen sich die Stellen nach dem Komma nach einem gewissen Muster (Periode).

Die reellen Zahlen sind alle Zahlen, die auf der reellen Achse der Zahlenebene (Gauß'sche Zahlenebene), der sogenannten Zahlengeraden, darstellbar sind.

Die reellen Zahlen setzen sich zusammen aus den rationalen Zahlen und den irrationalen Zahlen. Der Dezimalbruch einer irrationalen Zahl hat unendlich viele Stellen und keine Periode.

Man unterteilt die irrationalen Zahlen in algebraische irrationale Zahlen und transzendente Zahlen.

Eine algebraische irrationale Zahl ist eine irrationale Zahl, die Lösung (Wurzel) einer algebraischen Gleichung (Bestimmungsgleichung)

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0 = 0$$

mit rationalen Zahlen als Koeffizienten $a_{n-1}, a_{n-2}, \ldots, a_1, a_0$ ist, wobei n für eine natürliche Zahl größer Null steht.

Es gibt keine reelle Zahl, die Lösung der Gleichung $x^2 + 1 = 0$ ist. Deshalb werden die reellen Zahlen zu den komplexen Zahlen erweitert.

Komplexe Zahlen sind Zahlen der Form $z = a + bj$, wobei a und b reelle Zahlen sind und j die imaginäre Einheit, $j^2 = -1$ (j ist eine Lösung der algebraischen Gleichung $x^2 + 1 = 0$).

Eine komplexe Zahl z besteht also aus einem reellen Teil a (Realteil) und einem imaginären Teil b (Imaginärteil). Komplexe Zahlen z mit Realteil gleich 0 (also $a = 0$) heißen imaginäre Zahlen, die komplexen Zahlen z mit Imaginärteil gleich 0 (also $b = 0$) sind die reellen Zahlen.

Komplexe Zahlen lassen sich in der Zahlenebene darstellen.

Beispiele für ganze Zahlen

38; −700.632; 0; 105. ◀

Beispiele für rationale Zahlen

-2; $\frac{3}{2} = 1{,}5$; $\frac{4}{3} = 1{,}3333\ldots = 1{,}\overline{3}$; $-\frac{1}{8} = -0{,}125$; $-\frac{16}{11} = -1{,}454545\ldots = -1{,}\overline{45}$ (der periodische Teil wird überstrichen). ◀

Beispiele für reelle Zahlen

-4; $\frac{3}{4}$; $4-\pi$; e^3; $\sqrt{3}$; $\sin 5^0$. ◀

Beispiele für irrationale Zahlen

$\sqrt{3} = 1{,}732050808\ldots$; $\sqrt[3]{4} = 1{,}587401052\ldots$; $5-2\sqrt{3} = 1{,}535898385\ldots$; $-\pi = -3{,}141592654\ldots$; $e = 2{,}718281828\ldots$ ◀

Beispiele für algebraische irrationale Zahlen

$\sqrt{3}$ (denn $\sqrt{3}$ ist Lösung der Gleichung $x^2 - 3 = 0$); $\sqrt[3]{4}$ (denn $\sqrt[3]{4}$ ist Lösung der Gleichung $x^3 - 4 = 0$); $5 - 2\sqrt{3}$ (denn $5 - 2\sqrt{3}$ ist Lösung der Gleichung $x^2 - 10x + 13 = 0$. ◀

Beispiele für transzendente Zahlen

$-\pi$; e ◀

Beispiele für komplexe Zahlen

$3+\sqrt{2}j$; $-1+5j$; $e+\pi^2 j$; $-4j$ (imaginäre Zahl); $3\sqrt{2}$ (reelle Zahl). ◀

Ein hochgestellter Stern bedeutet die entsprechende Menge ohne die Null:

$\mathbb{N}^* = \{1, 2, 3, \ldots\}$:
Menge der natürlichen Zahlen ohne die Null

$\mathbb{Z}^* = \{\ldots, -3, -2, -1, 1, 2, 3, \ldots\} = \{x \,|\, x \in \mathbb{Z}, x \neq 0\}$:
Menge der ganzen Zahlen ohne die Null

$\mathbb{Q}^* = \left\{\frac{m}{n} \;\middle|\; m, n \in \mathbb{Z}^*\right\} = \{x \,|\, x \in \mathbb{Q}, x \neq 0\}$:
Menge der rationalen Zahlen ohne die Null

$\mathbb{R}^* = \{x \,|\, x \in \mathbb{R}, x \neq 0\}$:
Menge der reellen Zahlen ohne die Null

Ein hochgestelltes Plus bedeutet die Menge der entsprechenden positiven Zahlen:

$\mathbb{Z}^+ = \mathbb{N}^* = \{1, 2, 3, \ldots\} = \{x \,|\, x \in \mathbb{Z}, x > 0\}$:
Menge der positiven ganzen Zahlen

$\mathbb{Q}^+ = \left\{\frac{m}{n} \;\middle|\; m, n \in \mathbb{N}^*\right\} = \{x \,|\, x \in \mathbb{Q}, x > 0\}$:
Menge der positiven rationalen Zahlen

$\mathbb{R}^+ = \{x \,|\, x \in \mathbb{R}, x > 0\}$:
Menge der positiven reellen Zahlen

1.4 Grundrechenarten

Die vier Grundrechenarten sind die Addition, die Subtraktion, die Multiplikation und die Division.

Addition:	Summand plus Summand gleich Summe
Subtraktion:	Minuend minus Subtrahend gleich Differenz
Multiplikation:	Faktor mal Faktor gleich Produkt
Division:	Dividend geteilt durch Divisor gleich Quotient

Beispiele

$4+5 = 9$ (Addition); $7-2 = 5$ (Subtraktion); $3 \cdot 8 = 24$ (Multiplikation); $87 : 3 = 29$ (Division). ◀

Vereinbarung:

Der Multiplikationspunkt (Malpunkt) kann weggelassen werden zwischen zwei Variablen, zwischen einer Zahl und einer Variablen, zwischen einer Zahl und einer Klammer, zwischen einer Variablen und einer Klammer sowie zwischen zwei Klammern.

Beispiele

$a \cdot b = ab$, $4 \cdot a = 4a$, $2 \cdot (a+b) = 2(a+b)$, $a \cdot (3+a) = a(3+a)$, $(a+b) \cdot (c-d) = (a+b)(c-d)$. ◀

Achtung: Der Multiplikationspunkt zwischen zwei Zahlen darf nicht weggelassen werden.

Beispiel

$3 \cdot 4 \neq 34$. ◀

Eine Variable oder Veränderliche oder Platzhalter ist eine Größe, die in der Regel verschiedene Werte annehmen kann. Variable werden durch Symbole dargestellt (meist lateinische Buchstaben). Variable sind zum Beispiel die Platzhalter für die gesuchten Lösungen von einer oder mehreren gegebenen Gleichungen.

1.5 Grundlegende Rechenregeln

1.5.1 Buchstabenrechnen

Buchstabenrechnen ist das Rechnen mit unbestimmten Zahlen. Formuliert man eine mathematische Aussage, die nicht nur für eine bestimmte Zahl, sondern für einen ganzen Zahlbereich oder sogar für alle Zahlen gilt, dann benutzt man statt einer Zahl einen Buchstaben. Der Buchstabe heißt unbestimmte Zahl.

Beispiele

1. $(a+b)^2 = a^2+2ab+b^2$ (binomische Formel, sie gilt für alle reellen Zahlen a, b)
2. $(a{\cdot}b){\cdot}c = a{\cdot}(b{\cdot}c) = a{\cdot}b{\cdot}c$ (Assoziativgesetz bezüglich der Multiplikation, gilt für alle reellen Zahlen a, b, c)
3. $\sqrt[n]{a\cdot b} = \sqrt[n]{a}\cdot\sqrt[n]{b}$ (gilt für alle positiven reellen Zahlen a, b und alle natürlichen Zahlen $n \geq 2$). ◀

1.5.2 Kehrwert, Quersumme

Der Kehrwert einer Zahl $a \neq 0$ ist die Zahl $\frac{1}{a}$. Man sagt statt Kehrwert auch reziproker Wert.

So ist zum Beispiel der Kehrwert von 3 gleich $\frac{1}{3}$, der Kehrwert von -17 ist $-\frac{1}{17}$, der Kehrwert von $\frac{1}{4}$ ist 4.

Die Quersumme einer ganzen Zahl ist die Summe ihrer Ziffern.

So ist zum Beispiel die Quersumme der Zahl 239.503.618 gleich $2+3+9+5+0+3+6+1+8 = 37$, die Quersumme von 3.972.611.028 ist $3+9+7+2+6+1+1+0+2+8 = 39$, und die Quersumme der Zahl 209.334.042 ist $2+0+9+3+3+4+0+4+2 = 27$.

1.5.3 Teilbarkeitsregeln

Die einzelnen Zeichen einer Zahl sind ihre Ziffern. Aus Eigenschaften der Ziffern lassen sich Teilbarkeitseigenschaften der Zahlen ableiten.

Eine ganze Zahl ist teilbar durch

2, wenn die letzte Ziffer durch 2 teilbar ist
3, wenn die Quersumme der Zahl (also die Summe der Ziffern) durch 3 teilbar ist
4, wenn die Zahl aus den letzten beiden Ziffern durch 4 teilbar ist
5, wenn die letzte Ziffer durch 5 teilbar ist (also 0 oder 5 ist)
6, wenn die letzte Ziffer durch 2 und die Quersumme der Zahl durch 3 teilbar ist
8, wenn die Zahl aus den letzten drei Ziffern durch 8 teilbar ist
9, wenn die Quersumme der Zahl durch 9 teilbar ist
11, wenn die alternierende Quersumme der Zahl (also die Summe der Ziffern, die abwechselnd positives und negatives Vorzeichen erhalten) durch 11 teilbar ist.

Beispiele

1. 2486 ist teilbar durch 2, denn 6 ist teilbar durch 2
2. 263.451 ist teilbar durch 3, denn die Quersumme $2+6+3+4+5+1 = 21$ ist teilbar durch 3
3. 2.563.488 ist teilbar durch 4, denn 88 ist teilbar durch 4
4. 823.620 ist teilbar durch 5, denn 0 ist teilbar durch 5
5. 2.598.018 ist teilbar durch 6, denn 8 ist teilbar durch 2 und die Quersumme $2+5+9+8+0+1+8 = 33$ ist teilbar durch 3
6. 524.299.168 ist teilbar durch 8, denn 168 ist teilbar durch 8
7. 11.929.545 ist teilbar durch 9, denn die Quersumme $1+1+9+2+9+5+4+5 = 36$ ist teilbar durch 9
8. 14.739.296 ist teilbar durch 11, denn die alternierende Quersumme $1-4+7-3+9-2+9-6 = 11$ ist teilbar durch 11. ◀

1.5.4 Punktrechnung vor Strichrechnung

Die Rechenzeichen $\cdot$ und : binden stärker als $+$ und $-$, das heißt, Multiplikation und Division müssen vor Addition und Subtraktion ausgeführt werden.

$$a + b \cdot c = a + (b \cdot c)$$
$$a - b : c = a - (b : c)$$

Beispiele

$a : b{\cdot}c{-}d = (a : b){\cdot}c{-}d$; $3{+}4{\cdot}5{-}6 = 3{+}(4{\cdot}5){-}6 = 17$; $8{-}5{\cdot}4{\cdot}3{+}36 : 9{+}6{\cdot}(12{+}2{\cdot}7) = 8{-}60{+}4{+}6{\cdot}26 = 108$. ◀

Die Klammern geben an, welcher Teil der Rechnung zuerst ausgeführt wird.

1.5.5 Potenzrechnung vor Punktrechnung

Potenzieren bindet stärker als Multiplizieren und Dividieren.

$$a \cdot b^2 = a \cdot (b^2)$$

Es gilt also $ab^2 \neq (ab)^2$.

Beispiele

$a : b^2 - 3 \cdot 2^3 = \frac{a}{b^2} - 3(2^3) = \frac{a}{b^2} - 24$; $4 \cdot 5^3 - 7 \cdot 4^2 = 4 \cdot 125 - 7 \cdot 16 = 388$. ◀

1.5.6 Grundgesetze der Addition und Multiplikation

1. *Kommutativgesetz (Vertauschungsgesetz)*

Für reelle Zahlen gilt bezüglich der Addition und bezüglich der Multiplikation das Kommutativgesetz:

$$a + b = b + a \quad a \cdot b = b \cdot a$$

Bei der Addition kann man also die Summanden vertauschen, bei der Multiplikation kann man die Faktoren vertauschen.

Beispiele

$3 + 4 = 4 + 3; 3 \cdot 4 = 4 \cdot 3.$ ◀

2. *Assoziativgesetz (Verknüpfungsgesetz)*

Für reelle Zahlen gilt bezüglich der Addition und bezüglich der Multiplikation das Assoziativgesetz:

$$(a + b) + c = a + (b + c) = a + b + c$$
$$(a \cdot b) \cdot c = a \cdot (b \cdot c) = a \cdot b \cdot c$$

Bei der Addition kann man also Summanden beliebig verknüpfen (zusammenfassen), bei der Multiplikation kann man Faktoren beliebig verknüpfen.

Beispiele

$(3 + 4) + 7 = 3 + (4 + 7) = 3 + 4 + 7; (3 \cdot 4) \cdot 7 = 3 \cdot (4 \cdot 7) = 3 \cdot 4 \cdot 7.$ ◀

3. *Distributivgesetze (Zerlegungsgesetze)*

Für reelle Zahlen gelten die Distributivgesetze:

$$(a + b) \cdot c = a \cdot c + b \cdot c$$
$$a \cdot (b + c) = a \cdot b + a \cdot c$$

Beispiele

$(3 + 4) \cdot 7 = 3 \cdot 7 + 4 \cdot 7; 3 \cdot (4 + 7) = 3 \cdot 4 + 3 \cdot 7.$ ◀

Aus diesen Grundgesetzen ergeben sich die wichtigen Regeln der Klammerrechnung.

1.5.7 Indizes, Summenzeichen, Produktzeichen

Ein Index (Plural Indizes) ist ein Zeichen, das an Symbole für Variable, Funktionen oder Operationen angebracht wird.

Bezeichnet man zum Beispiel Variable mit x, dann kennzeichnet man verschiedene Variable dadurch, dass man an das x verschiedene tiefgestellte Indizes anhängt: $x_1, x_2, x_3, \ldots$ Ein Index ist meistens eine Zahl.

Das Summenzeichen $\sum$ (entstanden aus dem griechischen Buchstaben für S) dient zur vereinfachten Darstellung von Summen (gesprochen: Summe über a_k von $k = 1$ bis $k = n$).

$$\sum_{k=1}^{n} a_k = a_1 = a_2 + a_3 + \ldots + a_n$$

Man erhält alle Summanden der Summe, wenn man in a_k für den Index k zunächst 1, dann 2 usw. und schließlich n setzt. Dieser Buchstabe k heißt Summationsindex und kann durch einen beliebigen anderen Buchstaben ersetzt werden. Es gilt also zum Beispiel

$$\sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j.$$

Beispiele

1. $\sum_{k=1}^{6} k^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2$
2. $\sum_{i=1}^{3} \log(2i) = \log 2 + \log 4 + \log 6$
3. $\sum_{j=1}^{5} 6 = 6 + 6 + 6 + 6 + 6 = 5 \cdot 6 = 30.$ ◀

Das Produktzeichen $\prod$ dient zur vereinfachten Darstellung von Produkten (gesprochen: Produkt über a_k von $k = 1$ bis $k = n$).

$$\prod_{k=1}^{n} a_k = a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n$$

Man erhält alle Faktoren des Produkts, wenn man in a_k für den Index k zunächst 1, dann 2 usw. und schließlich n setzt. Der Index k kann durch einen beliebigen anderen Buchstaben ersetzt werden. Zum Beispiel gilt

$$\prod_{k=1}^{n} a_k = \prod_{i=1}^{n} a_i = \prod_{j=1}^{n} a_j.$$

Beispiele

1. $\prod_{k=1}^{7} k^2 = 1^2 \cdot 2^2 \cdot 3^2 \cdot 4^2 \cdot 5^2 \cdot 6^2 \cdot 7^2$
2. $\prod_{i=2}^{4} 3^i = 3^2 \cdot 3^3 \cdot 3^4 = 3^{2+3+4} = 3^9$
3. $\prod_{j=1}^{5} 4 = 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 4^5$ ◀

1.5.8 Binomische Formeln

Ein Binom ist ein zweigliedriger Ausdruck der Form $a + b$ oder $a - b$. Die Multiplikation von Binomen führt zu den binomischen Formeln (zwei Faktoren) und zum binomischen Lehrsatz (n Faktoren, $n \geq 1$ beliebig).

Die folgenden Rechenregeln heißen binomische Formeln oder binomische Gleichungen 2. Grades (a und b sind beliebige reelle Zahlen).

$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a-b)^2 = a^2 - 2ab + b^2$$
$$(a+b)(a-b) = a^2 - b^2$$

Beispiele

1. $21^2 = (20+1)^2 = 20^2 + 2 \cdot 20 \cdot 1 + 1^2 = 441$
2. $(2c+3)^2 = (2c)^2 + 2 \cdot 2c \cdot 3 + 3^2 = 4c^2 + 12c + 9$
3. $19^2 = (20-1)^2 = 20^2 - 2 \cdot 20 \cdot 1 + 1^2 = 361$
4. $(2x-5y)^2 = (2x)^2 - 2 \cdot 2x \cdot 5y + (5y)^2 = 4x^2 - 20xy + 25y^2$
5. $21 \cdot 19 = (20+1)(20-1) = 20^2 - 1^2 = 399$
6. $(3x+4y)(3x-4y) = (3x)^2 - (4y)^2 = 9x^2 - 16y^2$. ◀

1.5.9 Fakultäten, Binomialkoeffizienten und Pascal'sches Dreieck

Für eine natürliche Zahl $n \in \mathbb{N}^*$ ist $n!$ (gesprochen: n Fakultät) definiert als das Produkt der ersten n von Null verschiedenen natürlichen Zahlen.

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$$

Außerdem wird $0! = 1$ für $n = 0$ gesetzt.

Die Fakultät lässt sich mit der Formel $n! = n \cdot (n-1)!$, $n \in \mathbb{N}^*$ berechnen (Rekursionsformel).

Beispiele

$0! = 1$, $1! = 1$, $2! = 1 \cdot 2 = 2$,
$3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$,
$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$,
$6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$. ◀

Für alle natürlichen Zahlen n, k mit $1 \leq k \leq n$ ist der Binomialkoeffizient $\binom{n}{k}$ (gesprochen: n über k) definiert durch

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Man setzt außerdem $\binom{n}{0} = 1$ und $\binom{0}{0} = 1$.

Für die praktische Rechnung eignet sich die Formel

$$\binom{n}{k} = \frac{n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot k}$$

Zähler und Nenner bestehen jeweils aus k Faktoren. Dabei sind die k Faktoren im Nenner von 1 an um jeweils 1 aufsteigend und im Zähler von n an um jeweils 1 fallend.

Für die Binomialkoeffizienten gibt es eine Reihe von Rechenregeln. Für die praktische Berechnung sind wichtig:

$$\binom{n}{k} = \binom{n}{n-k} \quad \text{Symmetrie}$$
$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k} \quad \text{Additionssatz}$$

Beweis der Symmetrie:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
$$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!}$$
$$= \frac{n!}{(n-k)!k!}$$
$$= \frac{n!}{k!(n-k)!}$$

Beispiele

$\binom{8}{3} = \frac{8!}{3!5!} = \frac{8\cdot7\cdot6}{1\cdot2\cdot3} = 56$; $\binom{8}{3} = \binom{8}{5}$; $\binom{n}{n} = 1$; $\binom{n}{1} = n$; $\binom{n}{2} = \frac{n\cdot(n-1)}{2}$. ◄

Für jede natürliche Zahl n lassen sich die Binomialkoeffizienten $\binom{n}{k}$, $k = 0, 1, \ldots, n$ im Pascal'schen Dreieck darstellen. Wegen $\binom{n}{0} = \binom{n}{n} = 1$ sind die Zahlen der Ränder des Dreiecks gleich Eins. Jede nicht am Rand stehende Zahl des Dreiecks ist wegen des Additionssatzes gleich der Summe der beiden unmittelbar darüberstehenden Zahlen.

Pascal'sches Dreieck:

$$\begin{array}{ccccccccccc}
 & & & & & \binom{0}{0} & & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} \\
\binom{6}{0} & & \binom{6}{1} & & \binom{6}{2} & & \binom{6}{3} & & \binom{6}{4} & & \binom{6}{5} \quad \binom{6}{6} \\
 & & & & \cdots & & & & & &
\end{array}$$

Rechnet man die Binomialkoeffizienten aus, so lauten die ersten Zeilen des Pascal'schen Dreiecks:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 \quad 1 \\
 & & & & \cdots & & & & & &
\end{array}$$

Am Pascal'schen Dreieck erkennt man die Symmetrie der Binomialkoeffizienten.

Oftmals ist es einfacher, die Binomialkoeffizienten nicht direkt, sondern mit Hilfe des Pascal'schen Dreiecks zu berechnen.

1.5.10 Binomischer Lehrsatz

Für beliebige reelle Zahlen $a, b \neq 0$ und jede natürliche Zahl $n \geq 1$ gilt der binomische Lehrsatz (binomische Gleichung n-ten Grades)

$$\begin{aligned}
(a+b)^n &= a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 \\
&\quad + \binom{n}{3}a^{n-3}b^3 + \ldots + \binom{n}{k}a^{n-k}b^k \\
&\quad + \ldots + \binom{n}{n-1}ab^{n-1} + b^n \\
&= \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k
\end{aligned}$$

Die untere Form ist die abgekürzte Summenschreibweise (gesprochen: Summe über $\binom{n}{k}a^{n-k}b^k$ von $k = 0$ bis $k = n$). Man erhält also alle Summanden der Summe, wenn man für den Index k zunächst 0, dann 1, dann 2 usw. und schließlich n setzt. Die so erhaltenen Glieder werden dann addiert.

Man beachte, dass $\binom{n}{0} = \binom{n}{n} = 1$, $a^0 = b^0 = 1$ und $a^1 = a$, $b^1 = b$ gilt.

Für $n = 2$ ergeben sich die beiden ersten binomischen Formeln.

Spezialfälle:

$$\begin{aligned}
a = b = 1: \quad & \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} \\
&= \sum_{k=0}^{n}\binom{n}{k} = (1+1)^n = 2^n
\end{aligned}$$

Die Summe der Binomialkoeffizienten der n-ten Zeile des Pascal'schen Dreiecks ist also 2^n.

$$\begin{aligned}
a = 1, b = -1: \quad & \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - + \ldots + (-1)^n\binom{n}{n} \\
&= \sum_{k=0}^{n}(-1)^k\binom{n}{k} = (1-1)^n = 0
\end{aligned}$$

Die alternierende Summe (das heißt, abwechselnde Vorzeichen $+$ und $-$) der Binomialkoeffizienten der n-ten Zeile des Pascal'schen Dreiecks ist also 0.

Beispiele

1. $$(a+b)^3 = \binom{3}{0}a^3 + \binom{3}{1}a^2b + \binom{3}{2}ab^2 + \binom{3}{3}b^3$$
$$= a^3 + 3a^2b + 3ab^2 + b^3$$

2. $$(a-b)^6 = \binom{6}{0}a^6 - \binom{6}{1}a^5b + \binom{6}{2}a^4b^2 - \binom{6}{3}a^3b^3$$
$$+ \binom{6}{4}a^2b^4 - \binom{6}{5}ab^5 + \binom{6}{6}b^6$$
$$= a^6 - 6a^5b + 15a^4b^2 20a^3b^3$$
$$+ 15a^2b^4 - 6ab^5 + b^6$$

   Man vergleiche die Koeffizienten mit der letzten hingeschriebenen Zeile des Pascal'schen Dreiecks!

3. $$(x-3)^5 = \binom{5}{0}x^5 + \binom{5}{1}x^4(-3) + \binom{5}{2}x^3(-3)^2$$
$$+ \binom{5}{3}x^2(-3)^3 + \binom{5}{4}x(-3)^4 + \binom{5}{5}(-3)^5$$
$$= x^5 - 15x^4 + 90x^3 - 270x^2 + 405x - 243 \quad \blacktriangleleft$$

1.6 Potenz- und Wurzelrechnung

1.6.1 Definition der Potenz

Eine Zahl der Form a^x (gesprochen: a hoch x) heißt Potenz. Dabei ist a die Basis oder die Grundzahl und x der Exponent oder die Hochzahl der Potenz.

a^x	Potenz
a	Basis (Grundzahl)
x	Exponent (Hochzahl)

Ist die Basis a eine beliebige reelle Zahl und der Exponent x eine natürliche Zahl aus $\mathbb{N}^*$, dann steht a^x für die Vorschrift, die Basis a insgesamt x-mal mit sich selbst zu multiplizieren.

$$a^x = a \cdot a \cdot a \cdot \ldots \cdot a \quad (x \text{ Faktoren}, x \in \mathbb{N}^*)$$

Man spricht vom Potenzieren für diese algebraische Operation.

Ist der Exponent x eine natürliche Zahl $\neq 0$, dann kann die Basis a eine beliebige reelle Zahl sein.

Ist der Exponent x eine beliebige reelle Zahl, dann ist die Potenz nur für positive Basen a definiert ($a \in \mathbb{R}^+$). Dabei sind Potenzen mit irrationalen Exponenten mit Hilfe eines Grenzübergangs erklärt.

$$a^{\frac{p}{q}} = c \quad \Leftrightarrow \quad a^p = c^q$$

Für negative Exponenten gilt

$$a^{-m} = \frac{1}{a^m}$$

Für die Exponenten 1 und 0 gilt

$$a^1 = a \quad a^0 = 1 \quad (a \neq 0)$$

Eine spezielle Potenz für $k \in \mathbb{Z}$ ist

$$(-1)^k = \begin{cases} +1 & \text{falls } k \text{ gerade} \\ -1 & \text{falls } k \text{ ungerade} \end{cases}$$

Beispiele

1. $2^{\frac{3}{2}} = 2{,}8284\ldots$ Potenz mit Basis 2 und Exponent $\frac{3}{2} = 1{,}5$
2. $3^5 = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243$
3. $(-4)^3 = (-4) \cdot (-4) \cdot (-4) = -64$
4. $3^{-5} = \frac{1}{3^5} = \frac{1}{243}$
5. $6^0 = 1$
6. $(-\frac{1}{6})^0 = 1$
7. $8^{\frac{4}{3}} = c \Leftrightarrow 8^4 = c^3 \Leftrightarrow c = 16$
8. $8^{-\frac{4}{3}} = c \Leftrightarrow 8^{-4} = c^3 \Leftrightarrow c = \frac{1}{16}$
9. $4^1 = 4, (-12)^1 = -12$
10. $(-1)^5 = -1, (-1)^4 = 1, (-1)^0 = 1,$
$(-1)^{-11} = -1.$ $\blacktriangleleft$

1.6.2 Regeln der Potenzrechnung

1. *Potenzrechnung vor Punktrechnung*

$$ba^n = b \cdot (a^n)$$

Soll erst Punktrechnung erfolgen, dann muss ein Klammerpaar gesetzt werden.

Beispiele

$5 \cdot 3^4 = 5 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 5 \cdot (3^4)$;
$(5 \cdot 3)^4 = 5 \cdot 3 \cdot 5 \cdot 3 \cdot 5 \cdot 3 \cdot 5 \cdot 3 = 15^4$. ◀

Fehlerwarnung: $ba^n \neq (ba)^n$ für $a, b \neq 0, b, n \neq 1$.

2. *Addieren und Subtrahieren*

Potenzen kann man nur addieren oder subtrahieren, wenn sie in Basis *und* Exponent übereinstimmen.

$$pa^n + qa^n = (p+q)a^n$$

Beispiel

$2 \cdot 3^4 + 5 \cdot 3^4 = (2+5) \cdot 3^4 = 7 \cdot 3^4$. ◀

Fehlerwarnungen: $2^4 + 3^4 \neq 5^4$, $3^2 + 3^4 \neq 3^6$.

3. *Multiplizieren und Dividieren bei gleicher Basis*

Potenzen mit gleicher Basis werden multipliziert, indem man die Exponenten addiert (genauer: indem man die gemeinsame Basis mit der Summe der Exponenten potenziert).

$$a^n \cdot a^m = a^{n+m}$$

Potenzen mit gleicher Basis werden dividiert, indem man die Exponenten subtrahiert (genauer: indem man die gemeinsame Basis mit der Differenz der Exponenten potenziert).

$$\frac{a^n}{a^m} = a^{n-m}$$

Der Wert einer Potenz bleibt erhalten, wenn man gleichzeitig die Basis durch ihren Kehrwert und das Vorzeichen des Exponenten durch das entgegengesetzte ersetzt.

$$\left(\frac{a}{b}\right)^{-n} = \left(\frac{b}{a}\right)^n$$

Fehlerwarnungen: $3^6 \cdot 3^4 \neq 3^{6 \cdot 4} = 3^{24}$, $3^{12} : 3^4 \neq 3^{12:4} = 3^3$, $a^{-2} \neq -a^2$.

Beispiele

1. $3^6 \cdot 3^4 = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^{10} = 3^{6+4}$
2. $\frac{3^6}{3^4} = \frac{3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3}{3 \cdot 3 \cdot 3 \cdot 3} = 3 \cdot 3 = 3^2 = 3^{6-4}$
3. $3^5 : 3^4 = 3^{5-4} = 3^1 = 3$
4. $3^4 : 3^4 = 3^{4-4} = 3^0 = 1$
5. $\left(\frac{3}{2}\right)^4 : \left(\frac{3}{2}\right)^6 = \left(\frac{3}{2}\right)^{4-6} = \left(\frac{3}{2}\right)^{-2} = \left(\frac{2}{3}\right)^2$. ◀

4. *Multiplizieren und Dividieren bei gleichem Exponenten*

Potenzen mit gleichem Exponenten werden multipliziert, indem man die Basen multipliziert (genauer: indem man das Produkt der Basen mit dem gemeinsamen Exponenten potenziert).

$$a^n \cdot b^n = (a \cdot b)^n$$

Fehlerwarnung: $a^n \cdot b^n \neq (ab)^{2n}$ (außer in Sonderfällen wie $n = 0$), zum Beispiel: $2^3 \cdot 5^3 \neq 10^6$.

Potenzen mit gleichem Exponenten werden dividiert, indem man die Basen dividiert (genauer: indem man den Quotienten der Basen mit dem gemeinsamen Exponenten potenziert).

$$\frac{a^n}{b^n} = \left(\frac{a}{b}\right)^n$$

Fehlerwarnung: $\frac{a^n}{b^n} \neq \frac{a}{b}$ (außer in Sonderfällen) im Gegensatz zu $\frac{a \cdot n}{b \cdot n} = \frac{a}{b}$.

Umkehrungen

Ein Produkt wird potenziert, indem man die einzelnen Faktoren potenziert (genauer: indem man die Potenzen der einzelnen Faktoren miteinander multipliziert).

$$(a \cdot b)^n = a^n \cdot b^n$$

Ein Bruch wird potenziert, indem man Zähler und Nenner einzeln potenziert (genauer: indem man die Potenzen von Zähler und Nenner durcheinander dividiert).

$$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$$

Fehlerwarnung: Die Berechnung von $(a+b)^n$ darf nicht verwechselt werden mit der von $(a \cdot b)^n$. Es gilt $(a \cdot b)^n = a^n \cdot b^n$, aber $(a+b)^n \neq a^n + b^n$ (außer in Sonderfällen).

Beispiele

1. $2^4 \cdot 5^4 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 5 \cdot 5 \cdot 5 \cdot 5 = (2 \cdot 5) \cdot (2 \cdot 5) \cdot (2 \cdot 5) \cdot (2 \cdot 5)$
   $= (2 \cdot 5)^4 = 10^4 = 10.000$
2. $\frac{3^4}{5^4} = \left(\frac{3}{5}\right)^4 = 0{,}6^4 = 0{,}1296$
3. $30^5 = (3 \cdot 10)^5 = 3^5 \cdot 10^5 = 243 \cdot 100.000$
   $= 24.300.000 = 2{,}43 \cdot 10^7$
4. $0{,}3^5 = \left(\frac{3}{10}\right)^5 = \frac{3^5}{10^5} = \frac{243}{100000} = 0{,}00243$
   $= 2{,}43 \cdot 10^{-3}$. ◀

5. *Potenzieren einer Potenz*

Eine Potenz wird potenziert, indem man die Exponenten multipliziert (genauer: indem man die Basis mit dem Produkt der Exponenten potenziert).

$$(a^n)^m = a^{n \cdot m}$$

Bei der Potenz einer Potenz kann man die Exponenten miteinander vertauschen.

$$(a^n)^m = (a^m)^n$$

Fehlerwarnung: $(2^3)^4 \neq 2^7$.

Beispiele

1. $(2^3)^4 = (2 \cdot 2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) = 2^{3 \cdot 4} = 2^{12}$
2. $(3^4)^2 = 81^2 = 6561 = 9^4 = (3^2)^4$. ◀

Beispiele zur gesamten Potenzrechnung

1. $ar^p + bs^p - cr^p - ds^p = (a - c)r^p + (b + d)s^p$
2. $x^{3m-1}x^{m+1} = x^{3m-1+m+1} = x^{4m}$
3. $\frac{8^2 \cdot 5^2}{8^2 + 6^2} = \frac{(4 \cdot 2)^2 \cdot 5^2}{64 + 36} = \frac{4^2 \cdot (2 \cdot 5)^2}{100} = 4^2 = 16$
4. $\frac{5a^{x+y}b^{3u+v}}{7c^2} : \frac{5c^4}{28a^{y-x}b^{v-2u}} = \frac{5 \cdot 28 \cdot a^{x+y}a^{y-x}b^{3u+v}b^{v-2u}}{7 \cdot 5 \cdot c^2c^6}$
   $= \frac{4a^{2y}b^{u+2v}}{c^6}$ $(c \neq 0)$
5. $(-u^2)^3 = (-1)^3u^{2 \cdot 3} = -u^6$,
   $(-u^3)^2 = (-1)^2u^{3 \cdot 2} = u^6$,
   $((-u)^2)^3 = ((-1)^2u^2)^3 = (u^2)^3 = u^{2 \cdot 3} = u^6$
6. $\left(\frac{5a^{-1}}{-2^2b^{-3}}\right)^{-4} = \begin{cases} \left(-\frac{5b^3}{2^2a}\right)^{-4} = \left(-\frac{4a}{5b^3}\right)^4 = \frac{256a^4}{625b^{12}} \\ \frac{5^{-4}a^4}{(-1)^{-4}2^{-8}b^{12}} = \frac{2^8a^4}{5^4b^{12}} = \frac{256a^4}{625b^{12}} \end{cases}$
   $(a, b \neq 0)$
7. $\left(\frac{2}{3}x^{-1} - 3x\right)\left(3x^{-1} - \frac{2}{3}x\right) = 2x^{-2} - \frac{4}{9} - 9 + 2x^2$
   $= \frac{2}{x^2} - 9\frac{4}{9} + 2x^2$ $(x \neq 0)$. ◀

1.6.3 Definition der Wurzel

Eine Zahl der Form $\sqrt[n]{a} = a^{\frac{1}{n}}$ (gesprochen: n-te Wurzel aus a) heißt Wurzel. Dabei heißt a Radikand der Wurzel und ist eine reelle Zahl größer als 0, und n heißt Wurzelexponent und ist eine natürliche Zahl größer als 1.

$\sqrt[n]{a}$	Wurzel
$a > 0$	Radikand
$n > 1$	Wurzelexponent

Die Wurzel $\sqrt[n]{a}$ ist definiert als die eindeutig bestimmte Zahl $x \geq 0$ mit $x^n = a$. Die n-te Wurzel aus $a \geq 0$ ist also die nichtnegative reelle Zahl, deren n-te Potenz gleich a ist.

$$\sqrt[n]{a} = x \quad \Leftrightarrow \quad x^n = a$$

Man spricht vom Radizieren oder Wurzelziehen für diese algebraische Operation.

Ist der Wurzelexponent gleich 2, so heißt $\sqrt[2]{a} = \sqrt{a}$ (der Wurzelexponent 2 braucht nicht geschrieben zu werden) Quadratwurzel (oder einfach Wurzel) aus a; $\sqrt[3]{a}$ heißt Kubikwurzel aus a.

Bemerkungen:

1. Nach der Definition der Wurzel ist die Wurzel aus einer positiven Zahl wieder eine positive Zahl: $\sqrt{a^2} = |a|$ für jede reelle Zahl a (zum Absolutbetrag $|a|$ einer Zahl a vgl. Abschn. 1.9.2. Es gilt daher zum Beispiel nur $\sqrt{4} = 2$, nicht auch $\sqrt{4} = -2$. Dagegen hat die Gleichung $x^2 = 4$ die Lösungen $x_1 = +\sqrt{4} = +2$ und $x_2 = -\sqrt{4} = -2$.
2. Für ungerade n (zum Beispiel $n = 3$) kann die n-te Wurzel auch für negative Zahlen a eindeutig definiert werden, denn die n-te Potenz einer negativen Zahl ist selbst negativ. Zum Beispiel gilt $\sqrt[3]{-27} = -3$.
3. Wegen $\sqrt[n]{a} = a^{\frac{1}{n}}$ ergeben sich die Regeln der Wurzelrechnung aus den entsprechenden Regeln der Potenzrechnung. Wegen der besonderen Bedeutung werden die übertragbaren Regeln hier in Wurzelschreibweise wiederholt.

1.6.4 Regeln der Wurzelrechnung

1. *Addieren und Subtrahieren*

Wurzeln kann man nur addieren oder subtrahieren, wenn sie in Radikand *und* Wurzelexponent übereinstimmen.

$$p\sqrt[n]{a} + q\sqrt[n]{a} = (p + q)\sqrt[n]{a}$$

Beispiel

$2 \cdot \sqrt[4]{3} + 5 \cdot \sqrt[4]{3} = (2+5) \cdot \sqrt[4]{3} = 7 \cdot \sqrt[4]{3}.$ ◀

Fehlerwarnungen: $\sqrt[4]{2} + \sqrt[4]{3} \neq \sqrt[4]{5}$, $\sqrt[3]{2} + \sqrt[4]{2} \neq \sqrt[7]{2}$ (beides lässt sich nicht zusammenfassen).

2. *Multiplizieren und Dividieren bei gleichem Radikanden*

Wurzeln mit gleichem Radikanden und den Wurzelexponenten n, m werden multipliziert, indem man aus dem in die $(m+n)$-te Potenz erhobenen Radikanden die nm-te Wurzel zieht (denn $\sqrt[n]{a} \cdot \sqrt[m]{a} = a^{\frac{1}{n}} \cdot a^{\frac{1}{m}} = a^{\frac{1}{n}+\frac{1}{m}} = a^{\frac{m+n}{n \cdot m}} = \sqrt[n \cdot m]{a^{m+n}}$).

$$\sqrt[n]{a} \cdot \sqrt[m]{a} = \sqrt[n \cdot m]{a^{m+n}}$$

Beispiel

$\sqrt[3]{4096} \cdot \sqrt[4]{4096} = \begin{cases} 16 \cdot 8 = 128 \\ \sqrt[12]{4096^7} = 2^7 = 128 \end{cases}$ ◀

Fehlerwarnung: $\sqrt[3]{a} \cdot \sqrt[4]{a} \neq \sqrt[12]{a}$, $\sqrt[3]{a} \cdot \sqrt[4]{a} \neq \sqrt[12]{a^2}$, $\sqrt[3]{a} \cdot \sqrt[4]{a} \neq \sqrt[7]{a}$ (richtig: $\sqrt[3]{a} \cdot \sqrt[4]{a} = \sqrt[3 \cdot 4]{a^{4+3}} = \sqrt[12]{a^7}$).

Wurzeln mit gleichem Radikanden und den Wurzelexponenten n und m werden dividiert, indem man aus dem in die $(m-n)$-te Potenz erhobenen Radikanden die nm-te Wurzel zieht (denn $\frac{\sqrt[n]{a}}{\sqrt[m]{a}} = \frac{a^{\frac{1}{n}}}{a^{\frac{1}{m}}} = a^{\frac{1}{n}-\frac{1}{m}} = a^{\frac{m-n}{n \cdot m}} = \sqrt[n \cdot m]{a^{m-n}}$).

$$\frac{\sqrt[n]{a}}{\sqrt[m]{a}} = \sqrt[n \cdot m]{a^{m-n}}$$

Beispiel

$\frac{\sqrt[3]{4096}}{\sqrt[4]{4096}} = \begin{cases} 16 : 8 = 2 \\ \sqrt[3 \cdot 4]{4096^{4-3}} = \sqrt[12]{4096} = \sqrt[12]{2^{12}} = 2 \end{cases}$ ◀

3. *Multiplizieren und Dividieren bei gleichem Wurzelexponenten*

Wurzeln mit gleichem Wurzelexponenten werden multipliziert, indem man die Radikanden multipliziert (genauer: indem man die Wurzel aus dem Produkt der Radikanden zieht) (denn $\sqrt[n]{a} \cdot \sqrt[n]{b} = a^{\frac{1}{n}} \cdot b^{\frac{1}{n}} = (ab)^{\frac{1}{n}} = \sqrt[n]{ab}$).

$$\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$$

Beispiel

$\sqrt[3]{8} \cdot \sqrt[3]{27} = \begin{cases} 2 \cdot 3 = 6 \\ \sqrt[3]{8 \cdot 27} = \sqrt[3]{216} = 6 \end{cases}$ ◀

Fehlerwarnung: $\sqrt[n]{a} + \sqrt[n]{b} \neq \sqrt[n]{a+b}$ (außer in Sonderfällen); der gegebene Ausdruck lässt sich nicht vereinfachen.

Wurzeln mit gleichem Wurzelexponenten werden dividiert, indem man die Radikanden dividiert (genauer: indem man die Wurzel aus dem Quotienten der Radikanden zieht) (denn $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \frac{a^{\frac{1}{n}}}{b^{\frac{1}{n}}} = \left(\frac{a}{b}\right)^{\frac{1}{n}} = \sqrt[n]{\frac{a}{b}}$).

$$\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$$

Beispiel

$\sqrt[3]{8} : \sqrt[3]{27} = \begin{cases} 2 : 3 = \frac{2}{3} \\ \sqrt[3]{\frac{8}{27}} = \frac{2}{3} \end{cases}$ ◀

Umkehrungen

Man zieht die Wurzel aus einem Produkt, indem man die Wurzel aus den einzelnen Faktoren zieht (genauer: indem man die Wurzeln aus den einzelnen Faktoren miteinander multipliziert).

$$\sqrt[n]{ab} = \sqrt[n]{a} \cdot \sqrt[n]{b}$$

Beispiel

$\sqrt{36} = \begin{cases} 6 \\ \sqrt{4} \cdot \sqrt{9} = 2 \cdot 3 = 6 \end{cases}$ ◀

Fehlerwarnung: $\sqrt[n]{a+b} \neq \sqrt[n]{a} + \sqrt[n]{b}$ (außer in Sonderfällen).

Man zieht die Wurzel aus einem Bruch, indem man sie aus Zähler und Nenner einzeln zieht (genauer: indem man die Wurzeln aus Zähler und Nenner durcheinander dividiert).

$$\sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}$$

Spezialfall

$$\sqrt[n]{\frac{1}{a}} = \frac{1}{\sqrt[n]{a}}$$

Beispiele

1. $\sqrt[3]{\frac{8}{27}} = \sqrt[3]{8} : \sqrt[3]{27} = 2 : 3 = \frac{2}{3}$
2. $\sqrt[4]{\frac{1}{625}} = \frac{1}{\sqrt[4]{625}} = \frac{1}{5}$. ◀

4. *Radizieren und Potenzieren einer Wurzel*

Man zieht die Wurzel aus einer Wurzel, indem man die Wurzelexponenten multipliziert (genauer: indem man aus dem Radikanden die Wurzel mit dem aus dem Produkt beider Wurzelexponenten gebildeten neuen Wurzelexponenten zieht) (denn $\sqrt[n]{\sqrt[m]{a}} = (a^{\frac{1}{m}})^{\frac{1}{n}} = a^{\frac{1}{mn}} = \sqrt[m\cdot n]{a}$).

$$\sqrt[n]{\sqrt[m]{a}} = \sqrt[m\cdot n]{a}$$

Beispiel

$$\sqrt[4]{\sqrt[3]{4096}} = \begin{cases} \sqrt[4]{16} = 2 \\ \sqrt[12]{4096} = 2 \end{cases}$$ ◀

Fehlerwarnung: $\sqrt[3]{\sqrt[5]{7}} \neq \sqrt[8]{7}$ (richtig: $\sqrt[3]{\sqrt[5]{7}} = \sqrt[15]{7}$).

Bei der Wurzel aus einer Wurzel kann man die Wurzelexponenten miteinander vertauschen (denn $\sqrt[n]{\sqrt[m]{a}} = a^{\frac{1}{mn}} = a^{\frac{1}{nm}} = \sqrt[m]{\sqrt[n]{a}}$).

$$\sqrt[n]{\sqrt[m]{a}} = \sqrt[m]{\sqrt[n]{a}}$$

Beispiele

1. $\sqrt[4]{\sqrt[3]{4096}} = \begin{cases} \sqrt[4]{16} = 2 \\ \sqrt[3]{\sqrt[4]{4096}} = \sqrt[3]{8} = 2 \end{cases}$
2. $\sqrt[3]{\sqrt[4]{27}} = \sqrt[4]{\sqrt[3]{27}} = \sqrt[4]{3}$. ◀

Eine Wurzel wird potenziert, indem man den Radikanden potenziert (genauer: indem man die Wurzel aus der Potenz des Radikanden zieht) (denn $(a^{\frac{1}{n}})^m = a^{\frac{m}{n}} = (a^m)^{\frac{1}{n}}$).

$$(\sqrt[n]{a})^m = \sqrt[n]{a^m}$$

Beispiele

1. $(\sqrt[3]{8})^4 = \begin{cases} 2^4 = 16 \\ \sqrt[3]{8^4} = \sqrt[3]{4096} = 16 \end{cases}$
2. $(\sqrt[4]{4})^2 = \sqrt[4]{16} = 2$. ◀

Umkehrung

Man zieht die Wurzel aus einer Potenz, indem man die Wurzel aus der Basis in die entsprechende Potenz erhebt.

$$\sqrt[n]{a^m} = (\sqrt[n]{a})^m$$

Beispiel

$$\sqrt{4}^3 = \begin{cases} \sqrt{64} = 8 \\ (\sqrt{4})^3 = 2^3 = 8 \end{cases}$$ ◀

Spezialfall

$$(\sqrt[n]{a})^n = \sqrt[n]{a^n} = a$$

Beispiel

$(\sqrt[5]{3})^5 = \sqrt[5]{3^5} = 3$ ◀

Exponenten und Wurzelexponenten kann man gegeneinander kürzen.

$$\sqrt[np]{a^{nq}} = (\sqrt[np]{a})^{nq} = \sqrt[p]{a^q} = (\sqrt[p]{a})^q$$

Beispiel

$\sqrt[12]{2^8} = \sqrt[3]{2^2} = \sqrt[3]{4}$ ◀

5. *Rationalmachen des Nenners*

Zum Rationalmachen des Nenners eines Bruches erweitert man den Bruch so, dass die Wurzel im Nenner wegfällt.

$$\frac{a}{\sqrt{b}} = \frac{a\sqrt{b}}{b}$$

Beispiele

1. $\frac{a^2}{\sqrt{a}} = \frac{a^2 \cdot \sqrt{a}}{\sqrt{a} \cdot \sqrt{a}} = \frac{a^2 \cdot \sqrt{a}}{a} = a\sqrt{a}$
2. $\frac{a}{\sqrt[3]{a}} = \frac{a \cdot \sqrt[3]{a^2}}{\sqrt[3]{a} \cdot \sqrt[3]{a^2}} = \frac{a \cdot \sqrt[3]{a^2}}{a} = \sqrt[3]{a^2}$
3. $\frac{x+\sqrt{y}}{x-\sqrt{y}} = \frac{(x+\sqrt{y})(x+\sqrt{y})}{(x-\sqrt{y})(x+\sqrt{y})} = \frac{(x+\sqrt{y})^2}{x^2-y} = \frac{x^2+2x\sqrt{y}+y}{x^2-y}$. ◀

1.7 Dezimalzahlen und Dualzahlen

Es gibt verschiedene Möglichkeiten zur Darstellung von Zahlen. Die einzelnen Zeichen zur Darstellung von Zahlen sind die Zahlzeichen oder Ziffern. Grundsätzlich unterscheidet man zwischen sogenannten Positions- oder Stellenwertsystemen und Additionssystemen. Bei einem Positionssystem ist der Wert einer Ziffer abhängig von der Position dieser Ziffer innerhalb der Zahl. Bei Additionssystemen wird der Wert aller Zahlzeichen einfach addiert, um den Wert der Zahl festzulegen.

Ein Beispiel für ein Positionssystem ist unser Dezimalsystem, ein Beispiel für ein Additionssystem ist das römische Zahlensystem.

1.7.1 Dezimalsystem

Die heute übliche Schreibweise der Zahlen ist die Dezimalschreibweise, das heißt, es gibt 10 verschiedene Ziffern $(0, 1, 2, \ldots, 9)$ zur Darstellung der Zahlen. Jede Ziffer hat den zehnfachen Stellenwert der ihr rechts folgenden Ziffer. Im Dezimalsystem dargestellte Zahlen nennt man Dezimalzahlen.

Beispiel

$3607 = 3 \cdot 10^3 + 6 \cdot 10^2 + 0 \cdot 10^1 + 7 \cdot 10^0 = 3000 + 600 + 0 + 7.$ ◀

Der Wert einer Ziffer innerhalb der Zahl ergibt sich also folgendermaßen: Die Einerstelle bleibt unverändert (Multiplikation mit $10^0 = 1$), die Zehnerstelle wird mit $10^1 = 10$, die Hunderterstelle wird mit $10^2 = 100$, die Tausenderstelle wird mit $10^3 = 1000, \ldots$, die n-te Stelle wird mit 10^{n-1} multipliziert.

Das Dezimalsystem (auch Zehnersystem genannt) ist also ein Positionssystem zur Basis 10. Eine solche Schreibweise wurde erst möglich nach Einführung der Null (für „nichts“).

Unser Zahlensystem wurde im ersten Jahrtausend nach der Zeitenwende in Indien entwickelt. Es gelangte über den arabischen Raum zunächst nach Spanien und dann nach Mitteleuropa, wo noch bis zum 16. Jahrhundert mit dem römischen Zahlensystem gerechnet wurde. Wegen dieses Ursprungs nennt man unsere Ziffern auch indisch-arabische Ziffern.

Im Dezimalsystem lassen sich auch rationale und reelle Zahlen darstellen. Die Darstellung einer reellen (rationalen oder irrationalen) Zahl als Dezimalzahl nennt man auch Dezimalbruch (vgl. auch Abschn. 1.3).

Beispiel

$486{,}2545 = 4 \cdot 10^2 + 8 \cdot 10^1 + 6 \cdot 10^0 + 2 \cdot 10^{-1} + 5 \cdot 10^{-2} + 4 \cdot 10^{-3} + 5 \cdot 10^{-4} = 400 + 80 + 6 + 0{,}2 + 0{,}05 + 0{,}004 + 0{,}0005.$ ◀

Der Wert einer Ziffer innerhalb der Zahl ergibt sich dadurch, dass die n-te Stelle vor dem Komma mit 10^{n-1} und die m-te Stelle nach dem Komma mit 10^{-m} multipliziert wird.

Ist $a = a_n a_{n-1} \ldots a_2 a_1 a_0, a_{-1} a_{-2} \ldots a_{-m}$ eine Zahl mit den Ziffern $a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ vor dem Komma und den Ziffern $a_{-1}, a_{-2}, \ldots, a_{-m}$ nach dem Komma, dann gilt also

$$a = \sum_{i=-m}^{n} a_i 10^i \qquad (*)$$

Die Stellen mit $i \geq 0$ bilden den ganzen Teil, die mit $i < 0$ den gebrochenen Teil der Zahl.

Für andere Zahlensysteme, nämlich Positionssysteme zur Basis B, gilt $(*)$ ganz analog, wenn man 10 durch die entsprechende Basis B ersetzt (zum Beispiel $B = 2$ für das Dualsystem).

1.7.2 Dualsystem

Das Dualsystem ist ein System zur Darstellung der Zahlen, in dem es nur zwei Ziffern (0 und 1) gibt. Das Dualsystem wird deshalb manchmal auch Binärsystem oder Zweiersystem genannt. Es ist ein Positionssystem zur Basis 2. Der Wert einer Ziffer ist also abhängig von der Position innerhalb der Zahl. Jede Ziffer hat den doppelten Stellenwert der ihr rechts folgenden

Ziffer. Im Dualsystem dargestellte Zahlen nennt man Dualzahlen.

Beispiel

Der Dualzahl 1 001 101 entspricht die Dezimalzahl $1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 64 + 0 + 0 + 8 + 4 + 0 + 1 = 77$. ◀

Die Umrechnung von einem Zahlensystem in ein anderes wird als Konvertierung bezeichnet. Werden mehrere Zahlensysteme gleichzeitig benutzt, so ist es zur Vermeidung von Irrtümern üblich, die Basis als Index anzuhängen.

Beispiel

$1\,001\,101_2 = 77_{10}$. ◀

Die Darstellung reeller Zahlen im Dualsystem ist analog der Darstellung im Dezimalsystem. Der Wert einer Ziffer innerhalb der Zahl ergibt sich dadurch, dass die n-te Stelle vor dem Komma mit 2^{n-1} und die m-te Stelle nach dem Komma mit 2^{-m} multipliziert wird.

Beispiel

$101\,100\,011{,}1011_2 = 1 \cdot 2^8 + 1 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2 + 1 \cdot 1 + 1 \cdot 2^{-1} + 1 \cdot 2^{-3} + 1 \cdot 2^{-4} = 256 + 64 + 32 + 2 + 1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{16} = 355\frac{11}{16} = 355{,}6875_{10}$. ◀

Dualsysteme sind sehr bedeutend in Elektrotechnik und Datenverarbeitung. Computer sind zeichenverarbeitende Maschinen. Die externen Zeichen (Buchstaben, Ziffern, Sonderzeichen) werden intern im Binärcode in Form von Bitfolgen dargestellt. Ein Bit (Abkürzung von Binary Digit) ist die kleinste darstellbare Informationseinheit mit den Werten 0 und 1. Acht Bit werden zur nächsthöheren Einheit, dem Byte, zusammengefasst. Zahlen werden in Computern in mehreren aufeinanderfolgenden Bytes dargestellt. Die interne Durchführung arithmetischer Operationen erfolgt im Computer also im Dualsystem.

Andere Zahlensysteme, die im Zusammenhang mit der Nutzung von Computern eine Rolle spielen, sind das Oktalsystem (Positionssystem zur Basis 8) mit den Ziffern 0, 1, 2, 3, 4, 5, 6, 7 und das Hexadezimalsystem (Positionssystem zur Basis 16) mit den Ziffern 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F (die Buchstaben A, ..., F stehen für die Werte 10, ..., 15).

1.7.3 Runden

Als Dezimalstellen einer Dezimalzahl bezeichnet man die Stellen nach dem Komma. Runden ist das Verkürzen einer Dezimalzahl, also die Darstellung einer Dezimalzahl mit einer vorgegebenen Anzahl von Dezimalstellen.

Rundungsregel

Ist die erste weggelassene Ziffer 0, 1, 2, 3, 4, dann bleibt die letzte geschriebene Ziffer unverändert. Ist die erste weggelassene Ziffer 5, 6, 7, 8, 9, dann erhöht sich die letzte geschriebene Ziffer um 1.

Ist die gerundete Zahl kleiner als die ursprüngliche Zahl (die erste weggelassene Ziffer ist dann 0, 1, 2, 3 oder 4), spricht man von Abrunden. Ist die gerundete Zahl jedoch größer als die ursprüngliche Zahl (die erste weggelassene Ziffer ist dann 5, 6, 7, 8 oder 9), so spricht man von Aufrunden.

Beispiele

$3{,}456 \approx 3{,}46$ (aufgerundet);

$23{,}699 \approx 23{,}70$ (aufgerundet);

$14{,}3449 \approx 14{,}34$ (abgerundet);

$17{,}2496389 \approx 17{,}2496$ (auf 4 Dezimalstellen gerundet). ◀

1.8 Logarithmen

1.8.1 Definition des Logarithmus

Eine Zahl der Form $\log_a b$ (gesprochen: Logarithmus b zur Basis a) heißt Logarithmus. Dabei heißt b Numerus des Logarithmus und ist eine reelle Zahl größer als 0, und a heißt Basis des Logarithmus und ist eine positive reelle Zahl ungleich 1.

$\log_a b$	Logarithmus b zur Basis a
$b > 0$	Numerus
$a > 0, a \neq 1$	Basis

Der Logarithmus $\log_a b$ ist definiert als die eindeutige Lösung x der Gleichung $a^x = b$.

$$\log_a b = x \quad \Leftrightarrow \quad a^x = b$$

Der Logarithmus $x = \log_a b$ ist also der Exponent zu der Basis a, für den die Potenz a^x gleich dem Numerus b ist.

Aus der Definition folgt (denn $a^1 = a$ und $a^0 = 1$).

$$\log_a a = 1 \quad \log_a 1 = 0$$

Beispiele

1. $\log_2 8 = x \quad \Rightarrow x = 3 \quad \text{denn } 2^3 = 8$
2. $\log_4 16 = x \quad \Rightarrow x = 2 \quad \text{denn } 4^2 = 16$
3. $\log_4 4 = x \quad \Rightarrow x = 1 \quad \text{denn } 4^1 = 4$
4. $\log_5 1 = x \quad \Rightarrow x = 0 \quad \text{denn } 5^0 = 1$
5. $\log_3 81 = x \quad \Rightarrow x = 4 \quad \text{denn } 3^4 = 81$
6. $\log_5 5^7 = x \quad \Rightarrow x = 7 \quad \text{denn } 5^7 = 5^7$
7. $\log_{10} 1000 = x \quad \Rightarrow x = 3 \quad \text{denn } 10^3 = 1000$
8. $\log_2 0{,}5 = x \quad \Rightarrow x = -1 \quad \text{denn } 2^{-1} = \frac{1}{2} = 0{,}5$
9. $\log_a 8 = 1 \quad \Rightarrow a = 8 \quad \text{denn } 8^1 = 8$
10. $\log_a \frac{1}{125} = -3 \Rightarrow a = 5 \quad \text{denn } 5^{-3} = \frac{1}{125}$
11. $\log_a \frac{3}{5} = -1 \quad \Rightarrow a = \frac{5}{3} \quad \text{denn } (\frac{5}{3})^{-1} = \frac{3}{5}$
12. $\log_{10} b = -1 \quad \Rightarrow b = 0{,}1 \quad \text{denn } 10^{-1} = \frac{1}{10} = 0{,}1$
13. $\log_4 b = \frac{5}{2} \quad \Rightarrow b = 32$

    $\text{denn } 4^{\frac{5}{2}} = \sqrt{4}^5 = 2^5 = 32$ ◀

1.8.2 Spezielle Basen

Logarithmen zur Basis $a = 10$ heißen Zehnerlogarithmen oder dekadische Logarithmen oder Briggs'sche Logarithmen (nach dem englischen Mathematiker Henry Briggs, 1556–1630). Man schreibt statt $\log_{10} b$ auch einfach $\lg b$.

$$\log_{10} b = \lg b$$

Logarithmen mit der Euler'schen Zahl $e = 2{,}71828182\ldots$ als Basis werden natürliche Logarithmen oder Neper'sche Logarithmen (nach dem schottischen Mathematiker John Neper (Napier), 1550 bis 1617) genannt. Man schreibt $\ln b$ für $\log_e b$.

$$\log_e b = \ln b$$

Die Euler'sche Zahl e ist der Grenzwert der Folge $\left(1 + \frac{1}{n}\right)^n$ (vgl. Abschn. 1.8.1):

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2{,}7182818284\ldots$$

Sie hat ihren Namen nach dem schweizerischen Mathematiker Leonhard Euler (1707–1783). Die Euler'sche Zahl ist eine irrationale Zahl.

Logarithmen zur Basis $a = 2$ heißen Zweierlogarithmen oder binäre Logarithmen oder duale Logarithmen. Man schreibt statt $\log_2 b$ manchmal auch $\operatorname{ld} b$.

$$\log_2 b = \operatorname{ld} b$$

Beispiele

1. $\log_2 32 = \operatorname{ld} 32 = x$

   $\Rightarrow x = 5 \qquad \text{denn } 2^5 = 32$
2. $\operatorname{ld} \sqrt{2} = x$

   $\Rightarrow x = \frac{1}{2} \qquad \text{denn } 2^{\frac{1}{2}} = \sqrt{2}$
3. $\log_{10} 10.000 = \lg 10.000 = x$

   $\Rightarrow x = 4 \qquad \text{denn } 10^4 = 10.000$
4. $\lg 0{,}01 = x$

   $\Rightarrow x = -2 \qquad \text{denn } 10^{-2} = 0{,}01$
5. $\log_e 5 = \ln 5 = x$

   $\Rightarrow x = 1{,}6094\ldots \quad \text{denn } e^{1{,}6094\ldots} = 5$ ◀

1.8.3 Regeln der Logarithmenrechnung

1. Der Logarithmus eines Produkts ist gleich der Summe der Logarithmen der einzelnen Faktoren. Oder: Addiert man zum Logarithmus einer Zahl u den Logarithmus einer Zahl v, dann erhält man als Summe den Logarithmus des Produkts uv.

$$\log_a(u \cdot v) = \log_a u + \log_a v$$

Beweis:

Setze $\log_a u = x$, $\log_a v = y$

$\Rightarrow \log_a u + \log_a v = x + y$

$a^x = u$, $a^y = v \Rightarrow u \cdot v = a^x \cdot a^y = a^{x+y}$

$\Rightarrow \log_a(u \cdot v) = x + y.$

Beispiel

$\log_2 256 = \operatorname{ld} 256 = \operatorname{ld}(4{\cdot}64) = \operatorname{ld} 4 + \operatorname{ld} 64 = 2 + 6 = 8.$ ◀

2. Der Logarithmus eines Bruches (Quotienten) ist gleich der Differenz der Logarithmen von Zähler (Dividend) und Nenner (Divisor). Oder: Subtrahiert man vom Logarithmus einer

Zahl u den Logarithmus einer Zahl v, dann erhält man als Differenz den Logarithmus des Bruches (Quotienten) $\frac{u}{v}$.

$$\log_a \frac{u}{v} = \log_a u - \log_a v$$

Beweis:

Setze $\log_a u = x$, $\log_a v = y \Rightarrow \log_a u - \log_a v = x - y$

$a^x = u$, $a^y = v \Rightarrow \frac{u}{v} = \frac{a^x}{a^y} = a^{x-y}$

$\Rightarrow \log_a \frac{u}{v} = x - y$.

Beispiel

$\log_3 \frac{9}{243} = \log_3 9 - \log_3 243 = 2 - 5 = -3.$ ◀

3. Der Logarithmus einer Potenz ist gleich dem mit dem Exponenten multiplizierten Logarithmus der Basis. Oder: Multipliziert man den Logarithmus einer Zahl u mit einer Zahl r, dann erhält man den Logarithmus der Potenz u^r.

$$\log_a(u^r) = r \cdot \log_a u$$

Beweis:

Setze $\log_a u = x \Rightarrow a^x = u \Rightarrow u^r = (a^x)^r = a^{rx}$

$\Rightarrow rx = \log_a(u^r) = r \log_a u$.

Beispiele

1. $\log_2 8^3 = 3 \log_2 8 = 3 \log_2 2^3 = 3 \cdot 3 = 9$
2. $\lg 10.000 = \lg 10^4 = 4 \lg 10 = 4.$ ◀

4. Der Logarithmus einer Wurzel ist gleich dem durch den Wurzelexponenten dividierten Logarithmus des Radikanden. Oder: Dividiert man den Logarithmus einer Zahl u durch eine Zahl n, dann erhält man den Logarithmus der Wurzel $\sqrt[n]{u}$.

$$\log_a \sqrt[n]{u} = \frac{1}{n} \log_a u$$

Beweis:

Wegen $\sqrt[n]{u} = u^{\frac{1}{n}}$ folgt

$\log_a \sqrt[n]{u} = \log_a u^{\frac{1}{n}} = \frac{1}{n} \cdot \log_a u$

aus der Potenzregel.

Beispiele

1. $\log_5 \sqrt[3]{5} = \frac{1}{3} \log_5 5 = \frac{1}{3}$
2. $\log_2 \sqrt[3]{64} = \frac{1}{3} \cdot \log_2 64 = \frac{1}{3} \cdot \log_2 2^6 = \frac{1}{3} \cdot 6 = 2.$ ◀

5. Aus den Regeln ergeben sich die folgenden Spezialfälle (denn $a^r = a^r$ und $\log_a \frac{1}{v} = \log_a 1 - \log_a v = 0 - \log_a v$).

$$\log_a(a^r) = r \quad \text{und} \quad \log_a \frac{1}{v} = -\log_a v$$

1.8.4 Zusammenhang von Logarithmen mit verschiedenen Basen

Für Logarithmen mit verschiedenen Basen a und c gilt folgende Umrechnungsregel

$$\log_a u = \frac{\log_c u}{\log_c a}$$

Beweis:

Setze $\log_a u = x$. Es folgt $a^x = u$, also auch $\log_c a^x = \log_c u$. Nach der Potenzregel ergibt sich $x \log_c a = \log_c u$ und nach x aufgelöst $x = \frac{\log_c u}{\log_c a}$. Wegen $x = \log_a u$ folgt die Behauptung: $\log_a u = \frac{\log_c u}{\log_c a}$.

Logarithmen zu verschiedenen Basen (a und c) unterscheiden sich also nur durch einen konstanten Faktor $\left(\frac{1}{\log_c a}\right)$.

Für $c = u = b$ ergibt sich der Spezialfall

$$\log_a b = \frac{1}{\log_b a}$$

Beispiele

1. $\ln u = \frac{1}{\lg e} \cdot \lg u = \frac{1}{0{,}4342...} \cdot \lg u$
2. $\lg u = \frac{1}{\ln 10} \cdot \ln u = \frac{1}{2{,}3025...} \cdot \ln u$
3. $\log_5 u = \frac{1}{\lg 5} \cdot \lg u = \frac{1}{0{,}6989...} \cdot \lg u$
4. $\ln 1000 = \frac{1}{\lg e} \cdot \lg 1000 \approx \frac{1}{0{,}4342944819} \cdot 3 \approx 6{,}907755279.$ ◀

1.8.5 Dekadische Logarithmen

Die dekadischen Logarithmen (auch Briggs'sche Logarithmen oder Zehnerlogarithmen genannt) haben den Vorteil, dass man mit den Logarithmen der Dezimalzahlen zwischen 1 und 10 über die Logarithmen aller positiven reellen Zahlen verfügt.

Begründung: Jede reelle Zahl x lässt sich durch Abspalten einer Zehnerpotenz 10^k mit ganzzahligem k in der Form $x = 10^k \cdot \bar{x}$ mit $1 \le \bar{x} < 10$ schreiben. Dabei ist $\bar{x}$ durch die Ziffernfolge von x bestimmt, während 10^k die Größenordnung von x angibt. Logarithmieren ergibt $\lg x = \lg(10^k \cdot \bar{x}) = \lg(10^k) + \lg \bar{x} = k + \lg \bar{x}$ mit $0 \le \lg \bar{x} < 1$ (also $\lg \bar{x} = 0, \ldots$). Man nennt k die Kennzahl und die Ziffernfolge hinter dem Komma von $\lg \bar{x}$ die Mantisse des Logarithmus von x.

Von einer vor dem Komma n-stelligen Zahl ist die Kennzahl $n-1$ (also um 1 kleiner). Für Zahlen kleiner als 1 sind die Kennzahlen negativ.

Logarithmentafeln enthalten in der Regel nur die Mantisse.

Alle Zahlen, die sich nur durch Zehnerpotenzen unterscheiden, haben die gleiche Mantisse, aber unterschiedliche Kennzahlen.

Beispiele

1. $\lg 2250 = \lg(1000 \cdot 2{,}25) = \lg(10^3 \cdot 2{,}25) = \lg 10^3 + \lg 2{,}25 \approx 3 + 0{,}3522 = 3{,}3522$
2. $\lg 0{,}0315 = \lg(3{,}15 \cdot \frac{1}{100}) = \lg(3{,}15 \cdot 10^{-2}) = \lg 3{,}15 + \lg 10^{-2} \approx 0{,}4983 - 2$
3. $\lg 2000 = \lg(10^3 \cdot 2) = \lg 10^3 + \lg 2 \approx 3 + 0{,}3010 = 3{,}3010$, $\lg 200 \approx 2{,}3010$, $\lg 20 \approx 1{,}3010$, $\lg 2 \approx 0{,}3010$, $\lg 0{,}2 \approx 0{,}3010 - 1$, $\lg 0{,}02 \approx 0{,}3010 - 2$, $\lg 0{,}002 \approx 0{,}3010 - 3$. ◀

1.9 Ungleichungen

1.9.1 Definitionen und Rechenregeln

Zwischen zwei reellen Zahlen a und b besteht genau eine der drei Beziehungen: $a = b$ (a ist gleich b), $a < b$ (a ist kleiner als b), $a > b$ (a ist größer als b). Der Winkelhaken ist dabei immer nach der größeren Seite hin geöffnet.

$$a, b \in \mathbb{R}: \quad a = b \text{ oder } a < b \text{ oder } a > b$$

Wegen dieser Eigenschaft nennt man die Menge $\mathbb{R}$ der reellen Zahlen geordnet.

Im Falle $a \neq b$ (a ungleich b) gilt genau eine der beiden Ungleichungen $a < b$ oder $a > b$.

$$a \neq b \quad \Rightarrow \quad a < b \text{ oder } a > b$$

Die Ungleichung $a \le b$ bedeutet, dass a kleiner oder gleich b ist (a ist also nicht größer als b), und die Ungleichung $a \ge b$ bedeutet entsprechend, dass a größer oder gleich b ist (b ist also nicht größer als a).

Ist $a < b$ und $b < c$, dann kann man die beiden Ungleichungen fortlaufend schreiben: $a < b < c$. Man nennt dies fortlaufende Ungleichung oder Ungleichungskette.

$$a < b < c \quad \Leftrightarrow \quad a < b \text{ und } b < c$$

Entsprechend schreibt man zum Beispiel $a < x$ und $x \le b$ zusammenfassend als $a < x \le b$. Fortlaufende Ungleichungen werden oft benutzt, um einen Bereich oder ein Intervall anzugeben, aus dem eine Größe x gewählt werden darf oder gewählt werden soll.

Eigenschaften von Ungleichungen:

1. $a \le a$ (Reflexivität)
2. $a \le b$ und $b \le c \Rightarrow a \le c$ (Transitivität)
3. $a \le b$ und $b \le a \Rightarrow a = b$ (Antisymmetrie)

Rechenregeln für Ungleichungen:

Eine Ungleichung kann von beiden Seiten gelesen werden:

$$a < b \quad \Leftrightarrow \quad b > a \quad \text{für alle } a, b \in \mathbb{R}$$

Auf beiden Seiten einer Ungleichung darf dieselbe Zahl addiert werden:

$$a \le b \quad \Rightarrow \quad a + c \le b + c \quad \text{für alle } a, b, c \in \mathbb{R}$$

Zwei gleichgerichtete Ungleichungen dürfen addiert werden:

$$a \le b \text{ und } c \le d \quad \Rightarrow \quad a + c \le b + d$$

für alle $a, b, c, d \in \mathbb{R}$,

$$a < b \text{ und } c \le d \quad \Rightarrow \quad a + c < b + d$$

für alle $a, b, c, d \in \mathbb{R}$.

Eine Ungleichung darf mit einer nichtnegativen Zahl multipliziert werden:

$$a \le b \text{ und } c \ge 0 \quad \Rightarrow \quad ac \le bc \quad \text{für alle } a, b \in \mathbb{R}$$

Wird eine Ungleichung mit einer negativen Zahl multipliziert, so dreht sich das Ungleichheitszeichen um:

$$a \le b \text{ und } c \le 0 \quad \Rightarrow \quad ac \ge bc \quad \text{für alle } a, b \in \mathbb{R}$$

Bildet man auf beiden Seiten einer Ungleichung den Kehrwert, so dreht sich das Ungleichheitszeichen um:

$$a \le b \quad \Rightarrow \quad \frac{1}{a} \ge \frac{1}{b} \quad \text{für alle } a, b \in \mathbb{R}^*$$

Aus der Multiplikationsregel folgt im besonderen (Multiplikation mit -1), dass das Vertauschen der Vorzeichen auf beiden Seiten einer Ungleichung das Ungleichheitszeichen umdreht:

$$a < b \quad \Rightarrow \quad -a > -b$$

Beispiele

1. $3 \le 3; 5 \ge 5; -12{,}2 \le -12{,}2$
2. $3 \le 5, 5 \le 6 \Rightarrow 3 \le 6; -4 \ge -5, -5 \ge -7{,}1 \Rightarrow -4 \ge -7{,}1$
3. $3 \le 3, 3 \ge 3 \Rightarrow 3 = 3$
4. $3 < 5 \Leftrightarrow 5 > 3; -1 > -2 \Leftrightarrow -2 < -1$
5. $5 \le 7 \Leftrightarrow 5 + 2 \le 7 + 2; 4 > 1 \Leftrightarrow 4 - 3 > 1 - 3$
6. $3 < 4, 7 < 9 \Leftrightarrow 3 + 7 < 4 + 9; 4 \ge 3, -2 \ge -5 \Rightarrow 4 - 2 \ge 3 - 5$
7. $5 \le 7 \Rightarrow 5 \cdot 3 \le 7 \cdot 3; -3 > -4 \Rightarrow -3 \cdot 2 > -4 \cdot 2$
8. $4 > 2 \Rightarrow 4{\cdot}(-2) < 2{\cdot}(-2); 6 \le 7 \Rightarrow 6{\cdot}(-1) \ge 7{\cdot}(-1) \Rightarrow -6 \ge -7$
9. $2 < 3 \Rightarrow \frac{1}{2} > \frac{1}{3}; -5 > -6 \Rightarrow -\frac{1}{5} < -\frac{1}{6}$. ◀

1.9.2 Absolutbetrag

Der Betrag oder Absolutbetrag $|a|$ einer Zahl a stellt auf der Zahlengeraden den Abstand der Zahl a vom Nullpunkt dar. Da Abstände nicht negativ sind, gilt $|a| = a$ für $a \ge 0$ und $|a| = -a$ für $a < 0$.

$$|a| = \begin{cases} a & \text{für } a \ge 0 \\ -a & \text{für } a < 0 \end{cases}$$

Beträge sind nicht negativ.

Eigenschaften:

1. $|-a| = |a|$
2. $|a| \ge 0; |a| = 0 \Leftrightarrow a = 0$
3. $|a \cdot b| = |a| \cdot |b|$
4. $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$ für $b \ne 0$; $\left|\frac{1}{b}\right| = 1/|b|$ für $b \ne 0$
5. $|a^n| = |a|^n$ für $n \in \mathbb{N}$; $\left|\frac{1}{a^n}\right| = \frac{1}{|a|^n}$ für $n \in \mathbb{N}, a \ne 0$
6. $|a + b| \le |a| + |b|$ (sogenannte Dreiecksungleichung).

Beispiele

1. $|4{,}3| = 4{,}3; |-2| = 2; |-\pi| = \pi; |0| = 0$
2. $|3 \cdot 4| = |3| \cdot |4| = 3 \cdot 4 = 12; |(-3) \cdot 4| = |-3| \cdot |4| = 3 \cdot 4 = 12$
3. $\left|\frac{-2}{3}\right| = \frac{|-2|}{|3|} = \frac{2}{3}; \left|\frac{1}{-4}\right| = \frac{1}{|-4|} = \frac{1}{4}$
4. $|(-3)^5| = |-3|^5 = 3^5; \left|\frac{1}{(-2)^3}\right| = \frac{1}{|-2|^3} = \frac{1}{2^3}$
5. $|4{+}2| \le |4|{+}|2| \Rightarrow 6 \le 4{+}2 = 6; |5{-}3| \le |5|{+}|{-}3| \Rightarrow 2 \le 5 + 3 = 8.$ ◀

1.9.3 Intervalle

Es seien a und b zwei reelle Zahlen mit $a < b$. Die Menge aller reellen Zahlen x, die die fortlaufende Ungleichung $a < x < b$ erfüllen, heißt Intervall oder Zahlenintervall mit den Endpunkten oder Randpunkten a und b (genauer: offenes und beschränktes Intervall).

Gehört der Randpunkt nicht selbst zum Intervall, so spricht man von einem offenen Intervallende, im entgegengesetzten Fall von einem abgeschlossenen Intervallende.

Die Angabe eines Intervalls erfolgt durch seine Randpunkte a und b, indem diese in Klammern gesetzt werden. Eine eckige Klammer steht für ein abgeschlossenes Intervallende, eine runde für ein offenes Intervallende. Gehören beide Randpunkte zu dem Intervall, so heißt es abgeschlossen. Gehört nur einer der Randpunkte (also entweder a oder b) zum Intervall, so heißt es halboffen. Gehört keiner der Randpunkte zum Intervall, so heißt es offen.

Intervalle dienen der Beschreibung von Zahlenmengen. Man unterscheidet beschränkte und nicht beschränkte Intervalle.

Bei einem beschränkten Intervall sind die Intervallgrenzen a und b reelle Zahlen. Es besteht aus allen reellen Zahlen x, die zwischen diesen beiden Grenzen liegen.

Beschränkte Intervalle

$$[a, b] = \{x \,|\, x \in \mathbb{R} \text{ und } a \le x \le b\}$$

(abgeschlossenes Intervall)

$$(a, b) = \{x \,|\, x \in \mathbb{R} \text{ und } a < x < b\}$$

(offenes Intervall)

$$[a, b) = \{x \,|\, x \in \mathbb{R} \text{ und } a \le x < b\}$$

(halboffenes Intervall)

$$(a, b] = \{x \,|\, x \in \mathbb{R} \text{ und } a < x \le b\}$$

(halboffenes Intervall)

Das Symbol mit der Schreibweise ∞ heißt unendlich und steht für „beliebig groß“. Das Symbol $-\infty$ heißt entsprechend minus unendlich und steht für „beliebig klein“. Die Symbole ∞ und $-\infty$ sind keine reellen Zahlen; $-\infty$ ist kleiner als jede reelle Zahl, ∞ ist größer als jede reelle Zahl.

Bei einem nicht beschränkten Intervall ist mindestens eine der Intervallgrenzen $-\infty$ oder ∞. Solche Intervalle können durch eine Ungleichung beschrieben werden.

Nicht beschränkte Intervalle

$$[a,\infty) = \{x \,|\, x \in \mathbb{R} \text{ und } x \geq a\}$$

(halboffenes Intervall, nach rechts unbeschränkt)

$$(a,\infty) = \{x \,|\, x \in \mathbb{R} \text{ und } x > a\}$$

(offenes Intervall, nach rechts unbeschränkt)

$$(-\infty,a] = \{x \,|\, x \in \mathbb{R} \text{ und } x \leq a\}$$

(halboffenes Intervall, nach links unbeschränkt)

$$(-\infty,a) = \{x \,|\, x \in \mathbb{R} \text{ und } x < a\}$$

(offenes Intervall, nach links unbeschränkt)

$$(-\infty,\infty) = \{x \,|\, x \in \mathbb{R}\}$$

(offenes Intervall, nach links und nach rechts unbeschränkt)

Beispiele

1. $[3,4] = \{x \,|\, x \in \mathbb{R} \text{ und } 3 \leq x \leq 4 < 4\}$
   Alle reellen Zahlen zwischen 3 und 4; sowohl 3 als auch 4 gehören zum Intervall.
2. $[3,4) = \{x \,|\, x \in \mathbb{R} \text{ und } 3 \leq x < 4\}$
   Alle reellen Zahlen zwischen 3 und 4; 3 gehört zum Intervall, 4 jedoch nicht.
3. $(3,4] = \{x \,|\, x \in \mathbb{R} \text{ und } 3 < x \leq 4\}$
   Alle reellen Zahlen zwischen 3 und 4; 3 gehört nicht zum Intervall, aber 4.
4. $(3,4) = \{x \,|\, x \in \mathbb{R} \text{ und } 3 < x < 4\}$
   Alle reellen Zahlen zwischen 3 und 4; weder 3 noch 4 gehören zum Intervall.
5. $[3,\infty) = \{x \,|\, x \in \mathbb{R} \text{ und } x \geq 3\}$
   Alle reellen Zahlen größer oder gleich 3 gehören zum Intervall (3 gehört dazu).
6. $(3,\infty) = \{x \,|\, x \in \mathbb{R} \text{ und } x > 3\}$
   Alle reellen Zahlen größer als 3 gehören zum Intervall (3 gehört nicht dazu).
7. $(-\infty,4] = \{x \,|\, x \in \mathbb{R} \text{ und } x \leq 4\}$
   Alle reellen Zahlen kleiner oder gleich 4 gehören zum Intervall (4 gehört dazu).
8. $(-\infty,4) = \{x \,|\, x \in \mathbb{R} \text{ und } x < 4\}$
   Alle reellen Zahlen kleiner als 4 gehören zum Intervall (4 gehört nicht dazu).
9. $(-\infty,\infty) = \{x \,|\, x \in \mathbb{R}\}$
   Alle reellen Zahlen gehören zum Intervall. ◄

1.10 Komplexe Zahlen

1.10.1 Algebraische Form

Im Bereich der reellen Zahlen besitzt die Gleichung $x^2 + 1 = 0$ keine Lösung. Ebenso stellen $\sqrt{-3}$ oder $\sqrt[4]{-6}$ keine reellen Zahlen dar.

Falls eine quadratische Gleichung keine reelle Lösung besitzt, ist es trotzdem möglich, Lösungen anzugeben und zwar komplexe Zahlen als Lösungen. Zur Darstellung dieser komplexen Zahlen wird eine Erweiterung des Bereichs der reellen Zahlen vorgenommen.

Ausgangspunkt ist die imaginäre Einheit j, deren Quadrat gleich -1 ist: $j^2 = -1$. In der Mathematik wird für die imaginäre Einheit der Buchstabe i verwendet, in der Elektrotechnik nimmt man statt dessen jedoch den Buchstaben j, um Verwechslungen mit der Stromstärke i zu vermeiden.

Imaginäre Einheit j

$$j^2 = -1$$

Für die imaginäre Einheit gilt

$$j^2 = -1, \quad j^3 = -j, \quad j^4 = 1$$
$$j^{4n-3} = j, \quad j^{4n-2} = -1, \quad j^{4n-1} = -j, \quad j^{4n} = 1 \quad (n \in \mathbb{N}^*)$$

Die Zahlen j und $-j$ sind Lösungen der quadratischen Gleichung $x^2 + 1 = 0$.

Mit dieser imaginären Einheit j und zwei reellen Zahlen a und b stellt $z = a + bj$ eine komplexe Zahl dar.

$$z = a + bj, \quad a, b \in \mathbb{R}$$

Eine komplexe Zahl z besteht also aus einem reellen Teil a (Realteil) und einem imaginären Teil b (Imaginärteil).

Wenn a und b alle möglichen reellen Werte durchlaufen, dann werden alle möglichen komplexen Zahlen z erzeugt. Alle komplexen Zahlen bilden zusammen die Menge $\mathbb{C}$ der komplexen Zahlen.

$$\mathbb{C} = \{z = a + bj \,|\, a, b \in \mathbb{R}\}$$

Komplexe Zahlen z mit Realteil gleich 0 (also $a = 0$) heißen imaginäre Zahlen, die komplexen Zahlen z mit Imaginärteil

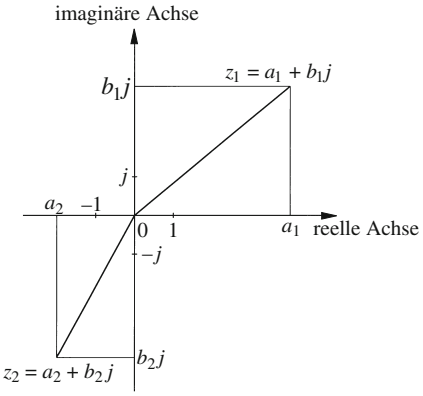


Abb. 1.1 Darstellung komplexer Zahlen in der Gauß'schen Zahlenebene

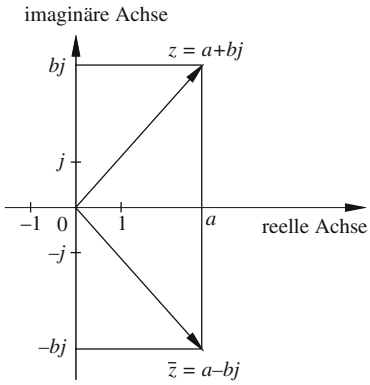


Abb. 1.2 Konjugiert komplexe Zahlen z und $\bar{z}$ in algebraischer Form

gleich 0 (also $b = 0$) sind die reellen Zahlen. Die komplexen Zahlen umfassen also die imaginären Zahlen und die reellen Zahlen.

$z = a + bj$		komplexe Zahlen
$z = bj$	$(a = 0)$	imaginäre Zahlen
$z = a$	$(b = 0)$	reelle Zahlen

Komplexe Zahlen $z = a + bj$ und $\bar{z} = a - bj$, also mit gleichem Realteil und entgegengesetzt gleichem Imaginärteil, heißen konjugiert komplex.

Komplexe Zahlen sind nicht mehr auf einer Zahlengeraden, sondern nur noch in einer Zahlenebene, der sogenannten Gauß'schen Zahlenebene, darstellbar (Name nach dem deutschen Mathematiker Carl Friedrich Gauß, 1777–1855).

Dabei wird in einem kartesischen Koordinatensystem der Ebene (siehe Abschn. 7.1.1) der Realteil a von z auf der Abszissenachse und der Imaginärteil b von z auf der Ordinatenachse abgetragen. Jeder komplexen Zahl entspricht ein Punkt der Ebene und umgekehrt. Die Zuordnung von Zahl und Punkt ist eineindeutig. Die reellen Zahlen liegen auf der Abszissenachse, die imaginären Zahlen liegen auf der Ordinatenachse. Deshalb nennt man die Abszissenachse auch reelle Achse und die Ordinatenachse imaginäre Achse (Abb. 1.1).

Die Darstellung einer komplexen Zahl in der Form $z = a + bj$, bei der kartesische Koordinaten verwendet werden, heißt algebraische Form. Daneben gibt es für die Darstellung der komplexen Zahlen die trigonometrische Form und die Exponentialform (Abb. 1.2 und 1.3).

1.10.2 Trigonometrische Form

Neben der Darstellung der komplexen Zahlen in algebraischer Form gibt es die Darstellung in trigonometrischer Form (vgl. Abschn. 6.1): $z = r(\cos\varphi + j\sin\varphi)$.

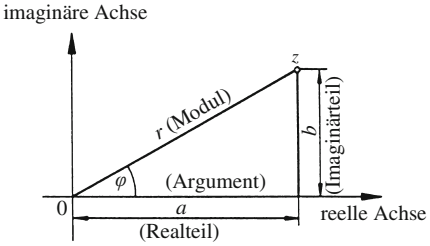


Abb. 1.3 Algebraische und trigonometrische Form einer komplexen Zahl z

Dabei heißt r Modul oder Absolutbetrag (also $r = |z|$) und φ Argument der komplexen Zahl z. Der (orientierte) Winkel φ wird im Bogenmaß (vgl. Abschn. 3.9.5) gemessen und ist nur bis auf Vielfache von 2π bestimmt. Deshalb wählt man meist für φ das halboffene Intervall $[0, 2\pi)$, also $0 \le \varphi < 2\pi$.

$$z = r(\cos\varphi + j\sin\varphi), \quad r \in \mathbb{R}, r \ge 0, 0 \le \varphi < 2\pi$$

Für $\varphi = 0$ ergeben sich die positiven reellen Zahlen, für $\varphi = \pi$ die negativen reellen Zahlen, für $\varphi = \frac{\pi}{2}$ die positiven imaginären Zahlen und für $\varphi = \frac{3}{2}\pi$ die negativen imaginären Zahlen.

Statt trigonometrischer Form sagt man mitunter auch goniometrische Form der komplexen Zahlen.

Für die Darstellung der komplexen Zahlen in der Ebene werden für die trigonometrische Form Polarkoordinaten (siehe Abschn. 7.1.2) verwendet, wohingegen für die algebraische Form kartesische Koordinaten (siehe Abschn. 7.1.1) benutzt werden.

Für den Zusammenhang zwischen algebraischer und trigonometrischer Form gilt

$$r = \sqrt{a^2 + b^2} \quad \tan\varphi = \frac{b}{a}$$
$$a = r\cos\varphi \quad b = r\sin\varphi$$

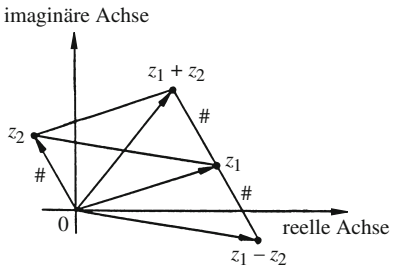


Abb. 1.4 Addition und Subtraktion komplexer Zahlen z_1 und z_2 (die mit # gekennzeichneten Strecken sind parallel und gleich lang)

Derselbe Zusammenhang gilt für die kartesischen Koordinaten und die Polarkoordinaten eines Punktes in der Ebene.

Multiplizieren, Dividieren, Potenzieren und Radizieren komplexer Zahlen lassen sich in der trigonometrischen Form einfacher durchführen.

1.10.3 Addieren und Subtrahieren komplexer Zahlen

Komplexe Zahlen $z_1 = a_1 + b_1 j$ und $z_2 = a_2 + b_2 j$ werden addiert, indem man die Realteile addiert und die Imaginärteile addiert (Abb. 1.4).

$$\begin{aligned} z_1 + z_2 &= (a_1 + b_1 j) + (a_2 + b_2 j) \\ &= (a_1 + a_2) + (b_1 + b_2) j \end{aligned}$$

Komplexe Zahlen $z_1 = a_1 + b_1 j$ und $z_2 = a_2 + b_2 j$ werden voneinander subtrahiert, indem man die Realteile subtrahiert und die Imaginärteile subtrahiert.

$$\begin{aligned} z_1 - z_2 &= (a_1 + b_1 j) - (a_2 + b_2 j) \\ &= (a_1 - a_2) + (b_1 - b_2) j \end{aligned}$$

Die Summe konjugiert komplexer Zahlen $z = a + bj$ und $\bar{z} = a - bj$ ist reell, die Differenz konjugiert komplexer Zahlen ist imaginär.

$$\begin{aligned} z + \bar{z} &= (a + bj) + (a - bj) = 2a \\ z - \bar{z} &= (a + bj) - (a - bj) = 2bj \end{aligned}$$

Beispiele

1. $z_1 + z_2 = (2{,}66 + 0{,}89j) + (-0{,}81 + 1{,}49j)$
   $= 1{,}85 + 2{,}38j$
2. $z_1 - z_2 = (2{,}66 + 0{,}89j) - (-0{,}81 + 1{,}49j)$
   $= 3{,}47 - 0{,}60j$
3. $z + \bar{z} = (2{,}4 + 0{,}9j) + (2{,}4 - 0{,}9j) = 4{,}8$
4. $z - \bar{z} = (2{,}4 + 0{,}9j) - (2{,}4 - 0{,}9j) = 1{,}8j.$ ◀

1.10.4 Multiplizieren komplexer Zahlen

Komplexe Zahlen $z_1 = a_1 + b_1 j$ und $z_2 = a_2 + b_2 j$ in algebraischer Form werden wie algebraische Summen multipliziert (denn $z_1 \cdot z_2 = (a_1 + b_1 j)(a_2 + b_2 j) = a_1 a_2 + a_1 b_2 j + b_1 a_2 j + b_1 b_2 j^2 = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) j$ wegen $j^2 = -1$).

$$\begin{aligned} z_1 \cdot z_2 &= (a_1 + b_1 j)(a_2 + b_2 j) \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) j \end{aligned}$$

Das Produkt konjugiert komplexer Zahlen ist reell.

$$z \cdot \bar{z} = (a + bj)(a - bj) = a^2 + b^2$$

Beispiele

1. $z_1 \cdot z_2 = (3 + 4j)(5 - 2j)$
   $= (3 \cdot 5 - 4 \cdot (-2)) + (3 \cdot (-2) + 5 \cdot 4) j = 23 + 14j$
2. $z \cdot \bar{z} = (2{,}4 + 0{,}9j)(2{,}4 - 0{,}9j) = (2{,}4)^2 + (0{,}9)^2$
   $= 5{,}76 + 0{,}81 = 6{,}57.$ ◀

Komplexe Zahlen $z_1 = r_1(\cos\varphi_1 + j\sin\varphi_1)$ und $z_2 = r_2(\cos\varphi_2 + j\sin\varphi_2)$ in trigonometrischer Form werden multipliziert, indem man die Moduln (r_1 und r_2) multipliziert und die Argumente (φ_1 und φ_2) addiert.

$$\begin{aligned} z_1 \cdot z_2 &= r_1(\cos\varphi_1 + j\sin\varphi_1) \cdot r_2(\cos\varphi_2 + j\sin\varphi_2) \\ &= r_1 r_2[cos(\varphi_1 + \varphi_2) + j\sin(\varphi_1 + \varphi_2)] \end{aligned}$$

Beweis:

$$\begin{aligned} z_1 \cdot z_2 &= r_1(\cos\varphi_1 + j\sin\varphi_1) \cdot r_2(\cos\varphi_2 + j\sin\varphi_2) \\ &= r_1 r_2[(\cos\varphi_1 \cos\varphi_2 - \sin\varphi_1 \sin\varphi_2) \\ &\quad + (\cos\varphi_1 \sin\varphi_2 + \sin\varphi_1 \cos\varphi_2) j] \\ &= r_1 r_2[\cos(\varphi_1 + \varphi_2) + j\sin(\varphi_1 + \varphi_2)] \end{aligned}$$

denn $\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2 = \cos(\varphi_1+\varphi_2)$ und $\cos\varphi_1\sin\varphi_2 + \sin\varphi_1\cos\varphi_2 = \sin(\varphi_1+\varphi_2)$.

Beispiele

1. $z_1 = 3(\cos 20° + j\sin 20°), z_2 = 7(\cos 65° + j\sin 65°)$ $\Rightarrow z_1 \cdot z_2 = 3(\cos 20° + j\sin 20°)\cdot 7(\cos 65° + j\sin 65°) = 21(\cos 85° + j\sin 85°)$
2. $z_1 = 5(\cos 30° + j\sin 30°) = \frac{5}{2}\sqrt{3} + \frac{5}{2}j$; $z_2 = 13(\cos 60° + j\sin 60°) = \frac{13}{2} + \frac{13}{2}\sqrt{3}j$ (denn $\sin 30° = \cos 60° = \frac{1}{2}$ und $\sin 60° = \cos 30° = \frac{1}{2}\sqrt{3}$). Es folgt $z_1\cdot z_2 = 5(\cos 30° + j\sin 30°)\cdot 13(\cos 60° + j\sin 60°) = 65(\cos 90° + j\sin 90°) = 65j$ oder $z_1\cdot z_2 = \left(\frac{5}{2}\sqrt{3} + \frac{5}{2}j\right)\left(\frac{13}{2} + \frac{13}{2}\sqrt{3}j\right)$ $= \frac{65}{4}\sqrt{3}\frac{65}{4}\sqrt{3} + \left(\frac{65\cdot 3}{4} + \frac{65}{4}\right)j = 65j.$ ◀

1.10.5 Dividieren komplexer Zahlen

Komplexe Zahlen $z_1 = a_1 + b_1 j$ und $z_2 = a_2 + b_2 j$ in algebraischer Form werden dividiert, indem man mit der konjugiert komplexen Zahl des Nenners (Divisors) erweitert.

$$\frac{z_1}{z_2} = \frac{a_1 + b_1 j}{a_2 + b_2 j} = \frac{a_1a_2 + b_1b_2}{a_2^2 + b_2^2} + \frac{b_1a_2 - a_1b_2}{a_2^2 + b_2^2} j \quad (z_2 \neq 0)$$

Beweis:

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{a_1 + b_1 j}{a_2 + b_2 j} = \frac{(a_1 + b_1 j)(a_2 - b_2 j)}{(a_2 + b_2 j)(a_2 - b_2 j)}\\ &= \frac{a_1a_2 + b_1b_2 + (b_1a_2 - a_1b_2)j}{a_2^2 + b_2^2}\\ &= \frac{a_1a_2 + b_1b_2}{a_2^2 + b_2^2} + \frac{b_1a_2 - a_1b_2}{a_2^2 + b_2^2} j\end{aligned}$$

Der Quotient konjugiert komplexer Zahlen ist wieder eine komplexe Zahl.

$$\frac{z}{\bar{z}} = \frac{a + bj}{a - bj} = \frac{a^2 - b^2}{a^2 + b^2} + \frac{2ab}{a^2 + b^2} j \quad (z \neq 0)$$

Beispiele

1. $\frac{z_1}{z_2} = \frac{3+4j}{5-2j} = \frac{3\cdot 5 + 4\cdot(-2)}{5^2+(-2)^2} + \frac{4\cdot 5 - 3\cdot(-2)}{5^2+(-2)^2} j = \frac{15-8}{25+4} + \frac{20+6}{25+4} j$ $= \frac{7}{29} + \frac{26}{29} j$
2. $\frac{z}{\bar{z}} = \frac{2{,}4+0{,}9j}{2{,}4-0{,}9j} = \frac{(2{,}4)^2-(0{,}9)^2}{(2{,}4)^2+(0{,}9)^2} + \frac{2\cdot 2{,}4\cdot 0{,}9}{(2{,}4)^2+(0{,}9)^2} j$ $= \frac{5{,}76-0{,}81}{5{,}76+0{,}81} + \frac{4{,}32}{5{,}76+0{,}81} j = \frac{4{,}95}{6{,}57} + \frac{4{,}32}{6{,}57} j$
3. $\frac{1}{j} = \frac{1\cdot(-j)}{j\cdot(-j)} = -j.$ ◀

Komplexe Zahlen $z_1 = r_1(\cos\varphi_1 + j\sin\varphi_1)$ und $z_2 = r_2(\cos\varphi_2 + j\sin\varphi_2)$ in trigonometrischer Form werden dividiert, indem man die Moduln (r_1 und r_2) dividiert und die Argumente (φ_1 und φ_2) subtrahiert.

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{r_1(\cos\varphi_1 + j\sin\varphi_1)}{r_2(\cos\varphi_2 + j\sin\varphi_2)}\\ &= \frac{r_1}{r_2}[\cos(\varphi_1 - \varphi_2) + j\sin(\varphi_1 - \varphi_2)]\end{aligned}$$

Beweis:

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{r_1(\cos\varphi_1 + j\sin\varphi_1)}{r_2(\cos\varphi_2 + j\sin\varphi_2)}\\ &= \frac{r_1}{r_2}\frac{(\cos\varphi_1 + j\sin\varphi_1)(\cos\varphi_2 - j\sin\varphi_2)}{(\cos\varphi_2 + j\sin\varphi_2)(\cos\varphi_2 - j\sin\varphi_2)}\\ &= \frac{r_1}{r_2}\left(\frac{\cos\varphi_1\cos\varphi_2 + \sin\varphi_1\sin\varphi_2}{\sin^2\varphi_2 + \cos^2\varphi_2} + \frac{\sin\varphi_1\cos\varphi_2 - \cos\varphi_1\sin\varphi_2}{\sin^2\varphi_2 + \cos^2\varphi_2} j\right)\\ &= \frac{r_1}{r_2}[\cos(\varphi_1\varphi_2) + j\sin(\varphi_1 - \varphi_2)]\end{aligned}$$

denn $\cos\varphi_1\cos\varphi_2 + \sin\varphi_1\sin\varphi_2 = \cos(\varphi_1 - \varphi_2)$, $\sin\varphi_1\cos\varphi_2 - \cos\varphi_1\sin\varphi_2 = \sin(\varphi_1 - \varphi_2)$ und $\sin^2\varphi + \cos^2\varphi = 1$.

Beispiele

4. $z_1 = 3(\cos 20° + j\sin 20°), z_2 = 7(\cos 65° + j\sin 65°)$ $\Rightarrow \frac{z_2}{z_1} = \frac{7(\cos 65° + j\sin 65°)}{3(\cos 20° + j\sin 20°)} = \frac{7}{3}(\cos 45° + j\sin 45°)$
5. $z_1 = 5(\cos 30° + j\sin 30°) = \frac{5}{2}\sqrt{3} + \frac{5}{2}j,$ $z_2 = 13(\cos 60° + j\sin 60°) = \frac{13}{2} + \frac{13}{2}\sqrt{3}j$ Es folgt

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{5(\cos 30° + j\sin 30°)}{13(\cos 60° + j\sin 60°)}\\ &= \frac{5}{13}(\cos(-30°) + j\sin(-30°))\\ &= \frac{5}{13}(\cos 30° - j\sin 30°)\\ &= \frac{5}{13}\left(\frac{1}{2}\sqrt{3} - \frac{1}{2}j\right) = \frac{5}{26}\sqrt{3} - \frac{5}{26}j\end{aligned}$$

oder

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{\frac{5}{2}\sqrt{3} + \frac{5}{2}j}{\frac{13}{2} + \frac{13}{2}\sqrt{3}j} = \frac{(\frac{5}{2}\sqrt{3} + \frac{5}{2}j)(\frac{13}{2} - \frac{13}{2}\sqrt{3}j)}{(\frac{13}{2} + \frac{13}{2}\sqrt{3}j)(\frac{13}{2} - \frac{13}{2}\sqrt{3}j)}\\ &= \frac{\frac{65}{4}\sqrt{3} - \frac{65}{4}\cdot 3j + \frac{65}{4}j + \frac{65}{4}\sqrt{3}}{169} = \frac{5}{26}\sqrt{3} - \frac{5}{26}j\end{aligned}$$

◀

1.10.6 Potenzieren komplexer Zahlen

Ist n eine natürliche Zahl, so wird die n-te Potenz z^n von z wie üblich durch $z^0 = 1$, $z^n = z^{n-1} \cdot z$ definiert.

Beispiele

1. $z^3 = (a+bj)^2(a+bj) = a^3 - 3ab^2 + (3a^2b - b^3)j$
2. $z^4 = (a+bj)^3(a+bj)$
   $= [a^3 - 3ab^2 + (3a^2b - b^3)j](a+bj)$
   $= a^4 - 6a^2b^2 + b^4 + (4a^3b - 4ab^3)j.$ ◀

Einfacher lässt sich das Potenzieren komplexer Zahlen in der trigonometrischen Form durchführen. Mit Hilfe der Additionstheoreme für die trigonometrischen Funktionen erhält man die Formel von Moivre.

$$z^n = [r(\cos\varphi + j\sin\varphi)]^n$$
$$= r^n(\cos n\varphi + j\sin n\varphi) \quad (n \in \mathbb{N}^*)$$

Eine komplexe Zahl in trigonometrischer Form wird also in die n-te Potenz erhoben, indem man den Modul (r) in die entsprechende Potenz r^n erhebt und das Argument (φ) mit dem Exponenten n multipliziert.

Beispiel

3. $z = 5(\cos 30° + j\sin 30°) = \frac{5}{2}\sqrt{3} + \frac{5}{2}j$

$$z^4 = \left[\frac{5}{2}\sqrt{3} + \frac{5}{2}j\right]^4$$
$$= \left(\frac{5}{2}\sqrt{3}\right)^4 - 6\left(\frac{5}{2}\sqrt{3}\right)^2\left(\frac{5}{2}\right)^2 + \left(\frac{5}{2}\right)^4$$
$$+ \left[4\left(\frac{5}{2}\sqrt{3}\right)^3 \cdot \frac{5}{2} - 4\frac{5}{2}\sqrt{3}\left(\frac{5}{2}\right)^3\right]j$$
$$= \frac{625\cdot 9}{16} - \frac{6\cdot 25\cdot 3\cdot 25}{4\cdot 4} + \frac{625}{16}$$
$$+ \left[\frac{4\cdot 125\cdot 3\cdot\sqrt{3}\cdot 5}{8\cdot 2} - \frac{4\cdot 5\cdot\sqrt{3}\cdot 125}{2\cdot 8}\right]j$$
$$= -\frac{625}{2} + \frac{625\cdot\sqrt{3}}{2}j$$

$$z^4 = [5(\cos 30° + j\sin 30°)]^4$$
$$= 5^4(\cos 120° + j\sin 120°)$$
$$= 5^4(-\sin 30° + j\cos 30°)$$
$$= 5^4\left(-\frac{1}{2} + \frac{1}{2}\sqrt{3}j\right) = -\frac{625}{2} + \frac{625\cdot\sqrt{3}}{2}j$$ ◀

Die Moivre'sche Formel lässt sich durch vollständige Induktion beweisen. Ihre Gültigkeit lässt sich schrittweise bis auf reelle Exponenten ausdehnen.

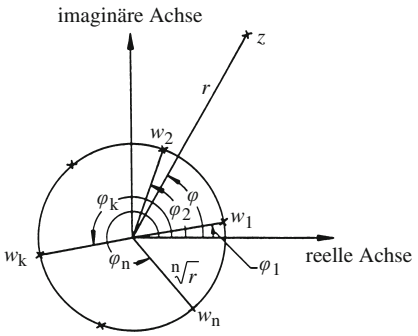


Abb. 1.5 Die n-ten Wurzeln $w_1, w_2, \ldots, w_k, \ldots, w_n$ einer komplexen Zahl z

1.10.7 Radizieren komplexer Zahlen

Die n-te Wurzel $\sqrt[n]{z}$ einer komplexen Zahl z ist definiert als eine komplexe Zahl w, deren n-te Potenz gleich z ist, also eine Lösung der Gleichung $w^n = z$.

Setzt man $z = r(\cos\varphi + j\sin\varphi)$ und $w = \varrho(\cos\psi + j\sin\psi)$, dann folgt mit der Formel von Moivre $w^n = \varrho^n(\cos n\psi + j\sin n\psi)$ und wegen $w^n = z = r(\cos\varphi + j\sin\varphi)$ weiter $\varrho^n = r$; $\cos n\psi = \cos\varphi$, $\sin n\psi = \sin\varphi$. Aus $\varrho^n = r$ ergibt sich $\varrho = \sqrt[n]{r}$, während es für $\cos n\psi = \cos\varphi$, $\sin n\psi = \sin\varphi$ wegen $\cos\varphi = \cos(\varphi + 2k\pi)$, $\sin\varphi = \sin(\varphi + 2k\pi)$ genau n verschiedene Lösungen $\psi_k = \frac{\varphi + 2(k-1)\pi}{n}$, $k = 1, 2, 3, \ldots, n$, gibt.

Somit gilt:

Für $n \in \mathbb{N}^*$ besitzt die Gleichung $w^n = z = r(\cos\varphi + j\sin\varphi)$ genau n verschiedene Lösungen $w_1, w_2, \ldots, w_n$ (die n-ten Wurzeln aus z).

$$w_k = \sqrt[n]{r}\left(\cos\frac{\varphi + 2(k-1)\pi}{n} + j\sin\frac{\varphi + 2(k-1)\pi}{n}\right)$$
$$k = 1, 2, \ldots, n$$

Die n-te Wurzel aus z ist also nicht eindeutig. Für $k = 1$ ergibt sich der sogenannte Hauptwert w_1 der n-ten Wurzel.

Hauptwert

$$w_1 = \sqrt[n]{r}\left(\cos\frac{\varphi}{n} + j\sin\frac{\varphi}{n}\right)$$

Stellt man die n-ten Wurzeln w_k, $k = 1, 2, 3, \ldots, n$ in der Gauß'schen Zahlenebene dar, so ergeben sich die Eckpunkte eines regelmäßigen n-Ecks mit dem Mittelpunkt im Koordinatenursprung. Die Punkte liegen auf einem Kreis mit dem Radius $\varrho = \sqrt[n]{r}$. Der Hauptwert w_1 besitzt das Argument $\frac{\varphi}{n}$. Durch wiederholte Drehung um den Winkel $\frac{2\pi}{n}$ erhält man die weiteren Lösungen (Abb. 1.5).

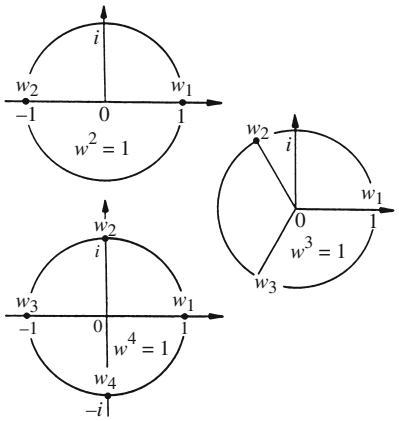


Abb. 1.6 Die n-ten Einheitswurzeln für $n = 2$, $n = 3$ und $n = 4$

Beispiel

$z = 2{,}985984(\cos 60° + j \sin 60°) = (1{,}2)^6(\cos 60° + j \sin 60°)$, $n = 6$

Wegen $\sqrt[n]{r} = \sqrt[6]{(1{,}2)^6} = 1{,}2$ und $\frac{\varphi}{n} = \frac{60°}{6} = 10°$ lauten die sechsten Wurzeln aus z:

$$w_1 = 1{,}2(\cos 10° + j \sin 10°),$$
$$w_2 = 1{,}2(\cos 70° + j \sin 70°)$$
$$w_3 = 1{,}2(\cos 130° + j \sin 130°),$$
$$w_4 = 1{,}2(\cos 190° + j \sin 190°)$$
$$w_5 = 1{,}2(\cos 250° + j \sin 250°),$$
$$w_6 = 1{,}2(\cos 310° + j \sin 310°)$$ ◀

Die n-ten Wurzeln aus $z = 1$ sind die sogenannten n-ten Einheitswurzeln (Abb. 1.6).

***n*-te Einheitswurzeln**

Lösungen von $w^n = z = 1$

Beispiele

1. $n = 2$: $z = w^2 = 1$
   $w_1 = 1(\cos 0° + j \sin 0°) = 1$;
   $w_2 = 1(\cos 180° + j \sin 180°) = -1$
2. $n = 3$: $z = w^3 = 1$
   $w_1 = 1(\cos 0° + j \sin 0°) = 1$
   $w_2 = 1(\cos 120° + j \sin 120°)$
   $= 1(-\cos 60° + j \sin 60°) = -\frac{1}{2} + \frac{1}{2}\sqrt{3}j$
   $w_3 = 1(\cos 240° + j \sin 240°)$
   $= 1(-\cos 60° - j \sin 60°) = -\frac{1}{2} - \frac{1}{2}\sqrt{3}j$
3. $n = 4$: $z = w^4 = 1$
   $w_1 = 1(\cos 0° + j \sin 0°) = 1$
   $w_2 = 1(\cos 90° + j \sin 90°) = j$
   $w_3 = 1(\cos 180° + j \sin 180°) = -1$
   $w_4 = 1(\cos 270° + j \sin 270°) = -j$ ◀

1.10.8 Euler'sche Formel

Die Euler'sche Formel für komplexe Zahlen z verknüpft die Exponentialfunktion und die trigonometrischen Funktionen miteinander (nach dem schweizerischen Mathematiker Leonhard Euler, 1707–1783). Dabei ist e die Euler'sche Zahl (vgl. Abschn. 1.7).

$$e^{jz} = \cos z + j \sin z, \quad z \in \mathbb{C}$$

Für reelle Zahlen x (die reellen Zahlen sind eine Teilmenge der komplexen Zahlen) gilt $e^{jx} = \cos x + j \sin x$.

Setzt man $x = \varphi$, dann erhält man die sogenannte Exponentialform der komplexen Zahlen.

$$z = r(\cos \varphi + j \sin \varphi) = re^{j\varphi}$$

Dabei ist r der Modul und φ das Argument der komplexen Zahl z.

Für das Produkt und den Quotienten zweier komplexer Zahlen $z_1 = r_1 \cdot e^{j\varphi_1}$ und $z_2 = r_2 \cdot e^{j\varphi_2}$ ergibt sich

$$z_1 \cdot z_2 = r_1 \cdot e^{j\varphi_1} \cdot r_2 \cdot e^{j\varphi_2} = r_1 \cdot r_2 \cdot e^{j(\varphi_1 + \varphi_2)}$$
$$\frac{z_1}{z_2} = \frac{r_1}{r_2} \frac{e^{j\varphi_1}}{e^{j\varphi_2}} = \frac{r_1}{r_2} e^{j(\varphi_1 - \varphi_2)} \quad (z_2 \neq 0)$$

Beispiel für eine komplexe Zahl in verschiedenen Formen

$$z = 2\left(\frac{1}{2} + \frac{\sqrt{3}}{2}j\right)$$
$$= 1 + \sqrt{3}j \qquad \text{(algebraische Form)}$$
$$= 2\left(\cos\frac{\pi}{3} + j \sin\frac{\pi}{3}\right) \qquad \text{(trigonometrische Form)}$$
$$= 2e^{j\frac{\pi}{3}} \qquad \text{(Exponentialform)}$$ ◀

2 Gleichungen

Arnfried Kemnitz



2.1 Gleichungsarten 30
2.2 Äquivalente Umformungen 31
2.3 Lineare Gleichungen 32
2.4 Proportionen 33
2.5 Quadratische Gleichungen 34
2.6 Algebraische Gleichungen höheren Grades 37
2.7 Auf algebraische Gleichungen zurückführbare Gleichungen 43
2.8 Transzendente Gleichungen 45
2.9 Lineare Gleichungssysteme 47

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_2

2.1 Gleichungsarten

Ein Term ist ein mathematischer Ausdruck, der aus Zahlen, Variablen, Rechenzeichen (mathematischen Operationen) und möglicherweise noch anderen mathematischen Symbolen (zum Beispiel Funktionswerten) besteht.

Will man ausdrücken, dass ein Term T_1 zu einem anderen Term T_2 äquivalent (gleichwertig) ist, so schreibt man

Gleichung

$$T_1 = T_2$$

Eine solche Darstellung heißt Gleichung. Die linke Seite der Gleichung ist T_1, die rechte Seite der Gleichung ist T_2.

Mit Hilfe von Gleichungen lassen sich quantitative Beziehungen in Natur und Technik beschreiben. Meist liegen jedoch in der Praxis auftretende Aufgaben nicht in Form von Gleichungen zwischen Termen vor, sondern sie werden als Textgleichungen mit Worten beschrieben. Daraus muss dann durch eine Übersetzung in die formale Sprache der Mathematik eine mathematische Beziehung hergestellt werden.

Die Überlegenheit der mathematischen Symbolik zeigt folgendes Beispiel:

Textgleichung: Den Umfang eines Kreises berechnet man, indem man das Produkt aus dem Verhältnis von Umfang eines beliebigen Kreises zu seinem Durchmesser ($\frac{U}{d} = \pi$) und dem Kreisradius (r) mit 2 multipliziert (vgl. Abschn. 3.9).

$$\text{Termgleichung:} \quad u_{\text{Kreis}} = 2\pi r$$

Man unterscheidet drei verschiedene Arten von Gleichungen:

1. Identische Gleichungen
2. Bestimmungsgleichungen
3. Funktionsgleichungen

Eine identische Gleichung oder Identität ist eine Gleichung zwischen zwei algebraischen Ausdrücken, die bei Einsetzen beliebiger Zahlenwerte anstelle der darin aufgeführten Buchstabensymbole erhalten bleibt.

Beispiele für identische Gleichungen (Identitäten)

1. $a(b+c) = ab + ac$
2. $(a+b)(c+d) = ac + ad + bc + bd$
3. $(a+b)^2 = a^2 + 2ab + b^2$
4. $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$
5. $a^n a^m = a^{n+m}$
6. $\sqrt[n]{c}\sqrt[n]{d} = \sqrt[n]{cd}$
7. $\log_a(xy) = \log_a x + \log_a y$
8. $e^{jx} = \cos x + j \sin x$ ◀

Eine Bestimmungsgleichung ist eine Gleichung, in der Variable (Unbekannte) auftreten, die durch eine Rechnung bestimmt werden sollen. Mit Hilfe zulässiger Rechenoperationen sollen alle Werte der Variablen aus dem zugrunde liegenden Zahlenbereich bestimmt werden, für die die Gleichung erfüllt ist. Man nennt diese Werte Lösungen oder auch Wurzeln der Gleichung. Alle Lösungen zusammen bilden die Lösungsmenge L der Bestimmungsgleichung. Eine Gleichung hat keine, eine oder mehrere Lösungen.

Eine Bestimmungsgleichung ist also nur für einige spezielle Werte der Variablen erfüllt.

Beispiele für Bestimmungsgleichungen

1. $x + 2 = 3$
   Lösung: $x = 1$
   Lösungsmenge: $L = \{1\}$
2. $x + 2 = x + 3$
   Keine Lösung
   Lösungsmenge: $L = \{\,\} = \emptyset$
3. $2x + 1 = x^2 - 2$
   Lösungen: $x = 3$ und $x = -1$
   Lösungsmenge: $L = \{-1, 3\}$
4. $5x^2 - 5 = x^3 - x$
   Lösungen: $x = 5, x = 1$ und $x = -1$
   Lösungsmenge: $L = \{-1, 1, 5\}$
5. $11 - \sqrt{x+3} = 6$
   Lösung: $x = 22$
   Lösungsmenge: $L = \{22\}$
6. $3^x = 4^{x-2} \cdot 2^x$
   Lösung: $x = \dfrac{4\log 2}{3\log 2 - \log 3} \approx 2{,}826780$
   Lösungsmenge: $L = \left\{\dfrac{4\log 2}{3\log 2 - \log 3}\right\}$
7. $\lg(6x + 10) - \lg(x - 3) = 1$
   Lösung: $x = 10$
   Lösungsmenge: $L = \{10\}$
8. $\sin^2 x - 1 = -0{,}5$
   Lösung: $x = 45° + k \cdot 180°, k \in \mathbb{Z}$
   Lösungsmenge: $L = \{x \,|\, x = 45° + k \cdot 180°, k \in \mathbb{Z}\}$ ◀

Die Bestimmungsgleichungen werden unterteilt in die algebraischen Gleichungen und in die transzendenten Gleichungen.

In einer algebraischen Gleichung werden mit der oder den Variablen nur algebraische Rechenoperationen vorgenommen; sie werden addiert, subtrahiert, multipliziert, dividiert, potenziert oder radiziert. Sowohl die auftretenden Zahlen (Koeffizienten

genannt) als auch die Lösungen können aber transzendente Zahlen sein. Jede algebraische Gleichung mit genau einer Variablen x lässt sich in der allgemeinen Form

$$a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0 = 0$$

schreiben. Die Zahlen $a_n, a_{n-1}, a_{n-2}, \ldots, a_1, a_0$ heißen Koeffizienten (Beizahlen) der Gleichung. Sie stehen für beliebige reelle oder komplexe Zahlen.

Ist x^n die höchste auftretende Potenz der Variablen x, so heißt die Gleichung vom Grad n.

Algebraische Gleichungen vom Grad 1 heißen auch lineare Gleichungen, Gleichungen vom Grad 2 quadratische Gleichungen und Gleichungen vom Grad 3 kubische Gleichungen.

Der sogenannte Fundamentalsatz der Algebra sagt aus, dass jede algebraische Gleichung n-ten Grades genau n (reelle oder komplexe) Lösungen (Wurzeln) besitzt.

Alle Bestimmungsgleichungen, die nicht algebraisch sind, heißen transzendent (deutsch: übersteigend). Sie haben ihren Namen daher, dass sie im Allgemeinen schwieriger aufzulösen sind als die algebraischen Gleichungen. Sie erfordern Auflösungsmethoden, die die Mittel der Algebra übersteigen.

Beispiele für transzendente Gleichungen sind Exponentialgleichungen, logarithmische Gleichungen und trigonometrische Gleichungen.

Bei den ersten fünf Beispielen handelt es sich um algebraische Bestimmungsgleichungen, bei den letzten drei Beispielen um transzendente Gleichungen.

Eine Funktionsgleichung dient dazu, eine Funktion zu definieren. Eine Funktion beschreibt den Zusammenhang zwischen verschiedenen veränderlichen Größen. Eine Funktionsgleichung enthält in der Regel zwei oder mehr Variable, die durch die Gleichung einander zugeordnet werden.

Funktionen werden ausführlich im Kap. 5 behandelt.

Beispiele für Funktionsgleichungen

1. $y = 2x + 1$
2. $y = -x^2 + x - 5$
3. $y = 2x^2 - x - 3\sqrt{x} + 4$
4. $y = \sin x$
5. $y = 2^x - 5x + 1$. ◀

2.2 Äquivalente Umformungen

Oft ist es möglich, eine gegebene Gleichung durch zulässige Rechenoperationen in eine Gleichung zu überführen, die die gleiche Lösungsmenge wie die Ausgangsgleichung besitzt, aber einfacher zu lösen ist. Eine solche Umformung heißt äquivalent. Man nennt auch die beiden Gleichungen äquivalent (gleichwertig).

Bei den zulässigen Rechenoperationen ist darauf zu achten, dass sie gleichzeitig auf beiden Seiten einer Gleichung durchgeführt werden, zum Beispiel die Addition einer Konstanten oder die Multiplikation mit einer Konstanten.

Grundregeln für äquivalente Umformungen:

Addition einer Zahl (hier a) auf beiden Seiten einer Gleichung

$$\begin{aligned} x - a &= b \quad |+a \\ x &= b + a \end{aligned}$$

Subtraktion einer Zahl (hier a) von beiden Seiten einer Gleichung

$$\begin{aligned} x + a &= b \quad |-a \\ x &= b - a \end{aligned}$$

Multiplikation beider Seiten einer Gleichung mit der gleichen Zahl (hier mit a); Bedingung: $a \neq 0$

$$\begin{aligned} \frac{x}{a} &= b \quad |\cdot a \\ x &= b \cdot a \end{aligned}$$

Division beider Seiten einer Gleichung durch die gleiche Zahl (hier durch a); Bedingung: $a \neq 0$

$$\begin{aligned} ax &= b \quad |:a \\ x &= \frac{b}{a} \end{aligned}$$

Beispiele

1.
$$\begin{aligned} 5x - 6 &= 29 \quad |+6 \quad \text{(Addition auf beiden Seiten)} \\ 5x &= 35 \quad |:5 \quad \text{(Division auf beiden Seiten)} \\ x &= 7 \end{aligned}$$

Alle Gleichungen sind äquivalent mit der Lösungsmenge $L = \{7\}$.

2.
$$\begin{aligned} 5x-20 &= 60-11x && |+11x && \text{(Addition auf beiden Seiten)}\\ 16x-20 &= 60 && |+20 && \text{(Addition auf beiden Seiten)}\\ 16x &= 80 && |:16 && \text{(Division auf beiden Seiten)}\\ x &= 5 \end{aligned}$$

Alle vier Gleichungen sind äquivalent mit der Lösungsmenge $L = \{5\}$.

3.
$$\frac{2}{x} - \frac{1}{x-1} = \frac{1}{x+2} \quad |\cdot x(x-1)(x+2) \quad (x \neq 0, 1, -2)$$
$$2(x-1)(x+2) - x(x+2) = x(x-1) \quad \text{(Hauptnenner)}$$
$$\begin{aligned} 2(x^2+x-2) - (x^2+2x) &= x^2 - x\\ x^2 - 4 &= x^2 - x && |-x^2+x\\ x-4 &= 0 && |+4\\ x &= 4 \end{aligned}$$

Alle Gleichungen sind äquivalent mit der Lösungsmenge $L = \{4\}$.

4.
$$\begin{aligned} 4x^2 - x + 3 &= 6x^2 + x - 1 && |-(4x^2 - x + 3)\\ 0 &= 6x^2 + x - 1 - 4x^2 + x - 3\\ 0 &= 2x^2 + 2x - 4 && |:2\\ 0 &= x^2 + x - 2 && |\text{ Vertauschen der Seiten}\\ x^2 + x - 2 &= 0 \end{aligned}$$

Berechnen der Lösungen der quadratischen Gleichung:
$$x_{1,2} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + 2} = -\frac{1}{2} \pm \frac{3}{2}$$
$$\Rightarrow \quad x_1 = 1, x_2 = -2$$

Alle Gleichungen sind äquivalent mit der Lösungsmenge $L = \{1, -2\}$.

5.
$$\begin{aligned} \sqrt{x+8} &= x+2 && |\text{ Quadrieren}\\ x+8 &= (x+2)^2 && |\text{ Klammer „beseitigen“}\\ x+8 &= x^2+4x+4 && |-(x+8)\\ 0 &= x^2+3x-4 && |\text{ Vertauschen der Seiten}\\ x^2+3x-4 &= 0 \end{aligned}$$

Berechnen der Lösungen der quadratischen Gleichung:
$$x_{1,2} = -\frac{3}{2} \pm \sqrt{\frac{9}{4} + 4} = -\frac{3}{2} \pm \frac{5}{2}$$
$$\Rightarrow \quad x_1 = 1, x_2 = -4$$

$x_1 = 1$ erfüllt die Ausgangsgleichung wegen $\sqrt{1+8} = 1+2$, dagegen ist $x_2 = -4$ keine Lösung der Ausgangsgleichung, denn es ist $\sqrt{-4+8} \neq -4+2$.
Somit: Das Quadrieren ist keine äquivalente Umformung! ◀

2.3 Lineare Gleichungen

Eine lineare Gleichung oder Gleichung ersten Grades ist eine algebraische Gleichung, in der die Variable x in keiner höheren als der ersten Potenz vorkommt.

Jede lineare Gleichung lässt sich durch äquivalente Umformungen überführen in die äquivalente Gleichung

$$ax + b = 0, \quad a \neq 0$$

Diese Gleichung heißt allgemeine Form der linearen Gleichung. Durch Division durch $a \neq 0$ erhält man die sogenannte Normalform der linearen Gleichung.

$$x + \frac{b}{a} = x + c = 0, \quad c = \frac{b}{a}$$

Die Lösung der linearen Gleichung ist $x = -c = -\frac{b}{a}$. Für die Lösungsmenge gilt also: $L = \{-c\} = \{-\frac{b}{a}\}$.

Allgemeines Verfahren zur Bestimmung der Lösung: Man „beseitigt“ zunächst alle Klammern und Brüche und ordnet dann die Glieder so, dass alle Glieder mit der Variablen x links vom Gleichheitszeichen und alle anderen rechts davon stehen:

$$\begin{aligned} a(bx+c) &= d(ex+f)\\ abx + ac &= dex + df\\ abx - dex &= df - ac\\ x(ab-de) &= df - ac\\ x &= \frac{df-ac}{ab-de} \end{aligned}$$

($ab \neq de$ ist Bedingung, denn durch 0 darf nicht dividiert werden.)

Beispiel

$$\begin{aligned} 3(x+2) &= 5(-2x+9)\\ 3x+6 &= -10x+45\\ 3x+10x &= 45-6\\ 13x &= 39\\ x &= 3 \end{aligned}$$

Probe:
$$\begin{aligned} 3(3+2) &= 5(-6+9)\\ 3\cdot 5 &= 5\cdot 3\\ 15 &= 15 \end{aligned}$$
◀

Grundsätzlich sollte eine Probe durchgeführt werden. Dabei ist jede Seite der Gleichung einzeln auszurechnen. Der berechnete Wert für x sollte stets in die Ausgangsgleichung eingesetzt werden.

Fehlerwarnung: Nach Einsetzen der Lösung sollen nicht die gleichen Umformungen wie bei der Hauptrechnung vorgenommen werden, da sonst leicht ein möglicher Fehler wiederholt werden kann.

2.4 Proportionen

Eine Sonderstellung unter den linearen Gleichungen mit einer Variablen nehmen die Proportionen wegen ihrer vielseitigen Anwendbarkeit ein.

Eine Proportion ist eine Verhältnisgleichung

$$a : b = c : x$$

oder mit $x = d$.

$$a : b = c : d$$

und in Bruchschreibweise

$$\frac{a}{b} = \frac{c}{d}$$

(gesprochen: a verhält sich zu b wie c zu d).

Treten in einer Proportion gleiche Innenglieder oder gleiche Außenglieder auf, so heißt die Proportion stetig. Im Fall gleicher Innenglieder, also $a : b = b : c$, nennt man b mittlere Proportionale.

Sind von den Gliedern einer Proportion drei bekannt, dann lässt sich die vierte Proportionale berechnen. Sind zum Beispiel a, b, c bekannt und d gesucht, so gilt $d = \frac{bc}{a}$.

Beispiele

1. Welche Kraft F dehnt eine Feder um 4 cm, wenn die Kraft 3 N (Newton) eine Dehnung um 2 cm bewirkt?
   Ansatz (*Hooke'sches Gesetz*): $F : 4\,\text{cm} = 3\,\text{N} : 2\,\text{cm}$
   Auflösung nach F: $F = \frac{4\,\text{cm}\cdot 3\,\text{N}}{2\,\text{cm}} = 6\,\text{N}$
   Antwort: Die Kraft 6 N bewirkt die Dehnung um 4 cm.
2. Wie weit kommt ein Flugzeug in $2\frac{1}{2}$ Stunden, wenn es 10 km in 45 s zurücklegt?
   Ansatz (*Bewegung mit konstanter Geschwindigkeit*):
   $x : 9000\,\text{s} = 10\,\text{km} : 45\,\text{s}$
   Auflösung nach x: $x = \frac{9000\,\text{s}\cdot 10\,\text{km}}{45\,\text{s}} = 2000\,\text{km}$
   Antwort: Das Flugzeug fliegt 2000 km weit. ◄

Die Proportion $a : b = c : d$ lässt sich verschieden umformen. Äquivalente Formen sind

$$d : c = b : a$$
$$a : c = b : d$$
$$d : b = c : a$$
$$b : a = d : c$$
$$a \cdot d = b \cdot c$$

Aus der Proportion $a : b = c : d$ lassen sich weitere Proportionen ableiten, etwa durch Addition oder Subtraktion von 1 auf beiden Seiten. Man nennt ein solches Umformen der Proportion korrespondierende Addition oder korrespondierende Subtraktion (Bedingung in allen Fällen: Nenner ungleich 0).

$$\frac{a+b}{b} = \frac{c+d}{d} \qquad \frac{a}{a+b} = \frac{c}{c+d}$$
$$\frac{a-b}{b} = \frac{c-d}{d} \qquad \frac{a}{a-b} = \frac{c}{c-d}$$

Dies sind Sonderfälle des allgemeinen Gesetzes der korrespondierenden Addition und Subtraktion. Aus $a : b = c : d$ folgt für beliebige reelle Zahlen p, q, r, s (r und s dürfen nicht gleichzeitig 0 sein)

$$\frac{pa+qb}{ra+sb} = \frac{pc+qd}{rc+sd}$$

Beispiele

3. $$\frac{5}{4} = \frac{5+x}{x}$$

   Nenner vom Zähler subtrahieren (korrespondierende Subtraktion):

   $$\frac{5-4}{4} = \frac{5+x-x}{x}$$

   Vereinfachen:

   $$\frac{1}{4} = \frac{5}{x}$$

   Nach x auflösen (Multiplikation der Gleichung mit $4x$):

   $$x = 20$$

   Die Lösung ist $x = 20$, wie die Probe bestätigt.

4.

$$5x : (4 - x) = 30 : 9$$

In Bruchschreibweise:

$$\frac{5x}{4-x} = \frac{30}{9}$$

Addition von $\frac{1}{5}$ des Zählers zum Nenner (korrespondierende Addition):

$$\frac{5x}{4 - x + \frac{5x}{5}} = \frac{30}{9 + \frac{30}{5}}$$

Vereinfachen:

$$\frac{5x}{4} = \frac{30}{15} \; (= 2)$$

Multiplikation mit $\frac{4}{5}$:

$$x = \frac{8}{5}$$

Die Lösung ist $x = \frac{8}{5}$.
Probe:

$$\begin{aligned} & \frac{5 \cdot \frac{8}{5}}{4 - \frac{8}{5}} = \frac{30}{9} \\ \Leftrightarrow \quad & \frac{8}{2\frac{2}{5}} = \frac{10}{3} \\ \Leftrightarrow \quad & 8 \cdot \frac{5}{12} = \frac{10}{3} \\ \Leftrightarrow \quad & \frac{10}{3} = \frac{10}{3} \end{aligned}$$

◀

2.5 Quadratische Gleichungen

2.5.1 Definitionen

Eine quadratische Gleichung oder Gleichung zweiten Grades ist eine algebraische Gleichung, in der die Variable x in keiner höheren als der zweiten Potenz vorkommt.

Jede quadratische Gleichung lässt sich durch äquivalente Umformungen überführen in die Gleichung

Allgemeine Form

$$ax^2 + bx + c = 0, \quad a \neq 0$$

Diese Gleichung heißt allgemeine Form der quadratischen Gleichung.

Durch Division durch $a \neq 0$ erhält man die sogenannte Normalform der quadratischen Gleichung (mit $p = \frac{b}{a}$, $q = \frac{c}{a}$).

Normalform

$$x^2 + px + q = 0$$

2.5.2 Lösungsverfahren

2.5.2.1 Normalform

Die Lösungen der Normalform $x^2 + px + q = 0$ der quadratischen Gleichung bestimmt man mit der Methode der „quadratischen Ergänzung“. Zunächst bringt man q auf die rechte Seite der Gleichung, das heißt, von beiden Seiten der Gleichung wird q subtrahiert. Auf beiden Seiten addiert man dann die quadratische Ergänzung $(\frac{p}{2})^2$ des Terms $x^2 + px$. Damit wird die linke Seite der Gleichung zu einem „vollständigen Quadrat“ (binomische Formel). Durch Radizieren und anschließender Subtraktion von $\frac{p}{2}$ ergeben sich dann die Lösungen x_1 und x_2 der Gleichung.

Bestimmung der Lösungen:

$$\begin{aligned} x^2 + px + q &= 0 \\ x^2 + px &= -q \\ x^2 + px + \left(\frac{p}{2}\right)^2 &= \left(\frac{p}{2}\right)^2 - q \\ \left(x + \frac{p}{2}\right)^2 &= \left(\frac{p}{2}\right)^2 - q \\ \left(x + \frac{p}{2}\right)^2 &= \frac{p^2}{4} - q \\ x + \frac{p}{2} &= \pm\sqrt{\frac{p^2}{4} - q} \\ x &= -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q} \end{aligned}$$

Daraus ergibt sich

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}, \quad x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

Normalform

Gleichung: $x^2 + px + q = 0$

Lösungen:

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

$$x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

Die Gleichung $x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$ für die Lösungen nennt man auch (p, q)-Formel.

(p, q)-Formel

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

Beispiele

1. (ausführlich)

$$\begin{aligned} x^2 + 12x + 35 &= 0 \\ x^2 + 12x &= -35 \\ x^2 + 12x + 6^2 &= -35 + 6^2 \\ (x+6)^2 &= 1 \\ x + 6 &= \pm 1 \\ x &= -6 \pm 1 \end{aligned}$$

Lösungen: $x_1 = -6 + 1 = -5, x_2 = -6 - 1 = -7$.
Es sind zwei Proben durchzuführen!

2. (mit (p, q)-Formel)

$$25x^2 + 13 = 70x$$

Sortieren und Dividieren durch 25 zum Beschaffen der Normalform:

$$x^2 - \frac{14}{5}x + \frac{13}{25} = 0 \quad \Rightarrow \quad p = -\frac{14}{5} \text{ und } q = \frac{13}{25}$$

Einsetzen in die (p, q)-Formel ergibt die Lösungen:

$$\begin{aligned} x_1 &= \frac{7}{5} + \sqrt{\left(\frac{7}{5}\right)^2 - \frac{13}{25}} = \frac{7}{5} + \sqrt{\frac{36}{25}} \\ &= \frac{7}{5} + \frac{6}{5} = \frac{13}{5}, \\ x_2 &= \frac{7}{5} - \frac{6}{5} = \frac{1}{5} \end{aligned}$$

Es sind zwei Proben durchzuführen! ◀

Den Radikanden in der (p, q)-Formel nennt man die Diskriminante D der Normalform der quadratischen Gleichung.

Diskriminante der Normalform

$$D = \frac{p^2}{4} - q$$

Die Lösungen der Normalform lassen sich auch mit Hilfe der Diskriminante schreiben.

Normalform

Gleichung: $x^2 + px + q = 0$

Lösungen: $x_{1,2} = -\frac{p}{2} \pm \sqrt{D}$

$$D = \frac{p^2}{4} - q$$

Der Wert der Diskriminante D bestimmt die Anzahl der reellen Lösungen der quadratischen Gleichung.

Für $D > 0$ existieren zwei reelle Lösungen x_1 und x_2, für $D = 0$ gibt es eine reelle Lösung (Doppellösung $x_1 = x_2$), für $D < 0$ hat die quadratische Gleichung keine reelle Lösung, es existieren zwei komplexe Lösungen x_1 und x_2 (x_1 und x_2 sind konjugiert komplex zueinander).

Beispiele

3.

$2x^2 - 10x + 12 = 0$ (allgemeineForm)

$x^2 - 5x + 6 = 0$ (Normalform)

$p = -5, q = 6 \quad \Rightarrow \quad D = \frac{p^2}{4} - q = \frac{25}{4} - 6 = \frac{1}{4}$

$D > 0 \quad \Rightarrow$ zwei reelle Lösungen

Lösungen:

$$\begin{aligned} x_{1,2} &= -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q} = -\frac{p}{2} \pm \sqrt{D} \\ &= -\frac{-5}{2} \pm \sqrt{\frac{1}{4}} = \frac{5}{2} \pm \frac{1}{2} \\ \Rightarrow \quad x_1 &= \frac{5}{2} + \frac{1}{2} = 3, \quad x_2 = \frac{5}{2} - \frac{1}{2} = 2 \end{aligned}$$

4.

$-9x^2 + 18x - 9 = 0$ (allgemeine Form)

$x^2 - 2x + 1 = 0$ (Normalform)

$p = -2, q = 1 \quad \Rightarrow \quad D = \frac{p^2}{4} - q = \frac{4}{4} - 1 = 0$

$D = 0 \quad \Rightarrow$ Doppellösung $x_1 = x_2$

Lösung:

$$x_1 = x_2 = -\frac{-2}{2} = 1$$

5.
$$3x^2 - 36x + 120 = 0 \quad \text{(allgemeine Form)}$$
$$x^2 - 12x + 40 = 0 \quad \text{(Normalform)}$$
$$p = -12, q = 40$$
$$\Rightarrow \quad D = \frac{p^2}{4} - q = \frac{(-12)^2}{4} - 40$$
$$= \frac{144}{4} 40 = -\frac{16}{4} = -4$$
$$D < 0 \quad \Rightarrow \quad \text{zwei konjugiert komplexe Lösungen}$$

Lösungen:

$$x_{1,2} = -\frac{-12}{2} \pm \sqrt{-4}$$
$$= 6 \pm \sqrt{-4} = 6 \pm 2\sqrt{-1} = 6 \pm 2j$$
$$\Rightarrow \quad x_1 = 6 + 2j, \quad x_2 = 6 - 2j \quad \blacktriangleleft$$

2.5.2.2 Allgemeine Form

Die Lösungen der allgemeinen Form $ax^2 + bx + c = 0$ erhält man durch Setzen von $p = \frac{b}{a}$, $q = \frac{c}{a}$ in der (p, q)-Formel.

Allgemeine Form

Gleichung: $ax^2 + bx + c = 0, \quad a \neq 0$

Lösungen:
$$x_1 = \frac{1}{2a}(-b + \sqrt{b^2 - 4ac}),$$
$$x_2 = \frac{1}{2a}(-b - \sqrt{b^2 - 4ac})$$

Beispiel

1. $2x^2 - 10x + 12 = 0$
$a = 2, b = -10, c = 12$
Lösungen:

$$x_1 = \frac{1}{4}(-(-10) + \sqrt{(-10)^2 - 4 \cdot 2 \cdot 12})$$
$$= \frac{1}{4}(10 + \sqrt{100 - 96}) = \frac{1}{4}(10 + 2) = 3,$$
$$x_2 = \frac{1}{4}(10 - 2) = 2 \quad \blacktriangleleft$$

Den Radikanden in der Lösungsformel nennt man die Diskriminante $\bar{D}$ der allgemeinen Form der quadratischen Gleichung.

Diskriminante der allgemeinen Form

$$\bar{D} = b^2 - 4ac$$

Die Lösungen der allgemeinen Form lassen sich auch mit Hilfe der Diskriminante schreiben.

Allgemeine Form

Gleichung: $ax^2 + bx + c = 0, \quad a \neq 0$

Lösungen:
$$x_{1,2} = \frac{1}{2a}(-b \pm \sqrt{\bar{D}}),$$
$$\bar{D} = b^2 - 4ac$$

Auch hier bestimmt der Wert der Diskriminante $\bar{D}$ die Anzahl der reellen Lösungen der quadratischen Gleichung.

Für $\bar{D} > 0$ gibt es zwei reelle Lösungen x_1 und x_2, für $\bar{D} = 0$ gibt es eine reelle Doppellösung ($x_1 = x_2$), und für $\bar{D} < 0$ gibt es keine reelle Lösung, sondern zwei konjugiert komplexe Lösungen x_1 und x_2.

Beispiel

2. $3x^2 - 18x + 42 = 0$
$a = 3, b = -18, c = 42 \Rightarrow \bar{D} = (-18)^2 - 4 \cdot 3 \cdot 42 = 324 - 504 = -180 < 0$
$\bar{D} < 0 \Rightarrow$ zwei konjugiert komplexe Lösungen
Lösungen:

$$x_{1,2} = \frac{1}{6}(18 \pm \sqrt{-180})$$
$$= \frac{1}{6}(18 \pm \sqrt{36 \cdot (-5)}) = 3 \pm \sqrt{-5}$$
$$\Rightarrow \quad x_1 = 3 + \sqrt{5}j, \quad x_2 = 3 - \sqrt{5}j \quad \blacktriangleleft$$

2.5.2.3 Zerlegung in Linearfaktoren

Sind x_1 und x_2 die (nicht unbedingt verschiedenen) Lösungen der quadratischen Gleichung $ax^2 + bx + c = 0$, $a \neq 0$, dann kann der quadratische Ausdruck in Linearfaktoren zerlegt werden.

$$ax^2 + bx + c = a(x - x_1)(x - x_2) = 0$$

Die Faktoren $x - x_1$ und $x - x_2$ heißen linear, weil die Variable x nur in erster Potenz, also linear auftritt.

Da ein Produkt genau dann gleich 0 ist, wenn mindestens einer der Faktoren gleich 0 ist, ergeben sich auch hieraus wieder die Lösungen x_1 und x_2.

Man nennt $a(x - x_1)(x - x_2) = 0$ auch Produktform der quadratischen Gleichung.

Beispiele

1. $2x^2 - 6x = 0$
   Zerlegung in Linearfaktoren: $2x^2 - 6x = 2x(x-3) = 0$
   Lösungen: $x_1 = 0, x_2 = 3$
2. $x^2 - x - 6 = 0$
   Zerlegung in Linearfaktoren:
   $x^2 - x - 6 = (x-3)(x+2) = 0$
   Lösungen: $x_1 = 3$ (denn $x - 3 = 0$ für $x = 3$),
   $x_2 = -2$ (denn $x + 2 = 0$ für $x = -2$). ◀

2.5.3 Satz von Viëta für quadratische Gleichungen

Die Produktform der quadratischen Gleichung $x^2 + px + q = 0$ in Normalform lautet $(x - x_1)(x - x_2) = 0$. Ausmultiplizieren und Vergleich ergibt

$$\begin{aligned}(x - x_1)(x - x_2) &= 0\\ x^2 - xx_2 - x_1x + x_1x_2 &= 0\\ x^2 - (x_1 + x_2)x + x_1x_2 &= 0\\ x^2 + px + q &= 0\end{aligned}$$

also die Beziehungen $p = -(x_1 + x_2)$, $q = x_1x_2$. Der Koeffizient p von x ist somit gleich der negativen Summe der beiden Lösungen, das Absolutglied q der quadratischen Gleichung ist gleich dem Produkt der Lösungen. Diese Beziehungen nennt man den Satz von Viëta für quadratische Gleichungen (nach dem französischen Mathematiker François Viëta, 1540–1603).

Satz von Viëta

$$p = -(x_1 + x_2), \quad q = x_1x_2$$

Beispiele

1. Die quadratische Gleichung $x^2 - 5x + 6 = 0$ mit $p = -5$, $q = 6$ hat die Lösungen $x_1 = 3$, $x_2 = 2$. Es gilt:

$$\begin{aligned}p &= -5 = -(3 + 2) = -(x_1 + x_2),\\ q &= 6 = 3 \cdot 2 = x_1x_2\end{aligned}$$

2. Die quadratische Gleichung $x^2 - 12x + 40 = 0$ mit $p = -12$, $q = 40$ hat die Lösungen $x_1 = 6 + 2j$, $x_2 = 6 - 2j$ (siehe oben). Es gilt:

$$\begin{aligned}p &= -12 = -(6 + 2j + 6 - 2j) = -(x_1 + x_2),\\ q &= 40 = 36 + 4\\ &= 36 - 4j^2 = (6 + 2j)(6 - 2j) = x_1x_2\end{aligned}$$

3. Welche quadratische Gleichung hat die Lösungen $x_1 = 5$ und $x_2 = -3$?
   Nach dem Satz von Viëta folgt:

$$\begin{aligned}p &= -(x_1 + x_2) = -(5 - 3) = -2,\\ q &= x_1x_2 = 5 \cdot (-3) = -15\end{aligned}$$

   Antwort: Die Normalform der gesuchten quadratischen Gleichung ist $x^2 - 2x - 15 = 0$. ◀

Hinweis: Der Satz von Viëta lässt sich auch für die Probe anwenden!

2.6 Algebraische Gleichungen höheren Grades

2.6.1 Kubische Gleichungen

Die allgemeine Form einer kubischen Gleichung lautet

Allgemeine Form

$$ax^3 + bx^2 + cx + d = 0, \quad a \neq 0$$

Die Normalform erhält man aus der allgemeinen Form durch Division durch $a \neq 0$ und Setzen von $\frac{b}{a} = r$, $\frac{c}{a} = s$, $\frac{d}{a} = t$.

Normalform

$$x^3 + rx^2 + sx + t = 0$$

Dabei sind a, b, c, d und somit auch r, s, t reelle (oder komplexe) Koeffizienten.

„Kubisch“ bedeutet, dass die Variable x in keiner höheren als der dritten Potenz vorkommt. Deshalb nennt man kubische Gleichungen auch Gleichungen dritten Grades.

Mit Hilfe der sogenannten Cardani'schen Formel lassen sich die Lösungen exakt berechnen. Es gibt entweder drei reelle Lösungen oder eine reelle Lösung und zwei konjugiert komplexe Lösungen.

In Spezialfällen führen oftmals einfachere Methoden zum Ziel.

Sind x_1, x_2, x_3 die Lösungen der kubischen Gleichung $ax^3 + bx^2 + cx + d = 0$, dann gilt

$$ax^3 + bx^2 + cx + d = a(x - x_1)(x - x_2)(x - x_3) = 0$$

Man nennt dies Produktform der kubischen Gleichung oder Zerlegung in Linearfaktoren.

Ist $t = 0$ in der Normalform (für die allgemeine Form ist die Methode ganz analog), also $x^3 + rx^2 + sx = 0$, dann erhält man durch Ausklammern von x die Gleichung $x(x^2 + rx + s) = 0$. Neben der reellen Lösung $x_1 = 0$ sind die Wurzeln der quadratischen Gleichung $x^2 + rx + s = 0$ die weiteren Lösungen.

Sonderfall $t = 0$

Gleichung: $x^3 + rx^2 + sx = 0$

Lösungen: $x_1 = 0$

$$x_2 = -\frac{r}{2} + \sqrt{\frac{r^2}{4} - s}$$

$$x^3 = -\frac{r}{2} - \sqrt{\frac{r^2}{4} - s}$$

Fehlerwarnung: Durch Division durch x geht die Lösung $x = 0$ verloren!

Beispiel

1. $x^3 - x^2 - 2x = 0$
   Ausklammern von x: $x(x^2 - x - 2) = 0$
   Erste Lösung: $x_1 = 0$
   Die quadratische Gleichung $x^2 - x - 2 = 0$ hat die Lösungen $x_{2,3} = -\frac{-1}{2} \pm \sqrt{\frac{1}{4} - (-2)} = \frac{1}{2} \pm \frac{3}{2}$, also $x_2 = 2, x_3 = -1$
   Lösungen der kubischen Gleichung $x^3 - x^2 - 2x = 0$ somit: $x_1 = 0, x_2 = 2, x_3 = -1$. ◀

Ist eine Lösung x_1 von $x^3 + rx^2 + sx + t = 0$ bekannt, dann lässt sich die kubische Gleichung durch Abspalten des Faktors $x - x_1$ reduzieren (auch diese Methode lässt sich für die allgemeine Form anwenden).

$$x^3 + rx^2 + sx + t = (x - x_1)(x^2 + ux + v) = 0$$

Dividiert man die linke Seite $x^3 + rx^2 + sx + t$ der kubischen Gleichung durch $x - x_1$, so erhält man einen quadratischen Term $x^2 + ux + v$. Die Wurzeln von $x^2 + ux + v = 0$ sind auch Lösungen der kubischen Gleichung.

Diese Methode heißt Reduktionsmethode, und die dabei durchgeführte Division nennt man Polynomdivision (vgl. auch nächsten Abschnitt).

Lösung x_1 bekannt

Gleichung: $x^3 + rx^2 + sx + t = (x - x_1)(x^2 + ux + v) = 0$

Lösungen: x_1

$$x_2 = -\frac{u}{2} + \sqrt{\frac{u^2}{4} - v}$$

$$x_3 = -\frac{u}{2} - \sqrt{\frac{u^2}{4} - v}$$

Beispiele

2. $x^3 - 6x^2 - x + 6 = 0$
   Eine Lösung dieser kubischen Gleichung ist $x_1 = 1$ (erhält man durch Probieren).
   Division von $x^3 - 6x^2 - x + 6$ durch $x - 1$:

$$\begin{array}{l} (x^3 - 6x^2 - x + 6) : (x - 1) = x^2 - 5x - 6 \\ \underline{-(x^3 - x^2)} \\ \quad -5x^2 - x \\ \quad \underline{-(-5x^2 + 5x)} \\ \quad\quad -6x + 6 \\ \quad\quad \underline{-(-6x + 6)} \\ \quad\quad\quad 0 \end{array}$$

   Die Lösungen der quadratischen Gleichung $x^2 - 5x - 6 = 0$ sind $x_2 = 6, x_3 = -1$.
   Lösungen der kubischen Gleichung $x^3 - 6x^2 - x + 6 = 0$ somit:
   $x_1 = 1, x_2 = 6, x_3 = -1$

3. $4x^3 - 12x^2 + 11x - 3 = 0$
   Eine Lösung dieser Gleichung ist $x_1 = 1$ (Probieren!).
   Division von $4x^3 - 12x^2 + 11x - 3$ durch $x - 1$:

$$\begin{array}{l} (4x^3 - 12x^2 + 11x - 3) : (x - 1) = 4x^2 - 8x + 3 \\ \underline{-(4x^3 - 4x^2)} \\ \quad -8x^2 + 11x \\ \quad \underline{-(-8x^2 + 8x)} \\ \quad\quad 3x - 3 \\ \quad\quad \underline{-(3x - 3)} \\ \quad\quad\quad 0 \end{array}$$

   Die Lösungen der quadratischen Gleichung $4x^2 - 8x + 3 = 0$ sind $x_2 = \frac{3}{2}, x_3 = \frac{1}{2}$.
   Lösungen der kubischen Gleichung $x^3 - 6x^2 - x + 6 = 0$ somit:
   $x_1 = 1, x_2 = \frac{3}{2}, x_3 = \frac{1}{2}$. ◀

2.6.2 Polynomdivision

Ein Ausdruck der folgenden Form mit $a_0, a_1, a_2, \ldots, a_{n-1}$, $a_n \in \mathbb{R}$, $a_n \neq 0$, $n \in \mathbb{N}^*$ heißt Polynom in x.

Polynom

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$$
$$= \sum_{k=0}^{n} a_k x^k$$

Das x in Klammern hinter P_n weist darauf hin, dass x die Variable ist. Die Koeffizienten $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$ dürfen dabei beliebige reelle (oder auch komplexe) Zahlen sein.

Ein Polynom ist also die linke Seite der allgemeinen Form einer algebraischen Gleichung.

Ist dabei x^n die höchste auftretende Potenz der Variablen x, so hat das Polynom den Grad n.

Zwei Polynome sind gleich, wenn sie vom gleichen Grad sind und die entsprechenden Koeffizienten übereinstimmen.

Beispiele für Polynome

1. $x^3 - 6x^2 - x + 6$
   (Polynom vom Grad 3)
2. $x - 1$
   (Polynom vom Grad 1)
3. $x^5 - \frac{1}{3}x^4 + 2x - 3$
   (Polynom vom Grad 5)
4. $0{,}34x^9 - 24{,}3x^6 + 22x^5 - \frac{1}{3}x^4 + 11$
   (Polynom vom Grad 9)
5. $x^{27} + 3$
   (Polynom vom Grad 27). ◀

Die Division zweier Polynome $P_n(x)$ und $P_m(x)$ mit $n \geq m$ verläuft ganz analog dem schriftlichen Dividieren von Dezimalzahlen. Dabei wird zuerst die höchste Potenz von x des Dividenden durch die höchste Potenz von x des Divisors geteilt.

Ist der Quotient der Polynomdivision wieder ein Polynom, gilt also $\frac{P_n(x)}{P_m(x)} = P_k(x)$ mit einem Polynom $P_k(x)$, dann heißt $P_m(x)$ Faktor des Polynoms $P_n(x)$ oder genauer Faktor m-ten Grades (und $P_k(x)$ ist ein Faktor k-ten Grades von $P_n(x)$).

Für $m = 1$ heißt $P_m(x)$ linearer Faktor, für $m = 2$ quadratischer Faktor und für $m = 3$ kubischer Faktor von $P_n(x)$.

Beispiele zur Polynomdivision

1. $P_3(x) = x^3 - 6x^2 - x + 6 = 0$, $P_1(x) = x - 1$

$$\begin{array}{l} (x^3 - 6x^2 - x + 6) : (x - 1) = x^2 - 5x - 6 \\ \underline{-(x^3 - x^2)} \\ \quad -5x^2 - x \\ \quad \underline{-(-5x^2 + 5x)} \\ \quad\quad -6x + 6 \\ \quad\quad \underline{-(-6x + 6)} \\ \quad\quad\quad 0 \end{array}$$

   Es gilt somit:

$$\frac{P_3(x)}{P_1(x)} = \frac{x^3 - 6x^2 - x + 6}{x - 1} = x^2 - 5x - 6 = P_2(x)$$

   $P_1(x) = x - 1$ ist also ein linearer Faktor und $P_2(x) = x^2 - 5x - 6$ ein quadratischer Faktor von $P_3(x) = x^3 - 6x^2 - x + 6$.

2. $P_3(x) = 4x^3 - 12x^2 + 11x - 3$, $P_1(x) = x - 1$

$$\begin{array}{l} (4x^3 - 12x^2 + 11x - 3) : (x - 1) = 4x^2 - 8x + 3 \\ \underline{-(4x^3 - 4x^2)} \\ \quad -8x^2 + 11x \\ \quad \underline{-(-8x^2 + 8x)} \\ \quad\quad 3x - 3 \\ \quad\quad \underline{-(3x - 3)} \\ \quad\quad\quad 0 \end{array}$$

   Es gilt $P_3(x) = 4x^3 - 12x^2 + 11x - 3 = (x-1)(4x^2 - 8x + 3) = P_1(x) \cdot P_2(x)$.
   $P_1(x) = x - 1$ ist ein linearer Faktor, $P_2(x) = 4x^2 - 8x + 3$ ist ein quadratischer Faktor von $P_3(x) = 4x^3 - 12x^2 + 11x - 3$.

3. $P_4(x) = 3x^4 - 10x^3 + 22x^2 - 24x + 10$, $P_2(x) = x^2 - 2x + 3$

$$\begin{array}{l} (3x^4 - 10x^3 + 22x^2 - 24x + 10) : (x^2 - 2x + 3) \\ \underline{-(3x^4 - 6x^3 + 9x^2)} \\ \quad -4x^3 + 13x^2 - 24x \\ \quad \underline{-(-4x^3 + 8x^2 - 12x)} \\ \quad\quad 5x^2 - 12x + 10 \\ \quad\quad \underline{-(5x^2 - 10x + 15)} \\ \quad\quad\quad -2x - 5 \\ \quad\quad\quad = 3x^2 - 4x + 5 \end{array}$$

   Es gilt somit:

$$(3x^4 - 10x^3 + 22x^2 - 24x + 10) : (x^2 - 2x + 3)$$
$$= 3x^2 - 4x + 5 + \frac{-2x - 5}{x^2 - 2x + 3}$$

   Also ist $P_2(x) = x^2 - 2x + 3$ kein Faktor von $P_4(x) = 3x^4 - 10x^3 + 22x^2 - 24x + 10$. ◀

Eine Zahl x_0 heißt Nullstelle des Polynoms $P_n(x)$, wenn $P_n(x_0) = 0$ gilt. Das Polynom $P_n(x)$ lässt sich dann durch $x-x_0$ dividieren, $x-x_0$ ist also ein linearer Faktor von $P_n(x)$. Es gibt dann ein Polynom $(n-1)$-ten Grades $P_{n-1}(x)$ mit $P_n(x) = (x-x_0)\cdot P_{n-1}(x)$.

Findet man eine Zerlegung $P_n(x) = (x-x_0)^m \cdot P_k(x)$, wobei $P_k(x)$ ein Polynom mit $P_k(x_0) \neq 0$ ist, dann heißt x_0 eine m-fache Nullstelle von $P_n(x)$, oder m heißt Vielfachheit der Nullstelle x_0.

$$P_n(x) = (x-x_0)^m \cdot P_k(x)$$

Ist x_0 eine reelle Zahl, dann nennt man x_0 eine reelle Nullstelle des Polynoms.

Beispiele

1. $P_3(x) = x^3 - 3x^2 + x - 3 = (x-3)(x^2+1) = (x-3)P_2(x)$
   Wegen $P_2(3) \neq 0$ ist $x_0 = 3$ eine einfache Nullstelle des Polynoms $P_3(x) = x^3 - 3x^2 + x - 3$.
2. $P_3(x) = x^3 - 3x^2 + 3x - 1 = (x-1)^3$
   Somit ist $x_0 = 1$ eine dreifache Nullstelle des Polynoms $P_3(x) = x^3 - 3x^2 + 3x - 1$.
3. $P_3(x) = x^3 - 3x + 2 = (x-1)^2(x+2)$
   Es ist also $x_0 = 1$ eine doppelte Nullstelle und $x_1 = -2$ eine einfache Nullstelle von $P_3(x) = x^3 - 3x + 2$. ◀

2.6.3 Gleichungen vierten Grades

Die allgemeine Form einer Gleichung vierten Grades lautet

Allgemeine Form

$$ax^4 + bx^3 + cx^2 + dx + e = 0, \quad a \neq 0$$

Die Normalform erhält man aus der allgemeinen Form durch Division durch $a \neq 0$ und Setzen von $\frac{b}{a} = r$, $\frac{c}{a} = s$, $\frac{d}{a} = t$, $\frac{e}{a} = u$.

Normalform

$$x^4 + rx^3 + sx^2 + tx + u = 0$$

Dabei sind a, b, c, d, e und somit auch r, s, t, u reelle (oder komplexe) Koeffizienten.

Auch für die Gleichungen vierten Grades existiert eine allgemeine Lösungsformel. Diese ist aber noch wesentlich komplizierter als die für die kubischen Gleichungen und wird deshalb hier weggelassen.

Eine Gleichung vierten Grades muss nicht unbedingt eine reelle Lösung besitzen. Im Bereich der komplexen Zahlen gibt es jedoch vier Lösungen. Dabei können Doppellösungen vorkommen. Komplexe Lösungen treten paarweise als konjugiert komplexe Zahlen auf.

In Spezialfällen führen manchmal einfachere Methoden zum Ziel.

Sind x_1, x_2, x_3, x_4 die Lösungen der Gleichung vierten Grades $ax^4 + bx^3 + cx^2 + dx + e = 0$, dann gilt

$$ax^4 + bx^3 + cx^2 + dx + e = a(x-x_1)(x-x_2)(x-x_3)(x-x_4) = 0$$

Man nennt dies wieder Produktform oder Zerlegung in Linearfaktoren.

Ist $u = 0$ in der Normalform (für die allgemeine Form ist die Methode analog), also $x^4 + rx^3 + sx^2 + tx = 0$, dann erhält man durch Ausklammern von x die Gleichung $x(x^3 + rx^2 + sx + t) = 0$. Eine Lösung ist $x_1 = 0$, die weiteren Lösungen sind die Wurzeln der kubischen Gleichung $x^3 + rx^2 + sx + t = 0$.

Beispiel

1. $x^4 - 6x^3 - x^2 + 6x = 0$
   Ausklammern von x: $x(x^3 - 6x^2 - x + 6) = 0$
   Erste Lösung: $x_1 = 0$
   Die kubische Gleichung $x^3 - 6x^2 - x + 6 = 0$ hat die Lösungen $x_2 = 1, x_3 = 6, x_4 = -1$ (vgl. Beispiel 2 im Abschn. 2.6.1).
   Lösungen der Gleichung vierten Grades $x^4 - 6x^3 - x^2 + 6x = 0$ somit: $x_1 = 0, x_2 = 1, x_3 = 6, x_4 = -1$. ◀

Ist eine Lösung x_1 von $x^4 + rx^3 + sx^2 + tx + u = 0$ bekannt, dann lässt sich die Gleichung durch Abspalten des Faktors $x-x_1$ durch Polynomdivision reduzieren (diese sogenannte Reduktionsmethode lässt sich auch wieder für die allgemeine Form anwenden):

$$x^4 + rx^3 + sx^2 + tx + u = (x-x_1)(x^3 + kx^2 + lx + m) = 0$$

Die Wurzeln der kubischen Gleichung $x^3 + kx^2 + lx + m = 0$ sind auch Lösungen der Gleichung vierten Grades $x^4 + rx^3 + sx^2 + tx + u = 0$.

Beispiele

2. $x^4 - 7x^3 + 5x^2 + 7x - 6 = 0$
Durch Probieren findet man die Lösung $x_1 = 1$.
Polynomdivision ergibt:

$$
\begin{array}{l}
(x^4 - 7x^3 + 5x^2 + 7x - 6) : (x-1) = x^3 - 6x^2 - x + 6 \\
\underline{-(x^4 - \ \ x^3)} \\
\qquad -6x^3 + 5x^2 \\
\qquad \underline{-(-6x^3 + 6x^2)} \\
\qquad\qquad -x^2 + 7x \\
\qquad\qquad \underline{-(-x^2 + \ \ x)} \\
\qquad\qquad\qquad 6x - 6 \\
\qquad\qquad\qquad \underline{-(6x-6)} \\
\qquad\qquad\qquad\quad 0
\end{array}
$$

Die kubische Gleichung $x^3 - 6x^2 - x + 6 = 0$ hat die Lösungen $x_2 = 1, x_3 = 6, x_4 = -1$ (vgl. Beispiel 2 im Abschn. 2.6.1).
Lösungen der Gleichung vierten Grades $x^4 - 7x^3 + 5x^2 + 7x - 6 = 0$ somit: $x_1 = x_2 = 1, x_3 = 6, x_4 = -1$.

3. $4x^4 - 8x^3 - x^2 + 8x - 3 = 0$
Durch Probieren erhält man die Lösung $x_1 = -1$.
Polynomdivision von $4x^4 - 8x^3 - x^2 + 8x - 3$ durch $x + 1$ ergibt:

$$
\begin{array}{l}
(4x^4 - 8x^3 - \ \ x^2 + \ \ 8x - 3) : (x+1) \\
\underline{-(4x^4 + 4x^3)} \\
\qquad -12x^3 - \ \ x^2 \\
\qquad \underline{-(-12x^3 - 12x^2)} \\
\qquad\qquad 11x^2 + \ \ 8x \\
\qquad\qquad \underline{-(11x^2 + 11x)} \\
\qquad\qquad\qquad -3x - 3 \\
\qquad\qquad\qquad \underline{-(-3x-3)} \\
\qquad\qquad\qquad\quad 0 \\
\qquad\qquad\qquad = 4x^3 - 12x^2 + 11x - 3
\end{array}
$$

Die kubische Gleichung $4x^3 - 12x^2 + 11x - 3 = 0$ hat die Lösungen $x_2 = 1, x_3 = \frac{3}{2}, x_4 = \frac{1}{2}$ (vgl. Beispiel 3 im Abschn. 2.6.1).
Lösungen der Gleichung vierten Grades $4x^4 - 8x^3 - x^2 + 8x - 3 = 0$ somit: $x_1 = -1, x_2 = 1, x_3 = \frac{3}{2}, x_4 = \frac{1}{2}$. ◀

Sind in einer Gleichung vierten Grades die Koeffizienten von x^3 und x gleich Null, dann heißt sie biquadratische Gleichung: $ax^4 + cx^2 + e = 0, a \neq 0$ (allgemeine Form) oder $x^4 + sx^2 + u = 0$ (Normalform).

Eine Substitution ist die Ersetzung eines algebraischen Ausdrucks durch einen anderen. Bei komplizierten Gleichungen können oftmals mit Hilfe einer geeigneten Substitution Lösungen gefunden werden.

Mit Hilfe der Substitution $x^2 = z$ wird aus einer biquadratischen Gleichung eine quadratische Gleichung mit der Variablen z: $az^2 + cz + e = 0$ oder $z^2 + sz + u = 0$. Aus deren Lösungen z_1, z_2 erhält man die Lösungen der biquadratischen Gleichung durch Radizieren (Wurzelziehen): $x_{1,2} = \pm\sqrt{z_1}, x_{3,4} = \pm\sqrt{z_2}$.

Biquadratische Gleichung

Gleichung (Normalform): $x^4 + sx^2 + u = 0$

Substitution $x^2 = z$: $z^2 + sz + u = 0$

Lösungen: $z_{1,2} = -\frac{s}{2} \pm \sqrt{\frac{s^2}{4} - u}$

Lösungen der biquadratischen Gleichung:

$x_{1,2} = \pm\sqrt{z_1}, \quad x_{3,4} = \pm\sqrt{z_2}$

Beispiele

4. $2x^4 - 6x^2 + 4 = 0$
Division durch 2 ergibt die Normalform:
$x^4 - 3x^2 + 2 = 0$
Substitution $x^2 = z$ ergibt eine quadratische Gleichung in z: $z^2 - 3z + 2 = 0$
Lösungen der quadratischen Gleichung: $z_1 = 1, z_2 = 2$
Lösungen der biquadratischen Gleichung durch Radizieren: $x_1 = \sqrt{z_1} = \sqrt{1} = 1, x_2 = -\sqrt{z_1} = -1, x_3 = \sqrt{z_2} = \sqrt{2}, x_4 = -\sqrt{z_2} = -\sqrt{2}$

5. $x^4 + 2x^2 - 15 = 0$
Substitution $x^2 = z$ ergibt eine quadratische Gleichung in z: $z^2 + 2z - 15 = 0$
Lösungen der quadratischen Gleichung: $z_1 = 3, z_2 = -5$
Lösungen der biquadratischen Gleichung durch Radizieren: $x_1 = \sqrt{z_1} = \sqrt{3}, x_2 = -\sqrt{z_1} = -\sqrt{3}, x_3 = \sqrt{z_2} = \sqrt{-5} = j\sqrt{5}, x_4 = -\sqrt{z_2} = -\sqrt{-5} = -j\sqrt{5}$

6. $x^4 - 4x^2 - 1 = 0$
Substitution $x^2 = z$ ergibt eine quadratische Gleichung in z: $z^2 - 4z - 1 = 0$
Lösungen der quadratischen Gleichung: $z_1 = 2 + \sqrt{5}, z_2 = 2 - \sqrt{5} = -(\sqrt{5} - 2)$
Lösungen der biquadratischen Gleichung durch Radizieren:
$x_1 = \sqrt{z_1} = \sqrt{2 + \sqrt{5}}, x_2 = -\sqrt{z_1} = -\sqrt{2 + \sqrt{5}}, x_3 = \sqrt{z_2} = \sqrt{-(\sqrt{5} - 2)} = j\sqrt{\sqrt{5} - 2}, x_4 = -\sqrt{z_2} = -\sqrt{-(\sqrt{5} - 2)} = -j\sqrt{\sqrt{5} - 2}$ ◀

2.6.4 Gleichungen *n*-ten Grades

Die allgemeine Form einer Gleichung n-ten Grades lautet

$$a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_2 x^2 + a_1 x + a_0 = 0, \quad a_n \neq 0$$

Die Normalform erhält man aus der allgemeinen Form durch Division durch $a_n \neq 0$ und Setzen von $\frac{a_i}{a_n} = b_i$, $i = 0, 1, 2, \ldots, n-1$.

Normalform

$$x^n + b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \ldots + b_2 x^2 + b_1 x + b_0 = 0$$

Dabei sind a_n, a_{n-1}, a_{n-2}, …, a_2, a_1, a_0 und b_{n-1}, b_{n-2}, …, b_2, b_1, b_0 reelle (oder komplexe) Koeffizienten, und n ist eine von Null verschiedene natürliche Zahl.

Nur für $n \leq 4$ (Gleichungen höchstens vierten Grades) gibt es allgemeine Lösungsformeln, in denen nur ineinander geschachtelte Wurzeln stehen. Für Gleichungen fünften und höheren Grades existieren solche Lösungsformeln nicht. In der Regel lassen sich dann die Lösungen außer in Spezialfällen nicht mehr exakt berechnen, man muss sich mit sogenannten Näherungslösungen begnügen. Besonders bei der Behandlung praktischer Probleme werden verschiedene Näherungsverfahren verwendet (wie zum Beispiel Regula falsi oder Newton'sches Verfahren, vgl. Abschn. 8.4.12).

In Spezialfällen führen die für Gleichungen dritten und vierten Grades vorgeführten Methoden wie zum Beispiel die Reduktionsmethode jedoch zu Lösungen.

Ist x_1 eine Lösung der Gleichung $a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_2 x^2 + a_1 x + a_0 = 0$, dann gilt

$$\begin{aligned} &a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_2 x^2 + a_1 x + a_0 \\ &= (x - x_1)(c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots + c_2 x^2 + c_1 x + c_0) \\ &= 0 \end{aligned}$$

mit $c_{n-1} = a_n$. Die weiteren Lösungen der Gleichung n-ten Grades $a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_2 x^2 + a_1 x + a_0 = 0$ sind die Lösungen der Gleichung $(n-1)$-ten Grades $c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots + c_2 x^2 + c_1 x + c_0 = 0$. Der Grad der zu lösenden Gleichung wird also um 1 reduziert.

Beispiel

$x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 = 0$

Sofort zu erkennen: $x_1 = 0, x_2 = 0$

Es folgt:

$$\begin{aligned} &x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 \\ &= x^2(x^4 - 6x^3 + 14x^2 - 14x + 5) \\ &= 0 \end{aligned}$$

Durch Probieren:

$x_3 = 1$ ist Lösung von $x^4 - 6x^3 + 14x^2 - 14x + 5 = 0$

Polynomdivision:

$(x^4 - 6x^3 + 14x^2 - 14x + 5) : (x - 1) = x^3 - 5x^2 + 9x - 5$

Es folgt:

$$\begin{aligned} &x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 \\ &= x^2(x-1)(x^3 - 5x^2 + 9x - 5) \\ &= 0 \end{aligned}$$

Durch Probieren: $x_4 = 1$ ist Lösung von $x^3 - 5x^2 + 9x - 5 = 0$

Polynomdivision:

$(x^3 - 5x^2 + 9x - 5) : (x - 1) = x^2 - 4x + 5$

Es folgt:

$$\begin{aligned} &x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 \\ &= x^2(x-1)^2(x^2 - 4x + 5) \\ &= 0 \end{aligned}$$

Die quadratische Gleichung $x^2 - 4x + 5 = 0$ hat die Lösungen $x_5 = 2 + j$, $x_6 = 2 - j$

Es folgt:

$$\begin{aligned} &x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 \\ &= x^2(x-1)^2(x-2-j)(x-2+j) \\ &= 0 \end{aligned}$$

Alle Lösungen von $x^6 - 6x^5 + 14x^4 - 14x^3 + 5x^2 = 0$ somit: $x_1 = x_2 = 0, x_3 = x_4 = 1, x_5 = 2 + j, x_6 = 2 - j$. ◀

2.6.5 Satz von Viëta für Gleichungen *n*-ten Grades

Wie für quadratische Gleichungen, so erhält man auch für beliebige Gleichungen n-ten Grades durch Vergleich der Normalform und der Produktform Beziehungen zwischen den Lösungen und den Koeffizienten der Potenzen von x in der Normalform. Ausmultiplizieren der Produktform, Gleichsetzen mit der Normalform und Koeffizientenvergleich (das heißt, Vergleich der Koeffizienten von $x^{n-1}, x^{n-2}, \ldots, x^2, x^1 = x, x^0 = 1$ auf beiden Seiten der Gleichung) ergibt die folgenden Beziehungen, die als Satz von Viëta bezeichnet werden.

Kubische Gleichungen:

Normalform: $x^3 + b_2x^2 + b_1x + b_0 = 0$,

Lösungen: x_1, x_2, x_3

Satz von Viëta:

$b_2 = -(x_1 + x_2 + x_3)$; $b_1 = x_1x_2 + x_1x_3 + x_2x_3$; $b_0 = -x_1x_2x_3$

Gleichungen vierten Grades:

Normalform: $x^4 + b_3x^3 + b_2x^2 + b_1x + b_0 = 0$,

Lösungen x_1, x_2, x_3, x_4

Satz von Viëta:

$$\begin{aligned} b_3 &= -(x_1 + x_2 + x_3 + x_4) \\ b_2 &= x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 \\ b_1 &= -(x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4) \\ b_0 &= x_1x_2x_3x_4 \end{aligned}$$

Gleichungen n-ten Grades:

Normalform:

$$x^n + b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \ldots + b_2x^2 + b_1x + b_0 = 0$$

Lösungen: $x_1, x_2, x_3, \ldots, x_n$

Satz von Viëta:

$$\begin{aligned} b_{n-1} &= -(x_1 + x_2 + \ldots + x_n), \\ b_{n-2} &= x_1x_2 + x_1x_3 + \ldots + x_1x_n + x_2x_3 + x_2x_4 + \ldots \\ &\quad + x_2x_n + \ldots + x_{n-1}x_n, \\ b_{n-3} &= -(x_1x_2x_3 + x_1x_2x_4 + \ldots + x_1x_2x_n + x_1x_3x_4 \\ &\quad + x_1x_3x_5 + \ldots + x_1x_3x_n + \ldots + x_2x_3x_4 \\ &\quad + x_2x_3x_5 + \ldots + x_2x_3x_n + x_2x_4x_5 + \ldots + x_{n-2}x_{n-1}x_n), \\ &\vdots \\ b_1 &= (-1)^{n-1}(x_1x_2x_3\ldots x_{n-1} + x_1x_2\ldots x_{n-2}x_n \\ &\quad + x_1x_2\ldots x_{n-3}x_{n-1}x_n + \ldots + x_1x_3x_4\ldots x_n + x_2x_3\ldots x_n), \\ b_0 &= (-1)^n x_1x_2x_3\ldots x_n \end{aligned}$$

Beispiele

1. Die kubische Gleichung $x^3 - 6x^2 - x + 6 = 0$ (also $b_2 = -6, b_1 = -1, b_0 = 6$) hat die Lösungen $x_1 = 1$, $x_2 = 6, x_3 = -1$ (vgl. Beispiel 2 im Abschn. 2.6.1).
   Es gilt nach dem Satz von Viëta:
   $b_2 = -(x_1 + x_2 + x_3) = -(1 + 6 - 1) = -6$,
   $b_1 = x_1x_2 + x_1x_3 + x_2x_3 = 1 \cdot 6 + 1 \cdot (-1) + 6 \cdot (-1)$
   $= 6 - 1 - 6 = -1$,
   $b_0 = -x_1x_2x_3 = -1 \cdot 6 \cdot (-1) = -(-6) = 6$
2. Für die Gleichung vierten Grades $x^4 - 7x^3 + 5x^2 + 7x - 6 = 0$ gilt: $b_3 = -7, b_2 = 5, b_1 = 7, b_0 = -6$.
   Die Gleichung hat die Lösungen $x_1 = x_2 = 1, x_3 = 6$, $x_4 = -1$ (vgl. Beispiel 2 im Abschn. 2.6.3).
   Nach dem Satz von Viëta gilt:
   $b_3 = -(x_1 + x_2 + x_3 + x_4) = -(1 + 1 + 6 - 1) = -7$,
   $b_2 = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4$
   $= 1 \cdot 1 + 1 \cdot 6 + 1 \cdot (-1) + 1 \cdot 6 + 1 \cdot (-1) + 6 \cdot (-1)$
   $= 1 + 6 - 1 + 6 - 1 - 6 = 5$,
   $b_1 = -(x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4)$
   $= -(1 \cdot 1 \cdot 6 + 1 \cdot 1 \cdot (-1) + 1 \cdot 6 \cdot (-1) + 1 \cdot 6 \cdot (-1))$
   $= -(6 - 1 - 6 - 6) = 7$,
   $b_0 = x_1x_2x_3x_4 = 1 \cdot 1 \cdot 6 \cdot (-1) = -6$.
3. Welche Normalform hat die kubische Gleichung mit den Lösungen $x_1 = -2, x_2 = -\frac{1}{2}, x_3 = \frac{1}{2}$?
   Normalform: $x^3 + b_2x^2 + b_1x + b_0 = 0$
   Nach dem Satz von Viëta gilt:
   $b_2 = -(x_1 + x_2 + x_3) = -(-2 - \frac{1}{2} + \frac{1}{2}) = 2$,
   $b_1 = x_1x_2 + x_1x_3 + x_2x_3$
   $= (-2)(-\frac{1}{2}) + (-2)\frac{1}{2} + (-\frac{1}{2})\frac{1}{2} = 1 - 1 - \frac{1}{4} = -\frac{1}{4}$,
   $b_0 = -x_1x_2x_3 = -((-2)(-\frac{1}{2})\frac{1}{2}) = -\frac{1}{2}$
   Die Normalform der kubischen Gleichung lautet somit $x^3 + 2x^2 - \frac{1}{4}x - \frac{1}{2} = 0$. ◀

2.7 Auf algebraische Gleichungen zurückführbare Gleichungen

2.7.1 Bruchgleichungen

Bestimmungsgleichungen mit Bruchtermen, bei denen die Variable (auch) im Nenner auftritt, heißen Bruchgleichungen.

Alle diese Bruchterme sind Quotienten $\frac{P(x)}{Q(x)}$, wo $P(x)$ und $Q(x)$ Polynome mit der gleichen Variablen x sind, also $P(x) = a_nx^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$ und $Q(x) = b_mx^m + b_{m-1}x^{m-1} + \ldots + b_2x^2 + b_1x + b_0$ mit $n \geq 0, m \geq 1$ (vgl. Abschn. 2.6.2). Multipliziert man eine solche Gleichung mit dem Hauptnenner aller auf beiden Seiten der Gleichung auftretenden Nenner durch, dann entsteht eine algebraische Gleichung (oder eine identische Gleichung).

Durch die Multiplikation mit dem Hauptnenner können Lösungen hinzukommen, es gehen aber keine Lösungen verloren. Deshalb ist es unbedingt erforderlich, die Lösungen der erhaltenen

algebraischen Gleichung in die Ausgangsgleichung einzusetzen. Dadurch werden hinzugekommene Lösungen, die nicht Lösungen der Ausgangsgleichung sind, ausgesondert, insbesondere solche Werte, für die ein Nenner der Ausgangsgleichung Null ist.

Beispiele

1. $\frac{x+2}{2x-1} = \frac{4x-1}{x}$
   Multiplikation mit dem Hauptnenner $x(2x-1)$:
   $(x+2)x = (4x-1)(2x-1)$
   Auflösen der Klammern: $x^2+2x = 8x^2-4x-2x+1$
   Sortieren nach Potenzen von x: $7x^2-8x+1=0$
   Quadratische Gleichung in allgemeiner Form mit $a=7, b=-8, c=1$ (vgl. Abschn. 2.5.2.2)
   Lösungen:
   $x_{1,2} = \frac{1}{14}(8 \pm \sqrt{64-4\cdot 7\cdot 1}) = \frac{1}{14}(8\pm 6) \Rightarrow x_1 = 1$, $x_2 = \frac{1}{7}$
   Einsetzen von x_1 in die Ausgangsgleichung:
   $\frac{1+2}{2\cdot 1-1} = \frac{4\cdot 1-1}{1} \Rightarrow 3=3$
   Einsetzen von x_2 in die Ausgangsgleichung:
   $\frac{\frac{1}{7}+2}{2\cdot\frac{1}{7}-1} = \frac{4\cdot\frac{1}{7}-1}{\frac{1}{7}} \Rightarrow \frac{\frac{15}{7}}{-\frac{5}{7}} = \frac{-\frac{3}{7}}{\frac{1}{7}} \Rightarrow -3=-3$
   Beide Werte sind also auch Lösungen der Ausgangsgleichung.
   Lösungsmenge: $L = \{\frac{1}{7}, 1\}$
2. $\frac{2x+1}{x-3} + \frac{3x-5}{x+3} = \frac{2x^2+2x+18}{x^2-9}$
   Multiplikation mit dem Hauptnenner $(x-3)(x+3) = x^2-9$:
   $(2x+1)(x+3)+(3x-5)(x-3) = 2x^2+2x+18$
   Auflösen der Klammern:
   $2x^2+6x+x+3+3x^2-9x-5x+15 = 2x^2+2x+18$
   Sortieren nach Potenzen von x: $3x^2-9x=0$
   Durch 3 dividieren und x ausklammern: $x(x-3)=0$
   Die Lösungen dieser quadratischen Gleichung sind $x_1=0, x_2=3$
   Einsetzen von x_1 in die Ausgangsgleichung:
   $\frac{1}{-3} + \frac{-5}{3} = \frac{18}{-9} \Rightarrow -2 = -2x_2 = 3$ ist keine Lösung der Ausgangsgleichung, da für $x=3$ zwei der Nenner gleich 0 sind.
   Lösungsmenge: $L = \{0\}$
3. $\frac{2x+3}{x-1} + \frac{4x+5}{x+1} = \frac{6x^2+6x-2}{x^2-1}$
   Multiplikation mit dem Hauptnenner $(x-1)(x+1) = x^2-1$: $(2x+3)(x+1)+(4x+5)(x-1) = 6x^2+6x-2$
   Auflösen der Klammern:
   $2x^2+2x+3x+3+4x^2-4x+5x-5 = 6x^2+6x-2$
   Zusammenfassen: $6x^2+6x-2 = 6x^2+6x-2$
   Dies ist eine identische Gleichung, die für alle reellen x erfüllt ist.
   Für $x=1$ und $x=-1$ ist die Ausgangsgleichung nicht erklärt, da Nenner gleich 0 sind.
   Lösungsmenge: $L = \{x \,|\, x \in \mathbb{R} \text{ und } x \neq 1, x \neq -1\}$
4. $\frac{3}{x-1} - \frac{2}{x-3} = -\frac{2x+1}{x^2-4x+3}$
   Multiplikation mit dem Hauptnenner $(x-1)(x-3) = x^2-4x+3$: $3(x-3)-2(x-1) = -(2x+1)$
   Auflösen der Klammern: $3x-9-2x+2 = -2x-1$
   Sortieren nach Potenzen von x: $3x-6=0$
   Die Lösung dieser linearen Gleichung ist $x=2$.
   Einsetzen in die Ausgangsgleichung:
   $\frac{3}{1}\frac{2}{-1} = -\frac{5}{-1} \Rightarrow 5=5$
   Lösungsmenge: $L=\{2\}$ ◀

2.7.2 Wurzelgleichungen

Bestimmungsgleichungen, bei denen die Variable (auch) unter einer Wurzel vorkommt, heißen Wurzelgleichungen.

Über die Existenz und die Anzahl von Lösungen bei Wurzelgleichungen lassen sich allgemein keine Aussagen machen. Ziel für die Bestimmung der Lösungen ist es, die in der Gleichung auftretenden Wurzeln zu beseitigen und die Gleichung in eine algebraische Gleichung zu überführen.

Oftmals gelingt es, die Wurzeln zu isolieren (das heißt, die Gleichung so umzuformen, dass die Wurzel allein auf einer Seite der Gleichung steht) und dann beide Seiten der Gleichung in die entsprechende Potenz zu erheben. Viele Wurzelgleichungen können auch schon durch ein- oder mehrmaliges Quadrieren in eine algebraische Gleichung überführt werden.

Durch das Quadrieren oder allgemeiner durch das Potenzieren können Lösungen hinzukommen, es gehen aber keine Lösungen verloren (Potenzieren ist keine äquivalente Umformung!). Durch die Probe mit Einsetzen der gefundenen Lösungen in die Ausgangsgleichung lassen sich diese zusätzlichen Werte aber leicht feststellen und aussortieren.

Beispiele

1. $11 - \sqrt{x+3} = 6$
   Isolieren der Wurzel ergibt $\sqrt{x+3} = 11-6$, also $\sqrt{x+3}=5$, woraus man durch Quadrieren der Gleichung $x+3=25$, also $x=22$ erhält.
   Einsetzen in die Ausgangsgleichung bestätigt $x=22$ als Lösung: $11-\sqrt{22+3}=6$
   Lösungsmenge: $L=\{22\}$
2. $5+\sqrt{9x^2-65} = 3x$
   Isolieren der Wurzel und Quadrieren der Gleichung ergibt
   $\sqrt{9x^2-65} = 3x-5 \Rightarrow 9x^2-65 = (3x-5)^2 = 9x^2-30x+25 \Rightarrow 30x = 90 \Rightarrow x=3$
   Einsetzen in die Ausgangsgleichung ergibt $x=3$ als Lösung. Lösungsmenge: $L=\{3\}$
3. $\sqrt{11-x} = x+1$
   Quadrieren ergibt $11-x = x^2+2x+1$, also die quadratische Gleichung $x^2+3x-10=0$, die die Lösungen $x_{1,2} = -\frac{3}{2} \pm \sqrt{\frac{9}{4}+10}$, also $x_1=2$ und $x_2=-5$ hat.
   Die Probe zeigt, dass x_1 die Wurzelgleichung erfüllt, x_2 jedoch nicht.
   Lösungsmenge: $L=\{2\}$

4. $\sqrt{x+5} = \sqrt{2x+3}+1$
   Quadrieren der Gleichung ergibt $x+5 = 2x+3+2\sqrt{2x+3}+1$.
   Durch Zusammenfassen und Isolieren der Wurzel erhält man $x+1 = 2\sqrt{2x+3}$.
   Erneutes Quadrieren ergibt $x^2-2x+1 = 4(2x+3)$ und somit die quadratische Gleichung $x^2-10x-11=0$.
   Diese Gleichung hat die Lösungen $x_{1,2} = 5 \pm \sqrt{25+11}$, also $x_1 = 11$ und $x_2 = -1$.
   Einsetzen in die Ausgangsgleichung $\sqrt{x+5} = \sqrt{2x+3}+1$ zeigt, dass x_2 eine Lösung ist, x_1 aber nicht.
   Lösungsmenge: $L = \{-1\}$
5. $\sqrt{x+2+\sqrt{2x+7}} = 4$
   Quadrieren: $x+2+\sqrt{2x+7} = 16$
   Wurzel isolieren: $\sqrt{2x+7} = 14-x$
   Quadrieren: $2x+7 = 196-28x+x^2$
   Sortieren nach Potenzen von x: $x^2-30x+189=0$
   Lösungen dieser quadratischen Gleichung:
   $x_{1,2} = 15 \pm \sqrt{225-189} = 15 \pm 6 \Rightarrow x_1 = 21, x_2 = 9$
   Einsetzen von x_1 in Ausgangsgleichung:
   $\sqrt{21+2+\sqrt{2\cdot 21+7}} = 4 \Rightarrow \sqrt{30} = 4$: unwahr!
   Somit ist x_1 keine Lösung der Ausgangsgleichung.
   Einsetzen von x_2 in Ausgangsgleichung:
   $\sqrt{9+2+\sqrt{2\cdot 9+7}} = 4 \Rightarrow 4 = 4$: wahr, $x_2 = 9$ ist also Lösung.
   Lösungsmenge: $L = \{9\}$
6. $\sqrt[3]{x+2} = 3$
   In die 3. Potenz erheben: $x+2 = 3^3 = 27 \Rightarrow x = 25$
   Einsetzen in die Ausgangsgleichung: $\sqrt[3]{25+2} = 3 \Rightarrow 3 = 3$
   Lösungsmenge: $L = \{25\}$
7. $x+5\sqrt[3]{x^2}-22\sqrt[3]{x}+16=0$
   Umformung der Wurzelexponenten: $x+5x^{\frac{2}{3}}-22x^{\frac{1}{3}}+16=0$
   Durch Potenzieren lassen sich die Wurzeln nicht beseitigen, stattdessen Substitution: $z = \sqrt[3]{x} = x^{\frac{1}{3}}$
   Es folgt: $z^3+5z^2-22z+16=0$
   Durch Probieren erhält man die Lösung $z=1$ dieser kubischen Gleichung.
   Polynomdivision:
   $(z^3+5z^2-22z+16) : (z-1) = z^2+6z-16$
   Lösungen der quadratischen Gleichung
   $z^2+6z-16=0$:
   $z_{2,3} = -3 \pm \sqrt{9+16} = -3 \pm 5 \Rightarrow z_2 = 2, z_3 = -8$
   Substitutionsgleichung in die dritte Potenz erheben: $x = z^3$
   Einsetzen der Lösungen: $x_1 = z_1^3 = 1^3 = 1$, $x_2 = z_2^3 = 2^3 = 8$, $x_3 = z_3^3 = (-8)^3 = -512$
   Einsetzen von x_1 in Ausgangsgleichung:
   $1+5\sqrt[3]{1^2}-22\sqrt[3]{1}+16=0 \Rightarrow 1+5-22+16=0 \Rightarrow 0=0$
   Einsetzen von x_2 in Ausgangsgleichung:
   $8+5\sqrt[3]{8^2}-22\sqrt[3]{8}+16=0 \Rightarrow 8+20-44+16=0 \Rightarrow 0=0$
   Einsetzen von x_3 in Ausgangsgleichung:
   $512+5\sqrt[3]{(-512)^2}22\sqrt[3]{-512}+16=0 \Rightarrow -512+5\cdot 64-22\cdot(-8)+16=0 \Rightarrow -512+320+176+16=0 \Rightarrow 0=0$
   Lösungsmenge: $L = \{-512, 1, 8\}$. ◀

2.8 Transzendente Gleichungen

2.8.1 Exponentialgleichungen

Bestimmungsgleichungen, bei denen die Variable (auch) im Exponenten einer Potenz steht, heißen Exponentialgleichungen.

Im Allgemeinen lassen sich in Exponentialgleichungen die Lösungen nur durch Näherungsverfahren (siehe Abschn. 8.4.12) bestimmen. Tritt in einer Exponentialgleichung die Variable jedoch nur in den Exponenten auf, so kann man sie oftmals lösen, und zwar durch Umformung mit Hilfe der Potenzgesetze und anschließendes Logarithmieren zu einer beliebigen Basis oder durch Überführung in eine algebraische Gleichung mit Hilfe einer geeigneten Substitution und anschließendem Logarithmieren.

Die Exponentialgleichung $a\cdot b^x = c$ geht durch Logarithmieren über in $\log a + x\cdot\log b = \log c$, woraus sich für $b \neq 1$ die Lösung $x = \frac{\log c - \log a}{\log b}$ ergibt.

Die Exponentialgleichung $a_n b^{nx} + a_{n-1}b^{(n-1)x} + \ldots + a_2 b^{2x} + a_1 b^x + a_0 = 0$ geht mit Hilfe der Substitution $z = b^x$ über in die algebraische Gleichung $a_n z^n + a_{n-1}z^{n-1} + \ldots + a_2 z^2 + a_1 z + a_0 = 0$. Ist $z > 0$ eine reelle Lösung dieser Gleichung, so ist $x = \frac{\log z}{\log b}$ eine Lösung der Exponentialgleichung.

Beispiele

1. $3^x = 4^{x-2}\cdot 2^x$
   Logarithmieren zu einer beliebigen Basis ergibt $x\log 3 = (x-2)\log 4 + x\log 2$.
   Auflösen nach x gibt die Lösung
   $x = \frac{2\log 4}{\log 4 - \log 3 + \log 2} = \frac{4\log 2}{3\log 2 - \log 3} \approx 2{,}826780$
   Bei der letzten Umformung wurde $\log 4 = 2\log 2$ gesetzt.
   Lösungsmenge: $L = \{2{,}826780\ldots\}$
2. $5\cdot 6^x = 2^x \cdot 7^{-x}$
   Durch Umformung ergibt sich $(\frac{2}{7\cdot 6})^x = (\frac{1}{21})^x = 5$, woraus man durch Logarithmieren die Lösung erhält:
   $x = \frac{\log 5}{\log(\frac{1}{21})} = \frac{\log 5}{-\log 21} \approx -0{,}528634$
   Lösungsmenge: $L = \{-0{,}528633\ldots\}$
3. $e^{2x+3} = e^{x-4}$
   Logarithmieren zur Basis e ergibt die Gleichheit der Exponenten: $2x+3 = x-4$, woraus sich unmittelbar die Lösung $x = -7$ ergibt.
   Lösungsmenge: $L = \{-7\}$

4. $2^{2x} - \frac{5}{2} \cdot 2^{x+1} + 6 = 0$
   Durch Umformung erhält man $(2^x)^2 - 5 \cdot 2^x + 6 = 0$, und mit der Substitution $z = 2^x$ ergibt sich die quadratische Gleichung in z: $z^2 - 5z + 6 = 0$, die die Lösungen $z_{1,2} = \frac{5}{2} \pm \sqrt{\frac{25}{4} - 6}$, also $z_1 = 2$ und $z_2 = 3$ hat.
   Durch Einsetzen dieser Werte in die Substitutionsgleichung ergeben sich die Lösungen x_1 und x_2 der Exponentialgleichung:
   $2^{x_1} = z_1 = 2 \Rightarrow x_1 = 1$;
   $2^{x_2} = z_2 = 3 \Rightarrow x_2 = \frac{\log 3}{\log 2} \approx 1{,}584963$
   Lösungsmenge: $L = \{1; 1{,}584962\ldots\}$
5. $3^x = x + 4$
   In dieser Gleichung steht x nicht nur im Exponenten, die Gleichung lässt sich deshalb nur mit einem Näherungsverfahren lösen.
   Man setzt $y = 3^x - x - 4$ und wendet etwa Regula falsi oder das Newton'sche Verfahren an (vgl. Abschn. 8.4.12). Mit den Startwerten $x_1 = 1{,}55$, $x_2 = 1{,}57$ ($\Rightarrow y_1 = -0{,}0604\ldots$, $y_2 = 0{,}0415\ldots$) erhält man zum Beispiel mit Regula falsi (Sekantenverfahren) in zwei Schritten die Näherungslösung: $x \approx 1{,}561919$.
   Lösungsmenge: $L = \{1{,}561918\ldots\}$. ◀

2.8.2 Logarithmische Gleichungen

Bestimmungsgleichungen, bei denen die Variable (auch) im Argument eines Logarithmus vorkommt, heißen logarithmische Gleichungen.

Einige dieser Gleichungen lassen sich mit Hilfe der Logarithmenrechnung auf die lösbare Form $\log_a x = b$ bringen. Die Lösung lautet dann $x = a^b$.

Spezielle logarithmische Gleichungen können mit Hilfe einer geeigneten Substitution in eine algebraische Gleichung umgewandelt werden.

Tritt die Variable nicht nur im Argument von Logarithmen auf, dann lassen sich die Lösungen von logarithmischen Gleichungen im Allgemeinen nur durch Näherungsverfahren bestimmen (siehe Abschn. 8.4.12).

Beispiele

1. $\log_7(x^2 + 19) = 3$
   Durch Potenzieren ergibt sich $7^3 = x^2 + 19$, also $x^2 = 324$.
   Daraus erhält man die Lösungen $x_1 = 18$ und $x_2 = -18$.
   Lösungsmenge: $L = \{18, -18\}$
2. $\lg(6x + 10) - \lg(x - 3) = 1$
   Es folgt $\lg \frac{6x+10}{x-3} = 1$ und durch Potenzieren $\frac{6x+10}{x-3} = 10^1 = 10$ und somit $6x + 10 = 10x - 30$, also $4x = 40$.
   Daraus ergibt sich $x = 10$ als Lösung der logarithmischen Gleichung.
   Lösungsmenge: $L = \{10\}$
3. $\lg(11x - 10) + [\lg(11x - 10)]^2 = 6$
   Durch die Substitution $z = \lg(11x - 10)$ erhält man die quadratische Gleichung $z^2 + z - 6 = 0$ mit den Lösungen $z_{1,2} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + 6}$, also $z_1 = -3$ und $z_2 = 2$.
   Durch Einsetzen dieser Werte in die Substitutionsgleichung ergeben sich die Lösungen x_1 und x_2 der logarithmischen Ausgangsgleichung:

$$z = \lg(11x - 10) \quad \Rightarrow \quad 10^z = 11x - 10$$
$$\Rightarrow \quad x = \frac{10^z + 10}{11}$$
$$z_1 = -3 \quad \Rightarrow \quad x_1 = \frac{10^{-3} + 10}{11} = 0{,}909\overline{18}$$
$$z_2 = 2 \quad \Rightarrow \quad x_2 = \frac{10^2 + 10}{11} = 10$$

   Lösungsmenge: $L = \{0{,}909\overline{18}, 10\}$. ◀

2.8.3 Trigonometrische Gleichungen

Bestimmungsgleichungen, in denen die Variable (auch) im Argument einer trigonometrischen Funktion auftritt, heißen trigonometrische Gleichungen oder goniometrische Gleichungen (siehe Kap. 6).

Trigonometrische Gleichungen sind ebenfalls transzendente Gleichungen. Sie lassen sich nur in Spezialfällen rechnerisch exakt lösen. Es existieren jedoch stets Näherungsverfahren, mit deren Hilfe sich die Lösungen mit beliebiger Genauigkeit angeben lassen (zum Beispiel Newton'sches Verfahren oder Regula falsi, vgl. Abschn. 8.4.12).

Tritt in der Gleichung nur eine trigonometrische Funktion auf, so erhält man mit den Arkusfunktionen die Lösungen (vgl. Abschn. 6.7). Auch für solche Gleichungen, in denen verschiedene trigonometrische Funktionen auftreten, die das gleiche Argument haben, kann man oft mit Hilfe der Arkusfunktionen die Lösungen berechnen. Die Lösungen können im Gradmaß oder im Bogenmaß angegeben werden.

Die Probe durch Einsetzen der gefundenen Werte ist wichtig, weil beim Lösen in der Regel auch nicht-äquivalente Umformungen vorgenommen werden.

Im Allgemeinen sind trigonometrische Gleichungen nicht eindeutig lösbar.

Beispiele

1. $\sin^2 x - 1 = -0{,}5$
   Man berechnet:
   $\sin^2 x = \frac{1}{2} \Rightarrow \sin x = \pm\sqrt{\frac{1}{2}} = \pm\frac{1}{2}\sqrt{2}$
   $\Rightarrow x = \arcsin(\pm\frac{1}{2}\sqrt{2}) = \pm 45°$
   Wegen $\sin x = \sin(180° - x)$ und $\sin x = \sin(x + k \cdot 360°)$ ergibt sich als Lösungsmenge der trigonometrischen Gleichung (im Gradmaß): $L = \{x \,|\, x = 45° + k \cdot 180°, k \in \mathbb{Z}\}$.
   Die Probe in der Ausgangsgleichung bestätigt diese Werte.
2. $\sin^2 x + 2\cos x = 1{,}5$
   Wegen $\sin^2 x + \cos^2 x = 1$ ergibt sich $1 - \cos^2 x + 2\cos x = 1{,}5 \Rightarrow \cos^2 x - 2\cos x + 0{,}5 = 0$.
   Dies ist eine quadratische Gleichung in $\cos x$. Mit der Substitution $z = \cos x$ ergibt sich die quadratische Gleichung $z^2 - 2z + 0{,}5 = 0$ mit den Lösungen $z = \cos x = 1 \pm \frac{1}{2}\sqrt{2}$. Wegen $\cos x \leq 1$ kommt $\cos x = 1 + \frac{1}{2}\sqrt{2}$ für eine Lösung nicht in Betracht. Somit folgt
   $\cos x = 1 - \frac{1}{2}\sqrt{2} \Rightarrow x = \arccos(1 - \frac{1}{2}\sqrt{2}) \approx 1{,}2735$.
   Wegen $\cos x = \cos(-x)$ und $\cos x = \cos(x + 2k\pi)$ ergibt sich als Lösungsmenge der Ausgangsgleichung (im Bogenmaß):
   $L = \{x \,|\, x \approx \pm 1{,}2735 + 2k\pi, k \in \mathbb{Z}\}$.
   Die Probe in der Ausgangsgleichung bestätigt diese Werte.
3. $\sin x + \cos x - 0{,}9x = 0$
   Diese Gleichung ist nicht geschlossen lösbar. Eine Näherungslösung erhält man zum Beispiel mit Regula falsi (vgl. Abschn. 8.4.12).
   Man setzt $y = \sin x + \cos x - 0{,}9x$ und erhält mit den Startwerten $x_1 = 77°$, $x_2 = 76{,}5°$ ($\Rightarrow y_1 = -0{,}0101\ldots$, $y_2 = 0{,}0041\ldots$) im ersten Schritt die Näherungslösung $x \approx 1{,}3378$. ◀

2.9 Lineare Gleichungssysteme

2.9.1 Definitionen

Die Schwierigkeiten beim Bestimmen der Lösungen von Gleichungen werden noch größer, wenn nicht nur eine Variable aus einer Bestimmungsgleichung errechnet werden soll, sondern wenn mehrere Variable mehrere Gleichungen gleichzeitig erfüllen sollen. Zum Beispiel sollen in den Gleichungen $x - 2y = 4$ und $2x + 5y = 35$ die Variablen x und y so berechnet werden, dass deren Werte beide Gleichungen erfüllen.

Sollen m Gleichungen von n Variablen gleichzeitig erfüllt sein, so spricht man von einem System von m Gleichungen mit n Variablen. Ein solches Gleichungssystem zu lösen, heißt, die Werte der Variablen zu bestimmen, die alle Gleichungen dieses Systems erfüllen. Eine Lösung eines Gleichungssystems mit m Gleichungen und n Variablen besteht also aus n Werten, einem sogenannten n-Tupel (für jede Variable ein Wert). So besteht eine Lösung eines Gleichungssystems mit zwei Variablen aus einem Paar (2-Tupel) von Werten, ein Gleichungssystem mit drei Variablen hat ein Lösungstripel (3-Tupel). In dem Beispiel ist das Paar $x = 10$, $y = 3$ Lösung. Man schreibt auch $(x, y) = (10, 3)$, oder die Lösungsmenge L der beiden Gleichungen ist $L = \{(10, 3)\}$ (die Reihenfolge ist zu beachten!).

Eine Gleichung in mehreren Variablen heißt linear, wenn alle Variablen höchstens in der ersten Potenz auftreten und nicht miteinander multipliziert werden. Die beiden Gleichungen des Beispiels sind lineare Gleichungen. Ein Gleichungssystem heißt linear, wenn alle Gleichungen des Systems lineare Gleichungen sind.

Im Allgemeinen ist die Bestimmung von Lösungen oder sogar aller Lösungen eines Gleichungssystems sehr schwierig oder auch nicht möglich. Für lineare Gleichungssysteme ist jedoch eine Methode entwickelt worden, die alle Lösungen des Systems liefert.

Ein lineares Gleichungssystem ist durch die Koeffizienten der Variablen und durch die Absolutglieder (die Terme, die die Variablen nicht enthalten) bestimmt.

2.9.2 Zwei lineare Gleichungen mit zwei Variablen

Die allgemeine Form eines linearen Gleichungssystems mit zwei Gleichungen und zwei Variablen x, y lautet:

$$a_1 x + b_1 y = c_1$$
$$a_2 x + b_2 y = c_2$$

a_1, a_2 sind die Koeffizienten von x, die Koeffizienten von y sind b_1, b_2, und c_1, c_2 sind die Absolutglieder des Gleichungssystems.

Es gibt verschiedene Methoden, solche linearen Gleichungssysteme zu lösen. Übliche Verfahren sind das Einsetzungsverfahren (Substitutionsverfahren), das Additionsverfahren und das Gleichsetzungsverfahren.

Einsetzungsverfahren

Beim Einsetzungsverfahren wird eine der beiden Gleichungen nach einer der Variablen aufgelöst, und der entsprechende Term wird in die andere Gleichung eingesetzt. Man erhält so eine lineare Gleichung mit einer Variablen, die gelöst werden kann. Durch Einsetzen dieses Wertes in eine der beiden Ausgangsgleichungen ergibt sich eine lineare Gleichung mit der anderen Variablen, die daraus dann auch berechnet werden kann.

Führt man das Verfahren mit der allgemeinen Form des Gleichungssystems durch, so ergibt sich

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \quad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

Fallunterscheidung:

1. Ist der Nenner $a_1b_2 - a_2b_1 \neq 0$, so erhält man genau eine Lösung, und zwar das Zahlenpaar

$$(x, y) = \left(\frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}\right)$$

2. Ist der Nenner $a_1b_2 - a_2b_1 = 0$, aber (mindestens) einer der Zähler $b_2c_1 - b_1c_2$ oder $a_1c_2 - a_2c_1$ ungleich 0, so gibt es keine Lösung.
3. Ist der Nenner gleich 0, und sind außerdem beide Zähler gleich 0, so gibt es unendlich viele Lösungen, und zwar jedes Paar (x, y), das die erste gegebene Gleichung $a_1x + b_1y = c_1$ und damit dann zugleich die zweite gegebene Gleichung $a_2x + b_2y = c_2$ erfüllt.

Der Graph (das Schaubild) zu jeder einzelnen gegebenen Gleichung ist eine Gerade (vgl. Abschn. 5.4.2). Im ersten Fall ist das die Lösung bildende Zahlenpaar das Koordinatenpaar des eindeutigen Schnittpunktes der beiden Geraden. Im zweiten Fall gibt es keine Lösung, die Graphen der beiden gegebenen Gleichungen sind parallele Geraden (Parallelen haben keinen Schnittpunkt). Im dritten Fall gibt es unendlich viele Lösungen, die Graphen der beiden gegebenen Gleichungen sind gleich, es handelt sich um ein und dieselbe Gerade (Abb. 2.1 und 2.2).

Wichtiger Hinweis: Man sollte stets die Probe in den noch nicht umgeformten Ausgangsgleichungen durchführen!

Beispiele

1. (I) $-x + y = 1$, (II) $2x + y = 4$
   (I) $\Rightarrow y = 1 + x$; Einsetzen in (II): $2x + 1 + x = 4 \Rightarrow 3x = 3 \Rightarrow x = 1$
   Einsetzen in (I) $\Rightarrow y = 2$
   Lösung: $(x, y) = (1, 2)$ (oder Lösungsmenge: $L = \{(1, 2)\}$)
   Berechnung der Lösung durch Einsetzen von $a_1 = -1$, $b_1 = 1$, $c_1 = 1$, $a_2 = 2$, $b_2 = 1$, $c_2 = 4$ in die Lösungsformel:
   $x = \frac{1\cdot1-1\cdot4}{(-1)\cdot1-2\cdot1} = \frac{-3}{-3} = 1$, $y = \frac{(-1)\cdot4-2\cdot1}{(-1)\cdot1-2\cdot1} = \frac{-6}{-3} = 2$
   Probe: (I) $-1 + 2 = 1 \Rightarrow 1 = 1$, (II) $2 \cdot 1 + 2 = 4 \Rightarrow 4 = 4$
2. (I) $2x + 3y = 8$, (II) $x + 4y = 9$ (Abb. 2.1)
   (II) $\Rightarrow x = 9 - 4y$; Einsetzen in (I): $2(9 - 4y) + 3y = 8 \Rightarrow 18 - 5y = 8 \Rightarrow 5y = 10 \Rightarrow y = 2$
   Einsetzen in (II) $\Rightarrow x = 9 - 4 \cdot 2 = 1$
   Lösung: $(x, y) = (1, 2)$ (oder Lösungsmenge: $L = \{(1, 2)\}$)

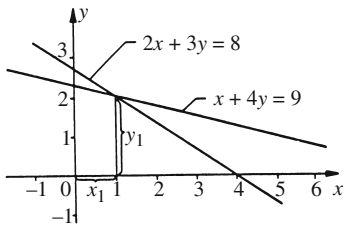


Abb. 2.1 Geraden von Beispiel 2

Berechnung der Lösung durch Einsetzen von $a_1 = 2$, $b_1 = 3$, $c_1 = 8$, $a_2 = 1$, $b_2 = 4$, $c_2 = 9$ in die Lösungsformel:
$x = \frac{4\cdot8-3\cdot9}{2\cdot4-1\cdot3} = \frac{5}{5} = 1$, $y = \frac{2\cdot9-1\cdot8}{2\cdot4-1\cdot3} = \frac{10}{5} = 2$
Probe: (I) $2 \cdot 1 + 3 \cdot 2 = 8 \Rightarrow 8 = 8$, (II) $1 + 4 \cdot 2 = 9 \Rightarrow 9 = 9$ ◀

Additionsverfahren

Beim Additionsverfahren werden beide Gleichungen jeweils so mit einem Faktor multipliziert, dass bei anschließender Addition der Gleichungen eine der Variablen wegfällt. Man erhält so eine lineare Gleichung mit einer Variablen, die gelöst werden kann. Durch Einsetzen dieses Wertes in eine der beiden Ausgangsgleichungen ergibt sich eine lineare Gleichung mit der anderen Variablen, die daraus dann auch berechnet werden kann.

Beispiele

3. (I) $-x + y = 1$, (II) $2x + y = 4$
   Multiplikation von (I) mit 2: (I′) $-2x + 2y = 2$
   Addition von (I′) und (II): $y + 2y = 4 + 2 \Rightarrow 3y = 6 \Rightarrow y = 2$
   Einsetzen in (I) $\Rightarrow x = 1$
   Lösung: $(x, y) = (1, 2)$ (oder Lösungsmenge: $L = \{(1, 2)\}$)
4. (I) $2x + 3y = 8$, (II) $x + 4y = 9$
   Multiplikation von (II) mit -2: (II′) $-2x - 8y = -18$
   Addition von (I) und (II′): $3y - 8y = 8 - 18 \Rightarrow 5y = 10 \Rightarrow y = 2$
   Einsetzen in (I) $\Rightarrow x = 1$
   Lösung: $(x, y) = (1, 2)$ ◀

Gleichsetzungsverfahren

Beim Gleichsetzungsverfahren löst man beide Gleichungen nach derselben Variablen (oder dem gleichen Vielfachen einer Variablen) auf und setzt die entsprechenden Terme gleich. Man erhält so eine lineare Gleichung mit einer Variablen, die gelöst werden kann. Durch Einsetzen dieses Wertes in eine der beiden Ausgangsgleichungen ergibt sich eine lineare Gleichung mit der anderen Variablen, die daraus dann auch berechnet werden kann.

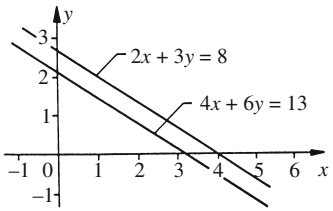


Abb. 2.2 Geraden von Beispiel 7

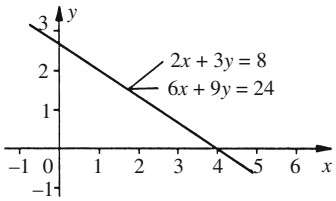


Abb. 2.3 Geraden von Beispiel 8

Beispiele

5. (I) $-x + y = 1$, (II) $2x + y = 4$
   (I) $\Rightarrow y = 1 + x$, (II) $\Rightarrow y = 4 - 2x$
   Gleichsetzen $\Rightarrow 1 + x = 4 - 2x \Rightarrow 3x = 3 \Rightarrow x = 1$
   Einsetzen in (I) $\Rightarrow y = 2$
   Lösung: $(x, y) = (1, 2)$ (oder Lösungsmenge: $L = \{(1, 2)\}$)
6. (I) $2x + 3y = 8$, (II) $x + 4y = 9$
   (I) $\Rightarrow$ (I′) $2x = 8 - 3y$
   Multiplikation von (II) mit 2: (II′) $2x + 8y = 18 \Rightarrow$ (II″) $2x = 18 - 8y$
   Gleichsetzen von (I′) und (II″) $\Rightarrow 8 - 3y = 18 - 8y \Rightarrow 5y = 10 \Rightarrow y = 2$
   Einsetzen in (I′): $2x = 2 \Rightarrow x = 1$
   Lösung: $(x, y) = (1, 2)$. ◀

Weitere Beispiele

7. (I) $2x + 3y = 8$, (II) $4x + 6y = 13$ (Abb. 2.2)
   Nach der Lösungsformel ist der Nenner $a_1b_2 - a_2b_1 = 2 \cdot 6 - 4 \cdot 3 = 0$. Der Zähler des x-Terms ist $b_2c_1 - b_1c_2 = 6 \cdot 8 - 3 \cdot 13 = 9 \neq 0$ (der Zähler des y-Terms ist $a_1c_2 - a_2c_1 = 2 \cdot 13 - 4 \cdot 8 = -6 \neq 0$). Es gibt also keine Lösung, die Graphen zu beiden gegebenen Gleichungen sind parallele Gleichungen.
   Anderer Nachweis, dass es keine Lösung gibt:
   Multiplikation von (I) mit -2: (I′) $-4x - 6y = -16$
   Addition von (I′) und (II): $0 = -3$
   Dies ist ein Widerspruch, also kann das Gleichungssystem keine Lösung haben.
8. (I) $2x + 3y = 8$, (II) $6x + 9y = 24$ (Abb. 2.3)
   Nach der Lösungsformel ist der Nenner $a_1b_2 - a_2b_1 = 2 \cdot 9 - 6 \cdot 3 = 0$. Der Zähler des x-Terms ist $b_2c_1 - b_1c_2 = 9 \cdot 8 - 3 \cdot 24 = 0$, der Zähler des y-Terms ist $a_1c_2 - a_2c_1 = 2 \cdot 24 - 6 \cdot 8 = 0$. Es gibt also unendlich viele Lösungen, die Graphen der beiden gegebenen Gleichungen sind dieselbe Gerade.
   Anderer Nachweis, dass es unendlich viele Lösungen gibt:
   Multiplikation von (I) mit 3 ergibt (II). Es handelt sich also um dieselbe Geradengleichung. Zu jedem x-Wert lässt sich ein y-Wert eindeutig berechnen. ◀

2.9.3 Drei lineare Gleichungen mit drei Variablen

Die allgemeine Form eines linearen Gleichungssystems mit drei Gleichungen und drei Variablen x, y, z lautet:

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ a_3x + b_3y + c_3z &= d_3 \end{aligned}$$

a_1, a_2, a_3 sind die Koeffizienten von x, die Koeffizienten von y sind b_1, b_2, b_3, die Koeffizienten von z sind c_1, c_2, c_3, und d_1, d_2, d_3 sind die Absolutglieder des Gleichungssystems.

Auch hier gibt es unterschiedliche Möglichkeiten, die Lösung oder die Lösungen eines solchen linearen Gleichungssystems zu bestimmen.

Einsetzungsverfahren

Beim Einsetzungsverfahren wird eine der drei Gleichungen nach einer Variablen aufgelöst und dann in die beiden anderen Gleichungen eingesetzt. Man erhält so ein Gleichungssystem mit zwei Gleichungen und zwei Variablen, das man entsprechend Abschn. 2.9.2 löst.

Beispiel

1.
$$\begin{aligned} &\text{(I)} && 3x - 9y - z = 5 \\ &\text{(II)} && 4x + 10y - 9z = 43 \\ &\text{(III)} && 5x - y - 2z = 36 \end{aligned}$$

Gleichung (I) nach z auflösen: $z = 3x - 9y - 5$
Einsetzen in die beiden anderen Gleichungen (II) und (III):

$$\begin{aligned} &\text{(II)} && 4x + 10y - 9(3x - 9y - 5) = 43 \\ &\text{(III)} && 5x - y - 2(3x - 9y - 5) = 36 \end{aligned}$$

Klammern auflösen und zusammenfassen:

$$\begin{aligned} \text{(II)} \quad & -23x + 91y = -2 \\ \text{(III)} \quad & -x + 17y = 26 \end{aligned}$$

Gleichung (III) nach x auflösen: $x = 17y - 26$.
Einsetzen in (II):
$-23(17y-26) + 91y = -2 \Rightarrow 300y = 600 \Rightarrow y = 2$
Einsetzen in (III): $x = 8$
Einsetzen von $x = 8$ und $y = 2$ in (I): $z = 3 \cdot 8 - 9 \cdot 2 - 5 = 1$
Lösung: $(x, y, z) = (8, 2, 1)$
(oder Lösungsmenge: $L = \{(8, 2, 1)\}$)
Probe:
(I) $3 \cdot 8 - 9 \cdot 2 - 1 = 5 \Rightarrow 24 - 18 - 1 = 5 \Rightarrow 5 = 5$
(II) $4 \cdot 8 + 10 \cdot 2 - 9 \cdot 1 = 43 \Rightarrow 32 + 20 - 9 = 43 \Rightarrow 43 = 43$
(III) $5 \cdot 8 - 2 - 2 \cdot 1 = 36 \Rightarrow 40 - 2 - 2 = 36 \Rightarrow 36 = 36$ ◀

Wichtiger Hinweis: Die Probe stets in allen noch nicht umgeformten Ausgangsgleichungen durchführen!

Additionsverfahren

Beim Additionsverfahren wird eine der Gleichungen jeweils so äquivalent umgeformt, dass bei Addition dieser umgeformten Gleichung mit einer der beiden anderen Gleichungen jeweils die gleiche Variable herausfällt. Auch dadurch erhält man wieder zwei Gleichungen mit zwei Variablen.

Beispiel

2.

$$\begin{aligned} \text{(I)} \quad & x - 2y - 3z = -5 \\ \text{(II)} \quad & 3x + 3y + z = 6 \\ \text{(III)} \quad & 2x + y - z = 0 \end{aligned}$$

Multiplikation von (II) mit 3: (II′) $9x + 9y + 3z = 18$
Addition von (I) und (II′): (IV) $10x + 7y = 13$
Addition von (II) und (III): (V) $5x + 4y = 6$
Multiplikation von (V) mit -2: (V′) $-10x - 8y = -12$
Addition von (IV) und (V′): $-y = 1 \Rightarrow y = -1$
Einsetzen in (V): $5x - 4 = 6 \Rightarrow x = 2$
Einsetzen von $x = 2$ und $y = -1$ in (III): $z = 2x + y = 2 \cdot 2 - 1 = 3$
Lösung: $(x, y, z) = (2, -1, 3)$ (oder Lösungsmenge: $L = \{(2, -1, 3)\}$) ◀

Weitere Lösungsverfahren wie die Cramer'sche Regel werden mit Hilfe der Determinantenrechnung im Abschn. 2.9.4 formuliert.

2.9.4 Matrizen und Determinanten

Eine Matrix (Plural Matrizen) ist ein System von $m \cdot n$ Größen, die in einem rechteckigen Schema von m (waagerechten) Zeilen und n (senkrechten) Spalten angeordnet sind. Die $m \cdot n$ Größen nennt man die Elemente der Matrix, es sind beliebige reelle (oder komplexe) Zahlen. Die Stellung eines Elementes, etwa a_{ij}, im Schema wird durch einen Doppelindex gekennzeichnet. Dabei gibt der erste Index i die Zeile und der zweite Index j die Spalte an, in der das Element steht. Die Nummerierungen der Zeilen verlaufen von oben nach unten, die der Spalten von links nach rechts. Das Element a_{ij} befindet sich also im Kreuzungspunkt der i-ten Zeile und der j-ten Spalte.

Eine Matrix mit m Zeilen und n Spalten nennt man (m, n)-Matrix. Meist kürzt man Matrizen durch große lateinische Buchstaben $A, B, \ldots$ ab. Man schreibt eine Matrix, indem man das Schema in eckige Klammern (oder auch in runde Klammern) setzt:

Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Abkürzend schreibt man dafür auch $A = (a_{ij})$.

Beispiel

$$A = \begin{bmatrix} 5 & -2 & 0 & 5 \\ 14 & 0 & -6 & 1 \\ -1 & 0 & 2 & 5 \end{bmatrix}$$

Dies ist eine $(3, 4)$-Matrix, also eine Matrix mit 3 Zeilen und 4 Spalten. Zum Beispiel ist $a_{12} = -2$ das Element, das in der ersten Zeile und zweiten Spalte steht. ◀

Achtung: Die Doppelindizes sind einzeln zu lesen, zum Beispiel wird a_{12} gesprochen: *a – eins – zwei*.

Quadratische Matrizen

Gilt $m = n$, also Zeilenanzahl gleich Spaltenanzahl, dann heißt A eine n-reihige quadratische Matrix oder eine quadratische Matrix der Ordnung n. Die Elemente einer quadratischen Matrix, für die $i = j$ gilt, bilden die sogenannte Hauptdiagonale der Matrix.

Beispiel

$$A = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{5}{2} \\ 1 & 1 & 1 \\ -1 & 10 & -\frac{4}{3} \end{bmatrix}$$

A ist eine quadratische 3-reihige Matrix. Die Hauptdiagonalelemente sind $a_{11} = \frac{4}{5}$, $a_{22} = 1$, $a_{33} = -\frac{4}{3}$. Alle Elemente der zweiten Zeile sind gleich 1: $a_{21} = a_{22} = a_{23} = 1$. ◀

Nullmatrix 0

Eine Matrix, deren Elemente alle gleich Null sind, also $a_{ij} = 0$ für $i = 1, \ldots, m$ und $j = 1, \ldots, n$, heißt eine Nullmatrix.

Einheitsmatrix E

Eine quadratische Matrix heißt Einheitsmatrix, falls

$$a_{ij} = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Diagonalmatrix

Eine quadratische Matrix, bei der für alle $i \neq j$ die Elemente a_{ij} gleich Null sind, heißt Diagonalmatrix.

Beispiel

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$ ◀

Obere Dreiecksmatrix

Eine quadratische Matrix, bei der für alle $i > j$ die Elemente a_{ij} gleich Null sind, heißt obere Dreiecksmatrix.

Beispiel

$$A = \begin{bmatrix} -1 & 6 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & -7 \end{bmatrix}$$ ◀

Untere Dreiecksmatrix

Eine quadratische Matrix, bei der für alle $i < j$ die Elemente a_{ij} gleich Null sind, heißt untere Dreiecksmatrix.

Matrizen vom gleichen Typ

Zwei Matrizen heißen vom gleichen Typ, wenn sie die gleiche Anzahl von Zeilen und die gleiche Anzahl von Spalten haben, wenn also beide (m, n)-Matrizen sind mit dem gleichen m und dem gleichen n.

Beispiel

$$A = \begin{bmatrix} -1 & 2 & 3 \\ 1 & 24 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 6 & 1 \end{bmatrix}; \quad C = \begin{bmatrix} -1 & 6 \\ 0 & 4 \end{bmatrix}$$

A und B sind vom gleichen Typ, C ist jedoch nicht vom gleichen Typ wie A und B. ◀

Gleichheit von Matrizen

Zwei Matrizen A und B heißen gleich, wenn beide vom gleichen Typ sind und wenn die entsprechenden Elemente übereinstimmen, wenn also $a_{ij} = b_{ij}$ für alle $i = 1, \ldots, m$ und $j = 1, \ldots, n$ gilt.

Beispiel

$$A = \begin{bmatrix} -1 & 2 & 3 \\ 1 & 24 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} -1 & 2 & 3 \\ 1 & 24 & 0 \end{bmatrix}: \quad A = B$$ ◀

Transponierte Matrix

Die transponierte oder gespiegelte Matrix A^T der Matrix A ist die Matrix, die durch Vertauschung von Zeilen und Spalten von A gebildet wird:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Beispiel

$$A = \begin{bmatrix} 2 & 3 & -1 \\ -5 & 0 & 4 \end{bmatrix}; \quad A^T = \begin{bmatrix} 2 & -5 \\ 3 & 0 \\ -1 & 4 \end{bmatrix}$$ ◀

Symmetrische Matrix

Eine quadratische Matrix A heißt symmetrisch, wenn $A = A^T$ ist, wenn also $a_{ij} = a_{ji}$ für alle i und j gilt.

Beispiel

$$A = \begin{bmatrix} -1 & 2 & 3 \\ 2 & -6 & 0 \\ 3 & 0 & 5 \end{bmatrix} = A^T$$

◀

Antisymmetrische Matrix

Eine quadratische Matrix A heißt antisymmetrisch oder schiefsymmetrisch, wenn $A^T = -A$ ist.

Addition und Subtraktion von Matrizen

Matrizen können nur dann addiert oder subtrahiert werden, wenn sie vom gleichen Typ sind.

Zwei Matrizen vom gleichen Typ werden addiert bzw. subtrahiert, indem man ihre korrespondierenden Elemente addiert bzw. subtrahiert:

$$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$$
$$A - B = (a_{ij}) - (b_{ij}) = (a_{ij} - b_{ij})$$

Eigenschaften der Addition:

1. $A + B = B + A$ (Kommutativgesetz)
2. $A + B) + C = A + (B + C) = A + B + C$ (Assoziativgesetz)
3. $(A + B)^T = A^T + B^T$.

Beispiel

$$A = \begin{bmatrix} 2 & -2 & 0 \\ -1 & 3 & 2 \end{bmatrix}; \quad B = \begin{bmatrix} 4 & 1 & -3 \\ 1 & 0 & -2 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 6 & -1 & -3 \\ 0 & 3 & 0 \end{bmatrix}$$

$$A - B = \begin{bmatrix} -2 & -3 & 3 \\ -2 & 3 & 4 \end{bmatrix}; \quad B - A = \begin{bmatrix} 2 & 3 & -3 \\ 2 & -3 & -4 \end{bmatrix}$$

◀

Multiplikation einer Matrix mit einer reellen Zahl

Man multipliziert eine Matrix A mit einer reellen Zahl k, indem man jedes Element der Matrix mit k multipliziert:

$$kA = k(a_{ij}) = (ka_{ij})$$

Eigenschaften:

Sind k und l zwei reelle Zahlen und A und B zwei Matrizen, so gilt:

1. $k(lA) = l(kA) = (kl)A$
2. $(k + l)A = kA + lA$
3. $k(A + B) = kA + kB$
4. $(kA)^T = kA^T$

Beispiel

$$3A = 3\begin{bmatrix} 2 & -2 & 0 \\ -1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 6 & -6 & 0 \\ -3 & 9 & 6 \end{bmatrix}$$

◀

Multiplikation von Matrizen

Das Produkt AB zweier Matrizen A und B kann nur dann gebildet werden, wenn die Spaltenanzahl von A gleich der Zeilenanzahl von B ist.

Ist $A = (a_{ij})$ eine (m, n)-Matrix und $B = (b_{jk})$ eine (n, r)-Matrix (Anzahl der Spalten von A = Anzahl der Zeilen von B), so ist die Produktmatrix $C = AB$ eine (m, r)-Matrix mit den Elementen $c_{ik} = \sum_{j=1}^{n} a_{ij} \cdot b_{jk}$. Das Element c_{ik} von $C = AB$ für ein festes i und ein festes k erhält man also, indem man das j-te Element der i-ten Zeile von A mit dem j-ten Element der k-ten Spalte von B multipliziert für $j = 1, \ldots, n$ und alle diese Produkte addiert.

$$A = (a_{ij}), \quad B = (b_{jk})$$
$$\Rightarrow \quad C = AB = (c_{ik}) \quad \text{mit } c_{ik} = \sum_{j=1}^{n} a_{ij} \cdot b_{jk}$$

Schematische Darstellung:

$$\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2k} & \cdots & b_{2r} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} & \cdots & b_{nr} \end{bmatrix} = B$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & c_{ik} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} = AB$$

Beispiel

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & -4 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & 2 & -1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & 2 & -1 \\ 1 & 1 & 0 & 1 \end{bmatrix} = B$$

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & -4 & 0 \end{bmatrix} \begin{bmatrix} 6 & 5 & 10 & 3 \\ 1 & -5 & -6 & 5 \end{bmatrix} = AB$$

$$AB = \begin{bmatrix} 6 & 5 & 10 & 3 \\ 1 & -5 & -6 & 5 \end{bmatrix}$$

BA existiert nicht. ◀

Eigenschaften der Matrizenmultiplikation:

1. $A(BC) = (AB)C$ (Assoziativgesetz)
2. $A(B + C) = AB + AC$ (Distributivgesetz)
3. $AB \neq BA$ (Kommutativgesetz gilt nicht)
4. $AE = EA = A$ (E Einheitsmatrix)
5. $AN = NA = N$ (N Nullmatrix)
6. $(AB)^T = B^T A^T$ (Reihenfolge ändert sich)

Orthogonale Matrix

Eine quadratische Matrix A heißt orthogonal, wenn $AA^T = A^T A = E$ (E Einheitsmatrix) ist.

Inverse Matrix

Eine Matrix B heißt Inverse der quadratischen Matrix A, wenn $AB = E$ (E Einheitsmatrix) gilt. Man schreibt dann $B = A^{-1}$. Existiert die Inverse einer Matrix, dann ist sie eindeutig.

Eine Matrix A, für die die Inverse A^{-1} existiert, heißt regulär, andernfalls heißt sie singulär.

Beispiel

Man berechne die Inverse der Matrix $A = \begin{bmatrix} -2 & 3 \\ 1 & -2 \end{bmatrix}$.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = A^{-1}$$

$$A = \begin{bmatrix} -2 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = E$$

Es ergibt sich das lineare Gleichungssystem $-2a + 3c = 1$, $-2b + 3d = 0$, $a - 2c = 0$, $b - 2d = 1$. Die Lösung des Gleichungssystems ist $a = -2$, $b = -3$, $c = -1$, $d = -2$.

Es folgt: Inverse $A^{-1} = \begin{bmatrix} -2 & -3 \\ -1 & -2 \end{bmatrix}$. ◀

Eine Determinante D ist ein algebraischer Ausdruck, der jeder n-reihigen quadratischen Matrix A mit reellen (oder komplexen) Elementen a_{ij} eindeutig zugeordnet wird. Dieser algebraische Ausdruck ist eine reelle (oder komplexe) Zahl. Die Determinante einer n-reihigen quadratischen Matrix nennt man n-reihige Determinante.

Man schreibt eine Determinante, indem man das quadratische Schema der Matrix zwischen senkrechte Striche setzt, oder in Kurzform $D = \det(A) = |A|$:

Determinante

$$D = \det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

Definition für zweireihige Determinanten ($n = 2$):

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Die Elemente a_{11}, a_{22} bilden die Hauptdiagonale, die Elemente a_{12}, a_{21} die sogenannte Nebendiagonale.

Merkregel zur Berechnung: Produkt der Hauptdiagonalelemente minus Produkt der Nebendiagonalelemente.

Beispiel

$$D = \begin{vmatrix} -1 & 1 \\ 2 & 1 \end{vmatrix} = (-1) \cdot 1 - 1 \cdot 2 = -3$$ ◀

Die allgemeine Lösungsformel

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \quad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

für ein lineares Gleichungssystem $a_1x + b_1y = c_1$, $a_2x + b_2y = c_2$ (vgl. Abschn. 2.9.2) lässt sich auch mit Hilfe von zweireihigen Determinanten schreiben:

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}$$

Die gemeinsame Nennerdeterminante wird aus den Koeffizienten von x und y der beiden Gleichungen in der gegebenen Anordnung gebildet. Die Nennerdeterminante heißt deshalb auch Koeffizientendeterminante. Man erhält die Zählerdeterminante von x, indem man die Koeffizienten von x durch die Absolutglieder ersetzt, und die Zählerdeterminante von y entsprechend durch Ersetzung der Koeffizienten von y durch die Absolutglieder (immer in der gleichen Reihenfolge, also Ersetzung von b_1 durch c_1 usw.).

Man nennt diese Methode Cramer'sche Regel zur Berechnung der Lösung eines linearen Gleichungssystems (nach dem schweizerischen Mathematiker Gabriel Cramer, 1704–1752).

Beispiel

(I) $x - 2y = 4$, (II) $2x + 5y = 35$

Einsetzen von $a_1 = 1, a_2 = 2, b_1 = -2, b_2 = 5, c_1 = 4, c_2 = 35$ in die Determinantengleichungen für x und y ergibt:

$$x = \frac{\begin{vmatrix} 4 & -2 \\ 35 & 5 \end{vmatrix}}{\begin{vmatrix} 1 & -2 \\ 2 & 5 \end{vmatrix}} = \frac{4 \cdot 5 - (-2) \cdot 35}{1 \cdot 5 - (-2)2}$$

$$= \frac{20 + 70}{5 + 4} = \frac{90}{9} = 10,$$

$$y = \frac{\begin{vmatrix} 1 & 4 \\ 2 & 35 \end{vmatrix}}{\begin{vmatrix} 1 & -2 \\ 2 & 5 \end{vmatrix}} = \frac{1 \cdot 35 - 4 \cdot 2}{1 \cdot 5 - (-2) \cdot 2} = \frac{35 - 8}{5 + 4} = \frac{27}{9} = 3$$

◀

Definition für dreireihige Determinanten ($n = 3$):

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22})$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}$$

Beispiel

$$D = \begin{vmatrix} 3 & 7 & -2 \\ 4 & 0 & 6 \\ -2 & -4 & 1 \end{vmatrix}$$

$$= 3 \begin{vmatrix} 0 & 6 \\ -4 & 1 \end{vmatrix} - 4 \begin{vmatrix} 7 & -2 \\ -4 & 1 \end{vmatrix} + (-2) \begin{vmatrix} 7 & -2 \\ 0 & 6 \end{vmatrix}$$

$$= 3(0 \cdot 1 - 6(-4)) - 4(7 \cdot 1 - (-2)(-4)) - 2(7 \cdot 6 - (-2)0)$$

$$= 3 \cdot 24 - 4(-1) - 2 \cdot 42 = -8$$

◀

Man nennt dies „Entwickeln" der dreireihigen Determinante nach der ersten Spalte. Dabei wird nacheinander jedes Element der ersten Spalte mit derjenigen zweireihigen Determinante multipliziert, die man erhält, wenn man in der dreireihigen Determinante die Zeile und die Spalte streicht, in der das Element steht. Die so gebildeten Produkte werden mit alternierenden (wechselnden) Vorzeichen versehen, angefangen mit einem +, und anschließend addiert.

Bezeichnet man die Determinante, die man durch Streichen der i-ten Zeile und der j-ten Spalte der Determinante D erhält, mit D_{ij}, so kann man das obige Entwickeln auch darstellen als

$$D = a_{11} \cdot D_{11} - a_{21} \cdot D_{21} + a_{31} \cdot D_{31}$$

Die mit dem Faktor $(-1)^{i+j}$ (dieser Faktor ist $+1$ oder -1) multiplizierte Determinante D_{ij} heißt Adjunkte oder algebraisches Komplement A_{ij} des Elements a_{ij}. Somit kann man für das obige Entwickeln auch schreiben

$$D = a_{11} \cdot A_{11} + a_{21} \cdot A_{21} + a_{31} \cdot A_{31}$$

Zur Berechnung kann man die Determinante nach einer beliebigen Zeile oder Spalte entwickeln.

Entwicklung nach einer beliebigen Zeile:

$$D = a_{i1} \cdot A_{i1} + a_{i2} \cdot A_{i2} + a_{i3} \cdot A_{i3} = \sum_{j=1}^{3} a_{ij}A_{ij}, \quad 1 \le i \le 3$$

Bei Entwicklung nach der ersten Zeile ist $i = 1$, bei Entwicklung nach der zweiten Zeile ist $i = 2$, und bei Entwicklung nach der dritten Zeile ist $i = 3$.

Entwicklung nach einer beliebigen Spalte:

$$D = a_{1j} \cdot A_{1j} + a_{2j} \cdot A_{2j} + a_{3j} \cdot A_{3j} = \sum_{i=1}^{3} a_{ij} A_{ij}, \quad 1 \le j \le 3$$

Bei Entwicklung nach der ersten Spalte ist $j = 1$, bei Entwicklung nach der zweiten Spalte ist $j = 2$, und bei Entwicklung nach der dritten Spalte ist $j = 3$.

Beispiel

$$D = \begin{vmatrix} 3 & 7 & -2 \\ 4 & 0 & 6 \\ -2 & -4 & 1 \end{vmatrix}$$

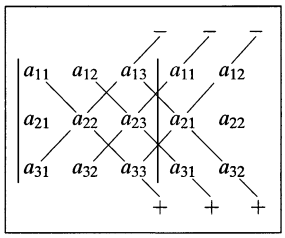
Entwicklung nach der zweiten Zeile:

$$\begin{aligned} D &= a_{21} \cdot A_{21} + a_{22} \cdot A_{22} + a_{23} \cdot A_{23} \\ &= 4 \cdot (-1)^{2+1} \begin{vmatrix} 7 & -2 \\ -4 & 1 \end{vmatrix} + 0 \cdot A_{22} \\ &\quad + 6 \cdot (-1)^{2+3} \begin{vmatrix} 3 & 7 \\ -2 & -4 \end{vmatrix} \\ &= -4[7 \cdot 1 - (-2)(-4)] + 0 - 6[3 \cdot (-4) - 7 \cdot (-2)] \\ &= -4 \cdot (-1) - 6 \cdot 2 = 4 - 12 = -8 \end{aligned}$$

◀

Dreireihige Determinanten können auch mit der Regel von Sarrus berechnet werden (nach dem französischen Mathematiker Pierre F. Sarrus, 1798 bis 1861).

Man fügt bei der Regel von Sarrus die ersten beiden Spalten der Determinante nochmals als 4. und 5. Spalte hinzu. Dann multipliziert man je drei diagonal aufeinanderfolgende Elemente und addiert (Hauptdiagonalen) bzw. subtrahiert (Nebendiagonalen) die so entstehenden sechs Produkte.



Die Regel ausgeführt ergibt

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{aligned}$$

Beispiel

$$\begin{aligned} \begin{vmatrix} 3 & 7 & -2 \\ 4 & 0 & 6 \\ -2 & -4 & 1 \end{vmatrix} &= 3 \cdot 0 \cdot 1 + 7 \cdot 6 \cdot (-2) + (-2) \cdot 4 \cdot (-4) \\ &\quad - (-2) \cdot 0 \cdot (-2) - 3 \cdot 6 \cdot (-4) - 7 \cdot 4 \cdot 1 \\ &= 0 - 84 + 32 - 0 + 72 - 28 = -8 \end{aligned}$$

◀

Fehlerwarnung: Die Regel von Sarrus gilt nur für dreireihige Determinanten!

Definition für n-reihige Determinanten ($n \ge 4$):

Auch für beliebige n-reihige Determinanten lässt sich der Wert mit Hilfe des Entwicklungssatzes definieren.

Entwicklung nach einer beliebigen Zeile:

$$D = a_{i1} \cdot A_{i1} + a_{i2} \cdot A_{i2} + \ldots + a_{in} \cdot A_{in} = \sum_{j=1}^{n} a_{ij} A_{ij}, \quad 1 \le i \le n$$

Bei Entwicklung nach der ersten Zeile ist $i = 1$, bei Entwicklung nach der zweiten Zeile ist $i = 2$, usw., und bei Entwicklung nach der n-ten Zeile ist $i = n$.

Entwicklung nach einer beliebigen Spalte:

$$D = a_{1j} \cdot A_{1j} + a_{2j} \cdot A_{2j} + \ldots + a_{nj} \cdot A_{nj} = \sum_{i=1}^{n} a_{ij} A_{ij}, \quad 1 \le j \le n$$

Bei Entwicklung nach der ersten Spalte ist $j = 1$, bei Entwicklung nach der zweiten Spalte ist $j = 2$, usw., und bei Entwicklung nach der n-ten Spalte ist $j = n$.

Die Cramer'sche Regel zur Berechnung der Lösung eines linearen Gleichungssystems ist immer dann anwendbar, wenn bei dem betrachteten linearen Gleichungssystem die Anzahl der Gleichungen und die Anzahl der Variablen übereinstimmen (und die Koeffizientendeterminante von Null verschieden ist).

Die Koeffizientendeterminante eines allgemeinen linearen Gleichungssystems mit drei Gleichungen und drei Variablen lautet

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

Ersetzt man die erste Spalte von D, also die Koeffizienten von x, durch die Absolutglieder des linearen Gleichungssystems, so ergibt sich die Determinante

$$D_x = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}$$

Durch Ersetzen der Koeffizienten von y und z erhält man analog die Matrizen

$$D_y = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}, \quad D_z = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}$$

Für $D \neq 0$ ergibt sich dann als eindeutige Lösung des linearen Gleichungssystems

$$x = \frac{D_x}{D} = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}},$$

$$y = \frac{D_y}{D} = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}},$$

$$z = \frac{D_z}{D} = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}$$

Ist jedoch $D = 0$, dann gibt es entweder keine oder unendlich viele Lösungen des linearen Gleichungssystems. In diesem Fall ist die Cramer'sche Regel nicht anwendbar.

Beispiel

Lineares Gleichungssystem:

$$\begin{aligned} 3x + 15y + 8z &= 10 \\ -5x + 10y + 12z &= -1 \\ 2x + 7y + z &= 1 \end{aligned}$$

Nennerdeterminante (Determinante der Koeffizientenmatrix):

$$D = \begin{vmatrix} 3 & 15 & 8 \\ -5 & 10 & 12 \\ 2 & 7 & 1 \end{vmatrix}$$
$$= 30 + 360 - 280 - 160 - 252 + 75 = -227$$

Zählerdeterminanten:

$$D_x = \begin{vmatrix} 10 & 15 & 8 \\ -1 & 10 & 12 \\ 1 & 7 & 1 \end{vmatrix}$$
$$= 100 + 180 - 56 - 80 - 840 + 15 = -681$$

$$D_y = \begin{vmatrix} 3 & 10 & 8 \\ -5 & -1 & 12 \\ 2 & 1 & 1 \end{vmatrix}$$
$$= -3 + 240 - 40 + 16 - 36 + 50 = 227$$

$$D_z = \begin{vmatrix} 3 & 15 & 10 \\ -5 & 10 & -1 \\ 2 & 7 & 1 \end{vmatrix}$$
$$= 30 - 30 - 350 - 200 + 21 + 75 = -454$$

Somit ergibt sich als Lösung des linearen Gleichungssystems:

$$x = \frac{D_x}{D} = \frac{-681}{-227} = 3,$$

$$y = \frac{D_y}{D} = \frac{227}{-227} = -1,$$

$$z = \frac{D_z}{D} = \frac{-454}{-227} = 2$$

Die Lösung des Gleichungssystems ist also das (geordnete) Zahlentripel $(x, y, z) = (3, -1, 2)$ (oder Lösungsmenge: $L = \{(3, -1, 2)\}$). ◀

Planimetrie

3

Arnfried Kemnitz



3.1 Geraden und Strecken 58
3.2 Winkel 58
3.3 Projektion 60
3.4 Geometrische Örter 60
3.5 Dreiecke 60
3.6 Vierecke 66
3.7 Reguläre n-Ecke 69
3.8 Polygone 70
3.9 Kreise 71
3.10 Symmetrie 73
3.11 Ähnlichkeit 73

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_3

Die Planimetrie (griech., Flächenmessung) ist ein Teilgebiet der Geometrie (griech., Erdmessung) und befasst sich mit höchstens zweidimensionalen Objekten. Dabei interessieren zum Beispiel Form, Größe und gegenseitige Lage solcher Objekte. Die Grundelemente dieser Geometrie der Ebene sind Punkte und Geraden.

3.1 Geraden und Strecken

Eine Gerade ist eine beidseitig unbegrenzte gerade Linie. Ein Punkt ist die Schnittstelle zweier Geraden.

Eine Gerade ist durch zwei voneinander verschiedene Punkte eindeutig bestimmt. Die kürzeste Verbindung zweier Punkte P_1 und P_2 liegt auf der Geraden durch P_1 und P_2. Zwei verschiedene Geraden in der Ebene sind parallel zueinander oder haben einen Punkt, ihren Schnittpunkt, gemeinsam. Die Gerade durch die Punkte P_1 und P_2 schreibt man P_1P_2 (gesprochen: Gerade P_1P_2) oder P_2P_1. Geraden kürzt man oft mit g ab, also $g = P_1P_2 = P_2P_1$.

$A \in g$ ist eine abkürzende Schreibweise dafür, dass der Punkt A ein Punkt der Geraden g ist, also auf der Geraden g liegt, und $B \notin g$ bedeutet, dass B außerhalb von g liegt.

$AB \parallel CD$ oder $g \parallel h$ (gesprochen: g parallel h) bedeutet, dass AB und CD zwei parallele Geraden sind, also keinen Schnittpunkt haben (Abb. 3.1).

Ein Strahl oder eine Halbgerade ist ein Teil einer Geraden, der von einem Punkt S einer Geraden aus in einer Richtung läuft. Der Punkt S heißt Anfangspunkt des Strahls. Jeder Punkt einer Geraden bestimmt zwei verschiedene Strahlen (Abb. 3.2).

Eine Strecke ist ein Abschnitt einer Geraden zwischen zwei Punkten. Eine Strecke ist also eine beidseitig begrenzte gerade Linie. Die Strecke zwischen den Punkten A und B schreibt man $\overline{AB}$ (gesprochen: Strecke AB). Die Punkte A und B heißen die Endpunkte der Strecke, alle anderen Punkte der Strecke bilden das Innere. Die Länge der Strecke wird mit $|\overline{AB}|$ bezeichnet (gesprochen: Länge oder Betrag der Strecke AB), (Abb. 3.3).

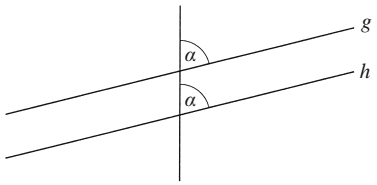


Abb. 3.1 Parallele Geraden g und h

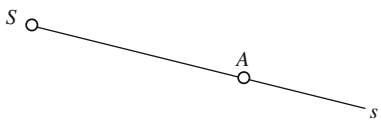


Abb. 3.2 Strahl s (A ist ein beliebiger Punkt von s)

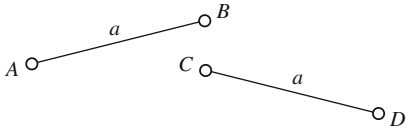


Abb. 3.3 Strecken $\overline{AB}$ und $\overline{CD}$ mit derselben Länge a

3.2 Winkel

Zwei Strahlen, die von demselben Punkt S ausgehen, können durch eine Drehung um S ineinander überführt werden, durch die der Winkel zwischen ihnen bestimmt wird. Die Strahlen heißen die Schenkel des Winkels, der Punkt S heißt Scheitelpunkt.

Sind g und h die beiden Strahlen und A ein Punkt von g und B ein Punkt von h, so bezeichnet man den Winkel mit $\angle(g,h)$ (gesprochen: Winkel zwischen g und h) oder mit $\angle ASB$ (gesprochen: Winkel ASB), (Abb. 3.4).

Winkel werden meist mit kleinen griechischen Buchstaben bezeichnet:

$$\alpha, \beta, \gamma, \delta, \ldots, \varphi, \ldots$$

Man unterscheidet in der Regel nicht zwischen Winkel und Größe (Maß, Betrag) eines Winkels.

Zur Winkelmessung unterscheidet man zwei verschiedene Winkelmaße: das Gradmaß und das Bogenmaß (Bogenmaß siehe Abschn. 3.9.5). Beide beruhen auf Kreisteilungen.

Beim Gradmaß wird ein Vollwinkel in 360 gleiche Teile eingeteilt (Sexagesimaleinteilung). Die Einheit des Gradmaßes ist Grad (°). 1° entspricht $\frac{1}{360}$ des Vollwinkels. Untereinheiten des Grads sind Minuten und Sekunden.

$$1^\circ \text{ (1 Grad)} = 60' \text{ (60 Minuten)}$$
$$1' \text{ (1 Minute)} = 60''(60 \text{ Sekunden})$$
$$1' = \left(\frac{1}{60}\right)^\circ, \quad 1'' = \left(\frac{1}{60}\right)'$$

In der Geodäsie wird eine Zentesimaleinteilung verwendet. Dabei wird der Vollwinkel in 400 gleiche Teile eingeteilt. Die Einheit ist gon (Gon). 1 gon entspricht also $\frac{1}{400}$ des Vollwinkels. Ältere, heute nicht mehr gebräuchliche Einheit ist Neugrad.

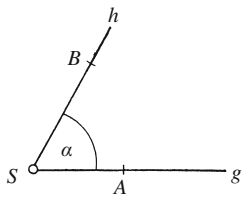


Abb. 3.4 Winkel $\alpha = \angle(g,h) = \angle ASB$

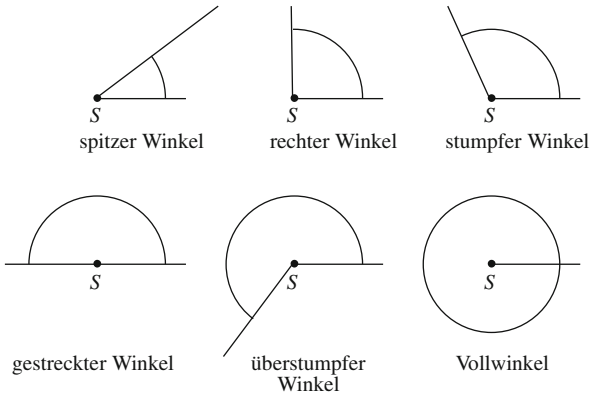


Abb. 3.5 Winkelbezeichnungen

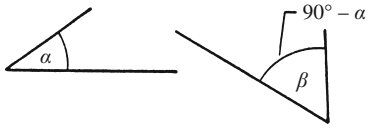


Abb. 3.6 Komplementwinkel

$$1 \text{ Vollwinkel} = 360° = 400 \text{ gon}$$

Für bestimmte Winkel gibt es besondere Bezeichnungen (Abb. 3.5):

Ein Winkel α mit $\alpha = 0°$ heißt Nullwinkel.

Ein Winkel α mit $\alpha = 90°$ heißt rechter Winkel.

Ein Winkel α mit $\alpha = 180°$ heißt gestreckter Winkel.

Ein Winkel α mit $\alpha = 360°$ heißt Vollwinkel.

Ein Winkel α, der größer als 0 und kleiner als ein rechter Winkel ist, heißt spitzer Winkel: $0° < \alpha < 90°$.

Ein Winkel α, der größer als ein rechter Winkel ist, heißt stumpfer Winkel: $\alpha > 90°$.

Ein Winkel α, der größer als ein gestreckter Winkel ist, heißt überstumpfer Winkel: $\alpha > 180°$.

In einer Figur kennzeichnet man einen rechten Winkel mit einem Punkt zwischen seinen Schenkeln und einem Winkelbogen.

Einige Paare von Winkeln haben bestimmte Namen:

1. *Komplementwinkel*
   Winkel, die sich zu 90° ergänzen. Der Komplementwinkel zu einem Winkel α ist der Winkel $\beta = 90° - \alpha$ (Abb. 3.6).

Beispiel

$\alpha = 32°$ und $\beta = 58°$ sind Komplementwinkel. ◀



Abb. 3.7 Supplementwinkel

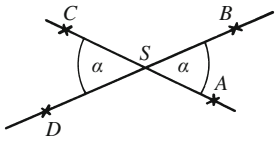


Abb. 3.8 Scheitelwinkel ($\angle ASB$ und $\angle CSD$)

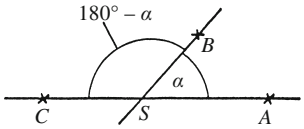


Abb. 3.9 Nebenwinkel ($\angle ASB$ und $\angle BSC$)

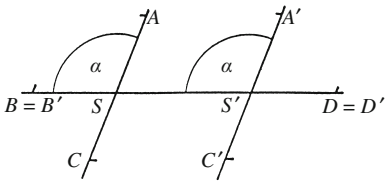


Abb. 3.10 Stufenwinkel ($\angle ASB$ und $\angle A'S'B'$)

2. *Supplementwinkel*
   Winkel, die sich zu 180° ergänzen. Der Supplementwinkel zu einem Winkel α ist der Winkel $\beta = 180° - \alpha$ (Abb. 3.7).

Beispiel

$\alpha = 62°$ und $\beta = 118°$ sind Supplementwinkel. ◀

3. *Scheitelwinkel*
   Gegenüberliegende Winkel an zwei sich schneidenden Geraden. Scheitelwinkel sind gleich groß (Abb. 3.8).

4. *Nebenwinkel*
   Benachbarte Winkel an zwei sich schneidenden Geraden. Nebenwinkel sind Supplementwinkel, sie ergänzen sich also zu 180° (Abb. 3.9).

5. *Stufenwinkel*
   Gleichliegende Winkel an von einer Geraden geschnittenen Parallelen. Stufenwinkel sind gleich groß (Abb. 3.10).

6. *Wechselwinkel*
   Entgegengesetzt liegende Winkel an von einer Geraden geschnittenen Parallelen. Wechselwinkel sind gleich groß (Abb. 3.11).

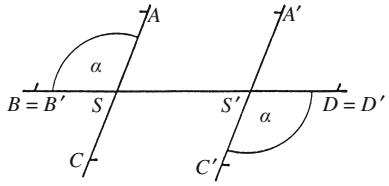


Abb. 3.11 Wechselwinkel ($\angle ASB$ und $\angle C'S'D'$)

3.3 Projektion

Man unterscheidet Parallelprojektion und Zentralprojektion.

Eine Parallelprojektion ist die Abbildung eines ebenen Gegenstandes durch parallele Strahlen auf eine Gerade (in der Stereometrie die Abbildung eines räumlichen Gegenstandes durch parallele Strahlen auf eine Ebene). Bei senkrechter Parallelprojektion stehen die projizierenden Strahlen senkrecht auf der Geraden (Stereometrie: auf der Ebene), bei schiefer Parallelprojektion nicht (Abb. 3.12).

Eine Zentralprojektion ist die Abbildung eines ebenen Gegenstandes durch Strahlen, die alle durch einen festen Punkt Z gehen (durch das Zentrum oder Projektionszentrum), auf eine Gerade (in der Stereometrie die Abbildung eines räumlichen Gegenstandes auf eine Ebene), (Abb. 3.13).

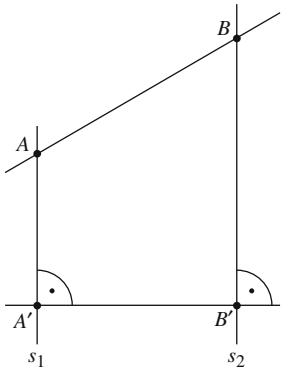


Abb. 3.12 Senkrechte Parallelprojektion einer Strecke $\overline{AB}$

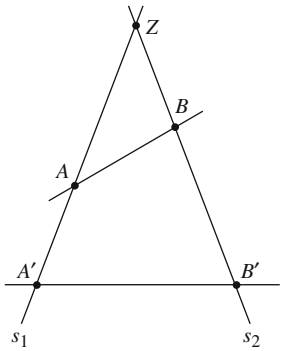


Abb. 3.13 Zentralprojektion von Z einer Strecke $\overline{AB}$

3.4 Geometrische Örter

Ein geometrischer Ort ist eine Punktmenge, die alle Elemente mit einer bestimmten geometrischen Eigenschaft (oder mit mehreren Eigenschaften) enthält.

In der Planimetrie sind die geometrischen Örter Linien, daher werden sie auch geometrische Ortslinien genannt. In der Stereometrie sind die geometrischen Örter Flächen.

Beispiele für geometrische Ortslinien

1. Der geometrische Ort aller Punkte, die von einem festen Punkt M die feste Entfernung r haben, ist der Kreis um M mit dem Radius r.
2. Der geometrische Ort aller Punkte, die von einer gegebenen Geraden g den festen Abstand d haben, ist eine Parallele zu g im Abstand d.
3. Geometrischer Ort aller Punkte, die von zwei festen Punkten A und B gleich weit entfernt sind, ist die Mittelsenkrechte auf der Strecke $\overline{AB}$.
4. Geometrischer Ort aller Punkte, die von zwei festen nicht parallelen Geraden g und h den gleichen Abstand haben, sind die beiden (zueinander senkrechten) Winkelhalbierenden zwischen g und h. ◀

3.5 Dreiecke

3.5.1 Allgemeine Dreiecke

Ein Dreieck besteht aus drei nicht auf einer Geraden liegenden Punkten A, B, C und den Strecken $\overline{AB}, \overline{AC}, \overline{BC}$.

Die Punkte A, B, C sind die Eckpunkte des Dreiecks, die Strecken $\overline{AB}, \overline{AC}, \overline{BC}$ sind die Seiten des Dreiecks, und ihre Längen $|\overline{AB}|, |\overline{AC}|, |\overline{BC}|$ sind die Seitenlängen des Dreiecks.

Meistens werden die Seitenlängen mit a, b, c und die Innenwinkel des Dreiecks mit α, β, γ bezeichnet, und zwar in der Weise, dass der Punkt A der Seite mit der Länge a, der Punkt B der Seite mit der Länge b, der Punkt C der Seite mit der Länge c gegenüberliegt und dass der Winkel α den Scheitelpunkt A, der Winkel β den Scheitelpunkt B und der Winkel γ den Scheitelpunkt C hat (Abb. 3.14).

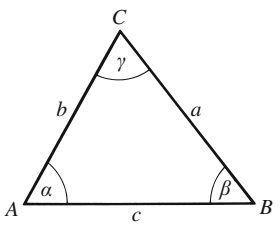


Abb. 3.14 Bezeichnungen im Dreieck

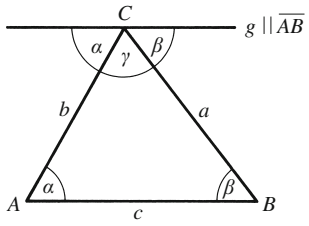


Abb. 3.15 Winkelsumme gleich 180°

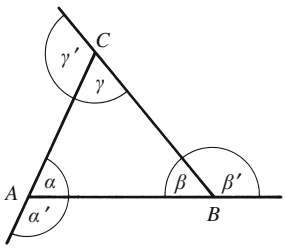


Abb. 3.16 Außenwinkelsumme gleich 360°

Abkürzend verwendet man für ein Dreieck den großen griechischen Buchstaben Δ, und für ein Dreieck mit den Eckpunkten A, B, C schreibt man $\Delta(ABC)$ (Abb. 3.15).

Die Winkelsumme in jedem Dreieck beträgt 180°.

$$\alpha + \beta + \gamma = 180^\circ$$

Im Dreieck ist die Summe zweier Seitenlängen stets größer als die dritte.

$$a + b > c, \quad a + c > b, \quad b + c > a$$

Diese drei Ungleichungen zusammen heißen Dreiecksungleichungen.

Die Supplementwinkel der Dreieckswinkel nennt man Außenwinkel des Dreiecks. Die Summe der Außenwinkel α', β', γ' in jedem Dreieck beträgt 360° (Abb. 3.16).

$$\alpha' + \beta' + \gamma' = 360^\circ$$

In einem Dreieck liegt der größeren Seite stets der größere Winkel gegenüber.

$$a > b > c \quad \Leftrightarrow \quad \alpha > \beta > \gamma$$

In einem spitzwinkligen Dreieck sind alle drei Innenwinkel kleiner als 90°, in einem rechtwinkligen Dreieck ist ein Winkel gleich 90°, in einem stumpfwinkligen Dreieck ist ein Winkel größer als 90°.

Der Umfang u eines Dreiecks ist die Summe der Seitenlängen.

$$u = a + b + c$$

Der Flächeninhalt A eines Dreiecks berechnet sich nach der Grundformel

$$A = \frac{1}{2} \cdot \text{Grundseite} \cdot \text{Höhe}$$

Daraus ergeben sich folgende Formeln für den Flächeninhalt A (vgl. Kap. 6)

$$\begin{aligned} A &= \frac{1}{2}a \cdot h_a = \frac{1}{2}b \cdot h_b = \frac{1}{2}c \cdot h_c \\ &= \frac{1}{2}a \cdot b \cdot \sin\gamma = \frac{1}{2}a \cdot c \cdot \sin\beta \\ &= \frac{1}{2}b \cdot c \cdot \sin\alpha \\ &= \sqrt{s(s-a)(s-b)(s-c)} \\ s &= \frac{1}{2}(a+b+c) \end{aligned}$$

Die letzte Formel ist die sogenannte Heronische Flächenformel (nach dem griechischen Mathematiker Heron von Alexandria, 1. Jahrhundert u. Z.), damit ist die Berechnung des Flächeninhalts eines Dreiecks allein mit den Seitenlängen möglich.

3.5.2 Gleichschenklige Dreiecke

Ein Dreieck mit zwei gleich langen Seiten heißt gleichschenklig (Abb. 3.17).

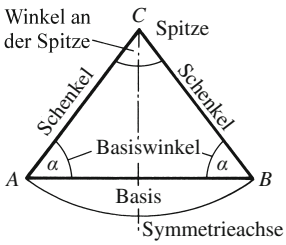


Abb. 3.17 Gleichschenkliges Dreieck

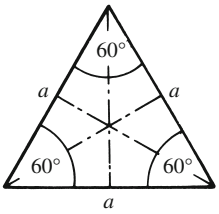


Abb. 3.18 Gleichseitiges Dreieck

Die gleich langen Seiten heißen Schenkel und die dritte Seite Basis des Dreiecks. Die Winkel an der Basis sind die Basiswinkel. Der der Basis gegenüberliegende Punkt heißt Spitze.

Die Basiswinkel sind gleich groß. Höhe, Seitenhalbierende, Mittelsenkrechte und Winkelhalbierende der Basis sind identisch (vgl. Abschn. 3.5.5). Sind umgekehrt in einem Dreieck je zwei dieser Strecken gleich, dann ist das Dreieck gleichschenklig.

3.5.3 Gleichseitige Dreiecke

Ein Dreieck mit drei gleich langen Dreiecksseiten heißt gleichseitig (Abb. 3.18).

In einem gleichseitigen Dreieck sind auch alle Winkel gleich groß, jeder Winkel beträgt also 60°.

Das gleichseitige Dreieck hat drei Symmetrieachsen. Die Höhen, Winkelhalbierenden, Seitenhalbierenden und Mittelsenkrechten fallen beim gleichseitigen Dreieck zusammen (vgl. Abschn. 3.5.5). Folglich fallen auch die Mittelpunkte des Inkreises und des Umkreises mit dem Schwerpunkt des Dreiecks zusammen.

Ein gleichseitiges Dreieck heißt auch reguläres oder regelmäßiges Dreieck.

3.5.4 Rechtwinklige Dreiecke

Ein Dreieck mit einem rechten Winkel, also mit einem Winkel von 90°, heißt rechtwinklig (Abb. 3.19).

Die Summe der beiden anderen (spitzen) Winkel in einem rechtwinkligen Dreieck ist ebenfalls 90°.

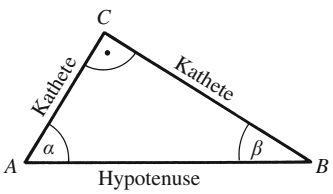


Abb. 3.19 Rechtwinkliges Dreieck ($\alpha + \beta = 90°$)

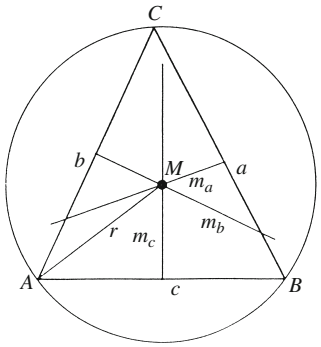


Abb. 3.20 Dreieck mit Umkreis

Die dem rechten Winkel gegenüberliegende Dreiecksseite ist die Hypotenuse, die beiden anderen Seiten (also die Schenkel des rechten Winkels) sind die Katheten des rechtwinkligen Dreiecks.

Für das rechtwinklige Dreieck gelten einige interessante Flächensätze (vgl. Abschn. 3.5.6).

3.5.5 Besondere Geraden, Strecken und Kreise

Der Umkreis eines Dreiecks ist der Kreis durch die drei Eckpunkte des Dreiecks, der dem Dreieck umbeschriebene Kreis (Abb. 3.20).

Der Inkreis eines Dreiecks berührt die drei Dreiecksseiten von innen, er hat die Dreiecksseiten also als Tangenten. Der Inkreis ist der dem Dreieck einbeschriebene Kreis.

Die Mittelsenkrechte einer Strecke ist die Senkrechte durch den Mittelpunkt der Strecke. Beim Dreieck schneiden sich die drei Mittelsenkrechten in einem Punkt M, dem Mittelpunkt des Umkreises. Bei spitzwinkligen Dreiecken liegt M innerhalb des Dreiecks, bei stumpfwinkligen Dreiecken außerhalb und bei rechtwinkligen Dreiecken auf dem Rand (Mittelpunkt der Hypotenuse) des Dreiecks.

Eine Höhe in einem Dreieck ist der Teil des Lotes von einem Eckpunkt auf die gegenüberliegende Seite, der von dem Eckpunkt und dieser Seite (bzw. ihrer Verlängerung) begrenzt wird.

Die drei Höhen eines Dreiecks (bzw. ihre Verlängerungen) schneiden sich in einem Punkt H. Der Höhenschnittpunkt des Dreiecks heißt Orthozentrum des Dreiecks (Abb. 3.21).

Bei spitzwinkligen Dreiecken liegt der Höhenschnittpunkt H im Innern des Dreiecks, bei einem rechtwinkligen Dreieck fällt H mit dem Scheitelpunkt des rechten Winkels zusammen (zwei Höhen fallen mit den Katheten zusammen), bei stumpfwinkligen Dreiecken liegt der Höhenschnittpunkt H außerhalb des Dreiecks (Abb. 3.22 und 3.23).

Die Längen der Höhen im Dreieck werden mit h_a, h_b, h_c bezeichnet.

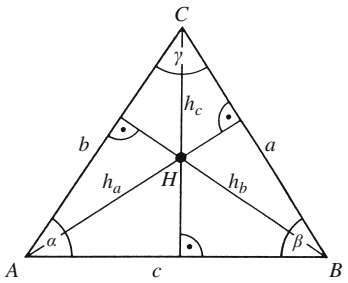


Abb. 3.21 Die Höhen eines Dreiecks schneiden sich in einem Punkt H

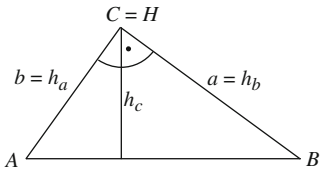


Abb. 3.22 In einem rechtwinkligen Dreieck fällt H mit dem Scheitelpunkt des rechten Winkels zusammen

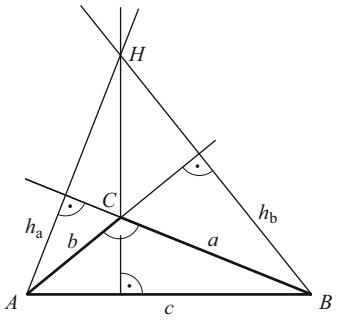


Abb. 3.23 In einem stumpfwinkligen Dreieck liegt H außerhalb des Dreiecks

Sind a, b, c die Dreiecksseiten und α, β, γ die Winkel des Dreiecks, so gilt (vgl. Kap. 6)

$$h_a = c \sin\beta = b \sin\gamma,$$
$$h_b = c \sin\alpha = a \sin\gamma,$$
$$h_c = a \sin\beta = b \sin\alpha$$

Die Längen der Höhen verhalten sich umgekehrt proportional wie die zugehörigen Seitenlängen.

$$h_a : h_b : h_c = \frac{1}{a} : \frac{1}{b} : \frac{1}{c}$$

Eine Winkelhalbierende ist eine Gerade durch den Scheitelpunkt eines Winkels, so dass die beiden Winkel zwischen Gerade und je einem Schenkel gleich sind (Abb. 3.24).

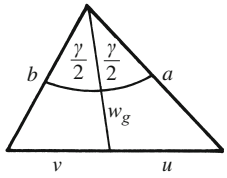


Abb. 3.24 Winkelhalbierende im Dreieck ($u : v = a : b$)

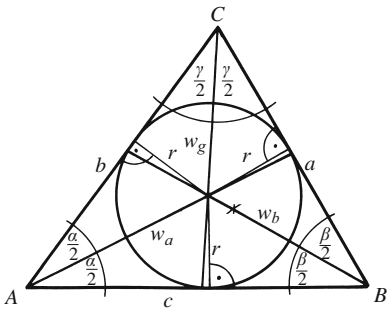


Abb. 3.25 Dreieck mit Inkreis

Im Dreieck sind die drei Winkelhalbierenden Strecken $\overline{PQ}$, wobei P ein Eckpunkt (Scheitelpunkt des entsprechenden Winkels) und Q der Schnittpunkt mit der gegenüberliegenden Seite ist. Die drei Winkelhalbierenden im Dreieck schneiden sich in einem Punkt, dem Mittelpunkt des Inkreises (Abb. 3.25).

Die Längen der Winkelhalbierenden im Dreieck werden mit $w_\alpha, w_\beta, w_\gamma$ bezeichnet.

Für die Winkelhalbierenden gilt (vgl. Kap. 6)

$$w_\alpha = \frac{2bc\cos\frac{\alpha}{2}}{b+c},$$
$$w_\beta = \frac{2ac\cos\frac{\beta}{2}}{a+c}$$
$$w_\gamma = \frac{2ab\cos\frac{\gamma}{2}}{a+b}$$

Eine Winkelhalbierende teilt die gegenüberliegende Seite im Verhältnis der Längen der anliegenden Seiten.

Eine Seitenhalbierende (auch Median genannt) in einem Dreieck ist die Verbindungsstrecke einer Ecke mit dem Mittelpunkt der gegenüberliegenden Seite.

Die drei Seitenhalbierenden eines Dreiecks schneiden sich in einem Punkt S, dem Schwerpunkt des Dreiecks. Der Schwerpunkt teilt die Seitenhalbierenden vom Eckpunkt aus im Verhältnis 2 : 1 (Abb. 3.26).

Die Längen der Seitenhalbierenden im Dreieck werden mit S_a, S_b, S_c bezeichnet.

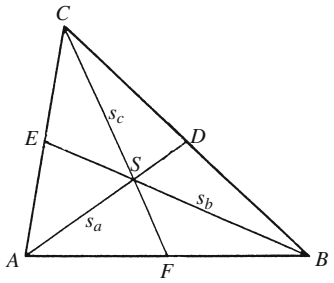


Abb. 3.26 Die drei Seitenhalbierenden eines Dreiecks schneiden sich im Schwerpunkt S ($|\overline{AS}| : |\overline{SD}| = |\overline{BS}| : |\overline{SE}| = |\overline{CS}| : |\overline{SF}| = 2 : 1$)

Für die Seitenhalbierenden gilt (vgl. Kap. 6)

$$s_a = \frac{1}{2}\sqrt{b^2 + c^2 + 2bc\cos\alpha}$$

$$s_b = \frac{1}{2}\sqrt{a^2 + c^2 + 2ac\cos\beta}$$

$$s_c = \frac{1}{2}\sqrt{a^2 + b^2 + 2ab\cos\gamma}$$

Mit Hilfe unterschiedlicher Größen des Dreiecks lassen sich die Radien des Inkreises und des Umkreises berechnen. Dabei ist s der halbe Umfang des Dreiecks, $s = \frac{1}{2}(a + b + c)$.

Radius ρ des Inkreises

$$\rho = \frac{(s-a)(s-b)(s-c)}{s}$$

$$= s \cdot \tan\frac{\alpha}{2} \cdot \tan\frac{\beta}{2} \cdot \tan\frac{\gamma}{2}$$

Radius r des Umkreises

$$r = \frac{a}{2\sin\alpha} = \frac{b}{2\sin\beta} = \frac{c}{2\sin\gamma}$$

$$= \frac{bc}{2h_a} = \frac{ac}{2h_b} = \frac{ab}{2h_c}$$

$$= \frac{abc}{4\sqrt{s(s-a)(s-b)(s-c)}}$$

3.5.6 Flächensätze im rechtwinkligen Dreieck

Kathetensatz

In einem rechtwinkligen Dreieck ist das Quadrat über einer Kathete gleich dem Rechteck aus Hypotenuse und zugehörigem Hypotenusenabschnitt (Abb. 3.27).

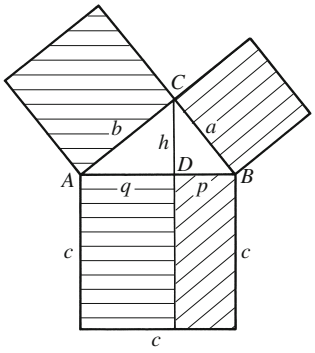


Abb. 3.27 Kathetensatz

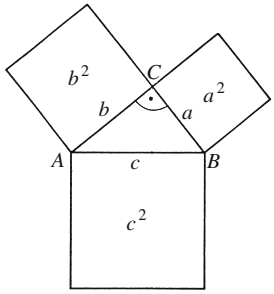


Abb. 3.28 Satz des Pythagoras: $a^2 + b^2 = c^2$

Der Hypotenusenabschnitt ist die Projektion (Parallelprojektion) der entsprechenden Kathete auf die Hypotenuse.

Sind a, b die Kathetenlängen, c die Hypotenusenlänge und p, q die zugehörigen Hypotenusenabschnitte des Dreiecks, so gilt

$$a^2 = pc, \quad b^2 = qc$$

Der Kathetensatz heißt auch erster Satz des Euklid (nach dem hellenistischen Mathematiker Euklid von Alexandria, ∼365–300 v. u. Z.).

Satz des Pythagoras

In einem rechtwinkligen Dreieck ist die Summe der Quadrate über den Katheten gleich dem Quadrat der Hypotenuse (Abb. 3.28).

Sind a und b die Kathetenlängen und c die Hypotenusenlänge, so gilt

$$a^2 + b^2 = c^2$$

Einfacher Beweis des Satzes von Pythagoras (Abb. 3.29):

In ein Quadrat Q_1 der Seitenlänge $a + b$ wird ein Quadrat Q_2 der Seitenlänge c so gelegt, dass die Eckpunkte von Q_2 die Seiten

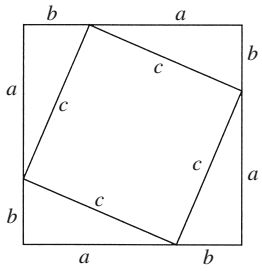


Abb. 3.29 Zum Beweis des Satzes von Pythagoras

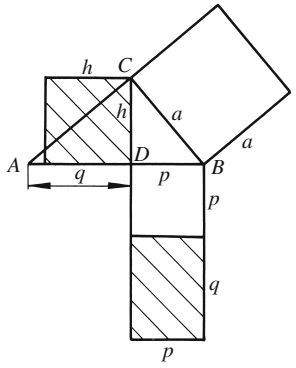


Abb. 3.30 Höhensatz

von Q_1 im Verhältnis $a : b$ teilen. Dann haben die vier innerhalb von Q_1 entstandenen Dreiecke alle den Flächeninhalt $\frac{1}{2}$ ab, und deshalb gilt:

$$\begin{aligned} (a+b)^2 &= c^2 + 4 \cdot \frac{1}{2}ab \\ \Rightarrow \quad a^2 + 2ab + b^2 &= c^2 + 2ab \\ \Rightarrow \quad a^2 + b^2 &= c^2 \end{aligned}$$

Höhensatz

In einem rechtwinkligen Dreieck ist das Quadrat über der Höhe auf der Hypotenuse gleich dem Rechteck aus den beiden durch die Höhe gebildeten Hypotenusenabschnitten (Abb. 3.30).

Bezeichnet man die Höhe mit h und die Hypotenusenabschnitte mit p und q, so gilt

$$h^2 = pq$$

Beweis des Höhensatzes:

Nach dem Satz des Pythagoras gilt $a^2 = h^2 + p^2$ und nach dem Kathetensatz $a^2 = pc = p(p+q)$.

Subtraktion der beiden Gleichungen ergibt $0 = h^2 + p^2 - (p^2 + pq) = h^2 - pq$, woraus $h^2 = pq$ folgt.

Der Höhensatz heißt auch zweiter Satz des Euklid.

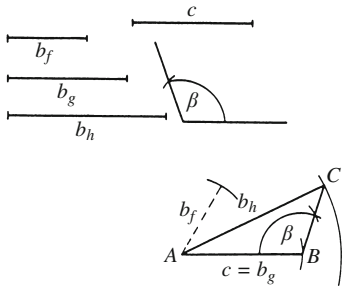


Abb. 3.31 Grundkonstruktion SSW (gegeben: $b, c, \beta \geq 90°$)

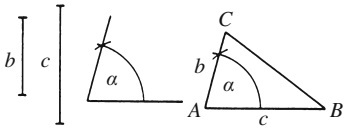


Abb. 3.32 Grundkonstruktion SWS

Grundkonstruktion SSW

Gegeben: $b, c, \beta \geq 90°$ (Seite, Seite, Winkel)

Konstruktion: Nach Konstruktion von $\overline{AB}$ und Antragen von β wie oben sind hier folgende Fälle möglich (Abb. 3.31):

f) Der Kreis um A mit dem Radius b_f schneidet den freien Schenkel des gegebenen Winkels β nicht; keine Lösung.
g) Der Kreis um A mit dem Radius b_g geht durch den Punkt B; „Lösung“ ist ein zur Strecke $(\overline{AB})$ entartetes Dreieck.
h) Der Kreis um A mit dem Radius b_h schneidet den freien Schenkel des gegebenen Winkels β in C; eine Lösung.

Der Kongruenzsatz SSW gilt hier nur für den Fall *h*).

Grundkonstruktion SWS

Gegeben: b, α, c (Seite, Winkel, Seite)

Konstruktion: Den Schenkeln des Winkels α gibt man die Längen $b = |\overline{AC}|$ und $c = |\overline{AB}|$. Man verbindet B und C (Abb. 3.32).

Bedingung: $\alpha < 180°$

Grundkonstruktion SSS

Gegeben: a, b, c (Seite, Seite, Seite).

Konstruktion: Man zeichnet $c = |\overline{AB}|$ und schlägt um A einen Kreisbogen mit dem Radius b und um B einen Kreisbogen mit dem Radius a. Der Eckpunkt C ist der Schnittpunkt der Kreisbögen, der „oberhalb“ von $\overline{AB}$ liegt (A, B, C folgen im mathematisch positiven Drehsinn, also entgegen dem Uhrzeigersinn aufeinander). Anschließend wird noch C mit den Eckpunkten A und B verbunden (Abb. 3.33).

Bedingungen: $a < b + c, b < a + c, c < a + b$ (Dreiecksungleichungen)

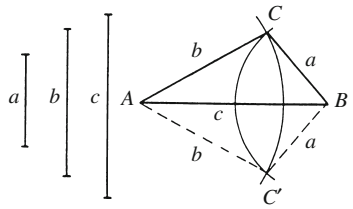


Abb. 3.33 Grundkonstruktion SSS

Anmerkung: Wenn sich, wie bei der letzten Konstruktion, zwei spiegelbildlich gleiche Lösungen ergeben (denn die Kreise schneiden sich zweimal, in C und C', falls die Dreiecksungleichungen erfüllt sind), wählt man diejenige aus, bei der die Punkte A, B, C entgegen dem Uhrzeigersinn, also im mathematisch positiven Sinn, aufeinander folgen.

3.6 Vierecke

3.6.1 Allgemeine Vierecke

Ein Viereck besteht aus vier Punkten A, B, C, D, von denen keine drei auf einer Geraden liegen, und den Strecken $\overline{AB}, \overline{BC}, \overline{CD}, \overline{DA}$ (Abb. 3.34).

Die Punkte A, B, C, D sind die Eckpunkte des Vierecks, die Strecken $\overline{AB}, \overline{BC}, \overline{CD}, \overline{DA}$ sind die Seiten des Vierecks, und ihre Längen $|\overline{AB}|, |\overline{BC}|, |\overline{CD}|, |\overline{DA}|$ sind die Seitenlängen des Vierecks.

Meistens werden die Seitenlängen mit a, b, c, d und die Innenwinkel des Vierecks mit $\alpha, \beta, \gamma, \delta$ bezeichnet, und zwar so, dass $a = |\overline{AB}|$, $b = |\overline{BC}|$, $c = |\overline{CD}|$, $d = |\overline{DA}|$ und dass der Winkel α den Scheitelpunkt A, der Winkel β den Scheitelpunkt B, der Winkel γ den Scheitelpunkt C und der Winkel δ den Scheitelpunkt D hat.

Die Strecken $\overline{AC}$ und $\overline{BD}$ heißen Diagonalen des Vierecks, ihre Längen werden meist mit e und f bezeichnet: $e = |\overline{AC}|, f = |\overline{BD}|$.

Abkürzend verwendet man für ein Viereck das Symbol $\square$, und für ein Viereck mit den Eckpunkten A, B, C, D schreibt man $\square(ABCD)$.

Die Winkelsumme in einem beliebigen Viereck beträgt 360°.

$$\alpha + \beta + \gamma + \delta = 360°$$

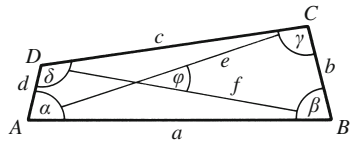


Abb. 3.34 Bezeichnungen im Viereck

Für die Längen der Seiten und der Diagonalen in einem Viereck gilt folgender Zusammenhang:

In einem Viereck ist das Produkt der Diagonalenlängen kleiner oder gleich der Summe der Produkte der Längen je zwei gegenüberliegender Seiten.

$$ef \leq ac + bd$$

Die Gleichheit gilt genau dann, wenn das Viereck ein Sehnenviereck ist (bei einem Sehnenviereck liegen alle vier Punkte auf einem Kreis).

Diese Aussage ist der verallgemeinerte Satz des Ptolemäus (vgl. Abschn. 3.6.8).

Der Umfang u eines Vierecks ist die Summe der Seitenlängen.

$$u = a + b + c + d$$

Für den Flächeninhalt A eines Vierecks gilt (vgl. Kap. 6):

$$\begin{aligned} A &= \frac{1}{2}(ad\sin\alpha + bc\sin\gamma) = \frac{1}{2}(ab\sin\beta + cd\sin\delta) \\ &= \frac{1}{2}ef\sin\varphi \\ &= \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2\frac{\alpha+\gamma}{2}} \\ &= \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2\frac{\beta+\delta}{2}} \end{aligned}$$

Dabei ist φ der Winkel zwischen den Diagonalen und s der halbe Umfang des Vierecks, also $s = \frac{1}{2}(a + b + c + d)$.

Die beiden letzten Formeln sind Verallgemeinerungen der Heronischen Flächenformel für Dreiecke.

3.6.2 Trapeze

Ein Trapez ist ein Viereck, bei dem zwei Seiten zueinander parallel sind (Abb. 3.35).

Die parallelen Seiten heißen Grundlinien und die anderen beiden Seiten Schenkel des Trapezes. Die Verbindungsstrecke der Mittelpunkte der Schenkel heißt Mittellinie, sie ist parallel zu den Grundlinien. Die Höhe eines Trapezes ist der Abstand der parallelen Grundlinien.

Sind die Schenkel gleich lang, so heißt das Trapez gleichschenklig.

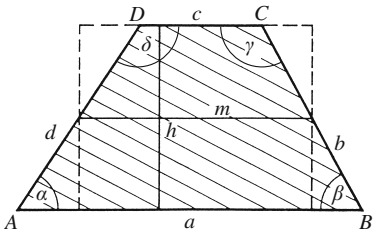


Abb. 3.35 Trapez

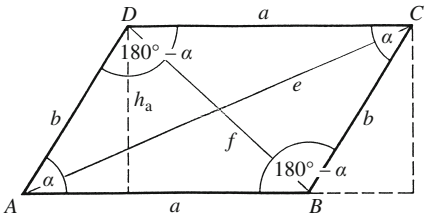


Abb. 3.36 Parallelogramm

Ist m die Länge der Mittellinie, so gilt $m = \frac{a+c}{2}$.

Das Trapez mit den Grundlinienlängen a und c und der Höhe h ist flächengleich einem Rechteck mit den Seitenlängen m und h.

Für den Flächeninhalt A des Trapezes gilt

$$A = mh = \frac{a+c}{2}h$$

3.6.3 Parallelogramme

Ein Parallelogramm ist ein Viereck, bei dem die beiden jeweils einander gegenüberliegenden Seiten parallel sind (Abb. 3.36).

Einander gegenüberliegende Seiten im Parallelogramm sind gleich lang; einander gegenüberliegende Winkel sind gleich groß; benachbarte Winkel ergänzen sich zu 180°; die Diagonalen halbieren sich in ihrem Schnittpunkt.

Das Parallelogramm mit den Seitenlängen a und b ist flächengleich einem Rechteck mit den Seitenlängen a und h_a (oder b und h_b).

Für den Umfang u und den Flächeninhalt A eines Parallelogramms mit den Seitenlängen a und b gilt

$$\begin{aligned} &\text{Umfang} && u = 2a + 2b \\ &\text{Flächeninhalt} && A = ah_a = bh_b = ab\sin\alpha \end{aligned}$$

Für die Längen der Seiten und der Diagonalen in einem Parallelogramm gilt folgender Zusammenhang: In einem Parallelogramm ist die Summe der Quadrate der Seitenlängen gleich der Summe der Quadrate der Diagonalenlängen, also

$$2a^2 + 2b^2 = e^2 + f^2$$

Diese Aussage ist der Satz von Apollonios (nach dem hellenistischen Geometer und Astronom Apollonios von Perge, ∼262–190 v. u. Z.).

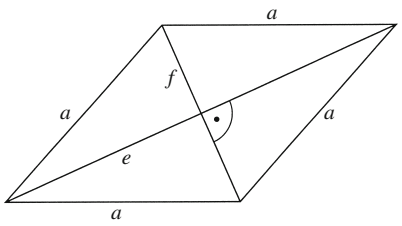


Abb. 3.37 Rhombus

3.6.4 Rhomben

Ein Rhombus ist ein Parallelogramm mit gleich langen Seiten.

Damit gelten für Rhomben alle Eigenschaften von Parallelogrammen. Darüber hinaus gilt: Die Diagonalen eines Rhombus halbieren sich in ihrem Schnittpunkt, sie halbieren alle Winkel, und sie stehen senkrecht aufeinander. Statt Rhombus sagt man auch Raute (Abb. 3.37).

Für den Umfang u und den Flächeninhalt A eines Rhombus der Seitenlänge a gilt (e und f sind die Längen der Diagonalen)

$$\begin{aligned} &\text{Umfang} && u = 4a \\ &\text{Flächeninhalt} && A = a^2\sin\alpha = \frac{1}{2}ef \end{aligned}$$

3.6.5 Rechtecke

Ein Rechteck ist ein Parallelogramm mit vier rechten Winkeln.

Die Diagonalen eines Rechtecks halbieren sich in ihrem Schnittpunkt und sind gleich lang, es gilt $e = f = \sqrt{a^2 + b^2}$ (Abb. 3.38).

Für den Umfang u und den Flächeninhalt A eines Rechtecks mit den Seitenlängen a und b gilt

$$\begin{aligned} &\text{Umfang} && u = 2a + 2b \\ &\text{Flächeninhalt} && A = ab \end{aligned}$$

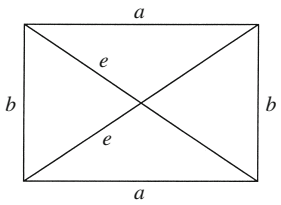


Abb. 3.38 Rechteck

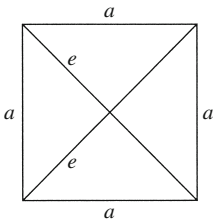


Abb. 3.39 Quadrat

3.6.6 Quadrate

Ein Quadrat ist ein Rechteck mit gleich langen Seiten. Die Diagonalen eines Quadrats sind gleich lang und stehen senkrecht aufeinander. Es gilt $e = f = a\sqrt{2}$ (Abb. 3.39).

Ein Quadrat ist auch ein Rhombus (eine Raute) mit vier rechten Winkeln.

Für den Umfang u und den Flächeninhalt A eines Quadrats der Seitenlänge a gilt

$$\text{Umfang} \qquad u = 4a$$

$$\text{Flächeninhalt} \qquad A = a^2 = \frac{1}{2}e^2$$

Ein Quadrat heißt auch reguläres oder regelmäßiges Viereck.

3.6.7 Drachen

Ein Drachen ist ein Viereck mit zwei Paaren gleich langer benachbarter Seiten (Abb. 3.40).

Statt Drachen sagt man auch Drachenviereck.

Die Diagonalen eines Drachens stehen senkrecht aufeinander.

Ein Drachen mit vier gleich langen Seiten ist ein Rhombus (eine Raute).

Sind e und f die Längen der Diagonalen, dann gilt für den Flächeninhalt A des Drachens

$$A = \frac{1}{2}ef$$

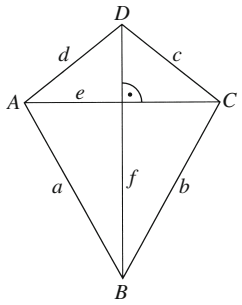


Abb. 3.40 Drachen

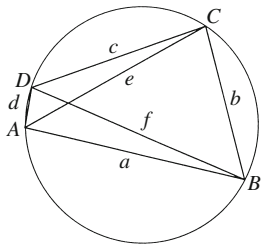


Abb. 3.41 Sehnenviereck

3.6.8 Sehnenvierecke

Ein Sehnenviereck ist ein Viereck, bei dem alle vier Eckpunkte auf einem Kreis liegen (Abb. 3.41).

Der Kreis heißt Umkreis des Vierecks, die Seiten sind Sehnen dieses Kreises.

Ein Viereck ist genau dann ein Sehnenviereck, wenn gegenüberliegende Winkel Supplementwinkel sind, sich also zu 180° ergänzen.

$$\alpha + \gamma = \beta + \delta = 180°$$

In einem Sehnenviereck ist das Produkt der Diagonalenlängen gleich der Summe der Produkte der Längen je zwei gegenüberliegender Seiten (Satz von Ptolemäus).

$$ef = ac + bd$$

Die Formel wurde hergeleitet und bewiesen von dem hellenistischen Geometer und Astronom Ptolemaios von Alexandria (~83–161 u. Z.).

In einem Sehnenviereck verhalten sich die Längen der Diagonalen wie die Summen der Produkte der Längen jener Seitenpaare,

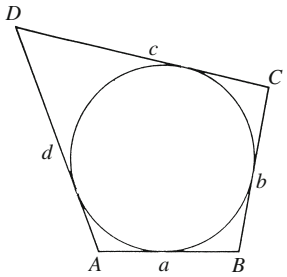


Abb. 3.42 Tangentenviereck

die sich in den Endpunkten der Diagonalen treffen (Satz von Brahmagupta).

$$\frac{e}{f} = \frac{ab + cd}{ad + bc}$$

Die Formel wurde von dem Inder Brahmagupta (6./7. Jahrhundert u. Z.) entdeckt, der erste Beweis stammt von dem deutschen Mathematiker Johannes Müller, genannt Regiomontanus (1436–1476).

Durch Multiplikation bzw. Division dieser beiden Formeln erhält man Ausdrücke für die Längen der beiden Diagonalen.

$$e = \sqrt{\frac{(ab + cd)(ac + bd)}{ad + bc}}$$

$$f = \sqrt{\frac{(ad + bc)(ac + bd)}{ab + cd}}$$

Für den Flächeninhalt A des Sehnenvierecks gilt, wenn s der halbe Umfang des Sehnenvierecks ist, also $s = \frac{1}{2}(a+b+c+d)$,

$$A = \sqrt{(s-a)(s-b)(s-c)(s-d)}$$

3.6.9 Tangentenvierecke

Ein Tangentenviereck ist ein Viereck, bei dem alle vier Seiten denselben Kreis berühren (Abb. 3.42).

Der Kreis heißt Inkreis des Vierecks, die Seiten sind Tangenten dieses Kreises.

Ein Viereck ist genau dann ein Tangentenviereck, wenn die Summe der Längen zweier gegenüberliegender Seiten gleich der Summe der Längen der beiden anderen Seiten ist.

$$a + c = b + d$$

3.7 Reguläre n-Ecke

Ein n-Eck besteht aus n Punkten, den Eckpunkten des n-Ecks, und n Seiten, den Strecken zwischen den Eckpunkten (Abb. 3.43).

Haben alle Seiten die gleiche Länge und sind alle Innenwinkel gleich groß, dann heißt das n-Eck regulär oder regelmäßig.

Bei einem regulären n-Eck liegen alle Eckpunkte auf einem Kreis, dem Umkreis des n-Ecks, und alle Seiten sind Tangenten eines einbeschriebenen Kreises, dem Inkreis des n-Ecks. Die Seiten sind Sehnen des Umkreises.

Ein reguläres Dreieck ist ein gleichseitiges Dreieck, ein reguläres Viereck ist ein Quadrat.

Die Summe der Innenwinkel in einem beliebigen n-Eck ist $(n-2)\cdot 180°$. Da alle n Innenwinkel γ_n gleich groß sind, gilt

Innenwinkel

$$\gamma_n = \frac{n-2}{n} \cdot 180°$$

Durch die Verbindungsstrecken der Eckpunkte mit dem Mittelpunkt des Umkreises wird das reguläre n-Eck in n kongruente Dreiecke zerlegt. Für die Basiswinkel β_n und die Zentriwinkel α_n gilt

$$\text{Basiswinkel} \quad \beta_n = \frac{1}{2}\gamma_n = \frac{n-2}{n} \cdot 90°$$

$$\text{Zentriwinkel} \quad \alpha_n = \frac{360°}{n}$$

Für die Seitenlänge a_n, den Umkreisradius r und den Inkreisradius ρ_n des regulären n-Ecks gilt (vgl. Kap. 6):

$$\text{Seitenlänge} \quad a_n = 2\sqrt{r^2 - \rho_n^2} = 2r \sin\frac{\alpha_n}{2}$$

$$= 2\rho_n \tan\frac{\alpha_n}{2}$$

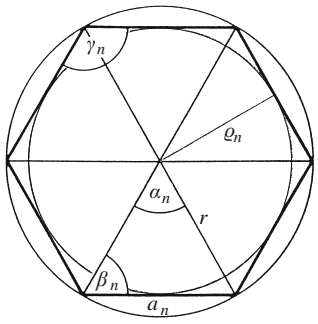


Abb. 3.43 Bezeichnungen im regulären n-Eck

Tab. 3.1 Übersicht über die regulären n-Ecke für kleine n (r = Umkreisradius)

n	Innenwinkel γ_n	Zentriwinkel α_n	Seitenlänge a_n	Umfang u_n	Flächeninhalt A_n
3	60°	120°	$r\sqrt{3}$	$2r \cdot 2{,}5980\ldots$	$\left(\frac{3}{4}\sqrt{3}\right) r^2$
4	90°	90°	$r\sqrt{2}$	$2r \cdot 2{,}8284\ldots$	$2r^2$
5	108°	72°	$\frac{r}{2}\sqrt{10-2\sqrt{5}}$	$2r \cdot 2{,}9389\ldots$	$\left(\frac{5}{8}\sqrt{10+2\sqrt{5}}\right) r^2$
6	120°	60°	r	$2r \cdot 3$	$\left(\frac{3}{2}\sqrt{3}\right) r^2$
8	135°	45°	$r\sqrt{2-\sqrt{2}}$	$2r \cdot 3{,}0614\ldots$	$\left(2\sqrt{2}\right) r^2$
10	144°	36°	$\frac{r}{2}(\sqrt{5}-1)$	$2r \cdot 3{,}0901\ldots$	$\left(\frac{5}{4}\sqrt{10-2\sqrt{5}}\right) r^2$
12	150°	30°	$r\sqrt{2-\sqrt{3}}$	$2r \cdot 3{,}1058\ldots$	$3r^2$

$$\text{Umkreisradius} \qquad r = \frac{a_n}{2\sin\frac{\alpha_n}{2}}$$

$$\text{Inkreisradius} \qquad \rho_n = \frac{a_n}{2}\cot\frac{\alpha_n}{2}$$

Für den Umfang u_n und den Flächeninhalt A_n des regulären n-Ecks ergibt sich dann

$$\text{Umfang} \qquad u_n = na_n$$

$$\text{Flächeninhalt} \qquad A_n = \frac{1}{2}na_n\rho_n = \frac{1}{2}nr^2\sin\alpha_n = \frac{1}{4}na_n^2\cot\frac{\alpha_n}{2}$$

Mit wachsendem n nähert der Umfang u_n sich dem Umfang $2r \cdot \pi = 2r \cdot 3{,}1415\ldots$ und der Flächeninhalt A_n sich dem Flächeninhalt πr^2 des Kreises mit dem Radius r an.

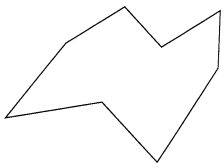


Abb. 3.44 Einfaches Polygon

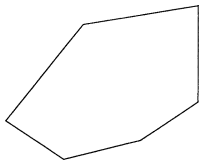


Abb. 3.45 Konvexes Polygon

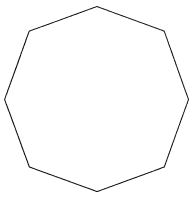


Abb. 3.46 Reguläres Polygon

3.8 Polygone

Ein Polygon ist ein geschlossener Streckenzug der Ebene.

Ein Polygon oder Vieleck mit n Eckpunkten ist ein n-Eck. Die Strecken zwischen den Eckpunkten sind die Seiten des Polygons.

Ein einfaches Polygon teilt die Ebene in zwei Gebiete, das Innere und das Äußere, die durch die Seiten des Polygons getrennt werden (Abb. 3.44).

Die Länge des geschlossenen Streckenzugs ist der Umfang des Polygons, und die Fläche des Inneren ist der Flächeninhalt des Polygons. Eine Verbindungsstrecke zweier nicht benachbarter Eckpunkte ist eine Diagonale des Polygons. (Zwei Eckpunkte heißen benachbart, wenn es zwischen ihnen eine Seite gibt.) Jeder Eckpunkt kann also mit $n-3$ anderen Eckpunkten durch eine Diagonale verbunden werden (denn jeder der n Eckpunkte hat $n-3$ andere Eckpunkte nicht als Nachbarn). Die Innenwinkel des Polygons haben als Scheitelpunkte die Eckpunkte des Polygons. In einem konvexen Polygon liegen alle Diagonalen im Innern des Polygons, und alle Innenwinkel sind kleiner als 180° (Abb. 3.45).

Da man das Innere eines n-Ecks durch $n-2$ sich nicht überschneidende Diagonalen in $n-2$ Dreiecke zerlegen kann und die Winkelsumme im Dreieck 180° ist, beträgt die Summe der Innenwinkel in einem beliebigen n-Eck $(n-2) \cdot 180°$.

In einem regulären (regelmäßigen) Polygon haben alle Seiten die gleiche Länge, und alle Innenwinkel sind gleich groß (Abb. 3.46).

3.9 Kreise

3.9.1 Definitionen

Ein Kreis ist der geometrische Ort aller Punkte der Ebene, die von einem festen Punkt M einen konstanten Abstand r haben.

Der Punkt M ist der Mittelpunkt und r der Radius des Kreises.

Zur Unterscheidung von der durch einen Kreis in der Ebene abgegrenzten Fläche, der Kreisfläche, wird der Kreis selbst auch als Kreisperipherie oder Kreisrand bezeichnet (Abb. 3.47).

Einen Kreis mit dem Radius $r = 0$ nennt man entartet. Ein Kreis mit dem Radius $r = 1$ heißt Einheitskreis.

Kreise kürzt man oft mit k ab, und für die Peripherie des Kreises mit dem Radius r und dem Mittelpunkt M schreibt man $k(M, r)$.

Jede Gerade durch zwei Punkte der Kreisperipherie nennt man Sekante. Der zwischen den Punkten gelegene Teil der Sekante heißt Sehne. Eine Sehne durch den Mittelpunkt heißt Durchmesser des Kreises. Durchmesser sind die größten Sehnen des Kreises. Für die Länge d eines Durchmessers gilt $d = 2r$.

Winkel, deren Scheitelpunkt ein Punkt der Kreisperipherie ist und deren Schenkel Sekanten des Kreises sind, heißen Peripherie- oder Umfangswinkel (Abb. 3.48).

Winkel, deren Scheitelpunkt der Kreismittelpunkt ist, nennt man Zentri- oder Mittelpunktswinkel. Der durch einen Peripherie- oder Zentriwinkel ausgeschnittene Teil der Kreisperipherie heißt Kreisbogen.

Für den Umfang u und die Fläche A eines Kreises mit dem Radius r und dem Durchmesser d gilt

$$\text{Kreisumfang} \quad u = 2\pi r = \pi d$$

$$\text{Kreisfläche} \quad A = \pi r^2 = \frac{\pi}{4} d^2$$

Das Verhältnis des Umfangs zum Durchmesser eines beliebigen Kreises definiert die so genannte Kreiszahl $\pi = 3{,}1415926535\ldots$ Die Zahl π (kleines „Pi“) ist transzendent (vgl. Abschn. 1.3).

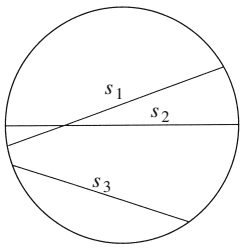


Abb. 3.47 Sehnen eines Kreises (Sehne s_2 ist auch Durchmesser)

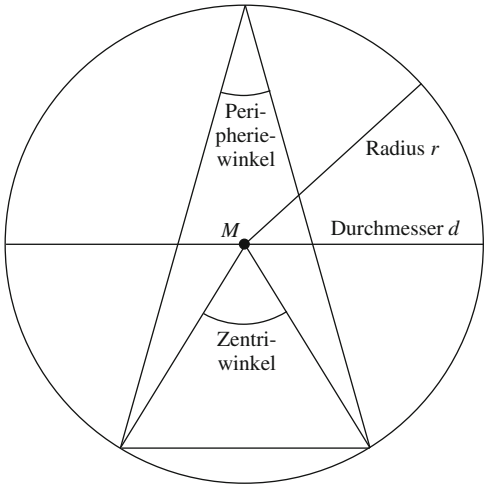


Abb. 3.48 Bezeichnungen am Kreis

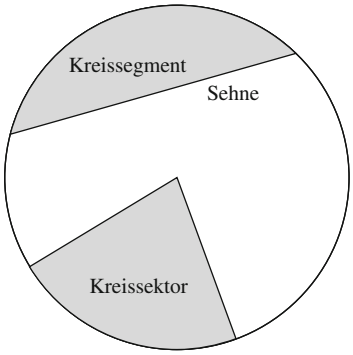


Abb. 3.49 Kreissektor und Kreissegment

Ein Kreis ist festgelegt durch den Mittelpunkt und einen weiteren Punkt oder durch drei Punkte (die nicht alle auf einer Geraden liegen).

Kreise, die den gleichen Mittelpunkt haben, heißen konzentrische Kreise. Zwei Kreise mit verschiedenen Mittelpunkten nennt man exzentrisch.

Die von zwei konzentrischen Kreisen begrenzte Fläche heißt Kreisring. Ist R der Radius des äußeren Kreises und r der Radius des inneren Kreises, dann gilt für den Flächeninhalt A des Kreisrings

$$A = \pi(R^2 - r^2)$$

3.9.2 Kreissektoren

Ein Kreissektor oder Kreisausschnitt ist der Teil der Fläche eines Kreises, der von den Schenkeln eines Zentriwinkels und dem zugehörigen Kreisbogen begrenzt wird (Abb. 3.49 und 3.50).

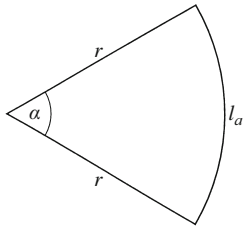


Abb. 3.50 Bezeichnungen Kreissektor

Ist α ein Zentriwinkel und hat der Kreis den Radius r, dann ergibt sich für die Länge l_α des Kreisbogens

$$l_\alpha = \frac{\alpha}{180°}\pi r$$

Für die Fläche A_α des Kreissektors gilt

$$A_\alpha = \frac{\alpha}{360°}\pi r^2 = \frac{1}{2} r l_\alpha$$

3.9.3 Kreissegmente

Ein Kreissegment oder Kreisabschnitt ist der Teil der Fläche eines Kreises, der von einer Sehne $\overline{AB}$ und einem der zugehörigen Kreisbögen $\overset{\frown}{AB}$ oder $\overset{\frown}{BA}$ begrenzt wird (Abb. 3.51).

Ein Kreissegment ist also der Teil eines Kreissektors, der zwischen dem Kreisbogen und der zugehörigen Sehne liegt. Ein Kreis wird von einer Sehne in zwei Segmente zerlegt.

Ist r der Radius eines Kreises, α ein Zentriwinkel, s die Länge der zugehörigen Sehne und h die Höhe des Kreissegments, dann berechnet man $s = 2r \sin\frac{\alpha}{2}$ und $h = 2r \sin^2\frac{\alpha}{4}$, woraus sich für die Länge l_α des Kreisbogens ergibt

$$l_\alpha = \frac{\alpha}{180°}\pi r$$

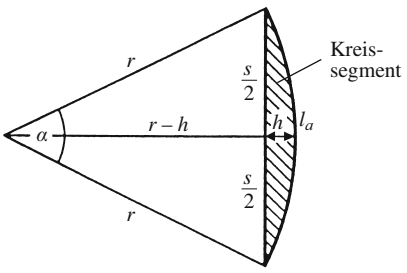


Abb. 3.51 Bezeichnungen Kreissegment

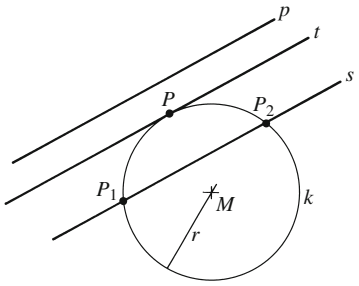


Abb. 3.52 Sekante s, Tangente t, Passante p

Für die Fläche A_α des Kreissegments gilt

$$A_\alpha = \frac{1}{2}[r l_\alpha - s(r-h)]$$

3.9.4 Kreise und Geraden

Ein Kreis und eine Gerade können drei grundsätzlich verschiedene Lagen zueinander haben (Abb. 3.52):

Die Gerade ist eine Passante p, sie hat mit dem Kreis keinen Punkt gemeinsam.

Die Gerade ist eine Tangente t, sie hat mit dem Kreis genau einen Punkt, den Berührungspunkt P, gemeinsam.

Die Gerade ist eine Sekante s, sie hat mit dem Kreis zwei Punkte, die Schnittpunkte P_1 und P_2, gemeinsam.

3.9.5 Bogenmaß

Neben dem Gradmaß gibt es das Bogenmaß zur Winkelmessung.

Beim Bogenmaß wird die Größe eines Zentriwinkels α in einem beliebigen Kreis durch das Verhältnis des zugehörigen Kreisbogens b zum Radius r des Kreises angegeben. Der Quotient $\frac{b}{r}$ heißt Bogenmaß des Winkels α (Abb. 3.53).

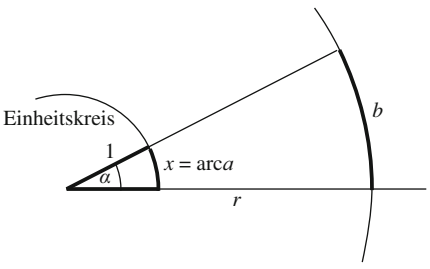


Abb. 3.53 Zusammenhang zwischen Gradmaß (α) und Bogenmaß ($x = \text{arc}\,\alpha$) eines Winkels

Die Einheit des Bogenmaßes ist der Radiant (rad), also der Zentriwinkel, dessen Bogen gleich dem Radius ist. Mitunter schreibt man arc α (Arcus α) für das Bogenmaß des Winkels α.

Bogenmaß

$$\operatorname{arc}\alpha = \frac{b}{r}$$

Da der Einheitskreis, also der Kreis mit dem Radius $r = 1$, den Umfang 2π hat, ist das Bogenmaß des Vollwinkels 2π.

$$2\pi \text{ rad} = 360^\circ \quad \text{oder} \quad 1 \text{ rad} = \frac{360^\circ}{2\pi} \approx 57{,}2958^\circ$$

Bezeichnet α den in Grad und $x = \operatorname{arc}\alpha$ den in Radiant gemessenen Winkel, so gilt für die Umrechnung von Gradmaß und Bogenmaß eines Winkels

$$x = \frac{\pi}{180^\circ} \cdot \alpha, \quad \alpha = \frac{180^\circ}{\pi} \cdot x$$

Beispiele zur Umrechnung

$$\begin{aligned}
1^\circ &= \frac{\pi}{180} \text{ rad} = 0{,}0174\ldots \text{ rad}\\
10^\circ &= \frac{10\pi}{180} \text{ rad} = \frac{\pi}{18} \text{ rad} = 0{,}1745\ldots \text{ rad}\\
30^\circ &= \frac{\pi}{6} \text{ rad}\\
57{,}2957\ldots^\circ &= 1 \text{ rad}\\
45^\circ &= \frac{\pi}{4} \text{ rad}\\
0{,}5 \text{ rad} &= \frac{0{,}5 \cdot 180^\circ}{\pi} = 28{,}6478\ldots^\circ\\
90^\circ &= \frac{\pi}{2} \text{ rad}\\
\frac{\pi}{3} \text{ rad} &= \frac{\pi}{3} \cdot \frac{180^\circ}{\pi} = 60^\circ
\end{aligned}$$

◀

3.10 Symmetrie

3.10.1 Punktsymmetrie

Eine ebene Figur F heißt punkt- oder zentralsymmetrisch, wenn sich in ihrer Ebene ein Punkt P angeben lässt, so dass F durch eine Spiegelung an P in sich übergeführt wird. Der Punkt P heißt dann Symmetriezentrum.

Beispiele

Folgende Figuren sind punktsymmetrisch:

1. Strecke mit ihrem Mittelpunkt als Symmetriezentrum
2. Rechteck mit seinem Mittelpunkt als Symmetriezentrum
3. Ellipse mit ihrem Mittelpunkt als Symmetriezentrum (vgl. Abschn. 7.5.1)

◀

3.10.2 Achsensymmetrie

Eine ebene Figur F heißt achsen- oder axialsymmetrisch, wenn sich in ihrer Ebene eine Gerade g angeben lässt, so dass F durch eine Spiegelung an g in sich übergeführt wird. Die Gerade g heißt dann Symmetrieachse.

Beispiele

Folgende Figuren sind achsensymmetrisch:

1. Gleichseitiges Dreieck mit einer der Winkelhalbierenden als Symmetrieachse
2. Rechteck mit einer Mittellinie als Symmetrieachse
3. Kreis mit einer beliebigen Geraden durch den Mittelpunkt als Symmetrieachse

◀

3.11 Ähnlichkeit

3.11.1 Zentrische Streckung

Die zentrische Streckung ist eine Abbildung, bei der für jedes Element Bild Q und Urbild P auf einem Strahl durch einen festen Punkt Z, dem Zentrum, liegen und für jedes Element das Verhältnis der Länge der Strecke vom Bild zum Zentrum zu der Länge der Strecke vom Urbild zum Zentrum konstant ist (Abb. 3.54).

$$\frac{|\overline{ZQ}|}{|\overline{ZP}|} = k \quad (k \text{ konstant})$$

Eigenschaften:

Die Bilder von Strecke, Strahl, Gerade sind wieder Strecke, Strahl, Gerade. Bild und Urbild von Strecke, Strahl, Gerade sind zueinander parallel.

Entsprechende Winkel von Bild und Urbild sind gleich.

Die Längen entsprechender Strecken von Bild und Urbild haben das gleiche Verhältnis, und zwar den Betrag des Streckungsfaktors k, also $|k|$.

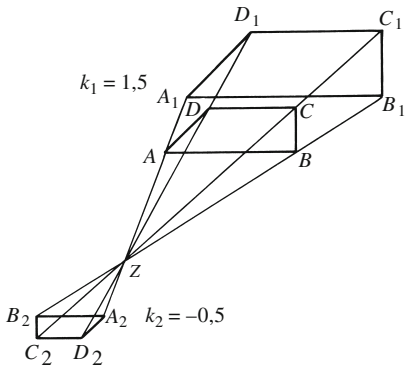


Abb. 3.54 Zentrische Streckung

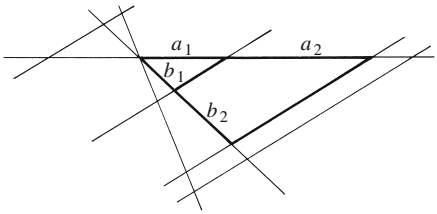


Abb. 3.55 Erster Strahlensatz: $a_1 : a_2 = b_1 : b_2$

3.11.2 Strahlensätze

Unmittelbare Anwendungen der zentrischen Streckung sind die Strahlensätze.

Erster Strahlensatz

Werden zwei Strahlen mit gleichem Anfangspunkt (Zentrum) von Parallelen geschnitten, so verhalten sich die Längen der Abschnitte eines Strahls wie die Längen entsprechender Abschnitte des anderen Strahls (Abb. 3.55).

$$a_1 : a_2 = b_1 : b_2$$

Zweiter Strahlensatz

Werden zwei Strahlen mit gleichem Anfangspunkt von Parallelen geschnitten, so verhalten sich die Längen der zwischen den Strahlen liegenden Abschnitte wie die Längen der zugehörigen vom Anfangspunkt aus gemessenen Abschnitte auf den Strahlen (Abb. 3.56).

$$c_1 : c_3 = a_1 : a_3$$

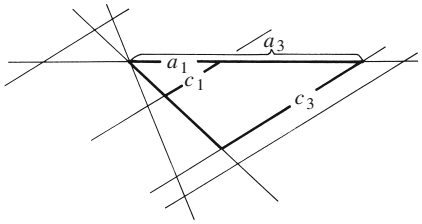


Abb. 3.56 Zweiter Strahlensatz: $c_1 : c_3 = a_1 : a_3$

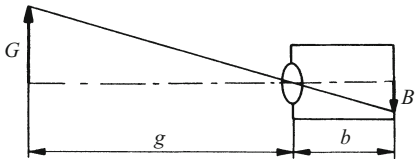


Abb. 3.57 Anwendung des zweiten Strahlensatzes

Beispiel

Welche Länge B ergibt sich für das Bild eines Gegenstands der Länge G bei einer Bildweite b und einer Gegenstandsweite g (Abb. 3.57)?

Nach dem zweiten Strahlensatz gilt

$$\frac{g}{b} = \frac{G}{B} \quad \Rightarrow \quad B = G\frac{b}{g}$$

Anmerkung: Ist $g \gg b$ (g sehr groß gegenüber b), dann kann man $b \approx f$ (f Brennweite) setzen und damit die zu erwartende Bildgröße abschätzen. ◀

3.11.3 Ähnliche Figuren

Geometrische Figuren heißen ähnlich, wenn sie nach geeigneter Parallelverschiebung, Drehung, Spiegelung durch zentrische Streckung zur Deckung gebracht werden können (Abb. 3.58, 3.59, 3.60, 3.61 und 3.62).

So sind zum Beispiel zwei Quadrate ähnlich (mit beliebigen Seitenlängen) oder zwei Kreise (mit beliebigen Radien und be-

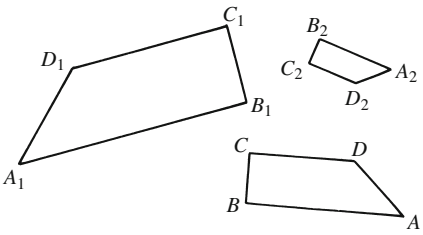


Abb. 3.58 Ähnliche Figuren

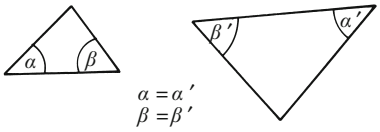


Abb. 3.59 Ähnliche Dreiecke

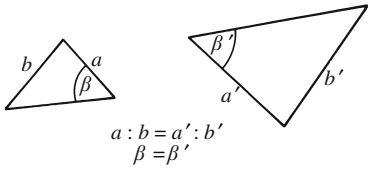


Abb. 3.60 Ähnliche Dreiecke

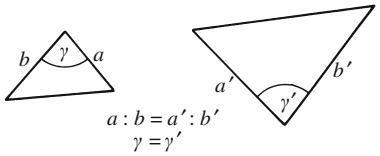


Abb. 3.61 Ähnliche Dreiecke

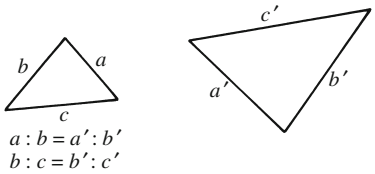


Abb. 3.62 Ähnliche Dreiecke

liebigen Mittelpunkten) oder zwei gleichseitige Dreiecke (mit beliebigen Seitenlängen).

Entsprechend den vier Kongruenzsätzen für Dreiecke gelten die folgenden Bedingungen für die Ähnlichkeit von Dreiecken:

1. Dreiecke sind ähnlich, wenn sie in zwei Winkeln übereinstimmen.
   Da im Dreieck die Winkelsumme gleich 180° ist, folgt, dass dann auch die jeweils dritten Winkel übereinstimmen.
2. Dreiecke sind ähnlich, wenn sie in dem Längenverhältnis eines Seitenpaares und dem Gegenwinkel der längeren Seite übereinstimmen.
3. Dreiecke sind ähnlich, wenn sie in dem Längenverhältnis eines Seitenpaares und dem eingeschlossenen Winkel übereinstimmen.

Bemerkungen:

1. Die Strahlensätze sind Anwendungen der Eigenschaften ähnlicher Dreiecke.
2. Nur bei Dreiecken folgt aus der Gleichheit der Winkel die Ähnlichkeit.
3. Dreiecke sind ähnlich, wenn sie in den Längenverhältnissen zweier Seitenpaare übereinstimmen.
   Zum Beispiel haben ein Rechteck mit den Seitenlängen a und $b \neq a$ und ein Quadrat gleich große Winkel, sind aber nicht ähnlich, denn ihre Seitenverhältnisse sind verschieden.
4. Da ein rechtwinkliges Dreieck durch seine Höhe in zwei untereinander und dem ganzen Dreieck ähnliche Teildreiecke geteilt wird (gleiche Winkel), folgen aus der Proportionalität der Längen entsprechender Seiten der Kathetensatz und der Höhensatz.

Stereometrie

4

Arnfried Kemnitz



4.1 Prismen 78
4.2 Zylinder 79
4.3 Pyramiden 80
4.4 Kegel 81
4.5 Cavalieri'sches Prinzip 82
4.6 Kugeln 82

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_4

Das Wort Stereometrie kommt aus dem Griechischen und bedeutet Körpermessung. Man beschäftigt sich in dieser Teildisziplin der Geometrie mit Form, gegenseitiger Lage, Größe und anderen Beziehungen geometrischer Objekte im Raum.

4.1 Prismen

4.1.1 Allgemeine Prismen

Gleitet eine Gerade, ohne ihre Richtung zu ändern, im Raum an den Begrenzungslinien eines ebenen n-Ecks ($n = 3, 4, \ldots$) entlang, so beschreibt sie eine prismatische Fläche. Schneiden zwei parallele Ebenen die prismatische Fläche, dann schließen sie zusammen mit dem zwischen ihnen liegenden Abschnitt der prismatischen Fläche einen Teil des Raums vollständig ein. Ein solcher Körper heißt Prisma (griech., das Gesägte) oder genauer n-seitiges Prisma (Abb. 4.1).

Die Schnitte der Ebenen mit der prismatischen Fläche sind kongruente n-Ecke. Diese n-Ecke heißen Grundfläche und Deckfläche des Prismas. Die Seitenflächen des Prismas heißen Mantelflächen. Die Kanten der Seitenflächen heißen Mantellinien. Die Mantelflächen sind Parallelogramme.

Bei einem Prisma sind alle Schnitte parallel zu Grund- und Deckfläche kongruent zu diesen Flächen. Ein Prisma ist also ein Körper mit einem gleichbleibenden Querschnitt.

Gleitet die Gerade senkrecht zur Ebene der Grundfläche, dann heißt das Prisma gerade. Bei einem geraden Prisma stehen die Mantellinien senkrecht auf der Grund- und Deckfläche, und die Mantelflächen sind Rechtecke. Ein nicht gerades Prisma nennt man auch schiefes Prisma.

Ein physikalisches Prisma ist mathematisch ein gerades dreiseitiges Prisma.

Das Volumen V eines Prismas ist der Inhalt A_G der Grundfläche multipliziert mit der Höhe h.

Volumen Prisma

$$V = A_G \cdot h$$

Die Oberfläche A_O eines Prismas ist die Summe der Mantelfläche A_M und der doppelten Grundfläche A_G (denn Grund- und Deckfläche sind kongruent, und damit ist ihr Flächeninhalt gleich).

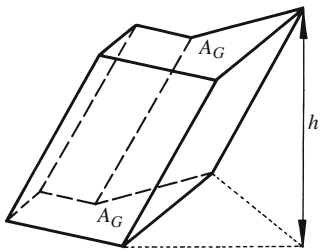


Abb. 4.1 Prisma

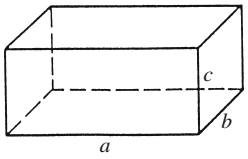


Abb. 4.2 Quader

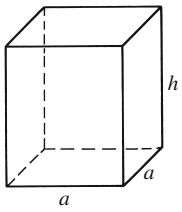


Abb. 4.3 Quadratische Säule

Oberfläche Prisma

$$A_O = A_M + 2A_G$$

4.1.2 Parallelepiped und Würfel

Ein Prisma mit einem Parallelogramm als Grundfläche heißt Parallelepiped oder Parallelflach oder Spat. Ein gerades Prisma mit einem Rechteck als Grundfläche heißt Quader (Abb. 4.2).

Sind a und b die Seitenlängen des Rechtecks und c die Höhe des Quaders, so gilt:

Quader

Volumen	$V = abc$
Oberfläche	$A_O = 2(ab + ac + bc)$
Gesamtkantenlänge	$l = 4(a + b + c)$

Ein Quader mit einem Quadrat als Grundfläche heißt quadratische Säule (Abb. 4.3).

Ist a die Seitenlänge des Quadrats und h die Höhe der quadratischen Säule, dann gilt:

Quadratische Säule

Volumen	$V = a^2h$
Oberfläche	$A_O = 2a^2 + 4ah$
Gesamtkantenlänge	$l = 8a + 4h$

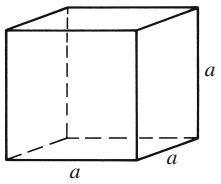


Abb. 4.4 Würfel

Ein Quader mit lauter gleich langen Kanten heißt Würfel (Abb. 4.4).

Ist a die Kantenlänge des Würfels, so gilt:

Würfel

Volumen	$V = a^3$
Oberfläche	$A_O = 6a^2$
Gesamtkantenlänge	$l = 12a$

Der Würfel ist einer der platonischen Körper. Er wird von sechs Quadraten begrenzt.

4.2 Zylinder

4.2.1 Allgemeine Zylinder

Wird eine Gerade (Erzeugende) im Raum längs einer ebenen geschlossenen Kurve (Leitkurve) parallel verschoben (also ohne ihre Richtung zu verändern), so entsteht eine Zylinderfläche. Ein Zylinder ist ein Körper, der von einer Zylinderfläche und zwei parallelen ebenen Flächenstücken begrenzt wird. Die ebenen Begrenzungsflächenstücke müssen nicht senkrecht auf der erzeugenden Gerade stehen (Abb. 4.5).

Ein Zylinder ist ein Körper mit gleichbleibendem Querschnitt.

Der Teil der Zylinderfläche zwischen den parallelen Begrenzungsflächenstücken heißt Mantelfläche des Zylinders, die parallelen Flächenstücke sind Grund- und Deckfläche des Zylinders. Grundfläche und Deckfläche sind zueinander kongruent. Die zwischen den Flächenstücken liegenden Strecken der Erzeugenden heißen Mantellinien, sie sind alle parallel und gleich lang. Der senkrechte Abstand zwischen Grund- und Deckfläche ist die Höhe des Zylinders.

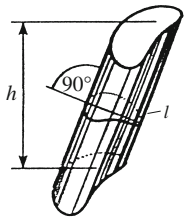


Abb. 4.5 Zylinder (l = Mantellinie)

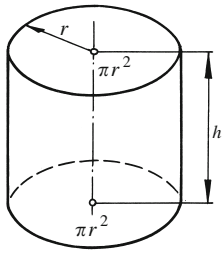


Abb. 4.6 Gerader Kreiszylinder

Prismen sind spezielle Zylinder, nämlich solche mit n-Ecken als Grundfläche.

Ein Zylinder heißt gerade, wenn die Mantellinien senkrecht auf Grund- und Deckfläche stehen. Ein nicht gerader Zylinder heißt schiefer Zylinder.

Ein Zylinder mit einer Kreisfläche als Grundfläche heißt Kreiszylinder.

Das Volumen V eines Zylinders ist der Inhalt A_G der Grundfläche multipliziert mit der Höhe h. Die Oberfläche A_O eines Zylinders ist die Summe der Mantelfläche A_M und der doppelten Grundfläche A_G.

Volumen	$V = A_G \cdot h$
Oberfläche	$A_O = A_M + 2A_G$

4.2.2 Gerade Kreiszylinder

Ein Zylinder mit senkrecht auf Grund- und Deckfläche stehenden Mantellinien und mit einer Kreisfläche als Grundfläche heißt gerader Kreiszylinder oder Walze (Abb. 4.6).

Die Mantelfläche eines geraden Kreiszylinders kann in ein Rechteck mit den Seitenlängen h und $2\pi r$ (Kreisumfang) abgewickelt werden, wobei h die Höhe des geraden Kreiszylinders ist und r der Radius des Kreises. Dies kann man sich dadurch veranschaulichen, dass man eine Dose ohne Deckel und Boden längs einer Mantellinie aufschneidet und in eine Ebene abwickelt.

Der Kreis als Grund- und Deckfläche (Grund- und Deckfläche sind kongruent) hat den Flächeninhalt πr^2.

Somit gilt für die Oberfläche insgesamt $A_O = 2\pi r^2 + 2\pi rh = 2\pi r(r + h)$.

Volumen	$V = \pi r^2 h$
Oberfläche	$A_O = 2\pi r(r + h)$

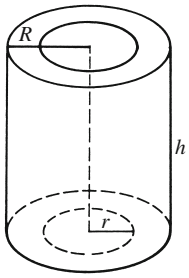


Abb. 4.7 Hohlzylinder

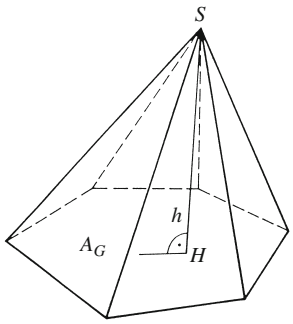


Abb. 4.8 Pyramide

4.2.3 Hohlzylinder

Ein Hohlzylinder ist ein gerader Kreiszylinder (Kreis mit Radius R), aus dem ein kleinerer gerader Kreiszylinder (konzentrischer Kreis mit Radius r, $r < R$) ausgeschnitten ist (Abb. 4.7).

Die Grundflächen der beiden Zylinder sind also konzentrische Kreise, das heißt, sie haben den gleichen Mittelpunkt.

Das Volumen des Hohlzylinders ist die Differenz der Volumina der beiden geraden Kreiszylinder. Die Oberfläche setzt sich aus der äußeren Mantelfläche $A_{M_a} = 2\pi Rh$ (h ist die Höhe des Hohlzylinders), aus der inneren Mantelfläche $A_{M_i} = 2\pi rh$, aus der Grundfläche und aus der Deckfläche zusammen. Grundfläche A_G und Deckfläche A_D sind gleich, sie ergeben sich aus der Differenz zweier Kreisflächen: $A_G = A_D = \pi(R^2 - r^2)$.

Für das Volumen und die Oberfläche des Hohlzylinders gilt somit

$$\begin{aligned} &\text{Volumen} && V = \pi h(R^2 - r^2) \\ &\text{Oberfläche} && A_O = 2\pi h(R + r) + 2\pi(R^2 - r^2) \\ & && \quad = 2\pi(R + r)(R - r + h) \end{aligned}$$

4.3 Pyramiden

4.3.1 Allgemeine Pyramiden

Gleitet ein von einem festen Punkt S des Raums ausgehender Strahl an den Begrenzungslinien eines ebenen n-Ecks ($n = 3, 4, \ldots$) entlang, in dessen Ebene der Anfangspunkt S des Strahls nicht liegt, so beschreibt der gleitende Strahl eine Pyramidenfläche. Das n-Eck schließt zusammen mit dem zwischen ihm und dem Punkt S liegenden Abschnitt der Pyramidenfläche einen Teil des Raums vollständig ein. Ein solcher Körper heißt Pyramide (Abb. 4.8).

Das n-Eck heißt Grundfläche, der Punkt S Spitze, der zum Körper gehörende Teil der Pyramidenfläche ist die Mantelfläche der Pyramide. Die Kanten der Grundfläche heißen Grundkanten, die Kanten der Mantelfläche Seitenkanten, und die ebenen Flächen der Mantelfläche sind die Seitenflächen.

Alle Seitenflächen einer Pyramide sind Dreiecke. Es gibt bei einem n-Eck als Grundfläche genau n Dreiecke als Seitenflächen. Deshalb nennt man solch eine Pyramide auch genauer n-seitige Pyramide.

Ist das n-Eck ein reguläres n-Eck, dann heißt die Pyramide reguläre (n-seitige) Pyramide.

Der Abstand der Spitze S von der Ebene der Grundfläche ist die Höhe der Pyramide. Man erhält die Höhe, indem man von S das Lot auf die Ebene der Grundfläche fällt. Das Lot durchstößt die Ebene der Grundfläche im Höhenfußpunkt H. Dieser kann auch außerhalb der Grundfläche liegen, dann liegt die Höhe außerhalb der Pyramide.

Fällt der Höhenfußpunkt mit dem Mittelpunkt der Grundfläche zusammen, so heißt die Pyramide gerade. Alle anderen Pyramidenformen nennt man schief. Die Höhe einer geraden Pyramide ist gleichzeitig ihre Achse.

Die Seitenflächen von regulären geraden Pyramiden sind kongruente gleichschenklige Dreiecke.

Für das Volumen V und die Oberfläche A_O einer beliebigen Pyramide gilt

$$\begin{aligned} &\text{Volumen} && V = \frac{1}{3} A_G \cdot h \\ &\text{Oberfläche} && A_O = A_M + A_G \end{aligned}$$

A_G ist der Flächeninhalt des n-Ecks, A_M der Inhalt der Mantelfläche, also die Summe der Flächeninhalte der Seitendreiecke, und h ist die Höhe der Pyramide.

Eine gerade reguläre dreiseitige Pyramide, bei der die Seitendreiecke kongruent zum Grunddreieck sind, heißt Tetraeder. Ein Tetraeder wird also von vier gleichseitigen Dreiecken begrenzt.

4.3.2 Gerade quadratische Pyramiden

Eine gerade quadratische Pyramide hat ein Quadrat als Grundfläche, und die Spitze der Pyramide steht senkrecht über dem Mittelpunkt des Quadrats, dem Diagonalenschnittpunkt. Die

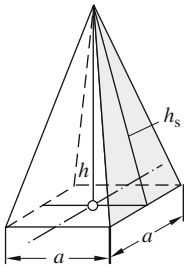


Abb. 4.9 Gerade quadratische Pyramide

Mantelfläche besteht aus vier kongruenten gleichschenkligen Dreiecken (Abb. 4.9).

Ist a die Kantenlänge des Quadrats der Grundfläche, h die Höhe der Pyramide und s die Kantenlänge der Seitenkanten, so folgt aus dem Satz des Pythagoras $s^2 = h^2 + \frac{1}{2}a^2$.

Bezeichnet man mit h_s die Höhe des gleichschenkligen Seitenflächendreiecks Δ_s, dann folgt ebenfalls mit dem Satz von Pythagoras $s^2 = h_s^2 + \frac{1}{4}a^2$, denn die Basis dieses Dreiecks hat die Länge a. Löst man nach h_s auf und ersetzt s^2 durch $h^2 + \frac{1}{2}a^2$, so ergibt sich $h_s = \sqrt{h^2 + \frac{1}{4}a^2}$. Daraus berechnet man den Flächeninhalt A_{Δ_s} von Δ_s zu $A_{\Delta_s} = \frac{1}{2}a \cdot h_s = \frac{1}{2}a\sqrt{h^2 + \frac{1}{4}a^2}$ und den Inhalt der Mantelfläche zu $A_M = 4A_{\Delta_s} = 2a\sqrt{h^2 + \frac{1}{4}a^2}$.

Für Volumen und Oberfläche einer geraden quadratischen Pyramide gilt somit

$$\text{Volumen} \qquad V = \frac{1}{3}a^2 \cdot h$$

$$\text{Oberfläche} \qquad A_O = 2a\sqrt{h^2 + \frac{1}{4}a^2} + a^2$$

Die Grabstätten altägyptischer Pharaone waren während des Alten und des Mittleren Reichs häufig gerade quadratische Pyramiden. Besonders beeindruckend sind die Pyramiden der Pharaonen Cheops, Chephren und Mykerinos in Gizeh am südlichen Rand von Kairo. Sie stammen aus dem Alten Reich und wurden in der Zeit zwischen 2600 und 2480 v. u. Z. erbaut, sie sind also rund 4500 Jahre alt. Die größte Pyramide ist die Cheopspyramide: Das Quadrat der Grundfläche hat eine Kantenlänge von 227,5 m (ursprünglich 230,38 m), und die Höhe ist 137 m (ursprünglich 146,6 m). Nimmt man die ursprünglichen Werte, so berechnet man für das Volumen:

$$V = \frac{1}{3}(230{,}38)^2 \cdot 146{,}6\,\text{m}^3 = 2.593.595{,}61\ldots\,\text{m}^3,$$

also mehr als 2,5 Millionen Kubikmeter! Für die Oberfläche ergibt sich:

$$A_O = \left(2 \cdot 230{,}38 \cdot \sqrt{(146{,}6)^2 + \frac{1}{4}(230{,}38)^2} + (230{,}38)^2\right)\,\text{m}^2$$
$$= 138.979{,}56\ldots\,\text{m}^2$$

Der Bau der Pyramiden war eine großartige ingenieurtechnische und logistische Leistung der Altägypter!

4.4 Kegel

4.4.1 Allgemeine Kegel

Wird eine Gerade (Erzeugende) im Raum längs einer ebenen geschlossenen Kurve (Leitkurve) so bewegt, dass sie durch einen festen Punkt, die Spitze S, geht, so entsteht eine Kegelfläche. Ein Kegel ist ein Körper, der von einer Kegelfäche und einem nicht durch deren Spitze gehenden ebenen Flächenstück begrenzt wird (Abb. 4.10).

Der Teil der Kegelfläche zwischen dem ebenen Flächenstück und der Spitze heißt Mantelfläche, das ebene Flächenstück Grundfläche des Kegels. Die zwischen Grundfläche und Spitze liegenden Strecken der Erzeugenden heißen Mantellinien. Der senkrechte Abstand der Spitze zur Ebene der Grundfläche ist die Höhe des Kegels.

Pyramiden sind spezielle Kegel, nämlich Kegel mit n-Ecken als Grundfläche.

Hat die Grundfläche einen Mittelpunkt (wie Kreis oder Ellipse), und liegt die Spitze senkrecht über diesem Mittelpunkt, so heißt der Kegel gerade, andernfalls schief.

Ein Kegel mit einer Kreisfläche als Grundfläche heißt Kreiskegel.

Das Volumen V eines Kegels ist ein Drittel des Inhalts A_G der Grundfläche multipliziert mit der Höhe h. Die Oberfläche A_O eines Kegels ist die Summe der Mantelfläche A_M und der Grundfläche A_G.

$$\text{Volumen} \qquad V = \frac{1}{3}A_G \cdot h$$

$$\text{Oberfläche} \qquad A_O = A_M + A_G$$

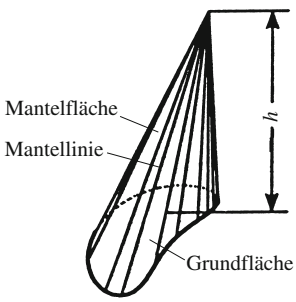


Abb. 4.10 Kegel

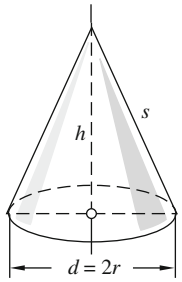


Abb. 4.11 Gerader Kreiskegel

4.4.2 Gerade Kreiskegel

Ein Kegel mit einer Kreisfläche als Grundfläche und der Spitze S senkrecht über dem Kreismittelpunkt heißt gerader Kreiskegel (Abb. 4.11).

Alle Mantellinien eines geraden Kreiskegels sind gleich lang. Ihre Länge ist $s = \sqrt{r^2 + h^2}$, wobei r der Radius des Kreises und h die Höhe des geraden Kreiskegels sind.

Die Mantelfläche kann in die Ebene abgewickelt werden. Dabei entsteht ein Kreissektor mit dem Radius s (Länge der Mantellinien) und der Kreisbogenlänge $2\pi r$ (Umfang des Kreises der Grundfläche). Der Flächeninhalt A_M dieses Kreissektors (= Mantelfläche) verhält sich zur gesamten Kreisfläche πs^2 wie die Kreisbogenlänge $2\pi r$ zum Gesamtkreisumfang $2\pi s$, woraus sich für die Mantelfläche $A_M = \pi rs$ ergibt. Der Kreis der Grundfläche, der Grundkreis, hat den Flächeninhalt πr^2.

Daraus folgt für die Oberfläche $A_O = \pi rs + \pi r^2 = \pi r(r+s) = \pi r(r + \sqrt{r^2 + h^2})$.

Somit gilt für den geraden Kreiskegel

Volumen	$V = \frac{1}{3}\pi r^2 h$
Oberfläche	$A_O = \pi r(r + s)$
Länge der Mantellinie	$s = \sqrt{r^2 + h^2}$

4.5 Cavalieri'sches Prinzip

Wesentlich zur Berechnung des Volumens von Prismen, Zylindern, Pyramiden und Kegeln ist das Cavalieri'sche Prinzip (nach dem italienischen Mathematiker Bonaventura Cavalieri, ~1591/98–1647, ein Schüler Galileis):

Körper mit inhaltsgleichem Querschnitt in gleichen Höhen haben gleiches Volumen.

Speziell gilt also: Prismen, Zylinder, Pyramiden und Kegel mit gleicher Grundfläche und gleicher Höhe haben gleiches Volumen.

4.6 Kugeln

4.6.1 Definitionen

Eine Kugel ist der geometrische Ort aller Punkte des Raumes, die von einem festen Punkt M einen konstanten Abstand r haben (Abb. 4.12).

Der Punkt M ist der Mittelpunkt und r der Radius der Kugel.

Zur Unterscheidung von dem durch eine Kugel abgegrenzten Raum nennt man die Kugel selbst auch Kugelfläche.

Eine Kugel mit dem Radius $r = 1$ heißt Einheitskugel.

Für das Volumen V und die Oberfläche A_O einer Kugel mit dem Radius r und dem Durchmesser d gilt

Volumen	$V = \frac{4}{3}\pi r^3 = \frac{\pi}{6}d^3$
Oberfläche	$A_O = 4\pi r^2 = \pi d^2$

Eine Kugel ist festgelegt durch den Mittelpunkt und einen weiteren Punkt oder durch vier Punkte (die nicht alle in einer Ebene liegen).

Kugeln mit gleichem Mittelpunkt heißen konzentrisch.

Jede die Kugel schneidende Ebene schneidet sie in einem Kreis.

Geraden haben mit einer Kugelfläche entweder zwei Punkte, einen Punkt oder keinen Punkt gemeinsam.

Eine Sekante s schneidet die Kugelfläche in zwei Punkten. Der zwischen den Punkten gelegene Teil der Sekante heißt Sehne. Eine Sehne durch den Mittelpunkt heißt Durchmesser der Kugel. Durchmesser sind die größten Sehnen der Kugel, für ihre Länge d gilt $d = 2r$.

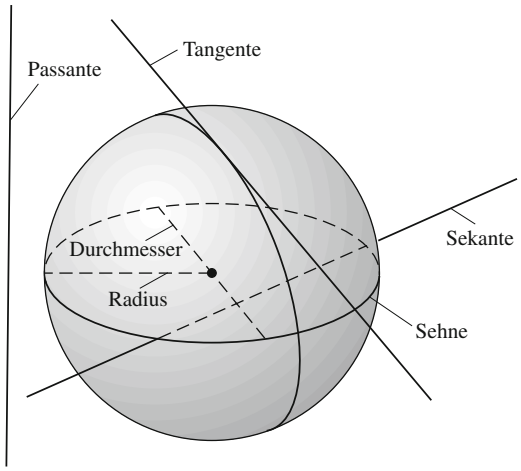


Abb. 4.12 Bezeichnungen an der Kugel

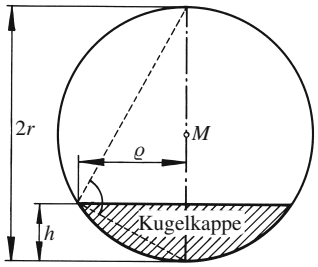


Abb. 4.13 Kugelsegment

Eine Tangente t berührt die Kugel in einem Punkt. Im Berührungspunkt sind beliebig viele Tangenten möglich: Alle Tangenten zusammen spannen die Tangentialebene auf.

Eine Passante p hat mit der Kugel keinen Punkt gemeinsam.

4.6.2 Kugelsegmente

Ein Kugelsegment oder Kugelabschnitt ist ein durch eine Ebene abgeschnittener Teil einer Kugel (Abb. 4.13).

Die Mantelfläche des Kugelsegments heißt Kugelkappe.

Ist r der Radius der Kugel, ρ der Radius des von der Ebene ausgeschnittenen Kreises und h die Höhe des Kugelsegments, dann gilt

Radius Schnittkreis	$\rho = \sqrt{h(2r-h)}$
Volumen Kugelsegment	$V = \frac{1}{6}\pi h(3\rho^2 + h^2)$
	$= \frac{1}{3}\pi h^2(3r-h)$
Flächeninhalt Kugelkappe	$A = 2\pi rh$
Oberfläche Kugelsegment	$A_O = 2\pi rh + \pi\rho^2$
	$= \pi(2rh + \rho^2)$

4.6.3 Kugelsektoren

Einem Kugelsegment (Kugelabschnitt) ist ein Kegel zugeordnet, dessen Grundfläche der Schnittkreis des Kugelsegments und dessen Spitze der Kugelmittelpunkt ist. Der Gesamtkörper aus Kugelsegment und zugeordnetem Kegel heißt Kugelsektor oder Kugelausschnitt (Abb. 4.14).

Das Volumen V des Kugelsektors setzt sich aus dem Volumen des Kugelabschnitts und dem des zugeordneten Kegels zusammen: $V = \pi\frac{h^2}{3}(3r-h) + \pi\frac{\rho^2}{3}(r-h)$.

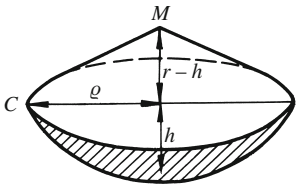


Abb. 4.14 Kugelsektor

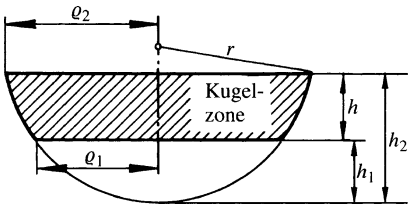


Abb. 4.15 Kugelschicht

Dabei ist r der Radius der Kugel, ρ der Radius des Schnittkreises und h die Höhe des Kugelsegments. Durch Einsetzen von $\rho^2 = h(2r-h)$ erhält man $V = \frac{2}{3}\pi r^2 h$.

Die Oberfläche A_O des Kugelsektors ist die Summe der Flächeninhalte von Kugelkappe und Kegelmantel: $A_O = 2\pi rh + \pi\rho r = \pi r(2h+\rho)$.

Volumen Kugelsektor	$V = \frac{2}{3}\pi r^2 h$
Oberfläche Kugelsektor	$A_O = \pi r(2h+\rho)$

4.6.4 Kugelschichten

Eine Kugelschicht ist der durch zwei zueinander parallelen Ebenen ausgeschnittene Teil einer Kugel (Abb. 4.15).

Die durch die beiden Ebenen ausgeschnittene Kugeloberfläche, also die Mantelfläche der Kugelschicht, heißt Kugelzone.

Ist r der Radius der Kugel, ρ_1 und ρ_2 die Radien der von den parallelen Ebenen ausgeschnittenen Kreise und h die Dicke der Kugelschicht, dann gilt

Radien Schnittkreise	$\rho_1 = \sqrt{h_1(2r-h_1)}$
	$\rho_2 = \sqrt{(h+h_1)(2r-h-h_1)}$
Volumen Kugelschicht	$V = \frac{1}{6}\pi h(3\rho_1^2 + 3\rho_2^2 + h^2)$
Flächeninhalt Kugelzone	$A = 2\pi rh$
Oberfläche Kugelschicht	$A_O = \pi(2rh + \rho_1^2 + \rho_2^2)$

5 Funktionen

Arnfried Kemnitz



5.1 Definition und Darstellungen von Funktionen 86
5.2 Verhalten von Funktionen 88
5.3 Einteilung der elementaren Funktionen 91
5.4 Ganze rationale Funktionen 93
5.5 Gebrochene rationale Funktionen 102
5.6 Irrationale Funktionen 106
5.7 Transzendente Funktionen 108

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_5

5.1 Definition und Darstellungen von Funktionen

5.1.1 Definitionen

Eine Abbildung oder Funktion f ist eine Zuordnung, die jeder Zahl x einer gegebenen Zahlenmenge D eine Zahl y einer Zahlenmenge W zuordnet. Die Zuordnung ist eindeutig, das heißt, jeder Zahl x wird genau eine Zahl y zugeordnet. Man schreibt dafür $y = f(x)$ oder manchmal auch $x \mapsto f(x)$. Man nennt $f(x)$ das Bild von x und umgekehrt x das Urbild von $f(x)$.

Die Menge D heißt Urbildmenge, Definitionsmenge oder Definitionsbereich. Die Menge W, aus der die Bilder stammen, heißt Wertemenge oder Wertebereich. Die Menge der Bilder (also alle y-Werte zusammen) heißt Bildmenge, bezeichnet mit $f(D)$.

D	Definitionsbereich
W	Wertebereich
$f(D)$	Bildmenge

Die Elemente der Bildmenge nennt man Funktionswerte. Die Bildmenge $f(D)$ ist eine Teilmenge des Wertebereichs W, und W ist eine Teilmenge der Menge $\mathbb{R}$ der reellen Zahlen.

$$f(D) \subseteq W \subseteq \mathbb{R}$$

Eine Funktion besteht aus drei Teilen: der Zuordnungsvorschrift f, dem Definitionsbereich D und dem Wertebereich W.

Zwei Funktionen sind genau dann gleich, wenn sowohl die Zuordnungsvorschriften als auch die Definitionsbereiche als auch die Wertebereiche übereinstimmen.

Beispiele

1. $y = f(x) = 5x, D = \mathbb{N}, W = \mathbb{N}$
   Die Zuordnungsvorschrift ist hier „5-mal“, das heißt, man muss jeden x-Wert mit 5 multiplizieren, um den zugehörigen Funktionswert y zu erhalten. Für $x = 3$ erhält man zum Beispiel $y = f(3) = 5 \cdot 3 = 15$.
   Sowohl der Definitionsbereich als auch der Wertebereich sind die natürlichen Zahlen. Für die Bildmenge ergibt sich $f(D) = \{0, 5, 10, 15, 20, \dots\}$.
2. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, \dots, 24, 25\}$,
   $f(-1) = 1, f(-2) = 4, f(-3) = 9, f(-4) = 16$,
   $f(-5) = 25$
3. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, 5\}$,
   $f(-1) = 1, f(-2) = 2, f(-3) = 3, f(-4) = 4$,
   $f(-5) = 5$
4. $y = f(x) = x + 2, D = \mathbb{R}, W = \mathbb{R}$ ◀

Bemerkung: Man kann Abbildungen (Funktionen) auch allgemeiner als eine Zuordnung zwischen beliebigen Mengen (also nicht eingeschränkt auf Zahlenmengen) definieren.

5.1.2 Funktionsgleichung

Explizite Darstellung der Funktionsgleichung

Die Zuordnungsvorschrift für eine Funktion ist im Regelfall eine Gleichung, die Funktionsgleichung $y = f(x)$ (gesprochen: y gleich f von x). Dabei heißt x unabhängige Variable und y abhängige Variable. Man nennt x auch das Argument der Funktion.

Die Form $y = f(x)$ heißt explizite Darstellung der Funktionsgleichung. Darüber hinaus gibt es die implizite Darstellung und die Parameterdarstellung der Funktionsgleichung (siehe unten).

Funktionen können aber zum Beispiel auch durch Tabellen, Schaubilder (Graphen), Pfeildiagramme oder geordnete Wertepaare (Wertetabelle) dargestellt werden.

Fehlt bei einer Funktion die Angabe des Definitionsbereichs, so gilt $D = \mathbb{R}$. Fehlt bei einer Funktion die Angabe des Wertebereichs, so gilt ebenfalls $W = \mathbb{R}$.

Die Schreibweise $y = f(x), f : D \to W$ für eine Funktion bedeutet, dass $y = f(x)$ die Funktionsgleichung ist, dass die Funktion den Definitionsbereich D und den Wertebereich W hat.

$$y = f(x), \quad f : D \to W$$

Beispiele

1. $y = f(x) = x^3 - 4x^2 - x + 4, f : \mathbb{R} \to \mathbb{R}$
2. $y = f(x) = x^2 - 1, f : \mathbb{R} \to \mathbb{R}$
3. $y = f(x) = \frac{x-3}{x^2-2}, f : [1, 1] \to \mathbb{R}$
   (also $D = [1, 1], W = \mathbb{R}$)
4. $y = f(x) = \sqrt{x}, \quad f : \mathbb{N} \to \mathbb{R}$
5. $y = f(x) = \begin{cases} -1 & \text{falls } x < 0 \\ 0 & \text{falls } x = 0, \\ +1 & \text{falls } x > 0, \end{cases} \quad f : \mathbb{R} \to \mathbb{R}.$ ◀

Eine Funktion mit der Funktionsgleichung $y = f(x)$, deren Definitions- und Wertemenge nur reelle Zahlen enthalten, nennt man eine reelle Funktion einer reellen Variablen.

Beispiele

6. $y = x^2, D = (-\infty, \infty), W = [0, \infty)$
7. $y = \sqrt{x}, D = [0, \infty), W = [0, \infty)$. ◀

Implizite Darstellung der Funktionsgleichung

Die Darstellung einer Funktion in der Form $F(x, y) = 0$ heißt implizit, falls sich diese Gleichung eindeutig nach y auflösen lässt.

Statt impliziter Darstellung der Funktion sagt man auch einfach nur implizite Funktion.

Beispiel

8. $F(x, y) = x^2 + y^2 - 1 = 0, D = [-1, 1], y \geq 0$
   Es handelt sich hierbei um die obere Hälfte des Einheitskreises mit dem Mittelpunkt im Koordinatenursprung (vgl. Abschn. 7.3.1).
   Man beachte, dass mit $x^2 + y^2 - 1 = 0$ keine reelle Funktion definiert wird, denn die Zuordnung ist nicht eindeutig, da jedem Element des Definitionsbereichs zwei Werte zugeordnet werden (einer auf dem oberen Halbkreis und einer auf dem unteren Halbkreis). ◀

Parameterdarstellung der Funktionsgleichung

Die Darstellung einer Funktion in der Form $x = \varphi(t), y = \psi(t)$ heißt Parameterdarstellung.

Die Werte von x und y werden dabei jeweils als Funktion einer Hilfsvariablen t angegeben, die Parameter genannt wird. Die Funktionen $\varphi(t)$ und $\psi(t)$ müssen denselben Definitionsbereich haben.

Beispiele

9. $x = 2t + 5, y = 8t + 4, t \in \mathbb{R}$
   Durch Elimination von t erhält man
   $4x - 20 = y - 4 \Rightarrow y = 4x - 16$,
   also eine Geradengleichung (in expliziter Form) (vgl. Abschn. 7.2.1).
10. $x = \varphi(t), y = \psi(t)$ mit $x = \cos t, y = \sin t$ und $0 \leq t \leq \pi$
    Hierbei handelt es sich um die obere Hälfte des Einheitskreises mit dem Mittelpunkt im Koordinatenursprung, denn Quadrieren und Addieren ergibt
    $x^2 + y^2 = \cos^2 t + \sin^2 t = 1$,
    und t durchläuft den ersten und den zweiten Quadranten (vgl. Kap. 6). ◀

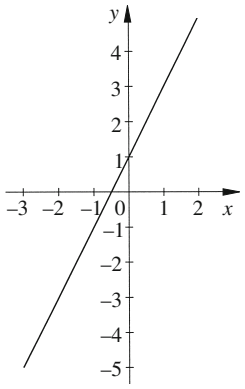


Abb. 5.1 Graph der Funktion mit der Gleichung $y = f(x) = 2x + 1$

5.1.3 Graph einer Funktion

Eine Möglichkeit der Funktionsdarstellung ist, den Graph der Funktion zu zeichnen (Abb. 5.1).

Der Graph einer Funktion f mit dem Definitionsbereich D ist das Bild, das man erhält, wenn man die geordneten Zahlenpaare $(x, y) = (x, f(x))$ mit $x \in D$ in ein Koordinatenkreuz einträgt. Geordnet bedeutet, dass in (x, y) die Reihenfolge von x und y wichtig ist: (x, y) ist verschieden von (y, x) (außer möglicherweise in Sonderfällen).

In einem kartesischen Koordinatensystem (siehe Abschn. 7.1) ist die waagerechte Achse die x-Achse oder Abszissenachse, die senkrechte Achse ist die y-Achse oder Ordinatenachse. Die Zahl x ist die Abszisse und y die Ordinate eines Punktes $(x|y)$ mit den Koordinaten x und y.

Statt Graph einer Funktion sagt man auch Schaubild oder Kurve der Funktion.

Bemerkung: Bei einem Zahlenpaar setzt man ein Komma oder ein Semikolon zwischen die beiden Komponenten: (x, y) oder $(x; y)$. Bei der Darstellung eines Punktes setzt man einen senkrechten Strich zwischen die beiden Koordinaten: $(x|y)$.

Beispiel

Graph der Funktion mit der Funktionsgleichung $y = f(x) = 2x + 1$ und dem Definitionsbereich $D = \mathbb{R}$: Abb. 5.1. ◀

5.1.4 Wertetabelle einer Funktion

Auch mittels einer Wertetabelle kann eine Funktion dargestellt werden.

In einer Wertetabelle werden für einige ausgewählte Argumente x die geordneten Zahlenpaare $(x, y) = (x, f(x))$ für eine Funktion $y = f(x)$ eingetragen. Dabei müssen die ausgewählten Werte für x Elemente des Definitionsbereichs D der Funktion sein.

Man stellt oftmals eine Wertetabelle auf, um den Graph einer Funktion zeichnen zu können.

Beispiel

Wertetabelle für die Funktion $y = -x^2 - 4x + 3$, $D = \mathbb{R}$:

x	-5	-4	-3	-2	-1	0	1	2
y	-2	3	6	7	6	3	-2	-9

◀

5.2 Verhalten von Funktionen

5.2.1 Monotone Funktionen

Eine Funktion mit der Gleichung $y = f(x)$ heißt in einem bestimmten Bereich B (B ist eine Teilmenge des Definitionsbereichs D)

- monoton wachsend, wenn aus $x_1 < x_2$ stets $f(x_1) \leq f(x_2)$ folgt,
- streng monoton wachsend, wenn aus $x_1 < x_2$ stets $f(x_1) < f(x_2)$ folgt,
- monoton fallend, wenn aus $x_1 < x_2$ stets $f(x_1) \geq f(x_2)$ folgt,
- streng monoton fallend, wenn aus $x_1 < x_2$ stets $f(x_1) > f(x_2)$ folgt.

Dabei sind x_1, x_2 beliebige Punkte aus diesem Bereich B.

Beispiele

1. $f(x) = 3x$, $D = \mathbb{R}$ ist streng monoton wachsend in D (Abb. 5.2).
2. $f(x) = x^2$, $D = \mathbb{R}$ ist in $B_1 = \{x \,|\, x \in D \text{ und } x \leq 0\}$ streng monoton fallend und in $B_2 = \{x \,|\, x \in D \text{ und } x \geq 0\}$ streng monoton wachsend (Abb. 5.3). ◀

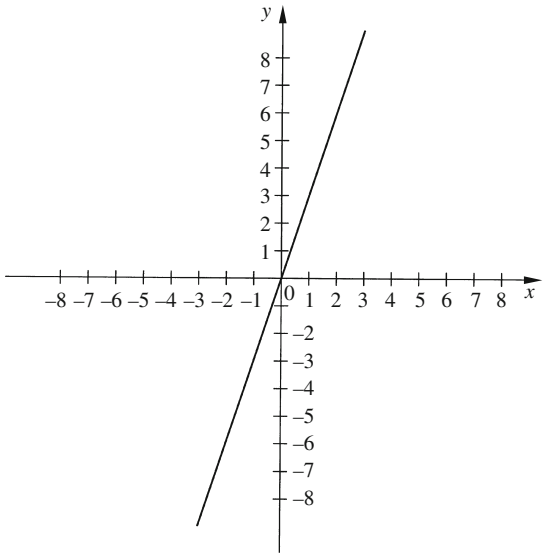


Abb. 5.2 Graph der Funktion mit der Gleichung $f(x) = 3x$

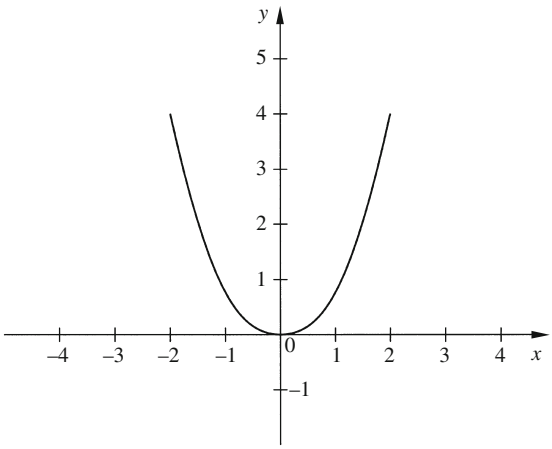


Abb. 5.3 Graph der Funktion mit der Gleichung $f(x) = x^2$

5.2.2 Symmetrische Funktionen

Der Graph einer Funktion mit der Gleichung $y = f(x)$ ist symmetrisch zur y-Achse, wenn $f(x) = f(-x)$ für alle $x \in D$ gilt. Eine solche Funktion heißt eine gerade Funktion.

Der Graph einer Funktion $y = f(x)$ ist symmetrisch zum Koordinatenursprung, wenn $f(-x) = -f(x)$ für alle $x \in D$ gilt. Eine solche Funktion heißt eine ungerade Funktion.

Beispiele

1. $f(x) = 2x^4 + 1$ (Abb. 5.4)
   Wegen $f(-x) = 2(-x)^4 + 1 = 2x^4 + 1 = f(x)$ ist $y = f(x)$ symmetrisch zur y-Achse, also eine gerade Funktion.
2. $f(x) = 2x^3 - 3x$ (Abb. 5.5)
   Wegen $f(-x) = 2(-x)^3 - 3(-x) = -2x^3 + 3x = -f(x)$ ist $y = f(x)$ symmetrisch zum Koordinatenursprung, also eine ungerade Funktion. ◀

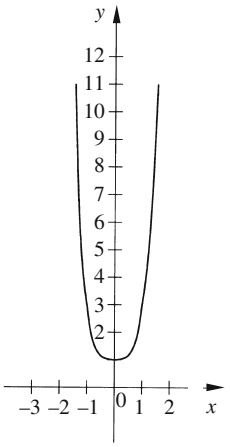


Abb. 5.4 Graph der Funktion mit der Gleichung $f(x) = 2x^4 + 1$

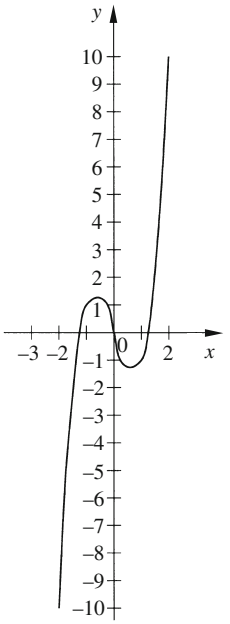


Abb. 5.5 Graph der Funktion mit der Gleichung $f(x) = 2x^3 - 3x$

5.2.3 Beschränkte Funktionen

Eine Funktion heißt nach oben beschränkt, wenn ihre Funktionswerte eine bestimmte Zahl nicht übertreffen, und nach unten beschränkt, wenn ihre Funktionswerte nicht kleiner als eine bestimmte Zahl sind. Eine Funktion, die sowohl nach oben als auch nach unten beschränkt ist, heißt beschränkt.

Bei einer beschränkten Funktion $y = f(x)$ existieren also reelle Zahlen a und b mit $a < b$, so dass gilt:

$$a \le f(x) \le b \quad \text{für alle } x \in D$$

Beispiele

1. $y = 1 - x^2$ ist nach oben beschränkt, denn $y \le 1$.
2. $y = e^x$ ist nach unten beschränkt, denn $y > 0$.
3. $y = \frac{4}{1+x^2}$ ist beschränkt, denn $0 < y \le 4$. ◀

5.2.4 Injektive Funktionen

Eine Funktion heißt injektiv, wenn jedes Bild genau ein Urbild besitzt.

Bei einer injektiven Funktion gehören zu verschiedenen Argumenten also stets verschiedene Bilder.

$$x_1 \neq x_2 \quad \Rightarrow \quad f(x_1) \neq f(x_2)$$

Beispiele

Folgende Funktionen sind injektiv:

1. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, \ldots, 24, 25\}$,
   $f(-1) = 1, f(-2) = 4, f(-3) = 9, f(-4) = 16$,
   $f(-5) = 25$
2. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, 5\}$,
   $f(-1) = 1, f(-2) = 2, f(-3) = 3, f(-4) = 4$,
   $f(-5) = 5$
3. $y = f(x) = x + 2, f : \mathbb{R} \to \mathbb{R}$ (also $D = W = \mathbb{R}$)
4. $y = f(x) = \sqrt{x}, f : \mathbb{N} \to \mathbb{R}$.

Folgende Funktionen sind nicht injektiv:

1. $y = f(x) = x^3 - 4x^2 - x + 4, f : \mathbb{R} \to \mathbb{R}$
2. $y = f(x) = x^2 - 1, f : \mathbb{R} \to \mathbb{R}$
3. $y = f(x) = \frac{x-3}{x^2-2}, f : [1, 1] \to \mathbb{R}$ (also $D = [1, 1]$, $W = \mathbb{R}$)
4. $y = f(x) = \begin{cases} -1 & \text{falls } x < 0, \\ 0 & \text{falls } x = 0, f : \mathbb{R} \to \mathbb{R}. \\ +1 & \text{falls } x > 0, \end{cases}$ ◀

5.2.5 Surjektive Funktionen

Eine Funktion heißt surjektiv, wenn ihre Bildmenge gleich dem Wertebereich ist.

$$f(D) = W$$

Beispiele

Folgende Funktionen sind surjektiv:

1. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, 5\}$,
   $f(-1) = 1, f(-2) = 2, f(-3) = 3, f(-4) = 4,$
   $f(-5) = 5$
2. $y = f(x) = x + 2, f : \mathbb{R} \to \mathbb{R}$
3. $y = f(x) = x^3 - 4x^2 - x + 4, f : \mathbb{R} \to \mathbb{R}$

Folgende Funktionen sind nicht surjektiv:

1. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, \ldots, 24, 25\}$,
   $f(-1) = 1, f(-2) = 4, f(-3) = 9, f(-4) = 16,$
   $f(-5) = 25$
2. $y = f(x) = x^2 - 1, f : \mathbb{R} \to \mathbb{R}$
3. $y = f(x) = \frac{x-3}{x^2-2}, f : [1, 1] \to \mathbb{R}$
4. $y = f(x) = \sqrt{x}, f : \mathbb{N} \to \mathbb{R}$
5. $y = f(x) = \begin{cases} -1 & \text{falls } x < 0, \\ 0 & \text{falls } x = 0, \\ +1 & \text{falls } x > 0, \end{cases} f : \mathbb{R} \to \mathbb{R}.$ ◀

5.2.6 Bijektive Funktionen

Eine Funktion heißt bijektiv, wenn sie sowohl injektiv als auch surjektiv ist.

Bei einer bijektiven Funktion ist also die Bildmenge gleich dem Wertebereich, und jedes Bild besitzt genau ein Urbild. Ist $y = f(x), f : D \to W$ eine bijektive Funktion, so sind die Mengen D und W gleich mächtig, das heißt, sie besitzen gleich viele Elemente.

Die bijektiven Funktionen besitzen eine Umkehrfunktion.

Beispiele

Folgende Funktionen sind bijektiv:

1. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, 5\}$,
   $f(-1) = 1, f(-2) = 2, f(-3) = 3, f(-4) = 4,$
   $f(-5) = 5$
2. $y = f(x) = x + 2, f : \mathbb{R} \to \mathbb{R}$

Folgende Funktionen sind nicht bijektiv:

1. $D = \{-1, -2, -3, -4, -5\}$,
   $W = \{1, 2, 3, 4, \ldots, 24, 25\}$,
   $f(-1) = 1, f(-2) = 4, f(-3) = 9, f(-4) = 16,$
   $f(-5) = 25$
2. $y = f(x) = x^3 - 4x^2 - x + 4, f : \mathbb{R} \to \mathbb{R}$
3. $y = f(x) = x^2 - 1, f : \mathbb{R} \to \mathbb{R}$
4. $y = f(x) = \frac{x-3}{x^2-2}, f : [1, 1] \to \mathbb{R}$
5. $y = f(x) = \sqrt{x}, f : \mathbb{N} \to \mathbb{R}$
6. $y = f(x) = \begin{cases} -1 & \text{falls } x < 0, \\ 0 & \text{falls } x = 0, \\ +1 & \text{falls } x > 0, \end{cases} f : \mathbb{R} \to \mathbb{R}.$ ◀

5.2.7 Periodische Funktionen

Eine Funktion, deren Funktionsgleichung die Bedingung $f(x + T) = f(x)$ erfüllt, wobei T eine Konstante (feste reelle Zahl) ist, heißt periodische Funktion. Die Gleichung $f(x + T) = f(x)$ gilt für alle x aus dem Definitionsbereich.

$$f(x + T) = f(x)$$

Die kleinste positive Zahl T mit dieser Eigenschaft heißt die Periode der Funktion. Den absolut größten Funktionswert nennt man Amplitude der periodischen Funktion.

Beispiele für periodische Funktionen sind die trigonometrischen Funktionen (vgl. Kap. 6).

5.2.8 Umkehrfunktionen

Die Funktion, die durch Vertauschen von x und y aus einer bijektiven Funktion $y = f(x)$ entsteht, heißt Umkehrfunktion oder inverse Funktion von $y = f(x)$.

Bei einer bijektiven Funktion $y = f(x), f : D \to W$ ist jedes Element $y \in W$ Bild von genau einem Element $x \in D$. Man kann eine neue Funktion definieren, die jedem $y \in W$ als Bild gerade das $x \in D$ zuordnet, das Urbild von y ist. Diese Funktion leistet das Umgekehrte wie f, ihr Definitionsbereich ist W, und ihr Wertebereich ist D. Man nennt diese Funktion daher die Umkehrfunktion von f und bezeichnet sie mit f^{-1}.

$$y = f^{-1}(x), \quad f^{-1} : W \to D$$

Versteht man unter der Schreibweise $g(f(x))$, dass man auf x die Zuordnungsvorschrift f und dann auf $f(x)$ die Vorschrift g anwendet, so gilt $f^{-1}(f(x)) = x$ und $f(f^{-1}(x)) = x$.

Zu einer streng monoton wachsenden oder streng monoton fallenden Funktion existiert die Umkehrfunktion.

Bestimmung der Umkehrfunktion:

1. Auflösen von $y = f(x)$ nach x: $x = f^{-1}(y)$
2. Vertauschen von x und y: $y = f^{-1}(x)$

Diesen Operationen entspricht die Spiegelung des Graphen der Funktion an der Winkelhalbierenden $y = x$.

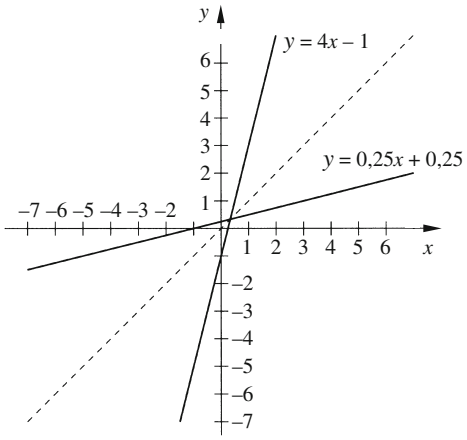


Abb. 5.6 Graphen der Funktionen von Beispiel 1

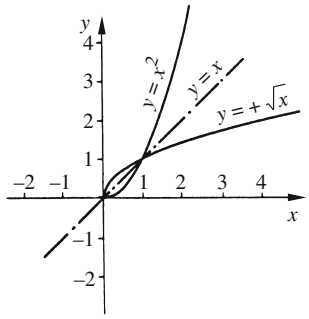


Abb. 5.7 Graphen der Funktionen von Beispiel 2

Beispiele

1. $y = f(x) = 4x - 1$ (Abb. 5.6), $D = W = \mathbb{R}$
   Umkehrfunktion: $y = f^{-1}(x) = \frac{1}{4}x + \frac{1}{4}$, $D = W = \mathbb{R}$
2. $y = f(x) = x^2$ (Abb. 5.7), $D = W = \{x \mid x \in \mathbb{R}, x \geq 0\}$
   Umkehrfunktion: $y = f^{-1}(x) = \sqrt{x}$, $D = W = \{x \mid x \in \mathbb{R}, x \geq 0\}$
3. $y = f(x) = e^x$ (Abb. 5.8), $D = \mathbb{R}$, $W = \mathbb{R}^+$
   Umkehrfunktion: $y = f^{-1}(x) = \ln x$, $D = \mathbb{R}^+$, $W = \mathbb{R}$
4. $y = f(x) = 2^x$, $D = \mathbb{R}$, $W = \mathbb{R}^+$
   Umkehrfunktion: $y = f^{-1}(x) = \log_2 x$, $D = \mathbb{R}^+$, $W = \mathbb{R}$ ◀

5.2.9 Reelle und komplexe Funktionen

Eine Funktion mit der Funktionsgleichung $y = f(x)$, deren Definitions- und Wertebereich nur reelle Zahlen enthalten, nennt man eine reelle Funktion einer reellen Variablen.

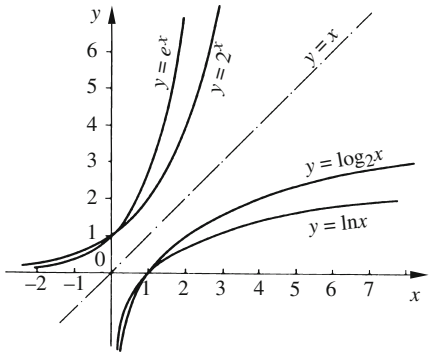


Abb. 5.8 Graphen der Funktionen von Beispiel 3 und Beispiel 4

Beispiele

1. $y = x^2$, $D = (-\infty, \infty)$, $W = [0, \infty)$
2. $y = \sqrt{x}$, $D = [0, \infty)$, $W = [0, \infty)$ ◀

Ist dagegen die unabhängige Variable einer Funktionsgleichung eine komplexe Zahl z, dann wird durch $w = f(z)$ eine komplexe Funktion einer komplexen Variablen beschrieben. Komplexe Funktionen werden in dem mathematischen Gebiet Funktionentheorie behandelt.

5.3 Einteilung der elementaren Funktionen

Eine elementare Funktion ist eine Funktion, deren Funktionsgleichung durch einen geschlossenen analytischen Ausdruck dargestellt werden kann.

Elementare Funktionen sind durch Formeln definiert, die nur endlich viele mathematische Operationen mit der unabhängigen Variablen x und den Koeffizienten enthalten.

Man teilt die elementaren Funktionen in algebraische Funktionen und transzendente Funktionen ein.

Bei algebraischen Funktionen lassen sich die Verknüpfung der unabhängigen Variablen x und der abhängigen Variablen y in einer algebraischen Gleichung folgender Form darstellen, wobei $p_0, p_1, \ldots, p_n$ Polynome in x beliebigen Grades sind.

$$p_0(x) + p_1(x)y + p_2(x)y^2 + \ldots + p_n(x)y^n = 0$$

Elementare Funktionen, die nicht algebraisch sind, heißen transzendent.

Beispiele für algebraische Funktionen

1. $y = 3x^2 + 4$
2. $y = \frac{2x}{x^3+2x-1}$
3. $y = \sqrt{2x+3}$
4. $3xy^3 - 4xy + x^3 - 1 = 0$ (hier also $p_0(x) = x^3 - 1$, $p_1(x) = -4x$, $p_2(x) = 0$, $p_3(x) = 3x$). ◀

Zu den transzendenten Funktionen gehören zum Beispiel die Exponentialfunktionen, die Logarithmusfunktionen und die trigonometrischen Funktionen.

Beispiele für transzendente Funktionen

1. $y = e^x$
2. $y = \sin x$
3. $y = \ln x$
4. $y = \frac{\ln x + \sqrt{\sin x}}{x^2+5}$. ◀

Die algebraischen Funktionen untergliedern sich in die rationalen Funktionen und in die irrationalen Funktionen.

Eine rationale Funktion ist eine algebraische Funktion, für die die Funktionsgleichung $y = f(x)$ als eine explizite Formel angegeben werden kann, in der auf die unabhängige Variable x nur endlich viele rationale Rechenoperationen (Addition, Subtraktion, Multiplikation und Division) angewandt werden.

Eine algebraische Funktion, die nicht rational ist, heißt irrational.

Beispiele für rationale Funktionen

1. $y = 3x^3 - \frac{1}{4}$
2. $y = \frac{2x^2-3x+5}{x^3+3x^2-2}$ ◀

Bei irrationalen Funktionen tritt die unabhängige Variable auch unter einem Wurzelzeichen auf.

Beispiele für irrationale Funktionen

1. $y = \sqrt{3x^2+4}$
2. $y = \sqrt[3]{(x^2+1)\sqrt{x}}$ ◀

Für rationale Funktionen ist $f(x)$ ein Polynom (dann ist $y = f(x)$ eine ganze rationale Funktion) oder ein Quotient aus Polynomen (dann heißt $y = f(x)$ eine gebrochene rationale Funktion).

Ganze rationale Funktionen lassen sich also darstellen in folgender Form mit $a_0, a_1, a_2, \ldots, a_{n-1}, a_n \in \mathbb{R}$, $a_n \neq 0$, $n \in \mathbb{N}$.

$$y = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0 = \sum_{k=0}^{n} a_k x^k$$

Ist n der Grad des Polynoms, so nennt man die Funktion ganze rationale Funktion n-ten Grades. Bei ganzen rationalen Funktionen werden auf die unabhängige Variable x nur die Operationen Addition, Subtraktion und Multiplikation angewandt.

Ganze rationale Funktionen vom Grad 0 ($y = a_0$) nennt man konstante Funktionen, vom Grad 1 ($y = a_1x + a_0$) lineare Funktionen, vom Grad 2 ($y = a_2x^2 + a_1x + a_0$) quadratische Funktionen und vom Grad 3 ($y = a_3x^3 + a_2x^2 + a_1x + a_0$) kubische Funktionen.

Konstante Funktionen	$y = a_0$
Lineare Funktionen	$y = a_1x + a_0$
Quadratische Funktionen	$y = a_2x^2 + a_1x + a_0$
Kubische Funktionen	$y = a_3x^3 + a_2x^2 + a_1x + a_0$

Beispiele für ganze rationale Funktionen

1. $y = 23x^4 - 12x + 4$
2. $y = \frac{11}{12}x^{23} - 11x^{17} - 12x^9 - \pi x^6 - \frac{12}{\sqrt{5}}x - 2$
3. $y = 1 - 3x + x^6 - 2x^2$
4. $y = -7$ (konstante Funktion)
5. $y = 3x - 4\pi$ (lineare Funktion)
6. $y = x^2 - x + 1$ (quadratische Funktion)
7. $y = 4x^3 - 2x + 5$ (kubische Funktion) ◀

Gebrochene rationale Funktionen sind Funktionen mit einer Funktionsgleichung $y = f(x)$, bei der $f(x)$ als Quotient zweier Polynome darstellbar ist. Sie besitzen also eine Darstellung folgender Form mit $a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_m \in \mathbb{R}$, $a_n, b_m \neq 0$, $m \neq 0$.

$$y = \frac{a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \ldots + b_2x^2 + b_1x + b_0} = \frac{\sum_{i=0}^{n} a_i x^i}{\sum_{k=0}^{m} b_k x^k}$$

Eine gebrochene rationale Funktion kann also immer als Quotient zweier ganzer rationaler Funktionen dargestellt werden. Bei gebrochenen rationalen Funktionen werden auf die unabhängige Variable x nur die Grundrechenarten (also die Operationen Addition, Subtraktion, Multiplikation und Division) angewandt. Die Definitionsmenge einer gebrochenen rationalen Funktion besteht aus denjenigen reellen Zahlen, für die der Nenner nicht Null wird.

Für $n < m$ heißt die Funktion echt gebrochene rationale Funktion, für $n \geq m$ heißt sie unecht gebrochene rationale Funktion.

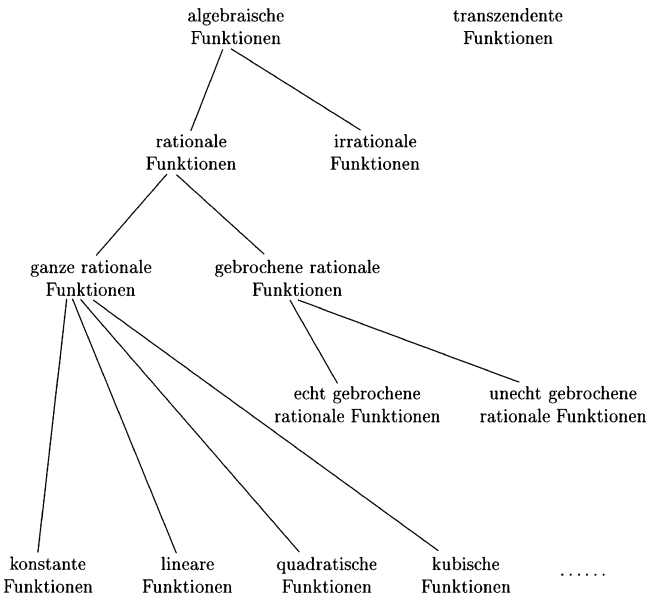
Gebrochene rationale Funktionen mit $n = 1$ und $m = 1$, also $y = \frac{a_1x+a_0}{b_1x+b_0}$, heißen gebrochene lineare Funktionen.

Beispiele für gebrochene rationale Funktionen

1. $y = \frac{2}{x}$
2. $y = \frac{x^4-22x^3+\frac{1}{3}x^2-12}{x^5-11x^3+x+1}$
3. $y = \frac{2x}{x^3-5x^2-2x+1}$
4. $y = \frac{2x+4}{x-3}$ (gebrochene lineare Funktion)
5. $y = \frac{x^5-2}{x^2+1}$
6. $y = x^2 + x + \frac{1}{x}$ $(= \frac{x^3+x^2+1}{x})$ ◀

Bei den ersten drei Beispielen handelt es sich um echt gebrochene rationale Funktionen, bei den letzten drei Beispielen um unecht gebrochene rationale Funktionen.

Zusammenfassende Übersicht über die elementaren Funktionen



5.4 Ganze rationale Funktionen

5.4.1 Konstante Funktionen

Funktionen mit einer Funktionsgleichung

$$y = f(x) = n \quad (n \in \mathbb{R})$$

Der Graph einer konstanten Funktion ist eine Parallele zur x-Achse, und zwar im Abstand n. Im Fall $n = 0$ ist die Gerade die x-Achse selbst. Die Geradengleichung der x-Achse ist also $y = 0$ (Abb. 5.9).

Um den Graph einer Funktion zu zeichnen, ist es sinnvoll, sich die Koordinaten von Punkten des Graphen in einer Wertetabelle aufzuschreiben. Da eine Gerade durch zwei auf ihr liegende Punkte festgelegt ist, reicht es im Prinzip, bei Geraden die Koordinaten von zwei Punkten zu berechnen.

Beispiel

Funktionsgleichung: $y = f(x) = 2$

Wertetabelle:

x	-1	0	1
y	2	2	2

◀

5.4.2 Lineare Funktionen

Funktionen mit einer Funktionsgleichung

$$y = f(x) = mx + n \quad (m, n \in \mathbb{R}, m \neq 0)$$

Eine lineare Funktion ist eine ganze rationale Funktion 1. Grades.

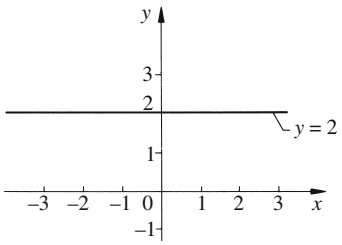


Abb. 5.9 Graph der konstanten Funktion $y = 2$

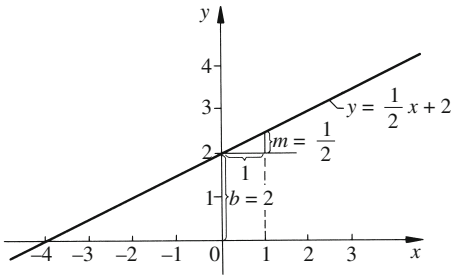


Abb. 5.10 Graph der linearen Funktion $y = \frac{1}{2}x + 2$

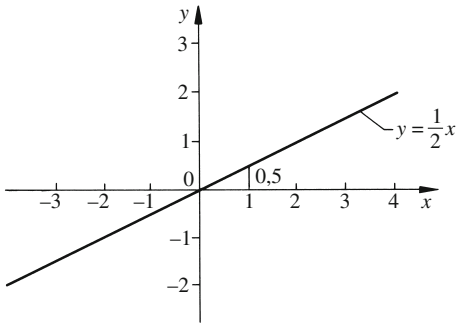


Abb. 5.11 Graph der Proportionalfunktion $y = \frac{1}{2}x$

Der Graph einer linearen Funktion ist eine Gerade (daher der Name lineare Funktion), und zwar die Gerade mit der Steigung m und dem Achsenabschnitt n auf der y-Achse (vgl. Abschn. 7.2.1). Die Steigung m einer Geraden ist der „Höhenzuwachs“ (die Differenz der y-Werte) bei einem Schritt um 1 nach rechts. Der Achsenabschnitt n ist der y-Wert, bei dem die Gerade die y-Achse schneidet.

Für $m > 0$ ist die Funktion streng monoton wachsend, für $m < 0$ ist sie streng monoton fallend.

Schnittpunkt des Graphen der Funktion mit der x-Achse: $S_x(\frac{-n}{m}|0)$, Schnittpunkt mit der y-Achse: $S_y(0|n)$.

Beispiel

Funktionsgleichung: $y = f(x) = \frac{1}{2}x + 2$ (Abb. 5.10)

Wertetabelle:

x	-6	-4	-2	0	2	4
y	-1	0	1	2	3	4

◀

Ist $n = 0$, so nennt man die lineare Funktion $y = mx$ ($m \in \mathbb{R}$, $m \neq 0$) auch Proportionalfunktion. Der Graph einer Proportionalfunktion ist eine Gerade durch den Koordinatenursprung, und zwar mit der Steigung m.

Man nennt m auch den Proportionalitätsfaktor der Gleichung, denn es gilt $m = \frac{y}{x}$.

Beispiel

Funktionsgleichung: $y = f(x) = \frac{1}{2}x$ (Abb. 5.11)

Wertetabelle:

x	-3	-2	-1	0	1	2	3
y	$\frac{3}{2} = -1{,}5$	-1	$\frac{1}{2} = -0{,}5$	0	$\frac{1}{2} = 0{,}5$	1	$\frac{3}{2} = 1{,}5$

◀

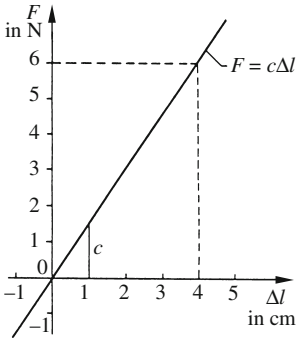


Abb. 5.12 Hooke'sches Gesetz

Anwendungsbeispiele

1. *Hooke'sches Gesetz* (Abb. 5.12)
   Funktionsgleichung: $F = c\Delta l$ (statt $y = mx$)
   Dabei bedeuten:
   F Federkraft, c Federkonstante (Federrate, Federsteifigkeit), Δl Längenveränderung der Feder
   Beispiel (vgl. Abschn. 2.4):
   Welche Kraft F dehnt eine Feder um 4 cm, wenn die Kraft 3 N (Newton) eine Dehnung um 2 cm bewirkt?

$$c = \frac{3\,\text{N}}{2\,\text{cm}} = 1{,}5\,\frac{\text{N}}{\text{cm}}, \quad \Delta l = 4\,\text{cm}$$

   Einsetzen in die Funktionsgleichung:

$$F = 1{,}5\,\frac{\text{N}}{\text{cm}} \cdot 4\,\text{cm} = 6\,\text{N}$$

   Antwort: Die Kraft 6 N bewirkt die Dehnung um 4 cm.
   Anmerkung: Es muss sichergestellt sein, dass sich die Werte im materialbedingten Gültigkeitsbereich des Hooke'schen Gesetzes bewegen.

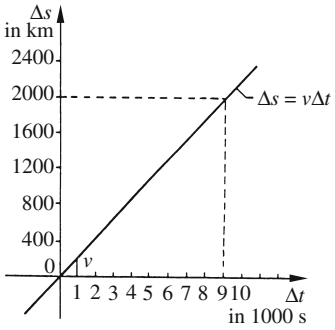


Abb. 5.13 Geschwindigkeit im Weg-Zeit-Diagramm

2. *Geschwindigkeit im Weg-Zeit-Diagramm* (Abb. 5.13)
Funktionsgleichung: $\Delta s = v\Delta t$ (statt $y = mx$)
Dabei bedeuten:
Δs zurückgelegter Weg, v Geschwindigkeit, Δt abgelaufene Zeit
Beispiel (vgl. Abschn. 2.4):
Wie weit kommt ein Flugzeug in $2\frac{1}{2}$ Stunden, wenn es 10 km in 45 s zurücklegt?

$$v = \frac{10.000\,\text{m}}{45\,\text{s}} = 222{,}\overline{2}\,\frac{\text{m}}{\text{s}},$$
$$\Delta t = 2{,}5\,\text{h} \cdot \frac{3600\,\text{s}}{\text{h}} = 9000\,\text{s}$$

Einsetzen in die Funktionsgleichung:

$$\Delta s = \frac{10.000\,\text{m}}{45\,\text{s}} \cdot 9000\,\text{s} = 2000\,\text{km}$$

Antwort: Das Flugzeug fliegt 2000 km weit.

3. *Gay-Lussac'sches Gesetz* (Abb. 5.14)
Funktionsgleichung: $V = V_0(1 + \gamma\delta)$ oder $V = V_0\gamma\delta + V_0$ (statt $y = mx + n$)
Dabei bedeuten:
V variables Gasvolumen, V_0 Volumen derselben Gasmenge bei 0 °C, $\gamma = \frac{1}{273}$ konstanter Volumenausdehnungskoeffizient, δ Maßzahl der in °C gemessenen variablen Temperatur
Ein Vergleich mit der linearen Funktion ergibt:

$$V = y, \quad V_0\gamma = \frac{V_0}{273} = m, \quad \delta = x, \quad V_0 = b$$

Der Definitionsbereich ist gegeben durch die Bedingung $-273 \le \delta < \infty$.
Beispiel (vgl. Abschn. 2.4):
Im Winderhitzer eines Hochofens werden stündlich 42.000 m³ Luft von 17 auf 800 °C erwärmt. Wie groß ist das Volumen der vom Winderhitzer pro Stunde gelieferten erhitzten Luft?
Lösung: Die Gleichung $V = V_0(1 + \gamma\delta)$ wird zweimal benutzt.

Wertetabelle:

	V
−273	0
0	V_0
100	$\frac{373}{273}V_0$
200	$\frac{473}{273}V_0$
273	$2\,V_0$

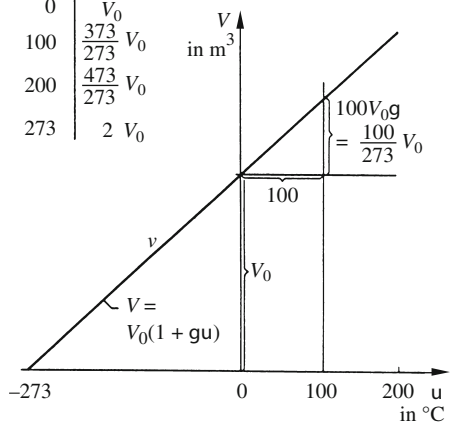


Abb. 5.14 Gay-Lussac'sches Gesetz

Zuerst Berechnung von V_0:

$$V_0 = V_1\frac{1}{1+\gamma\delta_1} = 42.000\,\text{m}^3\frac{1}{1+\frac{1}{273}\cdot 17}$$
$$= 39.537{,}9310\ldots\,\text{m}^3$$

Anschließend Berechnung des gesuchten V_2:

$$V_2 = V_0(1+\gamma\delta_2)$$
$$= 39.537{,}9310\ldots\,\text{m}^3\left(1+\frac{1}{273}\cdot 800\right)$$
$$= 155.400\,\text{m}^3$$

Antwort: Das Volumen beträgt 155.400 m³.
Wertetabelle:

ϑ	−273	0	100	200	300
V	0	V_0	$\frac{373}{273}V_0$	$\frac{473}{273}V_0$	$2V_0$

◀

5.4.3 Quadratische Funktionen

Funktionen mit einer Funktionsgleichung

$$y = f(x) = a_2x^2 + a_1x + a_0 \quad (a_2, a_1, a_0 \in \mathbb{R}, a_2 \neq 0)$$

Eine quadratische Funktion ist eine ganze rationale Funktion 2. Grades.

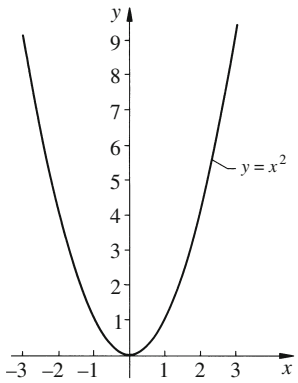


Abb. 5.15 Normalparabel

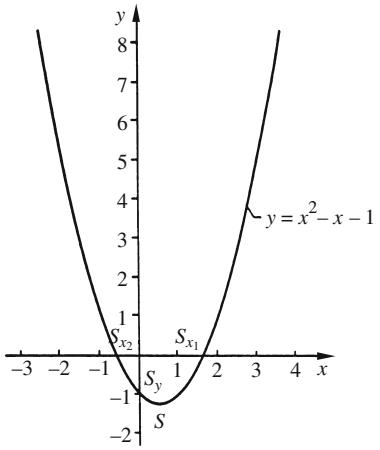


Abb. 5.16 Verschobene Normalparabel

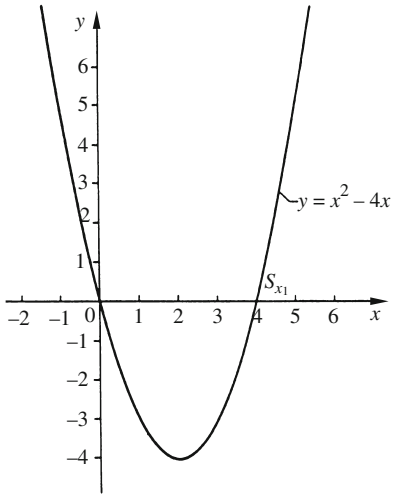


Abb. 5.17 Verschobene Normalparabel

Der Graph jeder quadratischen Funktion ist eine Parabel (vgl. auch Abschn. 7.5.3).

Für spezielle Koeffizienten a_2, a_1, a_0 in der Funktionsgleichung erhält man spezielle Parabeln.

Normalparabel

Mit den Koeffizienten $a_2 = 1$, $a_1 = 0$, $a_0 = 0$ in der Gleichung $y = a_2x^2 + a_1x + a_0$ der quadratischen Funktion erhält man die Gleichung $y = x^2$ der Normalparabel.

Normalparabel

$$y = x^2$$

Der Punkt (0|0), also der Koordinatenursprung, ist der Scheitelpunkt der Normalparabel. Die Normalparabel ist symmetrisch zur y-Achse und nach oben geöffnet (Abb. 5.15).

Der Definitionsbereich ist $D = \mathbb{R}$, der Wertebereich ist $W = \mathbb{R}$, und die Bildmenge $f(D)$ ist die Menge der nichtnegativen reellen Zahlen: $f(D) = R_0^+ = R^+ \cup \{0\}$.

Verschobene Normalparabel

Mit $a_2 = 1$ und beliebigen Werten für a_1 und a_0 (aber nicht beide gleich 0) in $y = a_2x^2 + a_1x + a_0$ ergibt sich die Gleichung $y = x^2 + a_1x + a_0$ einer verschobenen Normalparabel (Abb. 5.16).

Verschobene Normalparabel

$$y = x^2 + a_1x + a_0$$

Eine verschobene Normalparabel hat dieselbe Form wie die Normalparabel, der Scheitelpunkt liegt jedoch nicht im Koordinatenursprung (Abb. 5.17).

Berechnung des Scheitelpunkts S der verschobenen Normalparabel mit der Gleichung $y = x^2 + a_1x + a_0$: Subtraktion von a_0 ergibt $y - a_0 = x^2 + a_1x$.

Durch quadratische Ergänzung erhält man $y - a_0 + (\frac{a_1}{2})^2 = x^2 + a_1x + (\frac{a_1}{2})^2$, woraus $y - (a_0 - \frac{a_1^2}{4}) = (x + \frac{a_1}{2})^2$ folgt.

Die rechte Gleichungsseite und damit auch die linke ist ≥ 0. Der kleinste y-Wert ergibt sich, wenn beide Gleichungsseiten gleich 0 sind, also für $y = a_0 - \frac{a_1^2}{4}$. Dies ist der Wert der Ordinate des Scheitelpunkts.

Die rechte Gleichungsseite wird für $x = -\frac{a_1}{2}$ gleich 0, dem Wert der Abszisse des Scheitelpunkts. Man erhält somit als Scheitelpunkt der verschobenen Normalparabel

$$S(x_s|y_s) = S\left(-\frac{a_1}{2}\middle|a_0 - \frac{a_1^2}{4}\right)$$

Berechnung des Schnittpunkts S_y der verschobenen Normalparabel mit der y-Achse:

Die Ordinate des Schnittpunkts mit der y-Achse ergibt sich, wenn man in der Funktionsgleichung $x = 0$ setzt: $y = 0 + 0 + a_0 = a_0$. Somit ist $S_y = S_y(0|a_0)$.

Berechnung der Schnittpunkte S_{x_1} und S_{x_2} der verschobenen Normalparabel mit der x-Achse:

Die Abszissen der Schnittpunkte mit der x-Achse erhält man, wenn man in der Funktionsgleichung $y = 0$ setzt: $0 = x^2 + a_1x + a_0$. Dies ist eine quadratische Gleichung in x, deren Nullstellen bestimmt werden müssen:

$$x_{1,2} = -\frac{a_1}{2} \pm \sqrt{\left(\frac{a_1}{2}\right)^2 - a_0}$$

Die Schnittpunkte mit der x-Achse lauten also:

$$S_{x_1} = S_{x_1}\left(-\frac{a_1}{2} + \sqrt{\frac{a_1^2}{4} - a_0}\,\middle|\,0\right)$$
$$S_{x_2} = S_{x_2}\left(-\frac{a_1}{2} - \sqrt{\frac{a_1^2}{4} - a_0}\,\middle|\,0\right)$$

Beispiele

1. $y = x^2 - x - 1$ (also $a_1 = -1$ und $a_0 = -1$)
   Wegen $\frac{a_1^2}{4} = \frac{(-1)^2}{4} = \frac{1}{4}$ ergibt sich
   Scheitelpunkt: $S = S(\frac{1}{2}|-1-\frac{1}{4}) = S(\frac{1}{2}|-\frac{5}{4})$
   Schnittpunkt mit der y-Achse: $S_y = S_y(0|-1)$
   Schnittpunkte mit der x-Achse:

$$S_{x_1} = S_{x_1}\left(\frac{1}{2} + \sqrt{\frac{1}{4}+1}\,\middle|\,0\right) = S_{x_1}\left(\frac{1}{2}(1+\sqrt{5})\,\middle|\,0\right)$$
$$= S_{x_1}(1{,}6180\ldots|0),$$
$$S_{x_2} = S_{x_2}\left(\frac{1}{2} - \sqrt{\frac{1}{4}+1}\,\middle|\,0\right) = S_{x_2}\left(\frac{1}{2}(1-\sqrt{5})\,\middle|\,0\right)$$
$$= S_{x_2}(-0{,}6180\ldots|0)$$

2. $y = x^2 - 4x$ (also $a_1 = -4$ und $a_0 = 0$) (Abb. 5.17)
   Es ergibt sich $\frac{a_1^2}{4} = \frac{(-4)^2}{4} = 4$. Damit berechnet man den Scheitelpunkt: $S = S(-\frac{-4}{2}|-4) = S(2|-4)$
   Schnittpunkt mit der y-Achse: $S_y = S_y(0|0)$
   Schnittpunkte mit der x-Achse:

$$S_{x_1} = S_{x_1}\left(2 + \sqrt{4-0}\,\middle|\,0\right) = S_{x_1}(4|0),$$
$$S_{x_2} = S_{x_2}(2-2|0) = S_{x_2}(0|0)$$

   Die Parabel geht durch den Koordinatenursprung, deshalb fällt der Schnittpunkt mit der y-Achse mit dem einen Schnittpunkt mit der x-Achse zusammen. ◀

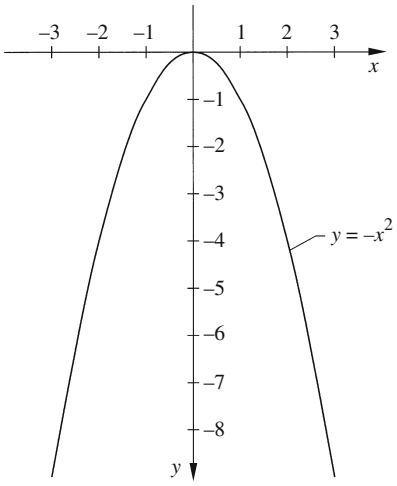


Abb. 5.18 Gespiegelte Normalparabel

Gespiegelte Normalparabel

Mit den Koeffizienten $a_2 = -1$, $a_1 = 0$, $a_0 = 0$ in $y = a_2x^2 + a_1x + a_0$ ergibt sich die Gleichung $y = -x^2$ der gespiegelten Normalparabel (Abb. 5.18).

Gespiegelte Normalparabel

$$y = -x^2$$

Die gespiegelte Normalparabel entsteht aus der Normalparabel durch Spiegelung an der x-Achse. Der Punkt $(0|0)$ ist der Scheitelpunkt der gespiegelten Normalparabel. Sie ist symmetrisch zur y-Achse und nach unten geöffnet.

Gespiegelte verschobene Normalparabel

Mit $a_2 = -1$ und beliebigen Werten für a_1 und a_0 (aber nicht beide gleich 0) ergibt sich die Gleichung $y = -x^2 + a_1x + a_0$ einer gespiegelten verschobenen Normalparabel.

Gespiegelte verschobene Normalparabel

$$y = -x^2 + a_1x + a_0$$

Berechnung des Scheitelpunkts S:

Subtraktion von a_0: $y - a_0 = -x^2 + a_1x$

Subtraktion von $(\frac{a_1}{2})^2$:

$$y - a_0 - \left(\frac{a_1}{2}\right)^2 = -x^2 + a_1x - \left(\frac{a_1}{2}\right)^2$$

Zusammenfassen auf der linken Seite des Gleichheitszeichens und Ausklammern von -1 auf der rechten Seite:

$$y - \left(a_0 + \left(\frac{a_1}{2}\right)^2\right) = -\left(x^2 - a_1x + \left(\frac{a_1}{2}\right)^2\right)$$

Anwenden der zweiten binomischen Formel auf der rechten Seite:

$$y-\left(a_0+\frac{a_1^2}{4}\right)=-\left(x-\frac{a_1}{2}\right)^2$$

Die rechte Gleichungsseite und damit auch die linke ist ≤ 0. Der größte y-Wert ergibt sich, wenn beide Gleichungsseiten gleich 0 sind, also für $y=a_0+\frac{a_1^2}{4}$. Dies ist der Wert der Ordinate des Scheitelpunkts.

Die rechte Gleichungsseite wird für $x=\frac{a_1}{2}$ gleich 0, dem Wert der Abszisse des Scheitelpunkts. Somit ergibt sich für den Scheitelpunkt der gespiegelten verschobenen Normalparabel

$$S(x_S|y_S)=S\left(\frac{a_1}{2}\middle|a_0+\frac{a_1^2}{4}\right)$$

Berechnung des Schnittpunkts S_y mit der y-Achse:

Durch Einsetzen von $x=0$ in die Funktionsgleichung erhält man $y=a_0$ als Ordinate des Schnittpunkts und als Schnittpunkt somit $S_y=S_y(0|a_0)$.

Berechnung der Schnittpunkte S_{x_1} und S_{x_2} mit der x-Achse:

Durch Einsetzen von $y=0$ in die Funktionsgleichung ergibt sich $0=-x^2+a_1x+a_0$ und nach Multiplikation der Gleichung mit -1 die quadratische Gleichung $x^2-a_1x-a_0=0$, die die Lösungen $x_{1,2}=\frac{a_1}{2}\pm\sqrt{(\frac{a_1}{2})^2+a_0}$ hat. Die Schnittpunkte mit der x-Achse lauten somit:

$$S_{x_1}=S_{x_1}\left(\frac{a_1}{2}+\sqrt{\frac{a_1^2}{4}+a_0}\middle|0\right)$$

$$S_{x_2}=S_{x_2}\left(\frac{a_1}{2}-\sqrt{\frac{a_1^2}{4}+a_0}\middle|0\right)$$

Beispiel

3. $y=-x^2-4x+3$ (also $a_1=-4$ und $a_0=3$) (Abb. 5.19)
   Man berechnet $\frac{a_1^2}{4}=\frac{(-4)^2}{4}=4$
   Scheitelpunkt: $S=S(-2|3+4)=S(-2|7)$
   Schnittpunkt mit der y-Achse: $S_y=S_y(0|3)$
   Schnittpunkte mit der x-Achse:

$$S_{x_1}=S_{x_1}(-2+\sqrt{4+3}|0)=S_{x_1}(-2+\sqrt{7}|0)$$
$$=S_{x_1}(0{,}6457\ldots|0),$$
$$S_{x_2}=S_{x_2}(-2-\sqrt{7}|0)=S_{x_2}(-4{,}6457\ldots|0)$$ ◀

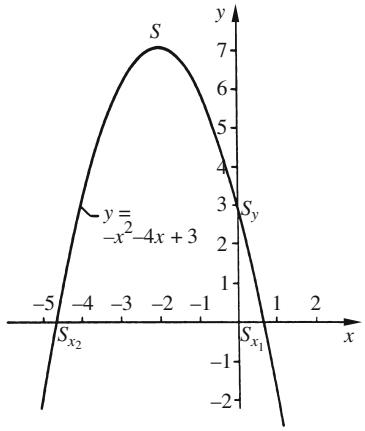


Abb. 5.19 Gespiegelte verschobene Normalparabel

Allgemeiner Fall

Parabel

$$y=a_2x^2+a_1x+a_0$$

Für $a_2>0$ ist die Parabel nach oben, für $a_2<0$ nach unten geöffnet.

Für $|a_2|>1$ ist die Parabel im Vergleich zur Normalparabel gestreckt und für $|a_2|<1$ gestaucht. Man nennt $|a_2|$ deshalb den Streckungsfaktor der Parabel.

Eine Änderung des Koeffizienten a_1 bewirkt eine Verschiebung der Parabel in x-Richtung, eine Änderung von a_0 bewirkt eine Verschiebung in y-Richtung.

Berechnung des Scheitelpunkts S:

Subtraktion von a_0: $y-a_0=a_2x^2+a_1x$

Ausklammern von a_2 auf der rechten Seite:

$$y-a_0=a_2\left(x^2+\frac{a_1}{a_2}x\right)$$

Quadratische Ergänzung in der Klammer auf der rechten Seite, also Addition von $\frac{a_1^2}{4a_2}$ auf beiden Seiten der Gleichung:

$$y-a_0+\frac{a_1^2}{4a_2}=a_2\left(x^2+\frac{a_1}{a_2}x+\left(\frac{a_1}{2a_2}\right)^2\right)$$

Anwenden der ersten binomischen Formel auf der rechten Seite:

$$y-\left(a_0-\frac{a_1^2}{4a_2}\right)=a_2\left(x+\frac{a_1}{2a_2}\right)^2$$

Hieraus liest man die Koordinaten des Scheitelpunkts ab, nämlich diejenigen Werte für x und y, für die beide Seiten der Gleichung gleich 0 werden:

$$x_S=-\frac{a_1}{2a_2},\quad y_S=a_0-\frac{a_1^2}{4a_2}$$

Scheitelpunkt S der Parabel:

$$S(x_S|y_S)=S\left(-\frac{a_1}{2a_2}\middle|a_0-\frac{a_1^2}{4a_2}\right)$$

Man nennt die Gleichung $y-y_S=a_2(x-x_S)^2$ Scheitelform der quadratischen Funktion, wohingegen $y=a_2x^2+a_1x+a_0$ Normalform der quadratischen Funktion heißt.

Scheitelform der quadratischen Funktion

$$y - y_S = a_2(x - x_S)^2$$

Berechnung des Schnittpunkts S_y mit der y-Achse:

Setzt man in der Funktionsgleichung $x = 0$ ein, so erhält man $y = a_0$ als Ordinate des Schnittpunkts S_y und als Schnittpunkt somit $S_y = S_y(0|a_0)$.

Der Wert $D = a_1^2 - 4a_2a_0$ heißt Diskriminante der quadratischen Funktion $y = a_2x^2 + a_1x + a_0$. Gilt $D > 0$, so hat die zugehörige Parabel zwei Schnittpunkte mit der x-Achse. Für $D = 0$ gibt es einen Schnittpunkt (der Schnittpunkt ist dann ein Berührpunkt). Für $D < 0$ gibt es keinen Schnittpunkt mit der x-Achse.

Berechnung des Schnittpunkts (für $D = 0$) bzw. der Schnittpunkte (für $D > 0$) der Parabel mit der x-Achse:

Durch Einsetzen von $y = 0$ in die Funktionsgleichung $y = a_2x^2 + a_1x + a_0$ erhält man die quadratische Gleichung $a_2x^2 + a_1x + a_0 = 0$, die die Lösungen $x_{1,2} = \frac{1}{2a_2}(-a_1 \pm \sqrt{a_1^2 - 4a_2a_0})$ hat. Daraus ergeben sich die Schnittpunkte mit der x-Achse:

$$S_{x_1}\left(\frac{1}{2a_2}(-a_1 + \sqrt{a_1^2 - 4a_2a_0})\middle|0\right),$$
$$S_{x_2}\left(\frac{1}{2a_2}(-a_1 - \sqrt{a_1^2 - 4a_2a_0})\middle|0\right)$$

Beispiele

4. $y = 4x^2$ (also $a_2 = 4$, $a_1 = a_0 = 0$) (Abb. 5.20)
   Wegen $|a_2| = |4| = 4 > 1$ ist die Parabel im Vergleich zur Normalparabel gestreckt, und wegen $a_2 = 4 > 0$ ist die Parabel nach oben geöffnet.
   Aus $a_1 = a_0 = 0$ folgt $S = S(0|0)$ und $S = S_x = S_y$.
5. $y = -3x^2 + 1{,}2x - 1{,}5$ (also $a_2 = -3$, $a_1 = 1{,}2$, $a_0 = -1{,}5$) (Abb. 5.21)
   Aus $|a_2| = |-3| = 3 > 1$ folgt, dass die Parabel im Vergleich zur Normalparabel gestreckt ist. Wegen $a_2 = -3 < 0$ ist die Parabel nach unten geöffnet.
   Berechnung des Scheitelpunkts S:

$$\begin{aligned} S(x_S|y_S) &= S\left(-\frac{1{,}2}{2\cdot(-3)}\middle|-1{,}5 - \frac{(1{,}2)^2}{4\cdot(-3)}\right) \\ &= S\left(\frac{1{,}2}{6}\middle|-1{,}5 + \frac{1{,}44}{12}\right) \\ &= S(0{,}2|-1{,}5 + 0{,}12) = S(0{,}2|-1{,}38) \end{aligned}$$

   Scheitelform der Parabelgleichung:
   $y + 1{,}38 = -3(x - 0{,}2)^2$
   Berechnung des Schnittpunkts S_y mit der y-Achse:
   $S_y = S_y(0|-1{,}5)$
   Berechnung der Diskriminante D:
   $D = a_1^2 - 4a_2a_0 = (1{,}2)^2 - 4\cdot(-3)\cdot(-1{,}5) = -16{,}56$.
   Wegen $D < 0$ gibt es keinen Schnittpunkt mit der x-Achse. ◀

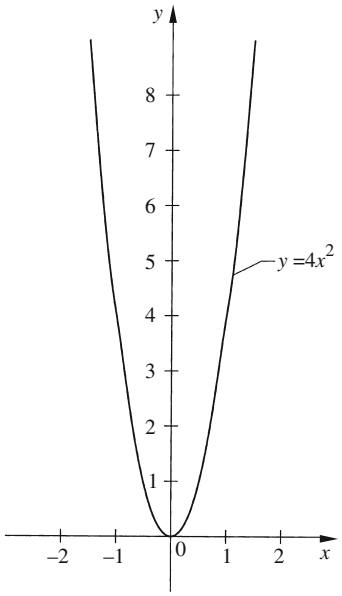


Abb. 5.20 Parabel

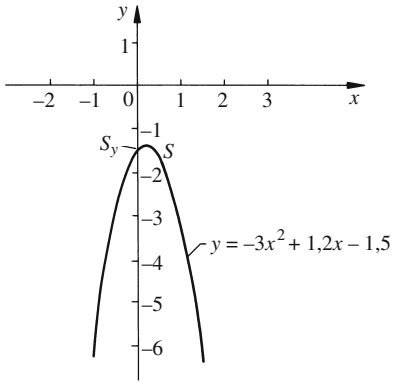


Abb. 5.21 Parabel

5.4.4 Kubische Funktionen

Funktionen mit einer Funktionsgleichung

$$y = f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$$
$$(a_3, a_2, a_1, a_0 \in \mathbb{R}, a_3 \neq 0)$$

Eine kubische Funktion ist eine ganze rationale Funktion 3. Grades.

Der Graph einer kubischen Funktion ist eine kubische Parabel (Abb. 5.22).

Das Verhalten der Funktion hängt wesentlich von dem Koeffizienten a_3 und der Diskriminante $D = 3a_3a_1 - a_2^2$ ab. Wenn $D \geq 0$ ist, dann ist die Funktion für $a_3 > 0$ monoton wachsend

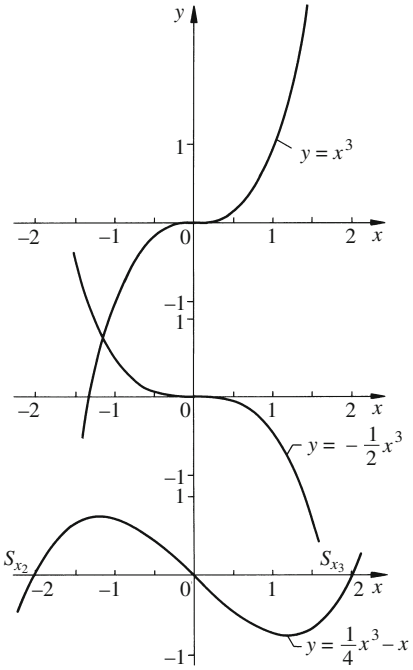


Abb. 5.22 Graphen der kubischen Funktionen $y = x^3$, $y = -\frac{1}{2}x^3$ und $y = \frac{1}{4}x^3 - x$

und für $a_3 < 0$ monoton fallend (vgl. Abschn. 5.2.1). Für $D < 0$ besitzt die Funktion ein Maximum und ein Minimum (siehe Abschn. 8.4.10). Für $a_3 > 0$ ist die Funktion dann von $-\infty$ bis zum Maximum monoton wachsend, monoton fallend vom Maximum bis zum Minimum und danach bis $+\infty$ wieder monoton wachsend. Für $a_3 < 0$ (und $D < 0$) ist die Funktion von $-\infty$ bis zum Minimum monoton fallend, vom Minimum bis zum Maximum monoton wachsend und danach bis $+\infty$ wieder monoton fallend.

Es gibt ein, zwei (dann ist ein Schnittpunkt ein Berührpunkt) oder drei Schnittpunkte mit der x-Achse (abhängig von den Koeffizienten a_3, a_2, a_1, a_0). Der Schnittpunkt mit der y-Achse ist $S_y(0|a_0)$.

Spezialfälle:

Kubische Normalparabel

Mit den Koeffizienten $a_3 = 1$, $a_2 = 0$, $a_1 = 0$, $a_0 = 0$ ergibt sich die kubische Normalparabel $y = x^3$. Sie schneidet sowohl die x- als auch die y-Achse im Ursprung (Abb. 5.22).

Kubische Normalparabel

$$y = x^3$$

Gespiegelte kubische Normalparabel

Mit den Koeffizienten $a_3 = -1$, $a_2 = 0$, $a_1 = 0$, $a_0 = 0$ erhält man die gespiegelte kubische Normalparabel $y = -x^3$.

Gespiegelte kubische Normalparabel

$$y = -x^3$$

Beispiele (Abb. 5.22)

1. $y = x^3$
2. $y = -\frac{1}{2}x^3$
3. $y = \frac{1}{4}x^3 - x$ ◀

5.4.5 Ganze rationale Funktionen n-ten Grades

Funktionen mit einer Funktionsgleichung folgender Art, wobei $a_0, a_1, a_2, \ldots, a_{n-1}, a_n \in \mathbb{R}$, $a_n \neq 0$, $n \in \mathbb{N}$.

$$y = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$$
$$= \sum_{k=0}^{n} a_k x^k$$

Die rechte Seite der Gleichung heißt auch Polynom n-ten Grades.

Der Graph einer ganzen rationalen Funktion n-ten Grades ist eine zusammenhängende Kurve, die von links aus dem Unendlichen kommt und nach rechts im Unendlichen verschwindet. Dabei hängt der Kurvenverlauf ganz wesentlich vom Grad n der Funktion und vom Vorzeichen von a_n ab. Es gilt:

n gerade ($n = 2, 4, 6, \ldots$) und $a_n > 0$:

$x \to -\infty \quad \Rightarrow \quad y \to +\infty$

$x \to +\infty \quad \Rightarrow \quad y \to +\infty$

n gerade ($n = 2, 4, 6, \ldots$) und $a_n < 0$:

$x \to -\infty \quad \Rightarrow \quad y \to -\infty$

$x \to +\infty \quad \Rightarrow \quad y \to -\infty$

n ungerade ($n = 1, 3, 5, \ldots$) und $a_n > 0$:

$x \to -\infty \quad \Rightarrow \quad y \to -\infty$

$x \to +\infty \quad \Rightarrow \quad y \to +\infty$

n ungerade ($n = 1, 3, 5, \ldots$) und $a_n < 0$:

$x \to -\infty \quad \Rightarrow \quad y \to +\infty$

$x \to +\infty \quad \Rightarrow \quad y \to -\infty$

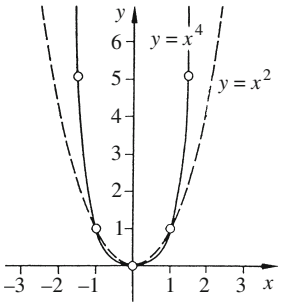


Abb. 5.23 Parabeln 2. und 4. Ordnung

Dabei bedeutet zum Beispiel $x \to -\infty$, dass x sich $-\infty$ nähert.

Ist von den Koeffizienten in der Funktionsgleichung nur $a_n \neq 0$, gilt also $a_0 = a_1 = a_2 = \ldots = a_{n-2} = a_{n-1} = 0$, dann nennt man die Funktion Potenzfunktion.

Potenzfunktionen

$$y = a_n x^n \quad (n \in \mathbb{N}^*, a_n \in \mathbb{R}, a_n \neq 0)$$

Die Graphen der Potenzfunktionen heißen für $n \geq 2$ Parabeln n-ter Ordnung.

Der Definitionsbereich der Potenzfunktionen ist $D = \mathbb{R}$. Für die Bildmenge gilt $f(D) = \{z \,|\, z \in \mathbb{R}, z \geq 0\}$ für gerade $n \geq 2$ und $a_n > 0$, $f(D) = \{z \,|\, z \in \mathbb{R}, z \leq 0\}$ für gerade $n \geq 2$ und $a_n < 0$ und $f(D) = \mathbb{R}$ für ungerade n.

Die Kurve der Funktion $y = ax^n$ ist im Vergleich zur Kurve der Funktion $y = x^n$ für $|a| < 1$ gestaucht, für $|a| > 1$ gestreckt und für $a < 0$ an der x-Achse gespiegelt.

Beispiele

1. $y = x^2$ und $y = x^4$ (Abb. 5.23)
   Die Graphen dieser Funktionen sind Parabeln 2. bzw. 4. Ordnung.
2. $y = x^3$ und $y = x^5$ (Abb. 5.24)
   Die Graphen dieser Funktionen sind Parabeln 3. bzw. 5. Ordnung.
3. $y = \frac{1}{100}x^6 + \frac{1}{100}x^5 - \frac{17}{100}x^4 - \frac{1}{20}x^3 + \frac{16}{25}x^2 + \frac{1}{25}x - \frac{12}{25}$ (Abb. 5.25)
   Das Polynom der rechten Seite lässt sich umformen:
   $\frac{1}{100}x^6 + \frac{1}{100}x^5 - \frac{17}{100}x^4 - \frac{1}{20}x^3 + \frac{16}{25}x^2 + \frac{1}{25}x - \frac{12}{25} = \frac{1}{100}(x^2-1)(x^2-4)(x^2+x-12)$
   Da ein Produkt genau dann gleich 0 ist, wenn mindestens einer der Faktoren gleich 0 ist, erhält man als Nullstellen der gegebenen Funktion die Lösungen der drei quadratischen Gleichungen $x^2-1=0$, $x^2-4=0$ und $x^2+x-12=0$:
   $x_1 = 1, x_2 = -1, x_3 = 2, x_4 = -2, x_5 = 3, x_6 = -4$
   Die Nullstellen sind die Abszissen der Schnittpunkte des Graphen der Funktion mit der x-Achse.

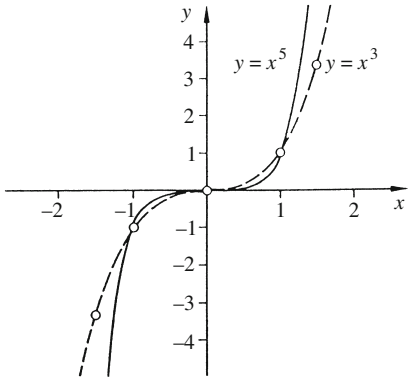


Abb. 5.24 Parabeln 3. und 5. Ordnung

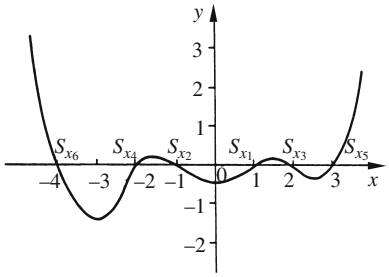


Abb. 5.25 Graph der Funktion zu der Gleichung aus Beispiel 3

Weil eine algebraische Gleichung n-ten Grades höchstens n reelle Wurzeln besitzt, hat die Kurve für den gegebenen Grad die Höchstzahl an Schnittpunkten mit der x-Achse, nämlich $n = 6$.
Da $a_n = \frac{1}{100} > 0$ und $n = 6$ geradzahlig ist, kommt die Kurve von links aus dem Positiv-Unendlichen und geht nach rechts ins Positiv-Unendliche.
Zur Berechnung des Schnittpunkts S_y mit der y-Achse setzt man in der Funktionsgleichung $x = 0$ ein und erhält $y = -\frac{12}{25}$ als Ordinate des Schnittpunkts und damit als Schnittpunkt mit der y-Achse:
$S_y = S_y(0|-\frac{12}{25})$ ◀

5.4.6 Horner-Schema

Das Horner-Schema ist ein Verfahren zur Berechnung von Funktionswerten ganzer rationaler Funktionen.

Ist eine Funktion $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 = \sum_{k=0}^{n} a_k x^k$ gegeben und der Funktionswert an der Stelle x_0 gesucht, so dividiert man das Polynom $\sum_{k=0}^{n} a_k x^k$ durch $(x - x_0)$:

$$(a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0) : (x - x_0)$$
$$= a_n x^{n-1} + c_1 x^{n-2} + \ldots + c_{n-2}x + c_{n-1} + \frac{c_n}{x - x_0}.$$

Für die Koeffizienten c_i gilt $c_1 = a_n x_0 + a_{n-1}$ und $c_i = c_{i-1}x_0 + a_{n-i}$ für $i = 2, 3, \ldots, n$. Damit kann die Funktion $f(x)$ auch durch die Gleichung $f(x) = (a_n x^{n-1} + c_1 x^{n-2} + \ldots + c_{n-2}x + c_{n-1})(x - x_0) + c_n$ beschrieben werden. Für $x = x_0$ ergibt sich dann $f(x_0) = c_n$. Die Berechnung des Funktionswertes $f(x_0)$ ist somit auf die Berechnung der Konstante c_n zurückgeführt worden, die man in n Schritten durch einander folgende Berechnung von $c_1, c_2, \ldots, c_n$ ermittelt. Man berechnet zuerst c_1 aus $c_1 = a_n x_0 + a_{n-1}$, dann c_2 aus $c_2 = c_1 x_0 + a_{n-2}$ und so weiter und schließlich c_n aus $c_n = c_{n-1}x_0 + a_0$.

Dieses Verfahren nennt man Horner-Schema (nach dem englischen Mathematiker William George Horner, 1786–1837). Es lässt sich folgendermaßen schematisch darstellen:

	a_n	a_{n-1}	a_{n-2}	$\cdots$	a_1	a_0
$+$		$a_n x_0$	$c_1 x_0$	$\cdots$	$c_{n-2}x_0$	$c_{n-1}x_0$
	a_n	c_1	c_2	$\cdots$	c_{n-1}	c_n

Beispiel

$f(x) = 2x^4 - 8x^3 + 2x^2 + 28x - 48$

Gesucht ist $f(-3)$, also der Funktionswert an der Stelle $x_0 = -3$.

Horner-Schema:

	2	-8	2	28	-48
$+$		-6	42	-132	312
		$(=2\cdot(-3))$	$(=(-14)\cdot(-3))$	$(=44\cdot(-3))$	$(=(-104)\cdot(-3))$
	2	-14	44	-104	264

Es gilt also $f(-3) = 264$. ◄

5.5 Gebrochene rationale Funktionen

5.5.1 Nullstellen, Pole, Asymptoten

Funktionen mit einer Funktionsgleichung folgender Art, wobei $a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_m \in \mathbb{R}$, $a_n, b_m \neq 0$, $m \neq 0$, heißen gebrochene rationale Funktionen.

$$y = \frac{a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \ldots + b_2 x^2 + b_1 x + b_0} = \frac{\sum_{i=0}^{n} a_i x^i}{\sum_{k=0}^{m} b_k x^k}$$

Eine gebrochene rationale Funktion $y = f(x)$ kann immer als Quotient zweier ganzer rationaler Funktionen dargestellt werden (sowohl Zähler als auch Nenner sind Polynome in x).

Gebrochene rationale Funktionen

$$y = \frac{P_n(x)}{P_m(x)}$$

Eine gebrochene rationale Funktion ist nicht für alle x definiert. Die Nullstellen des Nenners gehören nicht zum Definitionsbereich der Funktion.

Ist der Grad des Nennerpolynoms größer als der Grad des Zählerpolynoms ($n < m$), dann heißt die Funktion echt gebrochene rationale Funktion, andernfalls (also für $n \geq m$) heißt sie unecht gebrochene rationale Funktion.

Gebrochene rationale Funktionen, bei denen sowohl das Zählerpolynom als auch das Nennerpolynom den Grad 1 haben (also $n = 1$ und $m = 1$), heißen gebrochene lineare Funktionen.

Gebrochene lineare Funktionen

$$y = \frac{a_1 x + a_0}{b_1 x + b_0}$$

Die Graphen der gebrochenen rationalen Funktionen $y = \frac{a}{x^n}$, $n \in \mathbb{N}$, $n \geq 1$, $a \in \mathbb{R}$, $a \neq 0$ heißen Hyperbeln n-ter Ordnung (zu Hyperbeln vgl. auch Abschn. 7.5.2), (Abb. 5.28).

Durch Polynomdivision lässt sich jede unecht gebrochene rationale Funktion $y = f(x)$ darstellen als Summe einer ganzen rationalen Funktion $g(x)$ und einer echt gebrochenen rationalen Funktion $h(x)$: $y = f(x) = g(x) + h(x)$.

Beispiel

1. $\frac{2x^4+3x^3+5x^2-4x+1}{x^2-3x+1} = 2x^2 + 9x + 30 + \frac{77x-29}{x^2-3x+1}$ ◄

x_0 ist eine Nullstelle von $y = f(x) = \frac{P_n(x)}{P_m(x)} = \frac{P(x)}{Q(x)}$, wenn an der Stelle $x = x_0$ der Zähler Null ist und der Nenner von Null verschieden, also $P(x_0) = 0$, $Q(x_0) \neq 0$.

Eine Stelle $x = x_p$ heißt ein Pol der Funktion $y = \frac{P(x)}{Q(x)}$, wenn x_p eine Nullstelle des Nenners $Q(x)$ ist und der Zähler $P(x)$ an der Stelle x_p von Null verschieden ist, also $Q(x_p) = 0$, $P(x_p) \neq 0$. Ist $x = x_p$ eine k-fache Nullstelle des Nenners $Q(x)$ und gilt $P(x_p) \neq 0$, dann heißt x_p ein Pol k-ter Ordnung von $y = \frac{P(x)}{Q(x)}$.

Zwei Polynome $P(x)$ und $Q(x)$ heißen teilerfremd, wenn alle ihre Nullstellen verschieden sind. Gilt also für eine Stelle $x = x_1$, dass $P(x_1) = 0$, so folgt $Q(x_1) \neq 0$, und gilt umgekehrt für eine Stelle $x = x_2$, dass $Q(x_2) = 0$, so folgt $P(x_2) \neq 0$.

Jede gebrochene rationale Funktion lässt sich als Quotient zweier teilerfremder Polynome darstellen.

$$y = \frac{P(x)}{Q(x)}, \quad P(x) \text{ und } Q(x) \text{ teilerfremd}$$

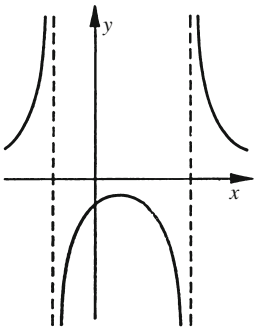


Abb. 5.26 Funktionsverlauf bei Polen ungerader Ordnung

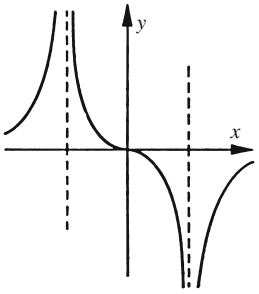


Abb. 5.27 Funktionsverlauf bei Polen gerader Ordnung

Eine solche Darstellung heißt Normalform der gebrochenen rationalen Funktion.

Die Nullstellen einer gebrochenen rationalen Funktion in Normalform sind die Nullstellen des Zählerpolynoms $P(x)$.

Ist $x = x_p$ ein Pol k-ter Ordnung der Funktion $y = \frac{P(x)}{Q(x)}$ mit teilerfremden $P(x)$ und $Q(x)$, dann lässt sich die Funktion in der Nähe des Pols darstellen durch

$$y = \frac{P(x)}{Q(x)} = \frac{1}{(x - x_p)^k} \cdot \frac{P(x)}{Q_1(x)}.$$

Dabei haben weder $P(x)$ noch $Q_1(x)$ in der Nähe von $x = x_p$ eine Nullstelle, sie ändern also ihr Vorzeichen nicht. Ihr Quotient hat deshalb einen von Null verschiedenen, beschränkten positiven oder negativen Wert. Die Funktion $\frac{1}{(x-x_p)^k}$ wächst aber, wenn sich x dem Pol x_p nähert, über alle Grenzen.

Nähert man sich dem Pol mit wachsenden x-Werten (also $x < x_p$), so ist $x - x_p$ negativ. Für ungerade k ($k = 1, 3, 5, \ldots$) geht dann $\frac{1}{(x-x_p)^k}$ gegen -∞, für gerade k ($k = 2, 4, 6, \ldots$) dagegen gegen $+\infty$ (Abb. 5.26 und 5.27).

Nähert man sich dem Pol mit abnehmenden x-Werten (also $x > x_p$), so ist $x - x_p$ positiv, $\frac{1}{(x-x_p)^k}$ geht dann also stets gegen $+\infty$.

Für negative Werte des Faktors $\frac{P(x)}{Q_1(x)}$ dreht sich das Vorzeichen der Funktion $y = f(x)$ um.

Die Gerade $x = x_p$ heißt Asymptote der gebrochenen rationalen Funktion $y = f(x)$. Asymptoten einer Funktion sind Geraden,

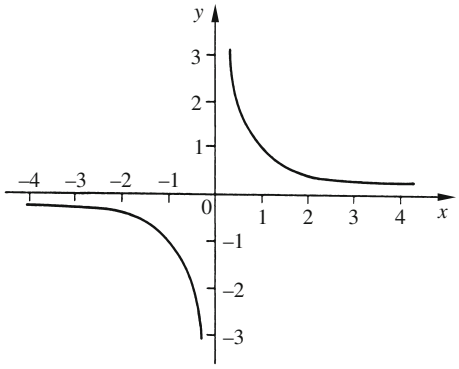


Abb. 5.28 Graph der Funktion mit der Gleichung $y = \frac{1}{x}$

denen sich der Graph der Funktion unbeschränkt nähert, ohne sie je zu erreichen (Asymptote = Nichtzusammenlaufende).

Das Verhalten einer gebrochenen rationalen Funktion $y = f(x) = \frac{P_n(x)}{P_m(x)}$ im Unendlichen:

Ist $y = f(x)$ eine echt gebrochene rationale Funktion, gilt also $n < m$, dann ist die x-Achse (Gerade mit der Gleichung $y = 0$) eine Asymptote.

Im Falle $n = m$ ist die zur x-Achse parallele Gerade mit der Gleichung $y = \frac{a_n}{b_m}$ eine Asymptote.

Ist $n > m$, so gilt $y = f(x) = g(x) + h(x)$, wobei $g(x)$ eine ganze rationale Funktion und $h(x)$ eine echt gebrochene rationale Funktion sind. Die Funktion $y = f(x)$ verhält sich dann im Unendlichen wie die rationale Funktion $y = g(x)$.

Beispiele

2. $y = \frac{1}{x}$ (Abb. 5.28)
   Zum Definitionsbereich gehören alle x außer $x = 0$.
   Wegen $f(-x) = \frac{1}{-x} = -\frac{1}{x} = -f(x)$ ist die Funktion ungerade, der Graph der Funktion ist also symmetrisch zum Nullpunkt (Koordinatenursprung).
   Die Funktion hat keine Nullstelle, denn der Zähler ist stets von Null verschieden ($P(x) = 1$).
   Die Stelle $x = 0$ ist ein Pol erster Ordnung der Funktion. Nähert man sich diesem Pol mit wachsenden x-Werten (also $x < 0$), dann geht y gegen $-\infty$. Nähert man sich dem Pol dagegen mit abnehmenden x-Werten (also $x > 0$), so geht y gegen $+\infty$.
   Die Geraden $x = 0$ (y-Achse) und $y = 0$ (x-Achse) sind Asymptoten der Funktion.
   Der Graph der Funktion ist eine Hyperbel.
   Wertetabelle (y-Werte auf zwei Stellen nach dem Komma gerundet):

x	-4	-3	-2	-1	$-0{,}5$	$-0{,}25$
y	$-0{,}25$	$-0{,}33$	$-0{,}50$	-1	-2	-4

x	$-0{,}1$	$-0{,}01$	1	2	3	4
y	-10	-100	1	0,50	0,33	0,25

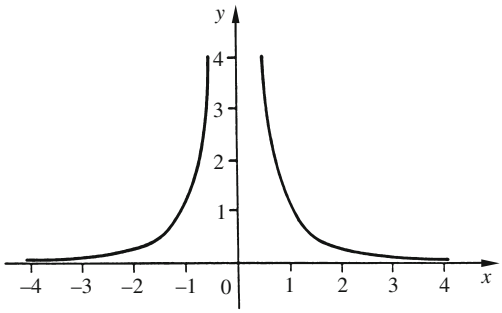


Abb. 5.29 Graph der Funktion mit der Gleichung $y = \frac{1}{x^2}$

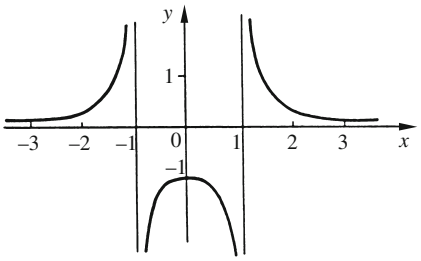


Abb. 5.30 Graph der Funktion mit der Gleichung $y = \frac{1}{x^2-1}$

3. $y = \frac{1}{x^2}$ (Abb. 5.29)
Die Funktion ist definiert für alle $x \neq 0$.
Es gilt $f(-x) = \frac{1}{(-x)^2} = \frac{1}{x^2} = f(x)$, die Funktion ist also gerade, ihr Graph ist symmetrisch zur y-Achse.
Da der Zähler konstant gleich 1 ist, besitzt die Funktion keine Nullstellen.
Die Stelle $x = 0$ ist ein Pol zweiter Ordnung. Nähert man sich dem Pol mit wachsenden x-Werten (also $x < 0$) oder mit abnehmenden x-Werten (also $x > 0$), dann geht y gegen $+\infty$. Die Geraden $x = 0$ (y-Achse) und $y = 0$ (x-Achse) sind Asymptoten der Funktion.
Wertetabelle:

x	4	3	2	1	0,5	0,25	0,1
y	0,06	0,11	0,25	1	4	16	100

x	-1	-2	-3	-4
y	1	0,25	0,11	0,06

Die y-Werte in der Wertetabelle sind auf zwei Stellen nach dem Komma gerundet.

4. $y = \frac{1}{x^2-1}$ (Abb. 5.30)
Die Funktion ist definiert für alle x, für die der Nenner ungleich 0 ist.
Die Nullstellen des Nenners berechnet man, indem man den Nenner (das Nennerpolynom) gleich Null setzt: $x^2 - 1 = 0$. Diese quadratische Gleichung hat die Lösungen $x_1 = 1$ und $x_2 = -1$.
Wegen $f(-x) = \frac{1}{(-x)^2-1} = \frac{1}{x^2-1} = f(x)$ ist die Funktion gerade, der Graph der Funktion ist also symmetrisch zur y-Achse.

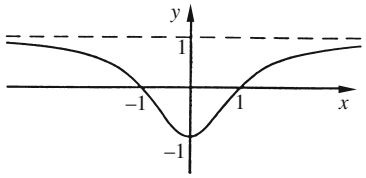


Abb. 5.31 Graph der Funktion mit der Gleichung $y = \frac{x^2-1}{x^2+1}$

Die Funktion hat keine Nullstellen (Schnittpunkte mit der x-Achse), denn der Zähler ist für alle x des Definitionsbereiches von Null verschieden.
Die Stellen $x_1 = 1$ und $x_2 = -1$ sind Pole erster Ordnung der Funktion.
Nähert man sich dem Pol x_2 mit wachsenden x-Werten (also $x < -1$), dann ist der Faktor $\frac{P(x)}{Q_1(x)} = \frac{1}{x-1}$ in der Zerlegung der Funktion
$y = \frac{P(x)}{Q(x)} = \frac{1}{x-x_2} \cdot \frac{P(x)}{Q_1(x)} = \frac{1}{x-(-1)} \cdot \frac{1}{x-1}$
negativ, das heißt, y geht gegen $+\infty$.
Nähert man sich entsprechend dem Pol x_2 mit abnehmenden x-Werten (also $x > -1$) oder dem Pol x_1 mit wachsenden x-Werten (also $x < 1$), so geht y gegen $-\infty$. Nähert man sich dagegen x_1 mit abnehmenden x-Werten (also $x > 1$), so geht y gegen $+\infty$.
Die Geraden $x_1 = 1$ und $x_2 = -1$ sowie $y = 0$ (x-Achse) sind Asymptoten der Funktion.
Funktionswerte für $-1 < y \leq 0$ gibt es nicht, da der Nenner nicht kleiner als -1 werden kann.
Wertetabelle (y-Werte auf drei Stellen nach dem Komma gerundet):

x	± 3	± 2	$\pm 1{,}5$	$\pm 1{,}1$	$\pm 0{,}9$	$\pm 0{,}5$	0
y	0,125	0,333	0,800	4,762	$-5{,}263$	$-1{,}333$	-1

5. $y = \frac{x^2-1}{x^2+1}$ (Abb. 5.31)
Die Funktion ist definiert für alle reellen x, denn es gibt keine Nullstellen des Nenners und damit keine Pole.
Die Nullstellen der Funktion sind $x_1 = 1$ und $x_2 = -1$.
Die zur x-Achse parallele Gerade mit der Gleichung $y = \frac{a_n}{b_m} = \frac{1}{1} = 1$ ist Asymptote.
Der Graph der Funktion verläuft überall unterhalb der Asymptote.
Wertetabelle:

x	± 4	± 3	± 2	$\pm 1{,}5$	± 1
y	0,87	0,80	0,60	0,38	0

x	$\pm 0{,}5$	$\pm 0{,}3$	$\pm 0{,}1$	0
y	$-0{,}60$	$-0{,}84$	$-0{,}98$	-1

Die y-Werte in der Wertetabelle sind auf zwei Stellen nach dem Komma gerundet.

6. $y = \frac{x^2-x-2}{2x-6}$ (Abb. 5.32)
Die Funktion hat die Nullstellen $x_1 = -1$ und $x_2 = 2$ und einen Pol bei $x = 3$.
Die Funktion ist also definiert für alle $x \neq 3$.

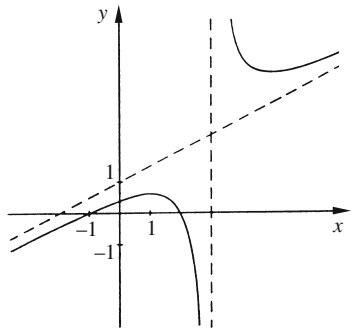


Abb. 5.32 Graph der Funktion mit der Gleichung $y = \frac{x^2-x-2}{2x-6}$

Die Funktion ist weder gerade noch ungerade, eine Symmetrie des Graphen bezüglich des Nullpunktes oder der y-Achse liegt also nicht vor.
Durch Polynomdivision erhält man die Darstellung $y = \frac{x^2-x-2}{2x-6} = \frac{1}{2}x + 1 + \frac{4}{2x-6}$ der Funktion als Summe einer ganzen rationalen Funktion und einer echt gebrochenen rationalen Funktion. Es ist also $y = \frac{1}{2}x + 1$ eine Asymptote der Funktion. Die Annäherung an die Asymptote erfolgt für $x \to -\infty$ von unten und für $x \to \infty$ von oben.
Wertetabelle (y-Werte auf zwei Stellen nach dem Komma gerundet):

x	−5	−3	−2	−1	0	1	1,5
y	−1,75	−0,83	−0,40	0	0,33	0,50	0,42

x	2	2,5	2,9	3,1	3,5	4	5	7
y	0	−1,75	−17,55	22,55	6,75	5	4,50	5

◀

5.5.2 Partialbruchzerlegung

Eine Partialbruchzerlegung ist die Zerlegung einer gebrochenen rationalen Funktion $y = f(x)$ mit

$$f(x) = \frac{a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \ldots + b_2x^2 + b_1x + b_0}$$

in eine Summe von Brüchen.

Durch eine Partialbruchzerlegung von $f(x)$ wird oftmals die Integration der Funktion einfacher oder überhaupt erst möglich (vgl. Abschn. 8.5.2).

Jede echt gebrochene rationale Funktion (also $n < m$) kann eindeutig in eine Summe von Partialbrüchen zerlegt werden.

Praktische Durchführung der Partialbruchzerlegung:

1. Im Falle $n \geq m$ Abspalten des ganzen rationalen Anteils mit Polynomdivision.
2. Kürzen des Bruches (also Division des Zählers und des Nenners) durch b_m, den Koeffizienten der höchsten Potenz des Nenners:
$$f(x) = \frac{c_n x^n + c_{n-1}x^{n-1} + \ldots + c_2x^2 + c_1x + c_0}{x^m + d_{m-1}x^{m-1} + \ldots + d_2x^2 + d_1x + d_0}$$
Es gilt also $\frac{a_i}{b_m} = c_i$ $(1 \leq i \leq n)$ und $\frac{b_j}{b_m} = d_j$ $(1 \leq j < m)$.
3. Bestimmung der Nullstellen $x_1, x_2, \ldots, x_r$ $(r \leq m)$ des Nennerpolynoms.
4. Zerlegung des Nennerpolynoms in die Form
$$\begin{aligned} x^m + d_{m-1}x^{m-1} + \ldots + d_2x^2 + d_1x + d_0 &= (x-x_1)^{k_1} \cdot (x-x_2)^{k_2} \cdot \ldots \cdot (x-x_r)^{k_r} \\ &\quad \cdot (x^2 + p_1x + q_1)^{l_1} (x^2 + p_2x + q_2)^{l_2} \\ &\quad \cdot \ldots \cdot (x^2 + p_sx + q_s)^{l_s} \end{aligned}$$
Eine solche Zerlegung ist immer möglich. Dabei sind $x_1, x_2, \ldots, x_r$ alle reellen Nullstellen mit den Vielfachheiten $k_1, k_2, \ldots, k_r$. Die restlichen quadratischen Faktoren ergeben die konjugierten Paare komplexer Nullstellen (also $p_i^2 - 4q_i < 0$).
5. Zerlegung von $f(x)$ in eine Summe von Brüchen:
$$\begin{aligned} f(x) = {} & \frac{A_{11}}{x-x_1} + \frac{A_{12}}{(x-x_1)^2} + \cdots + \frac{A_{1k_1}}{(x-x_1)^{k_1}} \\ & + \frac{A_{21}}{x-x_2} + \frac{A_{22}}{(x-x_2)^2} + \cdots + \frac{A_{2k_2}}{(x-x_2)^{k_2}} \\ & \vdots \\ & + \frac{A_{r1}}{x-x_r} + \frac{A_{r2}}{(x-x_r)^2} + \cdots + \frac{A_{rk_r}}{(x-x_r)^{k_r}} \\ & + \frac{B_{11} + C_{11}x}{x^2 + p_1x + q_1} + \frac{B_{12} + C_{12}x}{(x^2 + p_1x + q_1)^2} \\ & + \cdots + \frac{B_{1l_1} + C_{1l_1}x}{(x^2 + p_1x + q_1)^{l_1}} + \frac{B_{21} + C_{21}x}{x^2 + p_2x + q_2} \\ & + \frac{B_{22} + C_{22}x}{(x^2 + p_2x + q_2)^2} + \cdots + \frac{B_{2l_2} + C_{2l_2}x}{(x^2 + p_2x + q_2)^{l_2}} \\ & \vdots \\ & + \frac{B_{s1} + C_{s1}x}{x^2 + p_sx + q_s} + \frac{B_{s2} + C_{s2}x}{(x^2 + p_sx + q_s)^2} \\ & + \cdots + \frac{B_{sl_s} + C_{sl_s}x}{(x^2 + p_sx + q_s)^{l_s}} \end{aligned}$$
Dabei sind die Koeffizienten A_{ij}, B_{ij}, C_{ij} reelle Zahlen.
6. Bestimmung der Koeffizienten der Partialbrüche zum Beispiel mit der Methode des Koeffizientenvergleichs.

Die Brüche im Schritt 5 nennt man die Partialbrüche der gebrochenen rationalen Funktion $f(x)$.

Spezialfälle:

Wenn das Nennerpolynom nur reelle Nullstellen besitzt, dann fallen die Partialbrüche mit den nicht zerlegbaren quadratischen Funktionen im Nenner weg.

Besitzt das Nennerpolynom nur die einfachen reellen Nullstellen $x_1, x_2, \ldots, x_m$, dann lautet die Partialbruchzerlegung

$$f(x) = \frac{A_1}{x-x_1} + \frac{A_2}{x-x_2} + \cdots + \frac{A_m}{x-x_m}$$

Beispiele

1. $f(x)=\frac{6x^2-4}{2x^3+4x^2+4x+2}$
Division durch $b_3=2$: $f(x)=\frac{3x^2-2}{x^3+2x^2+2x+1}$
Nullstelle des Nennerpolynoms: $x_1=-1$
Zerlegung des Nennerpolynoms:
$x^3+2x^2+2x+1=(x+1)(x^2+x+1)$
Zerlegung von $f(x)$ in eine Summe von Partialbrüchen:

$$f(x)=\frac{3x^2-2}{x^3+2x^2+2x+1}=\frac{A}{x+1}+\frac{Bx+C}{x^2+x+1}$$

Bestimmung der Koeffizienten A, B, C durch Koffizientenvergleich:

$$\begin{aligned}f(x)&=\frac{3x^2-2}{(x+1)(x^2+x+1)}\\&=\frac{A(x^2+x+1)+(Bx+C)(x+1)}{(x+1)(x^2+x+1)}\\\Rightarrow\ 3x^2-2&=A(x^2+x+1)+(Bx+C)(x+1)\\&=(A+B)x^2+(A+B+C)x+(A+C)\end{aligned}$$

Vergleich der Koeffizienten von x^2, von x und der Absolutglieder links und rechts vom Gleichheitszeichen ergibt:
$A+B=3, A+B+C=0, A+C=-2\Rightarrow A=1$, $B=2, C=-3$
Lösung somit:

$$f(x)=\frac{6x^2-4}{2x^3+4x^2+4x+2}=\frac{1}{x+1}+\frac{2x-3}{x^2+x+1}$$

2. $f(x)=\frac{3x^2-x+1}{x^3-2x^2+x}$
Nullstellen des Nennerpolynoms: $x_1=0$, $x_2=1$ (x_2 ist doppelte Nullstelle)
Zerlegung des Nennerpolynoms:
$x^3-2x^2+x=x(x-1)^2$
Zerlegung von $f(x)$ in Partialbrüche:

$$f(x)=\frac{3x^2-x+1}{x^3-2x^2+x}=\frac{A}{x}+\frac{B}{x-1}+\frac{C}{(x-1)^2}$$

Bestimmung der Koeffizienten A, B, C durch Koeffizientenvergleich:

$$\begin{aligned}f(x)&=\frac{3x^2-x+1}{x(x-1)^2}\\&=\frac{A(x-1)^2+Bx(x-1)+Cx}{x(x-1)^2}\\\Rightarrow\ 3x^2-x+1&=A(x^2-2x+1)+B(x^2-x)+Cx\\&=(A+B)x^2+(C-2A-B)x+A\end{aligned}$$

Vergleich der Koeffizienten von x^2, von x und der Absolutglieder links und rechts vom Gleichheitszeichen ergibt:
$A+B=3, C-2A-B=-1, A=1\Rightarrow A=1$, $B=2, C=3$
Lösung somit:

$$f(x)=\frac{3x^2-x+1}{x^3-2x^2+x}=\frac{1}{x}+\frac{2}{x-1}+\frac{3}{(x-1)^2}$$

3. $f(x)=\frac{6x^2-x+1}{x^3-x}$
Nullstellen des Nennerpolynoms: $x_1=0$, $x_2=1$, $x_3=-1$
Zerlegung des Nennerpolynoms:
$x^3-x=x(x-1)(x+1)$
Zerlegung von $f(x)$ in Partialbrüche:

$$f(x)=\frac{6x^2-x+1}{x^3-x}=\frac{A}{x}+\frac{B}{x-1}+\frac{C}{x+1}$$

Bestimmung der Koeffizienten A, B, C durch Koeffizientenvergleich:

$$\begin{aligned}f(x)&=\frac{6x^2-x+1}{x(x-1)(x+1)}\\&=\frac{A(x-1)(x+1)+Bx(x+1)+Cx(x-1)}{x(x-1)(x+1)}\\\Rightarrow\ 6x^2-x+1&=A(x^2-1)+B(x^2+x)+C(x^2x)\\&=(A+B+C)x^2+(B-C)x-A\end{aligned}$$

Vergleich der Koeffizienten von x^2, von x und der Absolutglieder links und rechts vom Gleichheitszeichen ergibt:
$A+B+C=6, B-C=-1, -A=1\Rightarrow A=-1$, $B=3, C=4$
Lösung somit:

$$f(x)=\frac{6x^2-x+1}{x^3-x}=\frac{1}{x}+\frac{3}{x-1}+\frac{4}{x+1}$$ ◀

5.6 Irrationale Funktionen

Irrationale Funktionen sind algebraische Funktionen, die nicht rational sind. In der Funktionsgleichung $y=f(x)$ einer rationalen Funktion werden auf die unabhängige Variable x nur endlich viele rationale Rechenoperationen (Addition, Subtraktion, Multiplikation und Division) angewandt. Bei irrationalen Funktionen tritt die unabhängige Variable x auch unter einem Wurzelzeichen auf.

Beispiele

1. $y=x^2+x+\sqrt{x}$
2. $y=\sqrt{5x^3-2}$
3. $y=\sqrt[7]{(x^2-1)\sqrt[3]{5x+1}}$ ◀

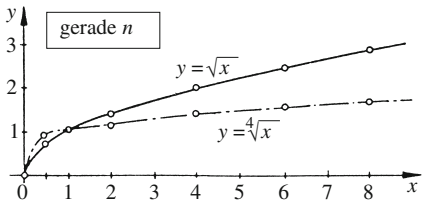


Abb. 5.33 Graph der Wurzelfunktionen $y = \sqrt{x}$ und $y = \sqrt[4]{x}$

Eine besonders wichtige Klasse von irrationalen Funktionen sind die sogenannten Wurzelfunktionen.

Wurzelfunktionen

$$y = \sqrt[n]{x} \quad (n \in N, n \geq 2)$$

Der Definitionsbereich der Wurzelfunktionen ist $D = \{x \,|\, x \in \mathbb{R}, x \geq 0\}$ für gerade n und $D = \mathbb{R}$ für ungerade n, die Bildmenge ist gleich dem Definitionsbereich, also $f(D) = D$.

Die Wurzelfunktionen sind im ganzen Definitionsbereich streng monoton wachsend.

Für ungerade n ist $y = \sqrt[n]{x}$ eine ungerade Funktion, der Graph der Funktion ist also punktsymmetrisch zum Koordinatenursprung (Abb. 5.34).

Die Graphen der Wurzelfunktionen gehen durch den Koordinatenursprung und durch den Punkt $P(1|1)$.

Für das Verhalten der Wurzelfunktionen im Unendlichen gilt (Abb. 5.33 und 5.34):

$n \in \mathbb{N}, n \geq 2$: $\quad x \to +\infty \;\Rightarrow\; y \to +\infty$

n ungerade $(n = 3, 5, 7, \ldots)$: $\quad x \to -\infty \;\Rightarrow\; y \to -\infty$

Die quadratische Funktion $y = x^2$ ist in den zwei getrennten Intervallen $0 \leq x < +\infty$ und $-\infty < x \leq 0$ jeweils monoton. Sie hat deshalb zwei Umkehrfunktionen, und zwar $y = +\sqrt{x}$ und $y = -\sqrt{x}$. Für beide Umkehrfunktionen ist der Definitionsbereich $0 \leq x < +\infty$ (entspricht $0 \leq y < +\infty$ der Funktion $y = x^2$), die Bildmenge ist $0 \leq y < +\infty$ bzw. $-\infty < y \leq 0$. Die Graphen der Umkehrfunktionen ergeben sich aus der Normalparabel durch Spiegelung an der Winkelhalbierenden $y = x$. Die (positive) Quadratwurzelfunktion $y = \sqrt{x}$ zum Beispiel ist also die Umkehrfunktion der Funktion des rechten Normalparabelastes (Abb. 5.35).

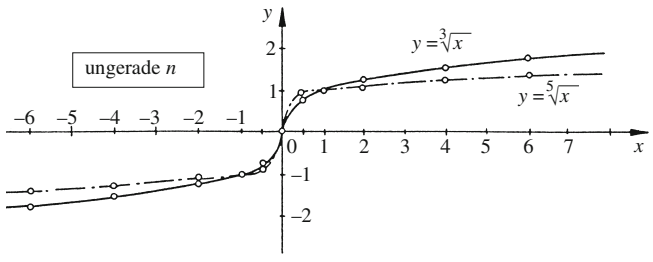


Abb. 5.34 Graph der Wurzelfunktionen $y = \sqrt[3]{x}$ und $y = \sqrt[5]{x}$

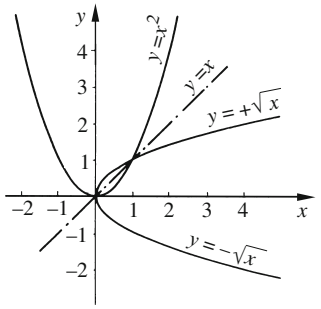


Abb. 5.35 Graphen von Funktionen und ihren Umkehrfunktionen

Die kubische Funktion $y = x^3$ ist in ihrem ganzen Definitionsbereich $D = (-\infty, \infty)$ monoton steigend. Ihre Umkehrfunktion ist $y = \sqrt[3]{x}$. Der Definitionsbereich der Umkehrfunktion ist $-\infty < x < \infty$, die Bildmenge $-\infty < y < \infty$. Der Graph der Umkehrfunktion ergibt sich aus der kubischen Normalparabel durch Spiegelung an der Winkelhalbierenden $y = x$.

Allgemein gilt:

Für ungerade n ist die Wurzelfunktion $y = f(x) = \sqrt[n]{x}$, $f : \mathbb{R} \to \mathbb{R}$ die Umkehrfunktion der Potenzfunktion $y = f(x) = x^n$, $f : \mathbb{R} \to \mathbb{R}$.

Für gerade n ist die Wurzelfunktion $y = f(x) = \sqrt[n]{x}$, $f : [0, \infty) \to [0, \infty)$ die Umkehrfunktion der Potenzfunktion $y = f(x) = x^n$, $f : [0, \infty) \to [0, \infty)$.

Man bezeichnet allgemeiner auch Funktionen $y = a\sqrt[n]{x}$, $a \in \mathbb{R}$, $a \neq 0$ als Wurzelfunktionen. Die Kurve der Funktion $y = a\sqrt[n]{x}$ ist im Vergleich zur Kurve der Funktion $y = \sqrt[n]{x}$ für $|a| < 1$ gestaucht, für $|a| > 1$ gestreckt und für $a < 0$ an der x-Achse gespiegelt.

Beispiele

4. $y = b + \sqrt{r^2 - (x-a)^2}$, $D = \{x \,|\, |x-a| \leq r\}$, $W = \mathbb{R}$
Der Graph dieser Funktion ist der obere Halbkreis des Kreises mit dem Mittelpunkt $M(a|b)$ und dem Radius r (Abb. 5.36).
Fehlerwarnung: Die Gleichung $(x-a)^2 + (y-b)^2 = r^2$ des Kreises mit dem Mittelpunkt $M(a|b)$ und dem Radius r (vgl. Abschn. 7.3.1) ist keine (implizite) Funktion, denn die Zuordnung einer Zahl y zu einer Zahl x ist nicht eindeutig, wie in der Definition einer Funktion gefordert (zu jedem x mit $|x - a| < r$ gibt es zwei y)!
Analog zu oben ist der Graph der Funktion $y = b - \sqrt{r^2 - (x-a)^2}$, $D = \{x \,|\, |x-a| \leq r\}$, $W = \mathbb{R}$ die untere Hälfte des Kreises mit dem Mittelpunkt $M(a|b)$ und dem Radius r.

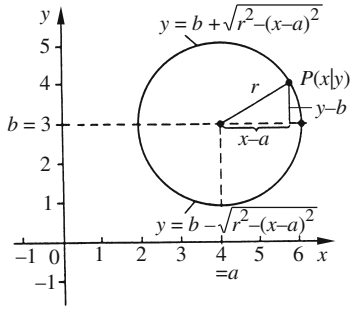


Abb. 5.36 Graphen der Funktionen $y = b + \sqrt{r^2 - (x-a)^2}$ und $y = b - \sqrt{r^2 - (x-a)^2}$ mit $a = 4$ und $b = 3$

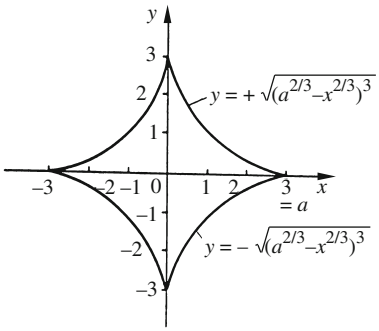


Abb. 5.37 Graphen der Funktionen $y = +\sqrt{(a^{\frac{2}{3}} - x^{\frac{2}{3}})^3}$ und $y = -\sqrt{(a^{\frac{2}{3}} - x^{\frac{2}{3}})^3}$ mit $a = 3$

5. Aus der Gleichung der Astroide $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$ $(a > 0)$ erhält man durch Auflösen nach y die Funktionen
$y = +\sqrt{(a^{\frac{2}{3}} - x^{\frac{2}{3}})^3}$, $D = \{x \,|\, |x| \le a\}$, $W = \mathbb{R}$
bzw.
$y = -\sqrt{(a^{\frac{2}{3}} - x^{\frac{2}{3}})^3}$, $D = \{x \,|\, |x| \le a\}$, $W = \mathbb{R}$.
Die Graphen dieser Funktionen sind der obere Teil $(y \ge 0)$ bzw. der untere Teil $(y \le 0)$ der Astroide (Abb. 5.37). ◀

5.7 Transzendente Funktionen

Elementare Funktionen, die nicht algebraisch sind, heißen transzendent.

Wichtige Klassen von transzendenten Funktionen sind die Exponentialfunktionen, die Logarithmusfunktionen sowie die trigonometrischen Funktionen und ihre Umkehrfunktionen, die Arkusfunktionen.

Die trigonometrischen Funktionen und die Arkusfunktionen werden in Kap. 6 behandelt.

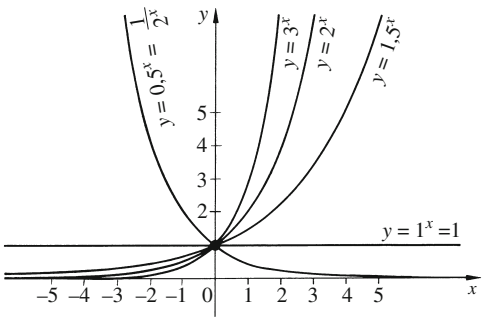


Abb. 5.38 Graphen von Exponentialfunktionen

5.7.1 Exponentialfunktionen

Bei einer Exponentialfunktion steht die unabhängige Variable x im Exponenten.

$$y = a^x, \quad a \in \mathbb{R}^+$$

Dabei ist die Basis a eine beliebige positive reelle Zahl.

Alle Exponentialfunktionen $y = a^x$, $a \in \mathbb{R}^+$ haben als Definitionsbereich $D = \mathbb{R}$ und, falls $a \neq 1$, als Bildmenge $W = f(D) = \mathbb{R}^+$. Alle Funktionswerte sind also positiv.

Wegen $a^0 = 1$ gehen die Graphen aller Funktionen durch den Punkt $P(0|1)$, (Abb. 5.38).

Für $a > 1$ ist die Funktion $y = a^x$ streng monoton wachsend mit $y \to 0$ für $x \to -\infty$ und $y \to \infty$ für $x \to \infty$. Die (negative) x-Achse ist also Asymptote.

Für $0 < a < 1$ ist die Funktion $y = a^x$ streng monoton fallend mit $y \to \infty$ für $x \to -\infty$ und $y \to 0$ für $x \to \infty$. Die (positive) x-Achse ist somit Asymptote.

Der Graph der Funktion nähert sich um so schneller der x-Achse, je größer $|\ln a|$ ist, für $a > 1$ also je größer a ist und für $a < 1$ je kleiner a ist.

Für $a = 1$ gilt $y = 1$, der Graph der Funktion ist also eine Parallele zur x-Achse.

Die Exponentialfunktionen $y = a^x$, $a > 0$ können wegen der Regeln der Logarithmen- und der Potenzrechnung auch in der Form

$$y = a^x = e^{\ln(a^x)} = e^{x \cdot \ln a}$$

dargestellt werden. Dabei ist $e = 2{,}7182818284\ldots$ die Euler'sche Zahl (vgl. Abschn. 8.4.5).

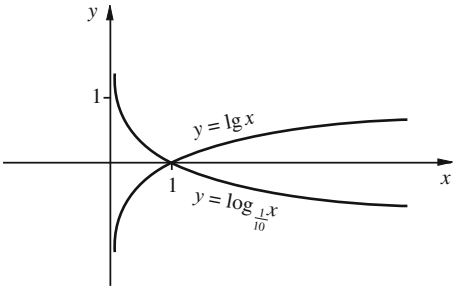


Abb. 5.39 Graphen der logarithmischen Funktionen $y = \lg x$ und $y = \log \frac{1}{10} x$

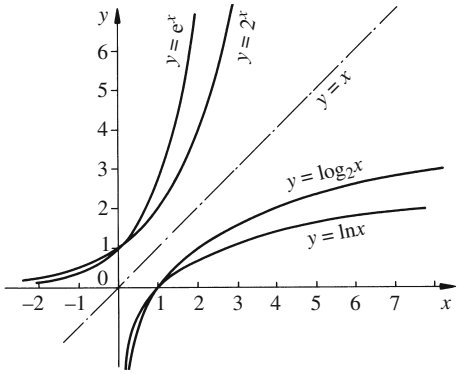


Abb. 5.40 Graphen der logarithmischen Funktionen $y = \ln x$ und $y = \log_2 x$ und ihrer Umkehrfunktionen $y = e^x$ und $y = 2^x$

Die Funktion $y = e^x$, $D = \mathbb{R}$, $W = f(D) = \mathbb{R}^+$, also die Exponentialfunktion mit der Basis $a = e$, heißt natürliche Exponentialfunktion oder e-Funktion.

$$y = e^x, \quad D = \mathbb{R}, W = f(D) = \mathbb{R}^+$$

Es handelt sich um eine spezielle Exponentialfunktion, die häufig als *die* Exponentialfunktion bezeichnet wird. Diese Funktion spielt bei vielen Wachstumsprozessen eine wichtige Rolle.

Noch allgemeiner bezeichnet man manchmal auch solche Funktionen, die eine algebraische Funktion des Arguments x im Exponenten haben, als Exponentialfunktionen, zum Beispiel $y = 2^{3x^2-7x}$.

Die Umkehrfunktionen der Exponentialfunktionen $y = a^x$ sind für $a \neq 1$ die Logarithmusfunktionen $y = \log_a x$. Die Umkehrfunktion der e-Funktion ist die natürliche Logarithmusfunktion $y = \ln x$.

5.7.2 Logarithmusfunktionen

Logarithmusfunktionen sind Funktionen der Form

$$y = \log_a x, \quad a \in \mathbb{R}^+, a \neq 1$$

Alle Logarithmusfunktionen $y = \log_a x$, $a \in \mathbb{R}^+$, $a \neq 1$ haben als Definitionsbereich $D = \mathbb{R}^+$ und als Bildmenge $W = f(D) = \mathbb{R}$.

Wegen $\log_a 1 = 0$ gehen die Graphen aller Funktionen durch den Punkt $P(1|0)$ (Abb. 5.39).

Für $a > 1$ ist die Funktion $y = \log_a x$ streng monoton wachsend mit $y \to \infty$ für $x \to \infty$ und $y \to -\infty$ für $x \to 0, x > 0$. Die (negative) y-Achse ist also Asymptote. Für $x > 1$ gilt $\log_a x > 0$, für $x = 1$ gilt $\log_a 1 = 0$, und für x mit $0 < x < 1$ gilt $\log_a x < 0$.

Für $0 < a < 1$ ist die Funktion $y = \log_a x$ streng monoton fallend mit $y \to -\infty$ für $x \to \infty$ und $y \to \infty$ für $x \to 0, x > 0$. Die (positive) y-Achse ist somit Asymptote.

Für $x > 1$ gilt $\log_a x < 0$, für $x = 1$ gilt $\log_a 1 = 0$, und für x mit $0 < x < 1$ gilt $\log_a x > 0$.

Der Graph der Funktion nähert sich für alle a um so schneller der y-Achse, je größer $|\ln a|$ ist, für $a > 1$ also je größer a ist und für $a < 1$ je kleiner a ist.

Die Logarithmusfunktionen $y = \log_a x$, $a > 0$, $a \neq 1$ können wegen der Regeln der Logarithmenrechnung auch in der Form

$$y = \log_a x = \frac{1}{\ln a} \cdot \ln x, \quad a \neq 1$$

dargestellt werden.

Dabei heißt die Logarithmusfunktion mit der Basis $a = e = 2{,}7182\ldots$ natürliche Logarithmusfunktion.

$$y = \ln x, \quad D = \mathbb{R}^+, W = f(D) = \mathbb{R}$$

Allgemeiner noch bezeichnet man auch solche Funktionen, die eine algebraische Funktion des Arguments x als Numerus haben, als Logarithmusfunktion, zum Beispiel $y = \log_2(5x^2 - 4x)x$.

Die Logarithmusfunktion $y = \log_a x$ ist für $a \neq 1$ die Umkehrfunktion der Exponentialfunktion $y = a^x$ und umgekehrt. Die natürliche Logarithmusfunktion $y = \ln x$ ist die Umkehrfunktion der e-Funktion $y = e^x$ und umgekehrt (Abb. 5.40).

Trigonometrie

6

Arnfried Kemnitz



6.1 Definition der trigonometrischen Funktionen 112
6.2 Trigonometrische Funktionen für beliebige Winkel 113
6.3 Beziehungen für den gleichen Winkel 114
6.4 Graphen der trigonometrischen Funktionen 114
6.5 Reduktionsformeln 116
6.6 Additionstheoreme 116
6.7 Sinussatz und Kosinussatz 117
6.8 Arkusfunktionen 118

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_6

Das Wort Trigonometrie kommt aus dem Griechischen und bedeutet Dreiecksmessung. Die Trigonometrie ist die Lehre von der Dreiecksberechnung mit Hilfe von Winkelfunktionen (trigonometrischen Funktionen).

6.1 Definition der trigonometrischen Funktionen

In einem rechtwinkligen Dreieck ist die Hypotenuse die dem rechten Winkel gegenüberliegende Dreiecksseite, die beiden anderen Seiten (also die Schenkel des rechten Winkels) sind die Katheten (vgl. Abschn. 3.6.4).

In einem rechtwinkligen Dreieck mit den Winkeln α, β und $\gamma = 90°$ gilt $\alpha + \beta = 90°$.

Die Ankathete eines Winkels α in einem rechtwinkligen Dreieck ist die Kathete, die auf einem Schenkel von α liegt. Die andere Kathete heißt Gegenkathete von α.

Das Verhältnis zweier beliebiger Seiten im rechtwinkligen Dreieck ist abhängig von dem Winkel α (und wegen $\beta = 90° - \alpha$ natürlich auch vom Winkel β), das heißt, das Verhältnis zweier Seiten ist eine Funktion des Winkels α (bzw. des Winkels β). Die trigonometrischen Funktionen sind definiert als das Verhältnis zweier Seiten im rechtwinkligen Dreieck.

In einem rechtwinkligen Dreieck ist (Abb. 6.1)

- $\sin\alpha$, der Sinus des Winkels α, das Verhältnis von Gegenkathete zu Hypotenuse,
- $\cos\alpha$, der Kosinus des Winkels α, das Verhältnis von Ankathete zu Hypotenuse,
- $\tan\alpha$, der Tangens des Winkels α, das Verhältnis von Gegenkathete zu Ankathete,
- $\cot\alpha$, der Kotangens des Winkels α, das Verhältnis von Ankathete zu Gegenkathete.

$$\text{Sinus:} \quad \sin\alpha = \frac{a}{c} = \frac{\text{Gegenkathete}}{\text{Hypotenuse}}$$

$$\text{Kosinus:} \quad \cos\alpha = \frac{b}{c} = \frac{\text{Ankathete}}{\text{Hypotenuse}}$$

$$\text{Tangens:} \quad \tan\alpha = \frac{a}{b} = \frac{\text{Gegenkathete}}{\text{Ankathete}}$$

$$\text{Kotangens:} \quad \cot\alpha = \frac{b}{a} = \frac{\text{Ankathete}}{\text{Gegenkathete}}$$

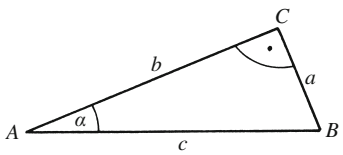


Abb. 6.1 $\sin\alpha = \frac{a}{c}$, $\cos\alpha = \frac{b}{c}$, $\tan\alpha = \frac{a}{b}$, $\cot\alpha = \frac{b}{a}$

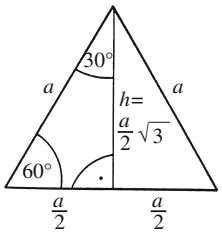


Abb. 6.2 Zur Berechnung von sin 30° und sin 60°

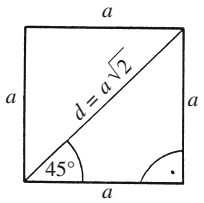


Abb. 6.3 Zur Berechnung von sin 45°

Andere, weniger gebräuchliche Namen für trigonometrische Funktionen sind Winkelfunktionen oder Kreisfunktionen oder goniometrische Funktionen.

In der folgenden Tabelle sind einige spezielle Werte der trigonometrischen Funktionen angegeben:

Gradmaß φ	0°	30°	45°	60°	90°
Bogenmaß b	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
sin	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
cos	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
tan	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	–
cot	–	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0

Merkregel:

Gradmaß φ	0°	30°	45°	60°	90°
$\sin\varphi$	$\frac{1}{2}\sqrt{0}$	$\frac{1}{2}\sqrt{1}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{4}$

Die meisten dieser Werte lassen sich mit Hilfe des Satzes von Pythagoras berechnen. Dies soll am Beispiel des Sinus vorgeführt werden (Abb. 6.2 und 6.3).

Für die Höhe h in einem gleichseitigen Dreieck mit der Seitenlänge a gilt nach dem Satz des Pythagoras $h^2 = a^2 - (\frac{a}{2})^2$, also $h = \frac{a}{2}\sqrt{3}$. Es folgt $\sin 30° = \frac{2}{a} = \frac{1}{2}$ und $\sin 60° = \frac{h}{a} = \frac{1}{2}\sqrt{3}$.

Für den Durchmesser d in einem Quadrat der Seitenlänge a gilt nach dem Satz des Pythagoras $d^2 = a^2 + a^2$, also $d = a\sqrt{2}$. Es folgt $\sin 45° = \frac{a}{d} = \frac{a\sqrt{2}}{a} = \frac{1}{\sqrt{2}} = \frac{1}{2}\sqrt{2}$.

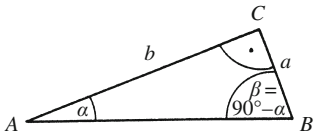


Abb. 6.4 Komplementwinkel

Die beiden spitzen Winkel α und β in einem rechtwinkligen Dreieck sind Komplementwinkel, es gilt also $\beta = 90° - \alpha$. Aus der Definition der trigonometrischen Funktionen folgt (Abb. 6.4)

$$\sin\beta = \frac{b}{c} \text{ und } \cos\alpha = \frac{b}{c} \quad \Rightarrow \quad \sin\beta = \cos\alpha$$

$$\cos\beta = \frac{a}{c} \text{ und } \sin\alpha = \frac{a}{c} \quad \Rightarrow \quad \cos\beta = \sin\alpha$$

$$\tan\beta = \frac{b}{a} \text{ und } \cot\alpha = \frac{b}{a} \quad \Rightarrow \quad \tan\beta = \cot\alpha$$

$$\cot\beta = \frac{a}{b} \text{ und } \tan\alpha = \frac{a}{b} \quad \Rightarrow \quad \cot\beta = \tan\alpha$$

Komplementwinkel

$$\sin(90° - \alpha) = \cos\alpha$$

$$\cos(90° - \alpha) = \sin\alpha$$

$$\tan(90° - \alpha) = \cot\alpha$$

$$\cot(90° - \alpha) = \tan\alpha$$

6.2 Trigonometrische Funktionen für beliebige Winkel

Die Definition der trigonometrischen Funktionen eines Winkels α im rechtwinkligen Dreieck ist nur für spitze Winkel möglich (also $0° < \alpha < 90°$). Am Einheitskreis (Kreis mit dem Radius $r = 1$) lassen sich die trigonometrischen Funktionen für beliebige Winkel definieren:

Der Mittelpunkt des Einheitskreises sei der Koordinatenursprung O eines kartesischen Koordinatensystems (vgl. Abschn. 7.1.1). Ein beliebiger Punkt $P = P(x|y)$ auf dem Einheitskreis legt einen Winkel α fest, nämlich den Winkel zwischen der x-Achse und der Geraden durch O und P. Dabei wird α in mathematisch positiver Richtung, also gegen den Uhrzeigersinn, gemessen.

Mit den vorzeichenbehafteten Koordinaten x und y des Punktes P werden die trigonometrischen Funktionen dann definiert durch (Abb. 6.5)

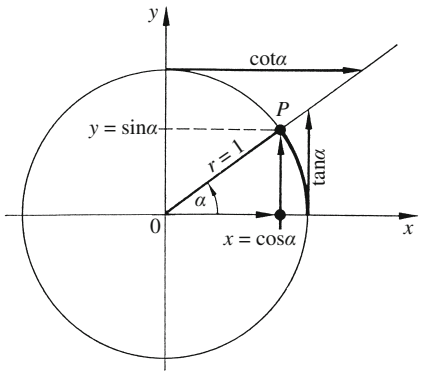


Abb. 6.5 Definition der trigonometrischen Funktionen für beliebige Winkel

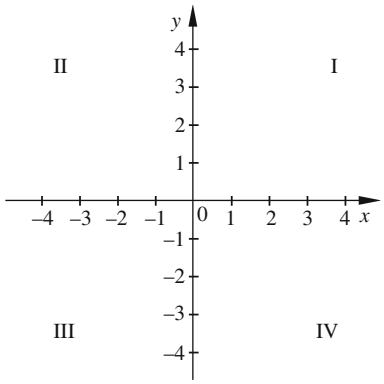


Abb. 6.6 Quadranten

$$\text{Sinus:} \quad \sin\alpha = y$$

$$\text{Kosinus:} \quad \cos\alpha = x$$

$$\text{Tangens:} \quad \tan\alpha = \frac{y}{x}$$

$$\text{Kotangens:} \quad \cot\alpha = \frac{x}{y}$$

Der Abschnitt des Einheitskreises zwischen der x-Achse und dem Punkt P ist das Bogenmaß b des Winkels α.

Durchläuft P den Einheitskreis im mathematisch positiven Drehsinn, dann sind α und b positiv. Durchläuft P den Einheitskreis jedoch im mathematisch negativen Drehsinn, dann sind α und b negativ.

Im Einheitskreis sind damit die trigonometrischen Funktionen für beliebige Winkel α im Gradmaß oder für beliebige reelle Zahlen b (Bogenmaß von α) definiert, für die die entsprechenden Nenner nicht verschwinden.

Bei der Berechnung von Funktionswerten muss beachtet werden, ob das Argument im Gradmaß oder im Bogenmaß angegeben ist.

Durch die beiden orientierten Achsen eines kartesischen Koordinatensystems wird die Ebene in vier Teile eingeteilt, die Quadranten (Abb. 6.6).

Die Punkte des ersten Quadranten haben sowohl positive x- als auch positive y-Koordinaten, die Punkte des zweiten Quadranten haben negative x- und positive y-Koordinaten, die Punkte des dritten Quadranten haben negative x- und negative y-Koordinaten und die Punkte des vierten Quadranten haben positive x- und negative y-Koordinaten.

Für die Vorzeichen der trigonometrischen Funktionen in den einzelnen Quadranten gilt:

Quadrant	sin	cos	tan	cot
I	+	+	+	+
II	+	−	−	−
III	−	−	+	+
IV	−	+	−	−

6.3 Beziehungen für den gleichen Winkel

Für beliebige Winkel α gelten folgende Umrechnungsformeln:

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha} = \frac{1}{\cot\alpha} \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha} = \frac{1}{\tan\alpha}$$

$$\sin^2\alpha + \cos^2\alpha = 1 \qquad \tan\alpha \cdot \cot\alpha = 1$$

$$1 + \tan^2\alpha = \frac{1}{\cos^2\alpha} \qquad 1 + \cot^2\alpha = \frac{1}{\sin^2\alpha}$$

Für Potenzen $(f(x))^k$ von Funktionswerten ist die Schreibweise $f^k(x)$ üblich, etwa $\sin^2\alpha$ (gesprochen: Sinus Quadrat Alpha) für $(\sin\alpha)^2$. Diese Beziehungen lassen sich im rechtwinkligen Dreieck leicht nachrechnen.

Beispiel

$$\sin^2\alpha + \cos^2\alpha = \left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = \frac{a^2+b^2}{c^2} = 1,$$

denn nach dem Satz des Pythagoras gilt im rechtwinkligen Dreieck $a^2 + b^2 = c^2$. ◀

Alle Beziehungen gelten auch allgemein, das heißt, für beliebige Winkel α.

Nach diesen Beziehungen lässt sich jede trigonometrische Funktion durch jede andere desselben Winkels ausdrücken.

Will man zum Beispiel $\sin\alpha$ durch $\cos\alpha$ ausdrücken, so folgt $\sin\alpha = \pm\sqrt{1-\cos^2\alpha}$ aus $\sin^2\alpha + \cos^2\alpha = 1$.

Für Winkel im ersten Quadranten, also für Winkel α mit $0° < \alpha < 90°$ gilt:

	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
$\sin\alpha =$	$\sin\alpha$	$\sqrt{1-\cos^2\alpha}$	$\frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}}$	$\frac{1}{\sqrt{1+\cot^2\alpha}}$
$\cos\alpha =$	$\sqrt{1-\sin^2\alpha}$	$\cos\alpha$	$\frac{1}{\sqrt{1+\tan^2\alpha}}$	$\frac{\cot\alpha}{\sqrt{1+\cot^2\alpha}}$
$\tan\alpha =$	$\frac{\sin\alpha}{\sqrt{1-\sin^2\alpha}}$	$\frac{\sqrt{1-\cos^2\alpha}}{\cos\alpha}$	$\tan\alpha$	$\frac{1}{\cot\alpha}$
$\cot\alpha =$	$\frac{\sqrt{1-\sin^2\alpha}}{\sin\alpha}$	$\frac{\cos\alpha}{\sqrt{1-\cos^2\alpha}}$	$\frac{1}{\tan\alpha}$	$\cot\alpha$

In den übrigen Quadranten sind die Vorzeichen der Wurzeln nach der Vorzeichentabelle (vgl. Abschn. 6.2) oder am Einheitskreis zu bestimmen.

Beispiel

Im dritten Quadranten sind sowohl $\sin\alpha$ als auch $\cos\alpha$ negativ. Deswegen gilt für Winkel α mit $180° < \alpha < 270°$ zum Beispiel $\sin\alpha = -\sqrt{1-\cos^2\alpha}$ und $\cos\alpha = -\sqrt{1-\sin^2\alpha}$. ◀

6.4 Graphen der trigonometrischen Funktionen

Ein anschauliches Bild von Eigenschaften der trigonometrischen Funktionen erhält man, wenn in einem kartesischen Koordinatensystem (vgl. Abschn. 7.1.1) als Abszissen (x-Werte) die Winkel (im Gradmaß oder im Bogenmaß) und als Ordinaten (y-Werte) die Werte der betreffenden trigonometrischen Funktionen eingetragen werden. Die Funktionswerte ergeben sich als vorzeichenbehaftete Längen der entsprechenden Strecken am Einheitskreis (Abb. 6.7 und 6.8).

Die Graphen der trigonometrischen Funktionen nennt man auch Kurven. So ist zum Beispiel die Sinuskurve der Graph der Sinusfunktion.

In der folgenden Aufzählung sind alle Winkel im Bogenmaß angegeben (Abb. 6.9).

1. *Sinusfunktion*
   Die Funktion $y = \sin x$ mit dem Definitionsbereich $D = \mathbb{R}$ und dem Wertebereich $W = [-1, 1]$. Die Sinusfunktion hat die Periode 2π, es gilt also $\sin(x + 2k\pi) = \sin x$ für $k = 0, \pm1, \pm2, \ldots$. Die Amplitude der Funktion ist 1, denn es gilt $|\sin x| \leq 1$ und $\sin\frac{\pi}{2} = 1$.
   Die Sinusfunktion ist wegen $\sin(-x) = -\sin x$ für alle x eine ungerade Funktion. Die Sinuskurve ist also symmetrisch zum Koordinatenursprung.
2. *Kosinusfunktion*
   Die Funktion $y = \cos x$ mit dem Definitionsbereich $D = \mathbb{R}$ und dem Wertebereich $W = [-1, 1]$. Die Kosinusfunktion hat ebenfalls die Periode 2π, es gilt $\cos(x + 2k\pi) = \cos x$ für $k = 0, \pm1, \pm2, \ldots$. Die Amplitude der Funktion ist 1, denn es gilt $|\cos x| \leq 1$ und $\cos 0 = 1$.
   Die Kosinusfunktion ist wegen $\cos(-x) = \cos x$ für alle x eine gerade Funktion. Die Kosinuskurve ist also symmetrisch zur y-Achse.

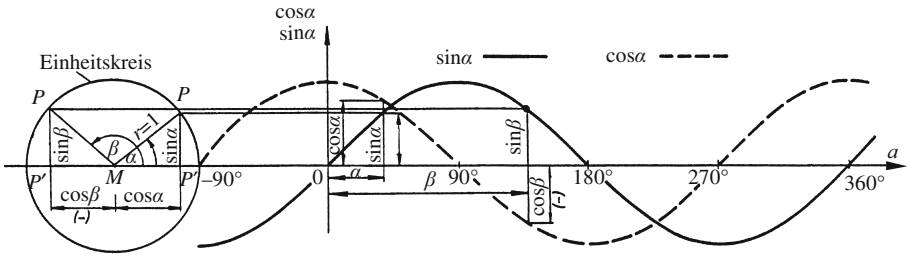


Abb. 6.7 Sinuskurve und Kosinuskurve

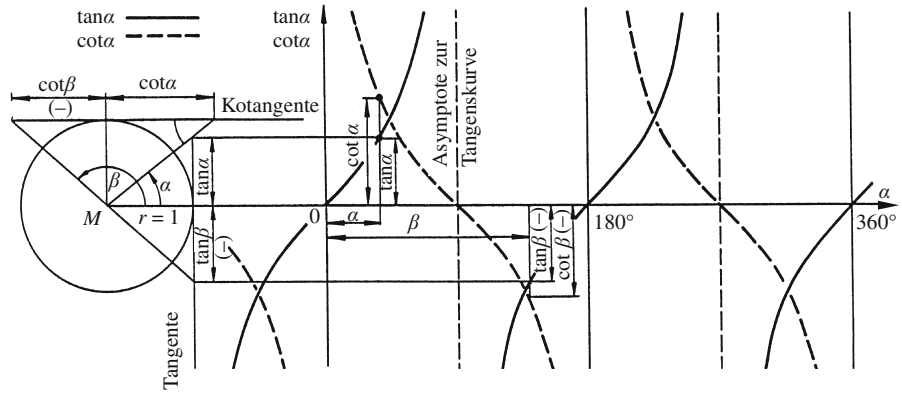


Abb. 6.8 Tangenskurve und Kotangenskurve

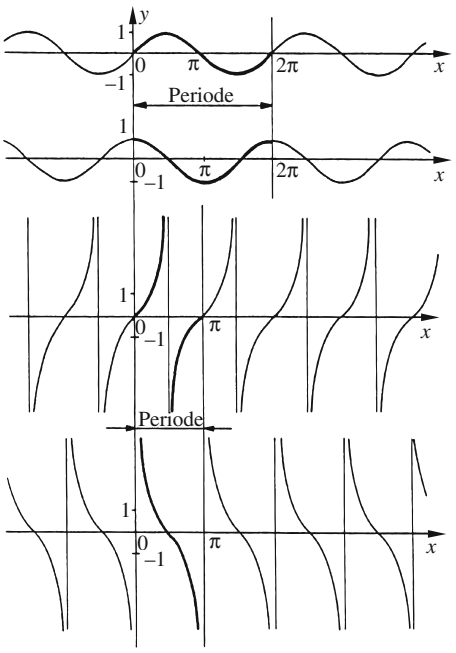


Abb. 6.9 Die Graphen der trigonometrischen Funktionen $y = \sin x$, $y = \cos x$, $y = \tan x$, $y = \cot x$ (von oben nach unten)

3. *Tangensfunktion*
   Die Funktion $y = \tan x$ mit dem Definitionsbereich $D = \mathbb{R}$, $x \neq \frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$ und dem Wertebereich $W = \mathbb{R}$.
   Die Stellen $x = \frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$ sind Pole der Funktion. Nähert man sich einem Pol $x = x_p$ mit wachsenden x-Werten (also $x < x_p$), dann geht $\tan x$ gegen $+\infty$. Nähert man sich dagegen einem Pol $x = x_p$ mit abnehmenden x-Werten (also $x > x_p$), so geht $\tan x$ gegen $-\infty$. Die Geraden $x = \frac{\pi}{2} + k\pi$ sind Asymptoten der Funktion.
   Die Tangensfunktion hat die Periode π, es gilt also $\tan(x + k\pi) = \tan x$ für $k = 0, \pm 1, \pm 2, \ldots$ Eine Amplitude besitzt die Funktion nicht (Pole!).
   Die Tangensfunktion ist wegen $\tan(-x) = -\tan x$ für alle x eine ungerade Funktion. Die Tangenskurve ist also symmetrisch zum Koordinatenursprung.
4. *Kotangensfunktion*
   Die Funktion $y = \cot x$ mit dem Definitionsbereich $D = \mathbb{R}$, $x \neq k\pi$, $k \in \mathbb{Z}$ und dem Wertebereich $W = \mathbb{R}$.
   Die Stellen $x = k\pi$, $k \in \mathbb{Z}$ sind Pole der Funktion. Nähert man sich einem Pol $x = x_p$ mit wachsenden x-Werten (also $x < x_p$), dann geht $\cot x$ gegen $-\infty$. Nähert man sich dagegen einem Pol $x = x_p$ mit abnehmenden x-Werten (also $x > x_p$), so geht $\cot x$ gegen $+\infty$. Die Geraden $x = k\pi$ sind Asymptoten der Funktion.
   Die Kotangensfunktion hat die Periode π, es gilt also $\cot(x + k\pi) = \cot x$ für $k = 0, \pm 1, \pm 2, \ldots$ Eine Amplitude besitzt die Funktion nicht (Pole!).
   Die Kotangensfunktion ist ungerade, denn es gilt $\cot(-x) = -\cot x$. Die Kotangenskurve ist also symmetrisch zum Koordinatenursprung.

6.5 Reduktionsformeln

Wegen der Periodizität können die trigonometrischen Funktionen für beliebige Winkel beim Sinus und Kosinus auf solche zwischen 0° und 360° und beim Tangens und Kotangens auf solche zwischen 0° und 180° zurückgeführt werden. Für beliebige ganze Zahlen k gilt:

$$\begin{aligned}\sin(360° \cdot k + \alpha) &= \sin\alpha\\ \cos(360° \cdot k + \alpha) &= \cos\alpha\\ \tan(180° \cdot k + \alpha) &= \tan\alpha\\ \cot(180° \cdot k + \alpha) &= \cot\alpha\end{aligned}$$

Wegen der Symmetrie der trigonometrischen Funktionen gilt für negative Winkel:

$$\begin{aligned}\sin(-\alpha) &= -\sin\alpha\\ \cos(-\alpha) &= \cos\alpha\\ \tan(-\alpha) &= -\tan\alpha\\ \cot(-\alpha) &= -\cot\alpha\end{aligned}$$

Jeder so reduzierte Winkel kann durch eine der folgenden Beziehungen auf einen Winkel zwischen 0° und 90° zurückgeführt werden:

	$\beta =$			
Funktion	$90° \pm \alpha$	$180° \pm \alpha$	$270° \pm \alpha$	$360° - \alpha$
$\sin\beta$	$+\cos\alpha$	$\mp\sin\alpha$	$-\cos\alpha$	$-\sin\alpha$
$\cos\beta$	$\mp\sin\alpha$	$-\cos\alpha$	$\pm\sin\alpha$	$+\cos\alpha$
$\tan\beta$	$\mp\cot\alpha$	$\pm\tan\alpha$	$\mp\cot\alpha$	$-\tan\alpha$
$\cot\beta$	$\mp\tan\alpha$	$\pm\cot\alpha$	$\mp\tan\alpha$	$-\cot\alpha$

6.6 Additionstheoreme

Die Additionstheoreme sind Formeln für die trigonometrischen Funktionen von Winkelsummen und Winkeldifferenzen.

Die meisten dieser Gleichungen lassen sich mit Hilfe der Euler'schen Formel $e^{jz} = \cos z + j\sin z$ für komplexe Zahlen (vgl. Abschn. 1.10.8) zusammen mit den Potenzgesetzen herleiten.

Beispiel

$$\begin{aligned}&\cos(\alpha+\beta) + j\sin(\alpha+\beta)\\ &= e^{j(\alpha+\beta)} = e^{j\alpha}\cdot e^{j\beta} = (\cos\alpha + j\sin\alpha)\cdot(\cos\beta + j\sin\beta)\\ &= (\cos\alpha\cos\beta - \sin\alpha\sin\beta) + j(\sin\alpha\cos\beta + \cos\alpha\sin\beta)\end{aligned}$$

Vergleich von Real- und Imaginärteil ergibt

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \quad \text{und}$$
$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta. \qquad \blacktriangleleft$$

Trigonometrische Funktionen der Summe und der Differenz zweier Winkel

$$\begin{aligned}\sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta\\ \sin(\alpha-\beta) &= \sin\alpha\cos\beta - \cos\alpha\sin\beta\\ \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta\\ \cos(\alpha-\beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta\\ \tan(\alpha+\beta) &= \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}\\ \tan(\alpha-\beta) &= \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta}\\ \cot(\alpha+\beta) &= \frac{\cot\alpha\cot\beta - 1}{\cot\alpha + \cot\beta}\\ \cot(\alpha-\beta) &= \frac{\cot\alpha\cot\beta + 1}{\cot\alpha - \cot\beta}\end{aligned}$$

Trigonometrische Funktionen für Winkelvielfache

$$\begin{aligned}\sin 2\alpha &= 2\sin\alpha\cos\alpha = \frac{2\tan\alpha}{1+\tan^2\alpha}\\ \sin 3\alpha &= 3\sin\alpha - 4\sin^3\alpha\\ \sin 4\alpha &= 8\sin\alpha\cos^3\alpha - 4\sin\alpha\cos\alpha\\ \cos 2\alpha &= \cos^2\alpha - \sin^2\alpha = 1 - 2\sin^2\alpha = 2\cos^2\alpha - 1\\ \cos 3\alpha &= 4\cos^3\alpha - 3\cos\alpha\\ \cos 4\alpha &= 8\cos^4\alpha - 8\cos^2\alpha + 1\\ \tan 2\alpha &= \frac{2\tan\alpha}{1-\tan^2\alpha} = \frac{2}{\cot\alpha - \tan\alpha}\\ \tan 3\alpha &= \frac{3\tan\alpha - \tan^3\alpha}{1 - 3\tan^2\alpha}\\ \tan 4\alpha &= \frac{4\tan\alpha - 4\tan^3\alpha}{1 - 6\tan^2\alpha + \tan^4\alpha}\\ \cot 2\alpha &= \frac{\cot^2\alpha - 1}{2\cot\alpha} = \frac{\cot\alpha - \tan\alpha}{2} = \frac{1}{\tan 2\alpha}\\ \cot 3\alpha &= \frac{\cot^3\alpha - 3\cot\alpha}{3\cot^2\alpha - 1}\\ \cot 4\alpha &= \frac{\cot^4\alpha - 6\cot^2\alpha + 1}{4\cot^3\alpha - 4\cot\alpha}\end{aligned}$$

Für allgemeine Winkelvielfache erhält man aus der Entwicklung der Formel von Moivre (vgl. Abschn. 1.10.6) durch Vergleich des Real- und Imaginärteils entsprechende Formeln:

$$\cos n\alpha + j \sin n\alpha = (\cos\alpha + j \sin\alpha)^n$$
$$= \sum_{k=0}^{n} j^k \cdot \binom{n}{k} \cdot \cos^{n-k}\alpha \cdot \sin^k \alpha$$

Summen und Differenzen zweier trigonometrischer Funktionen

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

$$\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

$$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$$

$$\tan\alpha + \tan\beta = \frac{\sin(\alpha+\beta)}{\cos\alpha\cos\beta}$$

$$\tan\alpha - \tan\beta = \frac{\sin(\alpha-\beta)}{\cos\alpha\cos\beta}$$

$$\cot\alpha + \cot\beta = \frac{\sin(\alpha+\beta)}{\sin\alpha\sin\beta}$$

$$\cot\alpha - \cot\beta = -\frac{\sin(\alpha+\beta)}{\sin\alpha\sin\beta}$$

Produkte trigonometrischer Funktionen

$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)]$$

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)]$$

$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha+\beta) + \sin(\alpha-\beta)]$$

$$\cos\alpha\sin\beta = \frac{1}{2}[\sin(\alpha+\beta) - \sin(\alpha-\beta)]$$

$$\tan\alpha\tan\beta = \frac{\tan\alpha+\tan\beta}{\cot\alpha+\cot\beta} = -\frac{\tan\alpha-\tan\beta}{\cot\alpha-\cot\beta}$$

$$\cot\alpha\cot\beta = \frac{\cot\alpha+\cot\beta}{\tan\alpha+\tan\beta} = -\frac{\cot\alpha-\cot\beta}{\tan\alpha-\tan\beta}$$

$$\tan\alpha\cot\beta = \frac{\tan\alpha+\cot\beta}{\cot\alpha+\tan\beta} = -\frac{\tan\alpha-\cot\beta}{\cot\alpha-\tan\beta}$$

Potenzen trigonometrischer Funktionen

$$\sin^2\alpha = \frac{1}{2}(1-\cos 2\alpha)$$

$$\sin^3\alpha = \frac{1}{4}(3\sin\alpha - \sin 3\alpha)$$

$$\sin^4\alpha = \frac{1}{8}(\cos 4\alpha - 4\cos 2\alpha + 3)$$

$$\cos^2\alpha = \frac{1}{2}(1+\cos 2\alpha)$$

$$\cos^3\alpha = \frac{1}{4}(3\cos\alpha + \cos 3\alpha)$$

$$\cos^4\alpha = \frac{1}{8}(\cos 4\alpha + 4\cos 2\alpha + 3)$$

Die Formeln für $\sin^n\alpha$ und $\cos^n\alpha$ erhält man, indem man die Formeln für $\cos n\alpha$ und $\sin n\alpha$ nacheinander anwendet.

6.7 Sinussatz und Kosinussatz

Sinussatz

In einem beliebigen Dreieck verhalten sich die Längen der Seiten wie die Sinuswerte der gegenüberliegenden Winkel.

Sinussatz

$$\sin\frac{\alpha}{a} = \sin\frac{\beta}{b} = \sin\frac{\gamma}{c} \quad \text{oder}$$
$$\sin\alpha : \sin\beta : \sin\gamma = a : b : c$$

Der Sinussatz lässt sich mit Hilfe der Höhen, also durch Zerlegung des Dreiecks in zwei rechtwinklige Dreiecke herleiten (Abb. 6.10):

$$\sin\alpha = \frac{h}{b} \quad \Rightarrow \quad h = b \cdot \sin\alpha;$$
$$\sin\beta = \frac{h}{a} \quad \Rightarrow \quad h = a \cdot \sin\beta$$

Es folgt $b \cdot \sin\alpha = a \cdot \sin\beta$ und daraus $a : b = \sin\alpha : \sin\beta$.

Die anderen Proportionen lassen sich analog herleiten.

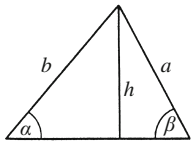


Abb. 6.10 Zur Herleitung des Sinussatzes

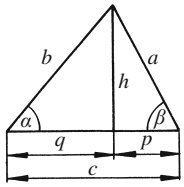


Abb. 6.11 Zur Herleitung des Kosinussatzes

Kosinussatz

In einem beliebigen Dreieck ist das Quadrat einer Seitenlänge gleich der Summe der Quadrate der beiden anderen Seitenlängen minus dem doppelten Produkt der Längen dieser beiden anderen Seiten und dem Kosinus des von ihnen eingeschlossenen Winkels.

Kosinussatz

$$a^2 = b^2 + c^2 - 2bc\cos\alpha$$
$$b^2 = a^2 + c^2 - 2ac\cos\beta$$
$$c^2 = a^2 + b^2 - 2ab\cos\gamma \quad \text{oder}$$
$$\cos\alpha = \frac{b^2 + c^2 - a^2}{2bc}$$
$$\cos\beta = \frac{a^2 + c^2 - b^2}{2ac}$$
$$\cos\gamma = \frac{a^2 + b^2 - c^2}{2ab}$$

Der Kosinussatz ist eine Verallgemeinerung des Satzes von Pythagoras, der für rechtwinklige Dreiecke gilt, auf beliebige Dreiecke. Gilt etwa $\gamma = 90°$, dann folgt $c^2 = a^2 + b^2$ wegen $\cos\gamma = \cos 90° = 0$.

Der Kosinussatz lässt sich durch Zerlegung des Dreiecks in zwei rechtwinklige Dreiecke und Anwendung des Satzes von Pythagoras herleiten (Abb. 6.11):

$$\cos\alpha = \frac{q}{b}, \quad q = c - p$$
$$\Rightarrow \quad \cos\alpha = \frac{c-p}{b} \quad \Rightarrow \quad p = c - b\cos\alpha$$
$$\sin\alpha = \frac{h}{b} \quad \Rightarrow \quad h = b\sin\alpha$$

Durch Einsetzen der Ausdrücke für p und für h ergibt sich:

$$a^2 = p^2 + h^2 \quad \Rightarrow \quad a^2 = (c - b\cos\alpha)^2 + (b\sin\alpha)^2$$
$$\Rightarrow \quad a^2 = c^2 - 2bc\cos\alpha + b^2\cos^2\alpha + b^2\sin^2\alpha$$
$$\Rightarrow \quad a^2 = c^2 - 2bc\cos\alpha + b^2(\cos^2\alpha + \sin^2\alpha)$$
$$\Rightarrow \quad a^2 = b^2 + c^2 - 2bc\cos\alpha$$

(denn $\sin^2\alpha + \cos^2\alpha = 1$).

Die anderen Gleichungen lassen sich ganz entsprechend herleiten.

6.8 Arkusfunktionen

Kennt man den Funktionswert einer trigonometrischen Funktion, etwa $y = \sin x$, und will man daraus den zugehörigen Winkel bestimmen, so muss man die Gleichung nach dem Winkel x auflösen, was mit Hilfe der Arkusfunktionen möglich ist: $x = \arcsin y$. Die Arkusfunktionen sind also die Umkehrfunktionen der trigonometrischen Funktionen.

Die Arkusfunktionen werden auch zyklometrische Funktionen oder inverse trigonometrische Funktionen genannt.

Zu ihrer eindeutigen Definition wird der Definitionsbereich der trigonometrischen Funktionen in Monotonieintervalle zerlegt, so dass für jedes Monotonieintervall eine Umkehrfunktion erhalten wird (vgl. Abschn. 5.2.8: Streng monotone Funktionen besitzen Umkehrfunktionen). Diese wird entsprechend dem zugehörigen Monotonieintervall mit dem Index k gekennzeichnet.

Die Vorgehensweise wird am Beispiel des Arkussinus gezeigt. Der Definitionsbereich von $y = \sin x$ wird in die Monotonieintervalle $k\pi - \frac{\pi}{2} \le x \le k\pi + \frac{\pi}{2}$ mit $k = 0, \pm1, \pm2, \ldots$ zerlegt. Durch Spiegelung von $y = \sin x$ an der Winkelhalbierenden $y = x$ erhält man die Umkehrfunktionen $y = \arc_k \sin x$ mit den Definitionsbereichen $D_k = [-1, 1]$ und den Wertebereichen $W_k = [k\pi - \frac{\pi}{2}, k\pi + \frac{\pi}{2}]$, wobei $k = 0, \pm1, \pm2, \ldots$. Die Schreibweise $y = \arc_k \sin x$ ist gleichbedeutend mit $x = \sin y$.

$$y = \arc_k \sin x \quad \Leftrightarrow \quad x = \sin y$$

Die übrigen Arkusfunktionen ergeben sich analog. In der Tabelle sind die Definitions- und Wertebereiche aller Arkusfunktionen zusammengestellt, die Abb. 6.12 bis 6.15 zeigen die Graphen der Arkusfunktionen.

Setzt man $k = 0$, dann erhält man jeweils den sogenannten Hauptwert der Arkusfunktion. Den Hauptwert schreibt man ohne den Index k, also zum Beispiel $\arcsin x = \arc_0 \sin x$. Für andere Werte von k erhält man Nebenwerte der entsprechenden Arkusfunktion. Den Hauptwert der Arkusfunktionen zeigt Abb. 6.16.

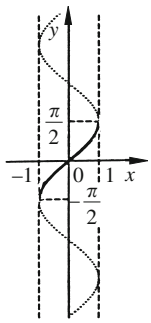


Abb. 6.12 Graph der Arkussinusfunktion

Tab. 6.1 Arkusfunktionen

Name	Schreibweise	Definitionsbereich	Wertebereich	Gleichbedeutende trigonometrische Funktion
Arkussinus	$y = \operatorname{arc}_k \sin x$	$-1 \le x \le 1$	$k\pi - \frac{\pi}{2} \le y \le k\pi + \frac{\pi}{2}$	$x = \sin y$
Arkuskosinus	$y = \operatorname{arc}_k \cos x$	$-1 \le x \le 1$	$k\pi \le y \le (k+1)\pi$	$x = \cos y$
Arkustangens	$y = \operatorname{arc}_k \tan x$	$-\infty < x < \infty$	$k\pi - \frac{\pi}{2} < y < k\pi + \frac{\pi}{2}$	$x = \tan y$
Arkuskotangens	$y = \operatorname{arc}_k \cot x$	$-\infty < x < \infty$	$k\pi < y < (k+1)\pi$	$x = \cot y$

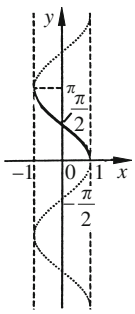


Abb. 6.13 Graph der Arkuskosinusfunktion

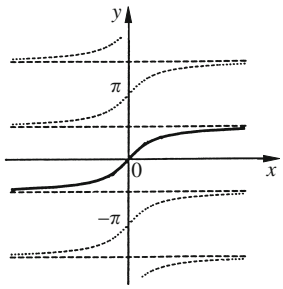


Abb. 6.14 Graph der Arkustangensfunktion

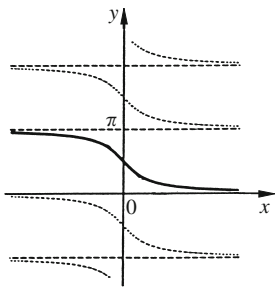


Abb. 6.15 Graph der Arkuskotangensfunktion

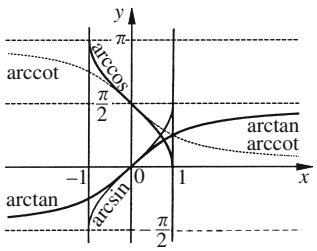


Abb. 6.16 Hauptwerte der Arkusfunktionen

Die Zurückführung von Nebenwerten auf die Hauptwerte der Arkusfunktionen erfolgt mit Hilfe der folgenden Formeln:

$$\operatorname{arc}_k \sin x = k\pi + (-1)^k \arcsin x$$

$$\operatorname{arc}_k \cos x = \begin{cases} (k+1)\pi - \arccos x & \text{falls } k \text{ ungerade} \\ k\pi + \arccos x & \text{falls } k \text{ gerade} \end{cases}$$

$$\operatorname{arc}_k \tan x = k\pi + \arctan x$$

$$\operatorname{arc}_k \cot x = k\pi + \operatorname{arccot} x$$

Taschenrechner geben immer die Hauptwerte der Arkusfunktionen an.

Beziehungen zwischen den Hauptwerten

$$\arcsin x = \frac{\pi}{2} - \arccos x = \arctan \frac{x}{\sqrt{1-x^2}}$$

$$\arccos x = \frac{\pi}{2} - \arcsin x = \operatorname{arccot} \frac{x}{\sqrt{1-x^2}}$$

$$\arctan x = \frac{\pi}{2} - \operatorname{arccot} x = \arcsin x \frac{x}{\sqrt{1+x^2}}$$

$$\operatorname{arccot} x = \frac{\pi}{2} \arctan x = \arccos \frac{x}{\sqrt{1+x^2}}$$

Formeln für negative Argumente

$$\arcsin(-x) = -\arcsin x$$

$$\arccos(-x) = \pi - \arccos x$$

$$\arctan(-x) = -\arctan x$$

$$\operatorname{arccot}(-x) = \pi - \operatorname{arccot} x$$

Beispiele

1. $\arcsin 0 = 0$; $\operatorname{arc}_k \sin 0 = k\pi$
2. $\arccos \frac{1}{2} = \frac{\pi}{3}$;
   $$\operatorname{arc}_k \cos \tfrac{1}{2} = \begin{cases} -\frac{\pi}{3} + (k+1)\pi & \text{falls } k \text{ ungerade} \\ \frac{\pi}{3} + k\pi & \text{falls } k \text{ gerade} \end{cases}$$
3. $\operatorname{arccot} 1 = \frac{\pi}{4}$; $\operatorname{arc}_k \cot 1 = \frac{\pi}{4} + k\pi$. ◀

Analytische Geometrie

7

Arnfried Kemnitz



7.1 Koordinatensysteme 122
7.2 Geraden 124
7.3 Kreise 128
7.4 Kugeln 130
7.5 Kegelschnitte 130
7.6 Vektoren 137

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_7

Der Grundgedanke der Analytischen Geometrie besteht darin, dass geometrische Untersuchungen mit rechnerischen Mitteln geführt werden. Geometrische Objekte werden dabei durch Gleichungen beschrieben und mit algebraischen Methoden untersucht.

7.1 Koordinatensysteme

Die Verbindung von Geometrie und Algebra wird dadurch erreicht, dass man die geometrischen Objekte als Punktmengen auffasst und jedem Punkt Zahlenwerte zuordnet, durch die er sich von anderen unterscheidet. Eine Kurve oder eine Gerade ist dann eine Menge von Punkten, für deren Zahlenwerte bestimmte Bedingungen gelten, die man Gleichungen dieser Objekte nennt, zum Beispiel Gleichung eines Kreises oder einer Geraden. Das geometrische Bild einer linearen Gleichung in zwei Variablen ist immer eine Gerade, das einer quadratischen Gleichung in zwei Variablen immer ein Kegelschnitt.

Die Grundlage für eine solche analytische Darstellung der Geometrie ist die Zuordnung zwischen Punkt und Zahl, die eindeutig sein muss. Auf einer Geraden oder allgemeiner auf einer Kurve genügt eine Zahl, auf einer Ebene oder einer Fläche ein Zahlenpaar und im Raum ein Zahlentripel (drei Zahlen), um einen Punkt eindeutig festzulegen. Umgekehrt bestimmt ein Punkt auf einer Kurve eindeutig eine Zahl, auf einer Fläche ein Zahlenpaar und im Raum ein Zahlentripel. Diese Zahlen werden Koordinaten des entsprechenden Punktes genannt. Die Koordinaten sind abhängig von dem zugrunde liegenden Koordinatensystem.

Es gibt verschiedene Möglichkeiten für Koordinatensysteme, von denen hier drei wichtige beschrieben werden. Allgemein kann man ein Koordinatensystem als ein System von geometrischen Objekten, mit deren Hilfe die Lage anderer geometrischer Objekte durch Zahlenwerte (Koordinaten) umkehrbar eindeutig beschrieben werden kann, bezeichnen.

Legt man auf einer Geraden g einen Anfangspunkt 0 (Nullpunkt), eine positive Richtung (Orientierung) und eine Längeneinheit l (Maßstab) fest, dann entspricht jeder reellen Zahl x ein bestimmter Punkt dieser Geraden, und umgekehrt entspricht jedem Punkt der Geraden eine reelle Zahl. Die Gerade g wird Zahlengerade genannt.

7.1.1 Kartesisches Koordinatensystem der Ebene

Um die Lage eines Punktes in der Ebene eindeutig festzulegen, sind zwei Zahlengeraden notwendig. Man ordnet die Zahlengeraden stets so an, dass ihre Nullpunkte zusammenfallen. Die Zahlengeraden werden Achsen des Koordinatensystems oder Koordinatenachsen genannt und als x- oder Abszissenachse und als y- oder Ordinatenachse bezeichnet. Der gemeinsame Nullpunkt, also der Schnittpunkt der beiden Geraden, heißt Koordinatenursprung oder Nullpunkt. Auf jeder der beiden Geraden wird vom Koordinatenursprung aus eine positive und eine negative Orientierung sowie ein Maßstab festgelegt (Abb. 7.1).

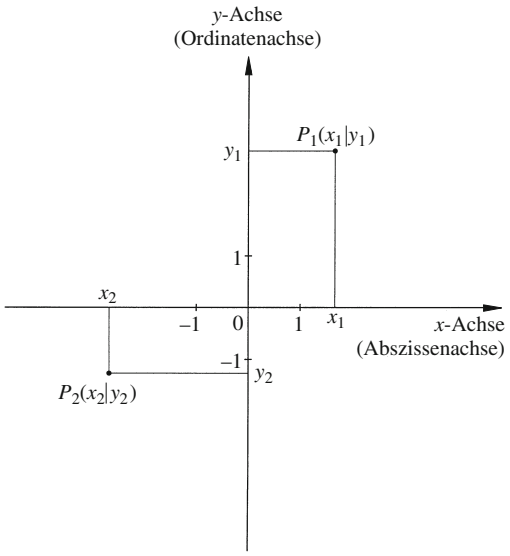


Abb. 7.1 Kartesisches Koordinatensystem der Ebene

In einem kartesischen (rechtwinkligen) Koordinatensystem stehen die Koordinatenachsen senkrecht aufeinander, die Achsen haben den gleichen Maßstab und bilden ein sogenanntes Rechtssystem: Die x-Achse geht durch Drehung um einen rechten Winkel im mathematisch positiven Sinne (linksdrehend, entgegen dem Uhrzeigersinn) in die y-Achse über.

Ein beliebiger Punkt P der Ebene kann dann durch seine kartesischen Koordinaten beschrieben werden: $P(x|y)$ mit x als Abszisse und y als Ordinate.

Dieses Koordinatensystem ist benannt nach dem französischen Mathematiker René Descartes, genannt Cartesius (1596–1650).

7.1.2 Polarkoordinatensystem der Ebene

Ein Polarkoordinatensystem der Ebene ist bestimmt durch einen festen Punkt, den Pol O, und einer von ihm ausgehenden fest gewählten Achse, der Polarachse, auf der wie bei einem Zahlenstrahl eine Orientierung und ein Maßstab festgelegt sind.

Ein beliebiger Punkt P der Ebene lässt sich dann durch seine Polarkoordinaten beschreiben: $P(r|\varphi)$, wobei r der Abstand des Punktes P vom Pol O ist und φ der Winkel, den der Strahl vom Pol O durch den Punkt P mit der Polarachse bildet (Abb. 7.2).

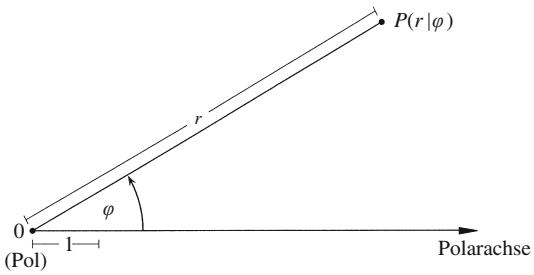


Abb. 7.2 Polarkoordinatensystem der Ebene

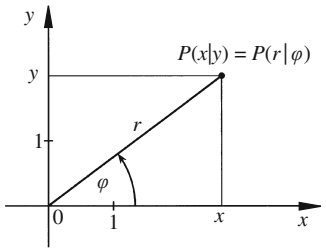


Abb. 7.3 Kartesische Koordinaten und Polarkoordinaten

Dabei wird der Winkel φ in mathematisch positiver Richtung (linksdrehend, entgegen dem Uhrzeigersinn) gemessen. Dieser Winkel φ ist nur bis auf ganzzahlige Vielfache von 2π bestimmt. Man nennt φ auch Polarwinkel des Punktes P.

7.1.3 Zusammenhang zwischen kartesischen und Polarkoordinaten

Ein beliebiges geometrisches Objekt kann in verschiedenen Koordinatensystemen beschrieben werden, zum Beispiel in einem kartesischen und in einem Polarkoordinatensystem. Für dieselben geometrischen Eigenschaften findet man dann zwei Gleichungen $f_1(x, y) = 0$ und $f_2(r, \varphi) = 0$. Durch Transformation (Überführung) des einen Koordinatensystems in das andere geht die eine Gleichung des geometrischen Objekts in die andere über (Abb. 7.3).

Die Transformationsgleichungen für den Übergang von Polarkoordinaten zu kartesischen Koordinaten und umgekehrt ergeben sich mit Hilfe der trigonometrischen und Arkusfunktionen. Zur Vereinfachung wird dabei vorausgesetzt, dass der Pol des Polarkoordinatensystems mit dem Koordinatenursprung des kartesischen Koordinatensystems und die Polarachse mit der x-Achse (Abszisse) zusammenfallen.

Transformationsgleichungen

$$x = r\cos\varphi \qquad r = \sqrt{x^2 + y^2}$$

$$y = r\sin\varphi \qquad \cos\varphi = \frac{x}{\sqrt{x^2 + y^2}}, \qquad \sin\varphi = \frac{y}{\sqrt{x^2 + y^2}}$$

Beispiele

1. Eine rechteckige Metallplatte soll zwei Bohrungen erhalten. Für die Mitten der Bohrungen soll gelten: Die erste Bohrung ist von einer Ecke der Platte 120 mm entfernt. Die Verbindungsstrecke zwischen dieser Ecke und der Bohrung soll mit der längeren Seite der Platte einen Winkel von 30° bilden. Die zweite Bohrung soll dreiviertel so weit von derselben Ecke entfernt sein: Die Verbindungsstrecke zwischen dieser Ecke und der zweiten Bohrung soll mit der ersten Verbindungsstrecke einen Winkel von 45° einschließen. Die Bohrungsmitten sind anzureißen (Abb. 7.4).

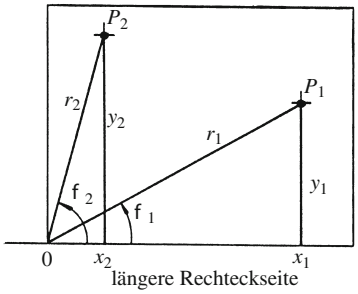


Abb. 7.4 Zu Beispiel 1

Mathematisch umgesetzt bedeutet die Aufgabe:
Die Polarkoordinaten $r_1 = 120\,\text{mm}$, $\varphi_1 = 30°$ und $r_2 = 90\,\text{mm}$, $\varphi_2 = 75°$ zweier Punkte $P_1(r_1|\varphi_1)$ und $P_2(r_2|\varphi_2)$ sind in kartesische Koordinaten umzurechnen ($\varphi_2 = \varphi_1 - 45°$ kommt nicht in Frage, da P_2 dann außerhalb der Platte läge).
Man berechnet:

$$x_1 = r_1\cos\varphi_1 = 120\,\text{mm}\cdot\cos 30° \approx 103{,}92\,\text{mm}$$
$$y_1 = r_1\sin\varphi_1 = 120\,\text{mm}\cdot\sin 30° = 60{,}00\,\text{mm}$$
$$x_2 = r_2\cos\varphi_2 = 90\,\text{mm}\cdot\cos 75° \approx 23{,}29\,\text{mm}$$
$$y_2 = r_2\sin\varphi_2 = 90\,\text{mm}\cdot\sin 75° \approx 86{,}93\,\text{mm}$$

Ergebnis:

$$x_1 \approx 130{,}92\,\text{mm}, \quad y_1 = 60{,}00\,\text{mm},$$
$$x_2 \approx 23{,}29\,\text{mm}, \quad y_2 \approx 86{,}93\,\text{mm}$$

2. Welche Polarkoordinaten haben die Ecken A, B, C des Dreiecks $A(2{,}9|2{,}3)$, $B(-3{,}0|-0{,}7)$, $C(1{,}8|-2{,}7)$?
Man berechnet (Abb. 7.5):

$$r_A = \sqrt{(2{,}9)^2 + (2{,}3)^2} \approx 3{,}7$$

$$\cos\varphi_A = \frac{2{,}9}{\sqrt{(2{,}9)^2 + (2{,}3)^2}} \approx 0{,}7835 \Rightarrow \varphi_A \approx 38{,}4°$$

$$r_B = \sqrt{(-3{,}0)^2 + (-0{,}7)^2} \approx 3{,}1$$

$\cos\varphi_B < 0$ und $\sin\varphi_B < 0 \Rightarrow 180° < \varphi_B < 270°$ (vgl. Vorzeichentabelle in Abschn. 6.2)

$$\cos\alpha = \frac{-3{,}0}{\sqrt{(-3{,}0)^2 + (-0{,}7)^2}} \approx -0{,}9738;$$

$$\cos\beta = 0{,}9738 \quad \Rightarrow \quad \beta \approx 13{,}1°$$

Wegen $\cos(180° + \alpha) = -\cos\alpha$ (vgl. Abschn. 6.5) folgt $\varphi_B \approx 180° + 13{,}1° = 193{,}1°$

$$r_C = \sqrt{(1{,}8)^2 + (-2{,}7)^2} \approx 3{,}2$$

$\cos\varphi_C > 0$ und $\sin\varphi_C < 0 \Rightarrow 270° < \varphi_C < 360°$ (vgl. Vorzeichentabelle in Abschn. 6.2)

$$\cos\alpha = \frac{1{,}8}{\sqrt{(1{,}8)^2 + (-2{,}7)^2}} \approx 0{,}5547 \Rightarrow \alpha \approx 56{,}3°$$

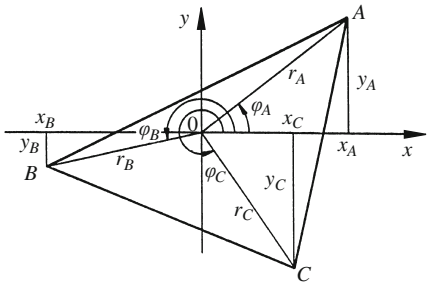


Abb. 7.5 Zu Beispiel 2

Wegen $\cos(360° - \alpha) = \cos\alpha$ (vgl. Abschn. 6.5) folgt $\varphi_C \approx 360° - 56{,}3° = 303{,}7°$.
Ergebnis: $A(3{,}7|38{,}4°)$, $B(3{,}1|193{,}1°)$, $C(3{,}2|303{,}7°)$ sind die drei Punkte in Polarkoordinaten. ◀

7.1.4 Kartesisches Koordinatensystem des Raums

Ein kartesisches Koordinatensystem des Raums besteht aus drei paarweise aufeinander senkrecht stehenden Geraden (Koordinatenachsen), die sich in einem Punkt, dem Koordinatenursprung, schneiden. Die drei Koordinatenachsen bilden ein Rechtssystem: Winkelt man Daumen, Zeigefinger und Mittelfinger der rechten Hand so ab, dass sie aufeinander senkrecht stehen, dann können diese Finger als positive Richtungen eines Rechtssystems aufgefasst werden. Man bezeichnet die Achsen in dieser Reihenfolge meist als x-Achse, y-Achse und z-Achse. Auf allen drei Achsen sind die Maßstäbe gleich.

Ein beliebiger Punkt P des Raums kann dann durch seine kartesischen Koordinaten beschrieben werden: $P(x|y|z)$, wobei x, y und z die senkrechten Projektionen des Punktes auf die drei Koordinatenachsen sind (Abb. 7.6).

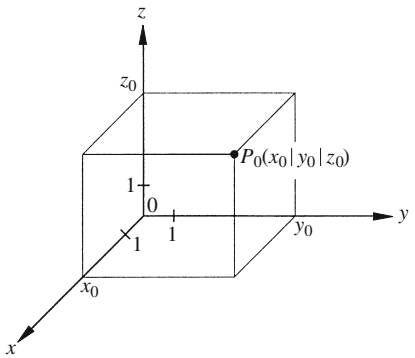


Abb. 7.6 Kartesisches Koordinatensystem des Raums

7.2 Geraden

7.2.1 Geradengleichungen

Eine Gerade ist die kürzeste Verbindung zweier Punkte. Eine Gerade ist durch zwei beliebige auf ihr liegende Punkte eindeutig bestimmt (vgl. Abschn. 3.1).

Für eine Gerade gibt es verschiedene Gleichungsformen.

1. Die Gleichung $ax + by + c = 0$ ist die allgemeine Geradengleichung, wobei die Koeffizienten a und b nicht gleichzeitig Null sein dürfen.

Allgemeine Geradengleichung

$$ax + by + c = 0$$

Die Variablen x und y sind die Koordinaten eines beliebigen Punktes der Geraden. Ein Punkt $P_0 = P(x_0|y_0)$ der Ebene liegt also genau dann auf der Geraden, wenn seine Koordinaten x_0 und y_0 die Gleichung erfüllen, wenn also $ax_0 + by_0 + c = 0$ gilt.

Die Koeffizienten a, b, c legen die Gerade eindeutig fest. Für $a = 0$ ist die Gerade eine Parallele zur x-Achse, für $b = 0$ eine Parallele zur y-Achse, und für $c = 0$ verläuft die Gerade durch den Koordinatenursprung (Nullpunkt).

2. Dividiert man die allgemeine Geradengleichung durch $b \neq 0$ (die Gerade ist also nicht parallel zur y-Achse), dann ergibt sich mit $m = -\frac{a}{b}$ und $n = -\frac{c}{b}$ die Hauptform oder Normalform der Geradengleichung (Abb. 7.7).

Hauptform oder Normalform

$$y = mx + n$$

Geraden, die Parallelen zur y-Achse sind, besitzen also keine Hauptform (Normalform).

Die Größe m wird Richtungskoeffizient oder Steigung der Geraden genannt. Die Steigung ist gleich dem Tangens des Winkels,

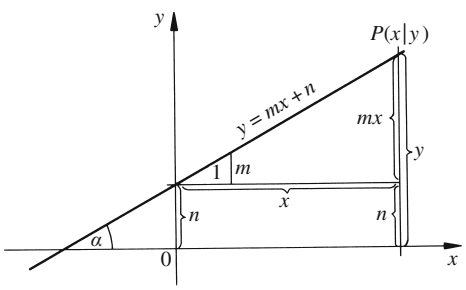


Abb. 7.7 Hauptform der Geradengleichung

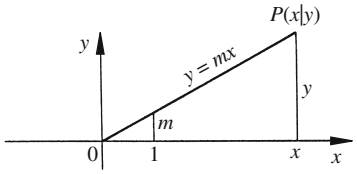


Abb. 7.8 Gerade mit der Gleichung $y = mx$

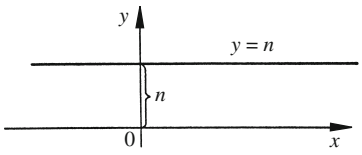


Abb. 7.9 Gerade mit der Gleichung $y = n$

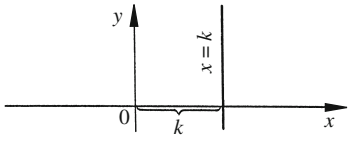


Abb. 7.10 Gerade mit der Gleichung $x = k$

den die Gerade mit der positiven Richtung der x-Achse einschließt. Die Strecke n wird von der Geraden auf der y-Achse abgeschnitten, deshalb heißt n auch Achsenabschnitt oder genauer y-Achsenabschnitt. Er kann ebenso wie der Tangens je nach Lage unterschiedliches Vorzeichen besitzen.

Sonderfälle:

2.1 $n = 0$: Die Gerade verläuft durch den Nullpunkt (Abb. 7.8).

Gerade durch Nullpunkt

$$y = mx$$

2.2 $m = 0$: Die Gerade ist eine Parallele zur x-Achse im Abstand n (Abb. 7.9).

Parallele zur x-Achse

$$y = n$$

2.3 Entsprechend ist $x = k$ die Gleichung einer Parallele zur y-Achse im Abstand k (Abb. 7.10).

Parallele zur y-Achse

$$x = k$$

3. Sind von einer Geraden ein Punkt $P_1 = P(x_1|y_1)$ und die Steigung m bekannt, dann lautet die Gleichung der Geraden

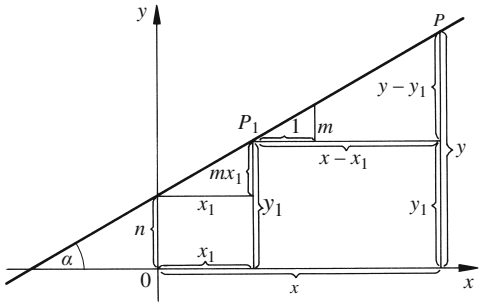


Abb. 7.11 Punktsteigungsform der Geradengleichung

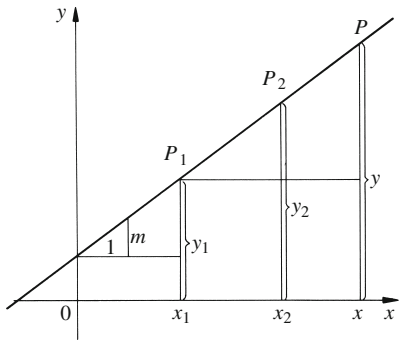


Abb. 7.12 Zweipunkteform der Geradengleichung

$y = m(x - x_1) + y_1$. Dies ist die Punktsteigungsform der Geradengleichung (Abb. 7.11).

Punktsteigungsform

$$y = m(x - x_1) + y_1$$

Wegen der Ähnlichkeit der rechtwinkligen Dreiecke mit den Katheten $y-y_1, x-x_1$ und mit den Katheten m, 1 gilt die Proportion

$$(y - y_1) : (x - x_1) = m : 1 \quad \text{oder} \quad \frac{y - y_1}{x - x_1} = m.$$

Auflösung nach y ergibt die Punktsteigungsform.

4. Die Gleichung einer Geraden durch zwei Punkte $P_1 = P(x_1 \mid y_1)$ und $P_2 = P(x_2 \mid y_2)$ mit $x_1 \neq x_2$ ergibt die Zweipunkteform der Geradengleichung (Abb. 7.12).

Zweipunkteform

$$y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1 \quad \text{oder} \quad \frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$$

Die Proportion ergibt sich aus der Ähnlichkeit der rechtwinkligen Dreiecke mit den Hypotenusen $\overline{P_1P}$ und $\overline{P_1P_2}$.

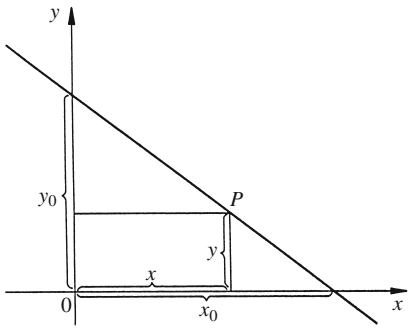


Abb. 7.13 Achsenabschnittsform der Geradengleichung

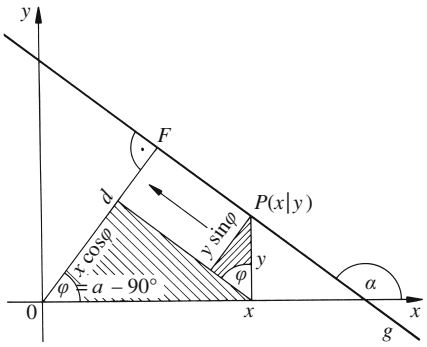


Abb. 7.14 Hesse'sche Normalform der Geradengleichung

5. Hat eine Gerade den Achsenabschnitt x_0 auf der x-Achse und den Achsenabschnitt y_0 auf der y-Achse, das heißt, die Gerade geht durch die Punkte $P_1(x_0|0)$ und $P_2(0|y_0)$, und gilt $x_0 \neq 0$ und $y_0 \neq 0$, dann lautet die Gleichung der Gerade $\frac{x}{x_0} + \frac{y}{y_0} = 1$. Dies ist die Achsenabschnittsform der Geradengleichung (Abb. 7.13).

Achsenabschnittsform

$$\frac{x}{x_0} + \frac{y}{y_0} = 1$$

Aus der allgemeinen Geradengleichung $ax + by + c = 0$ ergibt sich die Achsenabschnittsform durch Division durch $-c \neq 0$.

6. Die Hesse-Form oder Hesse'sche Normalform der Geradengleichung (nach dem deutschen Mathematiker Ludwig Otto Hesse, 1811–1874) lautet $x\cos\varphi + y\sin\varphi - d = 0$. Dabei ist $d \geq 0$ der Abstand des Koordinatenursprungs O von der Geraden g, also die Länge des Lotes von O auf die Gerade g (Fußpunkt F), und φ mit $0 \leq \varphi < 2\pi$ der Winkel zwischen der positiven x-Achse und dem Lot $\overline{OF}$ (Abb. 7.14).

Hesse'sche Normalform

$$x\cos\varphi + y\sin\varphi - d = 0$$

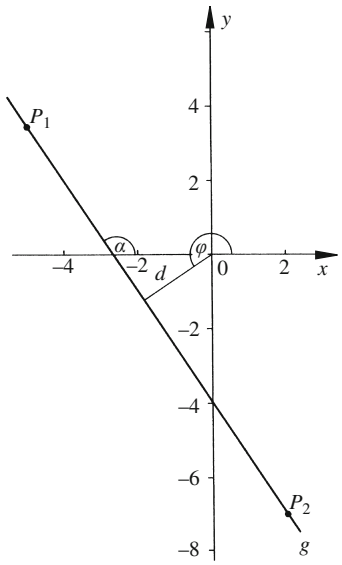


Abb. 7.15 Gerade durch die Punkte $P_1(-5|3{,}5)$ und $P_2(2|-7)$

Man kann die Hesse'sche Normalform aus der allgemeinen Geradengleichung $ax + by + c = 0$ durch Multiplikation mit dem Normierungsfaktor $\pm\frac{1}{\sqrt{a_2+b_2}}$ herleiten. Das Vorzeichen des Normierungsfaktors muss entgegengesetzt zu dem von c gewählt werden.

Beispiel

Gesucht ist die Gerade durch die Punkte $P_1(-5|3{,}5)$ und $P_2(2|-7)$ (Abb. 7.15).

Zweipunkteform:

$$\frac{y-3{,}5}{x+5} = \frac{-7-3{,}5}{2+5}$$

Punktsteigungsform:

Da die rechte Seite der Zweipunkteform die Steigung m angibt, also $m = \frac{-7-3{,}5}{2+5} = -1{,}5$ folgt

$$y = -1{,}5(x+5) + 3{,}5$$

Hauptform:

Aus der Punktsteigungsform ergibt sich $y = -1{,}5x - 7{,}5 + 3{,}5$ und somit

$$y = -1{,}5x - 4$$

Achsenabschnittsform:

Aus der Hauptform folgt $y = -4$ für $x = 0$, der y-Achsenabschnitt ist also $y_0 = -4$. Setzt man in die Hauptform $y = 0$ ein, so ergibt sich $1{,}5x = -4$, also

$x = -\frac{8}{3}$, der x-Achsenabschnitt ist somit $x_0 = -\frac{8}{3}$. Die Achsenabschnittsform lautet daher

$$\frac{x}{-\frac{8}{3}} + \frac{y}{-4} = 1$$

Man findet die Achsenabschnittsform auch direkt, indem man in der Hauptform durch Division durch -4 das Absolutglied zu 1 macht.

Hesse'sche Normalform:

Durch Umstellung der Hauptform ergibt sich $-1{,}5x - y - 4 = 0$. Durch Vergleich mit der allgemeinen Geradengleichung $ax + by + c = 0$ erhält man $a = -1{,}5$, $b = -1$, $c = -4$. Man erhält die Hesse'sche Normalform, indem man die Gleichung $-1{,}5x - y - 4 = 0$ durch $+\sqrt{a^2 + b^2} = +\sqrt{(-1{,}5)^2 + (-1)^2} = +\sqrt{3{,}25}$ dividiert:

$$\frac{1{,}5}{\sqrt{3{,}25}}x - \frac{1}{\sqrt{3{,}25}}y - \frac{4}{\sqrt{3{,}25}} = 0 \quad \blacktriangleleft$$

7.2.2 Abstände

Mit Hilfe der Hesse'schen Normalform der Geradengleichung lässt sich der Abstand zwischen einem Punkt und einer Geraden oder zwischen zwei parallelen Geraden berechnen. Zunächst werden jedoch Formeln zur Berechnung des Abstandes zwischen zwei Punkten hergeleitet.

1. *Punkt–Punkt*
   Der Abstand zweier Punkte P_1 und P_2 ist die Länge $|\overline{P_1P_2}|$ der Verbindungsstrecke $\overline{P_1P_2}$.
   Sind die Punkte im kartesischen Koordinatensystem dargestellt, also $P_1 = P_1(x_1|y_1)$, $P_2 = P_2(x_2|y_2)$, dann gilt für den Abstand $d(P_1, P_2)$ von P_1 und P_2 nach dem Satz des Pythagoras

   $$d(P_1, P_2) = |\overline{P_1P_2}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

   Sind die Punkte in Polarkoordinaten dargestellt, also $P_1 = P_1(r_1|\varphi_1)$, $P_2 = P_2(r_2|\varphi_2)$, dann folgt aus dem Kosinussatz

   $$d(P_1, P_2) = |\overline{P_1P_2}| = \sqrt{r_1^2 + r_2^2 - 2r_1r_2 \cdot \cos(\varphi_1 - \varphi_2)}$$

2. *Gerade–Gerade*
   Sind g_1: $y = mx + n_1$ und g_2: $y = mx + n_2$ zwei parallele Geraden (parallele Geraden haben gleiche Steigung), so ermittelt man die Hesse'sche Normalform der Geraden:

   $$g_1: \quad x\cos\varphi + y\sin\varphi - d_1 = 0$$
   $$g_2: \quad x\cos\varphi + y\sin\varphi - d_2 = 0$$

   Für den Abstand l der parallelen Geraden g_1 und g_2 voneinander gilt dann

   $l = |d_1 - d_2|$ wenn die Geraden auf der gleichen Seite des Koordinatenursprungs liegen,
   $l = d_1 + d_2$ wenn die Geraden auf verschiedenen Seiten des Koordinatenursprungs liegen.

3. *Punkt–Gerade*
   Ist $P_1(x_1|y_1)$ ein Punkt und g_1: $y = mx + n$ eine Gerade, dann ermittelt man zunächst die Hesse'sche Normalform von g_1:

   $$g_1: \quad x\cos\varphi + y\sin\varphi - d_1 = 0$$

   Durch den Punkt P_1 legt man eine zu g_1 parallele Gerade g_2:

   $$g_2: \quad x\cos\varphi + y\sin\varphi - d_2 = 0$$

   Ist l der Abstand zwischen P_1 und g_1, so ist l auch der Abstand zwischen den Geraden g_1 und g_2, und es gilt

   $$g_2: \quad x\cos\varphi + y\sin\varphi - (d_1 \mp l) = 0$$

   Da P_1 auf g_2 liegt, erfüllen seine Koordinaten die Geradengleichung

   $$x_1\cos\varphi + y_1\sin\varphi - (d_1 \mp l) = 0,$$

   woraus sich für den Abstand l ergibt

   $$l = |x_1\cos\varphi + y_1\sin\varphi - d_1|$$

Beispiele

1. Gegeben: Die Punkte $P_1(3|4)$ und $P_2(-2|6)$
   Gesucht: Der Abstand $d(P_1, P_2)$ von P_1 und P_2
   Es gilt

   $$\begin{aligned} d(P_1, P_2) &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ &= \sqrt{(-2 - 3)^2 + (6 - 4)^2} \\ &= \sqrt{5^2 + 2^2} = \sqrt{29} = 5{,}3851\ldots \end{aligned}$$

2. Gegeben: Die beiden parallelen Geraden

   $$g_1: \ 2x - 4y + 7 = 0, \quad g_2: \ -3x + 6y + 30 = 0$$

   Gesucht: Der Abstand l der beiden Geraden
   Hesse'sche Normalform von g_1: $-\frac{2}{\sqrt{20}}x + \frac{4}{\sqrt{20}}y - \frac{7}{\sqrt{20}} = 0$ (durch Multiplikation der allgemeinen Geradengleichung mit dem Normierungsfaktor $\frac{1}{\sqrt{a^2+b^2}} = -\frac{1}{\sqrt{2^2+(-4)^2}} = -\frac{1}{\sqrt{20}}$)
   Hesse'sche Normalform von g_2:
   $\frac{2}{\sqrt{20}}x - \frac{4}{\sqrt{20}}y - \frac{20}{\sqrt{20}} = 0$
   Entgegengesetzte Vorzeichen der x- und y-Glieder $\Rightarrow$ die Geraden liegen auf verschiedenen Seiten des Koordinatenursprungs
   Somit gilt für den Abstand l von g_1 und g_2:

   $$\begin{aligned} l &= d_1 + d_2 \\ &= \frac{7}{\sqrt{20}} + \frac{20}{\sqrt{20}} = \frac{27}{\sqrt{20}} = \frac{27}{2\sqrt{5}} = \frac{27 \cdot \sqrt{5}}{10} \end{aligned}$$

3. Gegeben: Punkt $P_1(5|10)$ und Gerade g_1: $3x - 4y + 10 = 0$
   Gesucht: Der Abstand l des Punktes P_1 von der Geraden g_1
   Hesse'sche Normalform von g_1: $-\frac{3}{5}x + \frac{4}{5}y - 2 = 0$ (durch Multiplikation der allgemeinen Geradengleichung mit dem Normierungsfaktor $\frac{1}{\sqrt{a^2+b^2}} = -\frac{1}{\sqrt{3^2+(-4)^2}} = -\frac{1}{5}$)
   Durch Einsetzen der Koordinaten von P_1 erhält man den gesuchten Abstand:

$$l = \left|-\frac{3}{5}\cdot 5 + \frac{4}{5}\cdot 10 - 2\right| = |-3 + 8 - 2| = 3 \quad \blacktriangleleft$$

7.3 Kreise

7.3.1 Kreisgleichungen

Der Kreis ist der geometrische Ort aller Punkte der Ebene, die von einem festen Punkt M (Mittelpunkt des Kreises) einen konstanten Abstand r (Radius des Kreises) haben (vgl. Abschn. 3.9).

Für einen Kreis gibt es verschiedene Gleichungsformen.

1. Liegt der Mittelpunkt eines Kreises mit dem Radius r im Koordinatenursprung, dann lautet die Gleichung des Kreises in kartesischen Koordinaten $x^2 + y^2 = r^2$. Dabei sind x und y die Koordinaten eines beliebigen Punktes $P(x|y)$ des Kreises. Die Gleichung ergibt sich nach dem Satz des Pythagoras (Abb. 7.16).

Mittelpunkt im Ursprung

$$x^2 + y^2 = r^2$$

2. Hat der Mittelpunkt allgemeiner die Koordinaten x_m und y_m, also $M = M(x_m|y_m)$, dann ergibt sich die Mittelpunktsform oder Hauptform der Kreisgleichung (Abb. 7.17).

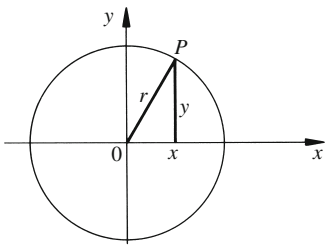


Abb. 7.16 Kreisgleichung $x^2 + y^2 = r^2$

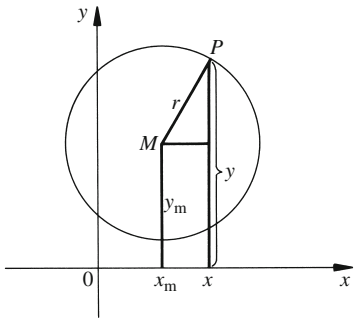


Abb. 7.17 Kreisgleichung $(x - x_m)^2 + (y - y_m)^2 = r^2$

Mittelpunktsform oder Hauptform

$$(x - x_m)^2 + (y - y_m)^2 = r^2$$

3. Löst man in der Mittelpunktsform die Klammern auf, dann ergibt sich die allgemeine Form der Kreisgleichung.

Allgemeine Form

$$x^2 + y^2 + 2ax + 2by + c = 0$$

Hierin bedeuten $a = -x_m$, $b = -y_m$, $c = x_m^2 + y_m^2 - r^2$. Aus der letzten Gleichung folgt $a^2 + b^2 - c = r^2 > 0$ als Bedingung dafür, dass es sich bei einer Gleichung der allgemeinen Form wirklich um eine Kreisgleichung handelt (für $c > a^2 + b^2$ liefert die Gleichung keine reelle Kurve, für $c = a^2 + b^2$ ergibt sich ein einziger Punkt $M(x_m|y_m)$).

4. Werden die beiden Koordinaten x und y jeweils als Funktion einer Hilfsvariablen t angegeben, so erhält man die Parameterdarstellung des Kreises mit dem Radius r und dem Mittelpunkt $M(x_m|y_m)$ (vgl. Abschn. 5.1.2).

Parameterdarstellung

$$x = x_m + r\cos t, \quad y = y_m + r\sin t, \quad 0 \le t < 2\pi$$

Beispiele

1. Welches geometrische Objekt beschreibt die Gleichung

$$1{,}5x^2 + 1{,}5y^2 + 3x - 6y + 4{,}5 = 0?$$

   Lösung:
   Division durch 1,5 ergibt $x^2 + y^2 + 2x - 4y + 3 = 0$, eine Kreisgleichung in allgemeiner Form.

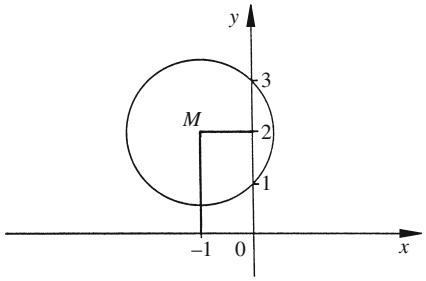


Abb. 7.18 Kreis mit der Gleichung $(x+1)^2+(y-2)^2=2$

Dabei ist $a=-x_m=1$, $b=-y_m=-2$, $c=3$. Die Bedingung $a^2+b^2-c>0$ ist erfüllt, denn $1+4-3=2>0$.
Die Koordinaten des Kreismittelpunktes sind $x_m=-1$, $y_m=2$, der Radius ist $r=\sqrt{a^2+b^2-c}=\sqrt{2}$.
Die Mittelpunktsform (Hauptform) dieses Kreises lautet somit $(x+1)^2+(y-2)^2=2$.
Die aus der gegebenen Gleichung abgeleitete Gleichung $x^2+y^2+2x-4y+3=0$ lässt sich auch ohne Benutzung der Formeln für die Mittelpunktskoordinaten und den Radius auf die Mittelpunktsform bringen, und zwar mit Hilfe von quadratischen Ergänzungen:

$$\begin{aligned}x^2+y^2+2x-4y+3&=0\\(x^2+2x)+(y^2-4y)&=-3\\(x^2+2x+1)+(y^2-4y+4)&=1+4-3\\(x+1)^2+(y-2)^2&=2\end{aligned}$$

(Abb. 7.18)

2. Welches geometrische Objekt beschreibt die Gleichung

$$1{,}5x^2+1{,}5y^2+3x-6y+9=0?$$

Lösung:
Die entsprechenden Rechnungen wie bei Beispiel 1 ergeben $(x+1)^2+(y-)^2=-1$.
Dies ist eine unerfüllbare Gleichung, denn der Term auf der linken Seite des Gleichheitszeichens ist als Summe zweier Quadrate nicht negativ (ein Quadrat ist nicht negativ, also ist auch eine Summe von Quadraten nicht negativ), während die rechte Seite gleich -1, also negativ ist. Man bestätigt, dass die Bedingung $a^2+b^2-c>0$ nicht erfüllt ist.

3. Welches geometrische Objekt beschreibt die Gleichung

$$1{,}5x^2+1{,}5y^2+3x-6y+7{,}5=0?$$

Lösung:
Man berechnet $a^2+b^2-c=1+4-5=0$ in der durch 1,5 dividierten Gleichung und somit $(x+1)^2+(y-2)^2=0$.
Dies ist die Gleichung eines entarteten Kreises, also eines Kreises mit dem Radius $r=0$. Die Gleichung wird nur von einem Koordinatenpaar, den Koordinaten des Mittelpunktes $M(-1|2)$, erfüllt. ◀

7.3.2 Berechnung von Kreisen

Ein Kreis ist festgelegt durch den Mittelpunkt und einen weiteren Punkt oder durch drei Punkte (die nicht alle auf einer Geraden liegen).

Die Gleichung eines Kreises kann also berechnet werden, wenn drei Punkte der Kreisperipherie gegeben sind oder der Mittelpunkt und ein Punkt der Peripherie.

Berechnung von Kreisen:

1. Gegeben: Mittelpunkt $M(x_m|y_m)$, Punkt $P_1(x_1|y_1)$
Gesucht: Kreis mit dem Mittelpunkt M durch den Punkt P_1
Der Radius r des gesuchten Kreises ergibt sich durch Einsetzen der Koordinaten des Punktes P_1 in die Mittelpunktsform der Kreisgleichung:

$$r=\sqrt{(x_1-x_m)^2+(y_1-y_m)^2}$$

Kreisgleichung somit:

$$(x-x_m)^2+(y-y_m)^2=(x_1-x_m)^2+(y_1-y_m)^2$$

2. Gegeben: Punkte $P_1(x_1|y_1)$, $P_2(x_2|y_2)$, $P_3(x_3|y_3)$
Gesucht: Kreis durch die Punkte P_1, P_2, P_3
Bestimmung der Koordinaten x_m, y_m des Mittelpunktes:
Einsetzen der Koordinaten der Punkte in die Mittelpunktsform der Kreisgleichung:

$$\begin{aligned}(x_1-x_m)^2+(y_1-y_m)^2&=(x_2-x_m)^2+(y_2-y_m)^2\\&=(x_3-x_m)^2+(y_3-y_m)^2=r^2\end{aligned}$$

Daraus erhält man ein lineares Gleichungssystem zur Bestimmung von x_m und y_m:

$$\begin{aligned}2(x_2-x_1)\cdot x_m+2(y_2-y_1)\cdot y_m&=x_2^2-x_1^2+y_2^2-y_1^2\\2(x_3-x_1)\cdot x_m+2(y_3-y_1)\cdot y_m&=x_3^2-x_1^2+y_3^2-y_1^2\end{aligned}$$

Bestimmung des Radius r:
Einsetzen von x_m, y_m und der Koordinaten eines Punktes in die Mittelpunktsform der Kreisgleichung:

$$r=\sqrt{(x_1-x_m)^2+(y_1-y_m)^2}$$

Einsetzen von x_m, y_m und r in die Mittelpunktsform der Kreisgleichung ergibt die Gleichung des gesuchten Kreises.

3. Gegeben: Mittelpunkt $M(-2|-1)$, Punkt $P_1(4|3)$
Gesucht: Kreis mit dem Mittelpunkt M durch den Punkt P_1
Berechnung des Radius:

$$r=\sqrt{(4-(-2))^2+(3-(-1))^2}=\sqrt{52}$$

Der gesuchte Kreis hat die Gleichung

$$(x+2)^2+(y+1)^2=52.$$

4. Gegeben: Punkte $P_1(6|7)$, $P_2(2|9)$, $P_3(-1|0)$
Gesucht: Kreis durch die Punkte P_1, P_2, P_3
Bestimmung der Koordinaten x_m, y_m des Mittelpunktes:

$$2(2-6)x_m + 2(9-7)y_m = 2^2 - 6^2 + 9^2 - 7^2$$
$$2(-1-6)x_m + 2(0-7)y_m = (-1)^2 - 6^2 + 0^2 - 7^2$$
$$\Rightarrow \quad \left.\begin{array}{r} -8x_m + 4y_m = 0 \\ -14x_m - 14y_m = -84 \end{array}\right\}$$
$$\Rightarrow \quad \left.\begin{array}{r} -2x_m + y_m = 0 \\ 2x_m + 2y_m = 12 \end{array}\right\}$$
$$\Rightarrow \quad y_m = 4 \quad \Rightarrow \quad x_m = 2$$

Bestimmung des Radius r:

$$r = \sqrt{(6-2)^2 + (7-4)^2} = \sqrt{25} = 5$$

Der gesuchte Kreis hat somit die Gleichung

$$(x-2)^2 + (y-4)^2 = 5^2.$$

7.4 Kugeln

Eine Kugel ist der geometrische Ort aller Punkte des Raumes, die von einem festen Punkt M (Mittelpunkt der Kugel) einen konstanten Abstand r (Radius der Kugel) haben (vgl. Abschn. 4.6).

Für eine Kugel gibt es verschiedene Gleichungsformen.

Liegt der Mittelpunkt einer Kugel mit dem Radius r im Ursprung eines (dreidimensionalen) kartesischen Koordinatensystems, dann lautet die Gleichung der Kugel $x^2 + y^2 + z^2 = r^2$. Dabei sind x, y und z die Koordinaten eines beliebigen Punktes $P(x|y|z)$ der Kugel (Kugeloberfläche) (Abb. 7.19).

Mittelpunkt im Ursprung

$$x^2 + y^2 + z^2 = r^2$$

Hat der Mittelpunkt allgemeiner die Koordinaten x_m, y_m und z_m, also $M = (x_m|y_m|z_m)$, dann ergibt sich die Mittelpunktsform oder Hauptform der Kugelgleichung.

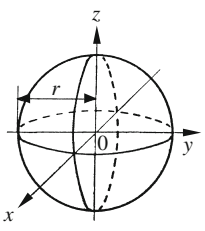


Abb. 7.19 Kugel mit der Gleichung $x^2 + y^2 + z^2 = r^2$

Mittelpunktsform oder Hauptform

$$(x-x_m)^2 + (y-y_m)^2 + (z-z_m)^2 = r^2$$

Eine Kugel ist festgelegt durch den Mittelpunkt und einen weiteren Punkt oder durch vier Punkte (die nicht alle in einer Ebene liegen).

Beispiele

1. Gegeben: Mittelpunkt im Koordinatenursprung, also $M = M(0|0|0)$, Punkt $P_1(4|3|1)$
Gesucht: Kugel mit dem Mittelpunkt M durch den Punkt P_1
Berechnung des Radius:

$$r = \sqrt{4^2 + 3^2 + 1^2} = \sqrt{26}$$

Die gesuchte Kugel hat die Gleichung

$$x^2 + y^2 + z^2 = 26$$

2. Gegeben: Mittelpunkt $M(2|-1|1)$, Punkt $P_1(0|4|-3)$
Gesucht: Kugel mit dem Mittelpunkt M durch den Punkt P_1
Berechnung des Radius:

$$r = \sqrt{(0-2)^2 + (4-(-1))^2 + (-3-1)^2} = \sqrt{45}$$

Die gesuchte Kugel hat die Gleichung

$$(x-2)^2 + (y+1)^2 + (z-1)^2 = 45. \quad \blacktriangleleft$$

7.5 Kegelschnitte

Ein Kegelschnitt ist die Schnittfigur einer Ebene und des Mantels eines geraden Doppelkreiskegels.

Ein gerader Kreiskegel entsteht durch Rotation einer Geraden (die Erzeugende oder Mantellinie) in einem festen Punkt (der Spitze) um eine vertikale Achse, wobei sich die rotierende Gerade entlang eines Kreises bewegt (also mit einem Kreis als Leitkurve), der in einer Ebene senkrecht zur Rotationsachse liegt (vgl. Abschn. 4.4).

Ein gerader Doppelkreiskegel besteht aus zwei geraden Kreiskegeln, die Spitze auf Spitze stehen und die gleiche Rotationsachse haben. Schneidet man einen geraden Doppelkreiskegel mit einer nicht durch die (gemeinsame) Spitze S gehenden Ebene E, dann entsteht als Kurve ein Kegelschnitt. Abhängig von der Lage der Ebene E zum Doppelkegel erhält man verschiedene Kurven.

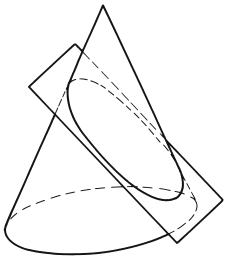


Abb. 7.20 Kegelschnitt Ellipse

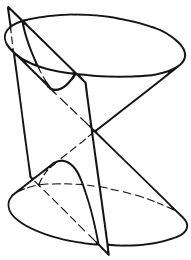


Abb. 7.21 Kegelschnitt Hyperbel

Kreis

Liegt die Ebene senkrecht zur Kegelachse (Rotationsachse), so schneidet sie aus der Mantelfläche des Kegels einen Kreis heraus.

Ellipse

Ist die Neigung der Ebene so, dass sie nur eine Hälfte des Doppelkegels schneidet und dass sie nicht parallel zu einer Mantellinie verläuft, so wird eine Ellipse ausgeschnitten (Abb. 7.20).

Parabel

Verläuft die Ebene parallel zu einer Mantellinie, so schneidet sie aus der Mantelfläche eine Parabel heraus.

Hyperbel

Trifft die Ebene beide Hälften des Doppelkegels (zum Beispiel wenn sie parallel zur Kegelachse steht), dann ist die Schnittfigur eine Hyperbel (es werden zwei Kurven ausgeschnitten, die beiden Äste einer Hyperbel) (Abb. 7.21).

Die Kegelschnitte lassen sich bezüglich der Lage der Ebene E zu den Mantellinien des Doppelkegels charakterisieren:

Beim Kreis und bei der Ellipse ist die Ebene zu keiner der Mantellinien parallel, bei der Parabel ist die Ebene zu einer Mantellinie parallel, und bei der Hyperbel ist die Ebene zu zwei Mantellinien des Doppelkegels parallel.

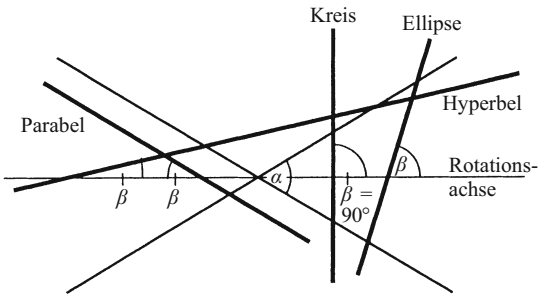


Abb. 7.22 Beziehung des Öffnungswinkels α des Kegels zum Neigungswinkel β der Schnittebene zur Rotationsachse

Die Kegelschnitte lassen sich auch durch die Beziehung des Öffnungswinkels α des Kegels zum Neigungswinkel β der Schnittebene E zur Rotationsachse beschreiben (Abb. 7.22):

$$\begin{aligned} &\text{Kreis:} && \beta = 90° \\ &\text{Ellipse:} && \frac{\alpha}{2} < \beta < 90° \\ &\text{Parabel:} && \beta = \frac{\alpha}{2} \\ &\text{Hyperbel:} && 0 \leq \beta < \frac{\alpha}{2} \end{aligned}$$

Der Kreis ist bezüglich der verschiedenen Lagen von Ebene und Doppelkegel ein Spezialfall der Ellipse.

Kreis und Ellipse sind beschränkt, nicht jedoch Parabel und Hyperbel. Die Parabel besteht aus einem einzigen Ast (sie ist also zusammenhängend), während die Hyperbel zwei getrennte symmetrische Äste besitzt.

Falls die Ebene E durch die Kegelspitze S geht, dann besteht die Schnittmenge entweder nur aus einem Punkt (dem Punkt S) oder aus einer oder zwei durch S gehenden Geraden (Mantellinien). Solche Schnittmengen heißen entartete Kegelschnitte.

Die nahe Verwandtschaft der Kegelschnitte zeigt sich auch in ihren Gleichungen. Jeder Kegelschnitt ist der Graph einer Funktion, die als Funktionsgleichung eine Gleichung zweiten Grades in x und y hat. In einer solchen Gleichung kommen x und y nur linear und quadratisch vor. Die allgemeine Gleichung eines Kegelschnitts lautet

$$Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F = 0$$

Diese Gleichung enthält als Sonderfälle auch Gleichungen von Punkten, Geraden, Geradenpaaren und imaginären Kurven.

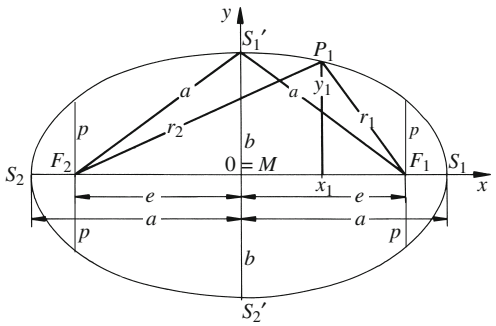


Abb. 7.23 Bezeichnungen für die Ellipse

Beispiele

1. $A = -1, B = C = D = 0, E = 1, F = 0$
   $\Rightarrow y = x^2$ Gleichung der Normalparabel
2. $A = 1, B = 0, C = 1, D = E = 0, F = -r^2$
   $\Rightarrow x^2 + y^2 = r^2$ Mittelpunktsform der Gleichung eines Kreises mit dem Mittelpunkt im Koordinatenursprung
3. $A = \frac{1}{a^2}, B = 0, C = \frac{1}{b^2}, D = E = 0, F = -1$
   $\Rightarrow \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ Mittelpunktsform der Gleichung einer Ellipse mit dem Mittelpunkt im Koordinatenursprung
4. $A = \frac{1}{a^2}, B = 0, C = -\frac{1}{b^2}, D = E = 0, F = -1$
   $\Rightarrow \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ Mittelpunktsform der Gleichung einer Hyperbel mit dem Mittelpunkt im Koordinatenursprung
5. $A = B = C = 0, D = -1, E = 1, F = 0$
   $\Rightarrow y = x$ Gleichung der Winkelhalbierenden (Gerade)

◀

7.5.1 Ellipsen

Eine Ellipse ist der geometrische Ort aller Punkte einer Ebene, für die die Summe der Abstände von zwei festen Punkten F_1 und F_2 konstant ist. Die Punkte F_1 und F_2 heißen Brennpunkte der Ellipse (Abb. 7.23).

Bezeichnet man den Abstand eines beliebigen Punktes P_1 der Ellipse zu F_1 mit r_1 und den Abstand von P_1 zu F_2 mit r_2, also $|\overline{P_1F_1}| = r_1$, $|\overline{P_1F_2}| = r_2$, dann gilt $r_1 + r_2 = 2a$ mit einer Konstanten a.

Bezeichnungen

$M(0\|0)$	Mittelpunkt
$F_1(e\|0), F_2(-e\|0)$	Brennpunkte
$S_1(a\|0), S_2(-a\|0)$	Hauptscheitelpunkte
$S_1'(0\|b), S_2'(0\|-b)$	Nebenscheitelpunkte
$\overline{S_1S_2}$	Hauptachse
$\overline{S_1'S_2'}$	Nebenachse
$\|\overline{S_1S_2}\| = 2a$	Länge der Hauptachse
$\|\overline{S_1'S_2'}\| = 2b$	Länge der Nebenachse ($b < a$)
$\|\overline{MF_1}\| = \|\overline{MF_2}\| = e$	Abstand der Brennpunkte vom Mittelpunkt
$p = \frac{b^2}{a}$	Halbparameter (die halbe Länge einer parallel zur Nebenachse gezogenen Sehne durch einen Brennpunkt)
$P_1(x_1\|y_1)$	beliebiger Punkt der Ellipse
$\|\overline{P_1F_1}\| = r_1, \|\overline{P_1F_2}\| = r_2$	Abstand von P_1 zu den Brennpunkten

Eigenschaften

$r_1 + r_2 = 2a$	Summe der Abstände ist konstant
$e^2 + b^2 = a^2$	gilt nach dem Satz des Pythagoras
$e = \sqrt{a^2 - b^2} > 0$	heißt lineare Exzentrizität der Ellipse
$\varepsilon = \frac{e}{a} < 1$	heißt numerische Exzentrizität der Ellipse

Bemerkungen

Eine der drei Größen a, b, e kann wegen $e^2 + b^2 = a^2$ aus den beiden anderen berechnet werden.

Im Falle $a = b$ entartet die Ellipse zu einem Kreis. Die beiden Brennpunkte F_1, F_2 fallen dann mit dem Kreismittelpunkt zusammen.

Ellipsengleichungen

1. Fallen die Koordinatenachsen mit den Ellipsenachsen zusammen und ist der Koordinatenursprung der Mittelpunkt der Ellipse, dann lautet die Gleichung der Ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Dies ist die Normalform der Ellipsengleichung.

Normalform

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

2. Ist $M(x_m|y_m)$ der Mittelpunkt der Ellipse, und sind die Ellipsenachsen parallel zu den Koordinatenachsen, dann erhält man die Mittelpunktsform der Ellipsengleichung.

Mittelpunktsform

$$\frac{(x - x_m)^2}{a^2} + \frac{(y - y_m)^2}{b^2} = 1$$

3. Werden die beiden Koordinaten x und y jeweils als Funktion einer Hilfsvariablen t angegeben, so erhält man die Parameterdarstellung einer Ellipse, deren Achsen mit den Koordinatenachsen zusammenfallen (vgl. Abschn. 5.1.2).

Parameterdarstellung

$$x = a\cos t, \quad y = b\sin t, \quad 0 \le t < 2\pi$$

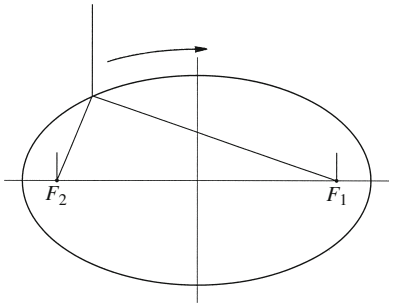


Abb. 7.24 Gärtnerkonstruktion einer Ellipse

Für die Fläche A und den Umfang u einer Ellipse gilt

Ellipsenfläche $A = \pi ab$

Ellipsenumfang $u \approx \pi[1{,}5(a+b) - \sqrt{ab}]$

Der Wert für den Umfang ist nur eine Näherung, eine exakte Formel gibt es nicht.

Beispiele

1. Gegeben: Ellipsengleichung $\frac{x^2}{6{,}25} + \frac{y^2}{4} = 1$
   Gesucht: Länge der Achsen, Brennpunkte, numerische Exzentrizität
   Länge der Hauptachse $= 2a = 2 \cdot 2{,}5 = 5$ (denn $a = \sqrt{6{,}25} = 2{,}5$)
   Länge der Nebenachse $= 2b = 2 \cdot 2 = 4$ (denn $b = \sqrt{4} = 2$)
   Berechnung der Brennpunkte:
   $e^2 = a^2 - b^2 = 6{,}25 - 4 = 2{,}25 = 1{,}5^2 \Rightarrow F_1(1{,}5|0)$, $F_2(-1{,}5|0)$
   Numerische Exzentrizität: $\varepsilon = \frac{e}{a} = \frac{1{,}5}{2{,}5} = 0{,}6$
2. Gegeben: Brennpunkte $F_1(4|0)$, $F_2(-4|0)$, halbe Länge der Hauptachse $a = 5$
   Gesucht: Ellipsengleichung, numerische Exzentrizität
   Berechnung der halben Länge der Nebenachse:
   $b^2 = a^2 - e^2 = 25 - 16 = 9 \Rightarrow b = 3$
   Ellipsengleichung: $\frac{x^2}{25} + \frac{y^2}{9} = 1$
   Numerische Exzentrizität: $\varepsilon = \frac{e}{a} = \frac{4}{5} = 0{,}8$ ◀

Es gibt etliche Konstruktionsmöglichkeiten für eine Ellipse. Eine davon ist die sogenannte Faden- oder Gärtnerkonstruktion (Abb. 7.24):

Die Endpunkte eines Fadens der Länge $2a$ werden mit Hilfe zweier Pfähle an zwei Punkten F_1 und F_2 fest verankert, und zwar so, dass der Faden nicht straff ist. Ein dritter Pfahl, der den Faden spannt und ringsum geführt wird, bewegt sich dann auf einer Ellipse.

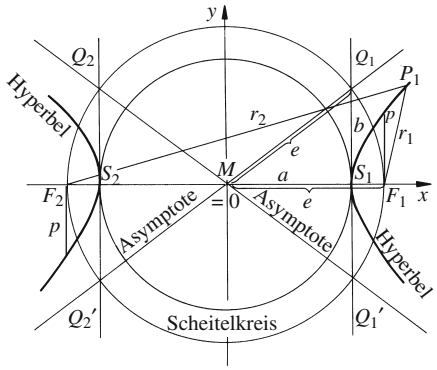


Abb. 7.25 Bezeichnungen für die Hyperbel

7.5.2 Hyperbeln

Eine Hyperbel ist der geometrische Ort aller Punkte einer Ebene, für die der Betrag der Differenz der Abstände von zwei festen Punkten F_1 und F_2 konstant ist. Die Punkte F_1 und F_2 heißen Brennpunkte der Hyperbel (Abb. 7.25).

Bezeichnet man den Abstand eines beliebigen Punktes P_1 der Hyperbel zu F_1 mit r_1 und den Abstand von P_1 zu F_2 mit r_2, also $|\overline{P_1F_1}| = r_1$, $|\overline{P_1F_2}| = r_2$, dann gilt $|r_1 - r_2| = 2a$ mit einer Konstanten a.

Die Hyperbel ist nicht zusammenhängend, sie besteht aus zwei getrennten symmetrischen Ästen (die Hyperbel hat also auch keinen endlichen Flächeninhalt). Sie besitzt zwei Asymptoten.

Bezeichnungen

$M(0\|0)$	Mittelpunkt
$F_1(e\|0)$, $F_2(-e\|0)$	Brennpunkte
$S_1(a\|0)$, $S_2(-a\|0)$	Scheitelpunkte
$\overline{S_1S_2}$	Hauptachse (Hyperbelachse)
$\overline{S_1S_2} = 2a$	Länge der Hauptachse
$\|\overline{MF_1}\| = \|\overline{MF_2}\| = e$	Abstand der Brennpunkte vom Mittelpunkt
Q_1, Q_1', Q_2, Q_2'	Schnittpunkte der Asymptoten mit den Senkrechten zur Hauptachse durch die Scheitelpunkte
$\|\overline{S_1Q_1}\| = \|\overline{S_1Q_1'}\| = \|\overline{S_2Q_2}\| = \|\overline{S_2Q_2'}\| = b$	Abstand der Schnittpunkte zu den Scheitelpunkten
$p = \frac{b^2}{a}$	Halbparameter (die halbe Länge einer senkrecht zur Hauptachse gezogenen Sehne durch einen Brennpunkt)
$P_1(x_1\|y_1)$	beliebiger Punkt der Hyperbel
$\|\overline{P_1F_1}\| = r_1$, $\|\overline{P_1F_2}\| = r_2$	Abstand von P_1 zu den Brennpunkten

Eigenschaften

$\|r_1 - r_2\| = 2a$	Betragsdifferenz der Abstände ist konstant
$a^2 + b^2 = e^2$	gilt nach dem Satz des Pythagoras

$e = \sqrt{a^2 + b^2} > 0$ heißt lineare Exzentrizität der Hyperbel

$\varepsilon = \frac{e}{a} > 1$ heißt numerische Exzentrizität der Hyperbel

Bemerkung

Eine der drei Größen a, b, e kann wegen $a^2 + b^2 = e^2$ aus den beiden anderen berechnet werden.

Hyperbelgleichungen

1. Scheitelpunkte auf der x-Achse, Mittelpunkt im Koordinatenursprung:

Normalform

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

Beide Koordinatenachsen sind Symmetrieachsen der Hyperbel. Die Hyperbel ist nach rechts und nach links geöffnet. Diese Gleichung nennt man auch die Normalform der Hyperbelgleichung.

Gleichungen der Asymptoten

$$y = \pm\frac{b}{a}x$$

Nur im Falle $a = b$ stehen die Asymptoten senkrecht aufeinander. Solche Hyperbeln heißen gleichseitige Hyperbeln.

2. Hauptachse parallel zur x-Achse, Mittelpunkt $M(x_m \mid y_m)$:

Mittelpunktsform

$$\frac{(x - x_m)^2}{a^2} - \frac{(y - y_m)^2}{b^2} = 1$$

Die Hyperbel ist nach rechts und nach links geöffnet. Diese Gleichung heißt auch Mittelpunktsform der Hyperbelgleichung.

Gleichungen der Asymptoten

$$y = \pm\frac{b}{a}(x - x_m) + y_m$$

3. Koordinatenachsen als Asymptoten, Mittelpunkt im Koordinatenursprung:

$$x \cdot y = c \quad \text{oder} \quad y = \frac{c}{x} \quad (c \neq 0)$$

Für $c > 0$ ist die Winkelhalbierende $y = x$ die Hauptachse, die Hyperbeläste liegen im ersten und im dritten Quadranten. Im Falle $c < 0$ ist die Winkelhalbierende $y = -x$ die Hauptachse, die Hyperbeläste liegen im zweiten und im vierten Quadranten.

Gleichungen der Asymptoten

$$x = 0, \quad y = 0$$

Beispiele

1. Gegeben: Hyperbelgleichung $\frac{x^2}{16} - \frac{y^2}{20} = 1$
Gesucht: Brennpunkte, numerische Exzentrizität
Berechnung der Brennpunkte:

$$e^2 = a^2 + b^2$$
$$= 16 + 20 = 36 = 6^2 \quad \Rightarrow \quad F_1(6|0), F_2(-6|0)$$

Numerische Exzentrizität: $\varepsilon = \frac{e}{a} = \frac{6}{4} = 1{,}5$

2. Gegeben: Punkte $P_1(2|1)$, $P_2(3|2)$ einer Hyperbel
Gesucht: Normalform der Hyperbelgleichung durch P_1 und P_2
Berechnung von a und b:
Koordinaten von P_1 in die Normalform $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ der Hyperbelgleichung einsetzen:

$$\frac{4}{a^2} - \frac{1}{b^2} = 1 \quad \Rightarrow \quad 4b^2 - a^2 = a^2b^2$$

Kooordinaten von P_2 in die Normalform einsetzen:

$$\frac{9}{a^2} - \frac{4}{b^2} = 1 \quad \Rightarrow \quad 9b^2 4a^2 = a^2b^2$$

Gleichsetzen der Gleichungen ergibt

$$4b^2 - a^2 = 9b^2 - 4a^2 \quad \Rightarrow \quad 3a^2 = 5b^2$$
$$\Rightarrow \quad a^2 = \frac{5}{3}b^2$$

Durch Einsetzen errechnet man

$$4b^2 - \frac{5}{3}b^2 = \frac{5}{3}b^4 \quad \Rightarrow \quad \frac{7}{3}b^2 = \frac{5}{3}b^4$$
$$\Rightarrow \quad b^2 = \frac{7}{5}$$

Daraus ergibt sich schließlich

$$a^2 = \frac{5}{3} \cdot \frac{7}{5} = \frac{7}{3}$$

Die Hyperbelgleichung lautet

$$\frac{x^2}{\frac{7}{3}} - \frac{y^2}{\frac{7}{5}} = 1$$

◀

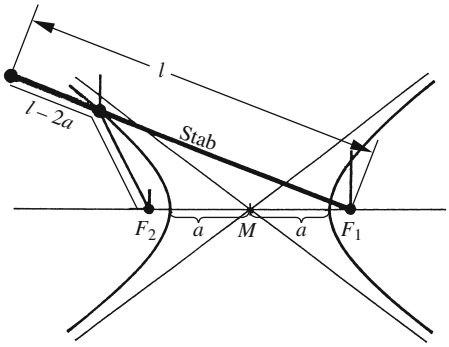


Abb. 7.26 Fadenkonstruktion einer Hyperbel

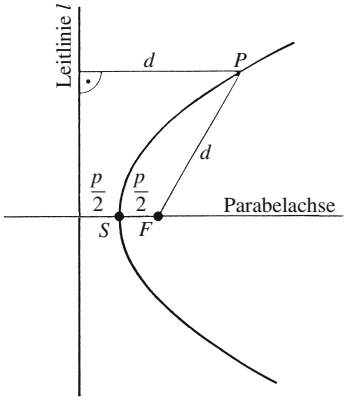


Abb. 7.27 Parabel mit Brennpunkt F und Scheitelpunkt S

Es gibt verschiedene Konstruktionsmöglichkeiten für eine Hyperbel. Eine davon ist die sogenannte Fadenkonstruktion:

In einem der beiden Brennpunkte wird ein Stab der Länge $l > 2a$ drehbar befestigt. Die Enden eines Fadens der Länge $l - 2a$ werden am freien Stabende und am anderen Brennpunkt befestigt. Mit einem Stift wird der Faden am Stab gestrafft. Wird der Stab um den Brennpunkt gedreht, dann beschreibt der Stift einen Teil eines Hyperbelastes (Abb. 7.26).

7.5.3 Parabeln

Eine Parabel ist der geometrische Ort aller Punkte einer Ebene, die von einem festen Punkt F (Brennpunkt) und einer festen Geraden l (Leitlinie) den gleichen Abstand besitzen.

Der Punkt, der in der Mitte zwischen dem Brennpunkt F und der Leitlinie l liegt, ist der Scheitelpunkt S. Die Gerade durch die Punkte F und S heißt Parabelachse. Sie ist Symmetrieachse für die Parabel und steht senkrecht auf der Leitlinie l. Der Abstand p des Brennpunkts F von der Leitlinie l heißt Parameter der Parabel.

Der Brennpunkt hat die Eigenschaft, alle innen an der Parabel reflektierten achsenparallelen Strahlen in sich zu vereinigen (Anwendung: Parabolspiegel) (Abb. 7.27).

Parabelgleichungen

1. x-Achse ist Parabelachse, Scheitelpunkt im Koordinatenursprung, Parabel nach rechts geöffnet:

Normalform

$$y^2 = 2px, \quad p > 0$$

Der Brennpunkt ist $F(\frac{p}{2}|0)$, die Gleichung der Leitlinie ist $x = -\frac{p}{2}$. Diese Gleichung nennt man auch die Normalform der Parabelgleichung.

2. Parabelachse parallel zur x-Achse, Scheitelpunkt $S(x_s \mid y_s)$, Parabel nach rechts geöffnet:

Scheitelpunktsform

$$(y - y_s)^2 = 2p(x - x_s), \quad p > 0$$

Der Brennpunkt ist $F(\frac{p}{2} + x_s|y_s)$, die Gleichung der Leitlinie ist $x = x_s - \frac{p}{2}$. Diese Gleichung heißt auch Scheitelpunktsform der Parabelgleichung.

1. y-Achse ist Parabelachse, Scheitelpunkt im Koordinatenursprung, Parabel nach oben geöffnet:

$$x^2 = 2py \quad \text{oder} \quad y = \frac{1}{2p}x^2 \quad (p > 0)$$

Der Brennpunkt ist $F(0|\frac{p}{2})$, die Gleichung der Leitlinie ist $y = -\frac{p}{2}$. Eine Parabel in dieser Lage ist der Graph einer quadratischen Funktion (vgl. Abschn. 5.4.3).

Beispiele

1. Gegeben: Parabelgleichung $y^2 = 6x$
   Gesucht: Brennpunkt, Gleichung der Leitlinie
   Parameter: $p = 3$
   Brennpunkt: $F(\frac{p}{2}|0) = F(\frac{3}{2}|0)$
   Gleichung der Leitlinie: $x = -\frac{p}{2} = -\frac{3}{2}$
2. Eine nach rechts offene Parabel hat die Gerade $x = -2$ als Tangente im Scheitelpunkt, geht durch den Punkt $P_1(2|7)$ und hat in diesem Punkt die Tangentensteigung $m_1 = \frac{1}{2}$. Wie lautet die Gleichung der Parabel?
   Lösung:
   Berechnung der Tangentengleichung durch Einsetzen in die Punktsteigungsform und Umrechnung auf Normalform:

$$\frac{y-7}{x-2} = \frac{1}{2} \quad \Rightarrow \quad y = \frac{1}{2}x + 6$$

Eine nach rechts geöffnete Parabel mit Scheitelpunkt $S(x_s|y_s)$ erfüllt die Gleichung $(y-y_s)^2 = 2p(x-x_s)$ und eine Tangente an diese Parabel auch die Gleichung $(y_1 - y_s)(y - y_s) = p(x + x_1 - 2x_s)$. Setzt man die Koordinaten des Punktes P_1 ein und bringt auch diese Tangentengleichung auf Normalform, so ergibt sich

$$(7 - y_s)(y - y_s) = p(x + 2 - 2 \cdot (-2))$$
$$\Rightarrow \quad y = \frac{p}{7 - y_s}x + \frac{6p}{7 - y_s} + y_s$$

Vergleich dieser Gleichung mit der obigen Normalform der Gleichung für dieselbe Tangente ergibt mit der Methode des Koeffizientenvergleichs die folgenden Bedingungen für die Koeffizienten:

$$m_1 = \frac{p}{7 - y_s} = \frac{1}{2} \quad \text{und} \quad n = \frac{6p}{7 - y_s} + y_s = 6$$

Dies ist ein System aus zwei Gleichungen zur Bestimmung der beiden Unbekannten p und y_s.
Setzt man zur Lösung des Systems zum Beispiel die Bedingung $\frac{p}{7-y_s} = \frac{1}{2}$ in die Gleichung für n ein, so folgt

$$6 \cdot \frac{1}{2} + y_s = 6 \quad \Rightarrow \quad y_s = 3$$

Durch Einsetzen von $y_s = 3$ in die Gleichung für m_1 folgt $\frac{p}{7-3} = \frac{1}{2} \Rightarrow p = 2$
Die gesuchte Parabelgleichung lautet

$$(y - 3)^2 = 4(x + 2). \qquad \blacktriangleleft$$

Es gibt einige Konstruktionsmöglichkeiten für eine Parabel. Eine davon ist die sogenannte Fadenkonstruktion (Abb. 7.28):

Ein rechtwinkliges Dreieck wird entlang der Leitlinie verschoben. Ein Faden mit der Länge der Kathete $\overline{AC}$ wird mit den Enden in A und dem Brennpunkt F befestigt. Mit einem Stift wird der Faden an der Kathete $\overline{AC}$ gestrafft. Gleitet das Dreieck entlang der Leitlinie, dann beschreibt der Stift ein Parabelstück.

7.5.4 Anwendungen

In diesem Abschnitt werden in einigen Beispielen verschiedene Anwendungen der Kegelschnitte aus Technik und Mathematik angegeben.

Beispiel 1

Ein parabelförmiger Brückenbogen (Achse vertikal und Parabel nach unten geöffnet) hat zwischen den in gleicher Höhe liegenden Lagern (Enden) des Bogens L und

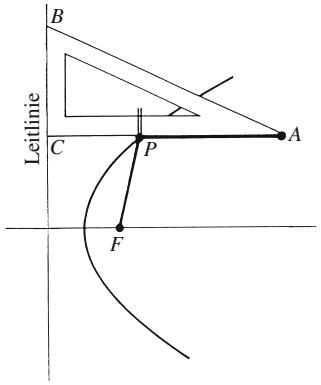


Abb. 7.28 Fadenkonstruktion einer Parabel

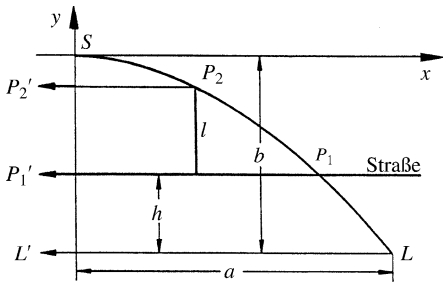


Abb. 7.29 Zu Beispiel 1

L' die Spannweite $2a = |\overline{LL'}| = 32\,\text{m}$. Die Scheitelhöhe (Höhe des Scheitelpunktes S über LL') beträgt $b = 10\,\text{m}$. Die horizontal verlaufende Straße liegt $h = 4\,\text{m}$ über LL' und schneidet den Brückenbogen in P_1 und P_1', den Befestigungspunkten des Straßenkörpers. Der Straßenkörper wird außer von einem Vertikalstab im Scheitelpunkt S (Länge $b - h = 6\,\text{m}$) noch von zwei weiteren Vertikalstäben gehalten, die in der Mitte des horizontalen Abstandes von S und P_1 sowie von S und P_1' in den Punkten P_2 und P_2' am Brückenbogen angebracht sind (Abb. 7.29).

Wie groß ist die Länge l dieser Vertikalstäbe? Wie groß sind $|\overline{P_1P_1'}|$ und $|\overline{P_2P_2'}|$?

Die Skizze veranschaulicht nur eine Hälfte der symmetrischen Straßenbrücke.

Zur Lösung der Aufgabe denkt man sich ein Koordinatenkreuz gelegt, so dass die Parabel des Brückenbogens die Gleichung $y = -px^2$ hat.

Setzt man die Koordinaten des Lagerpunktes L ein, so ergibt sich $-b = -pa^2 \Rightarrow p = \frac{b}{a^2}$

Der Befestigungspunkt P_1 hat nach Aufgabenstellung die Ordinate $y_1 = -(b - h)$. Mit Hilfe der Parabelgleichung $y_1 = -px_1^2$ erhält man seine Abszisse x_1 durch Auflösen

nach x_1 und Einsetzen von y_1 und p:

$$x_1 = \sqrt{\frac{y_1}{p}} = \sqrt{\frac{-(b-h)}{-\frac{b}{a^2}}} = a\sqrt{\frac{b-h}{b}}$$

Der Befestigungspunkt P_2 soll die Abszisse $x_2 = \frac{1}{2}x_1 = \frac{a}{2}\sqrt{\frac{b-h}{b}}$ haben, also ist seine Ordinate $y_2 = -px_2^2 = -\frac{b}{a^2}(\frac{a}{2}\sqrt{\frac{b-h}{b}})^2 = -\frac{b-h}{4}$.

Die gesuchte Vertikalstablänge l ist $l = y_2 - y_1 = \frac{3}{4}(b-h)$.

Die Strecken $|\overline{P_1P_1'}|$ und $|\overline{P_2P_2'}|$ haben die Längen

$$2x_1 = 2a\sqrt{\frac{b-h}{b}} \quad \text{und} \quad 2x_2 = a\sqrt{\frac{b-h}{b}}.$$

Mit den gegebenen Abmessungen ergibt sich für die gesuchten Längen

$$l = \frac{3}{4}(10-4) = 4{,}50\,\text{m},$$

$$2x_1 = 2 \cdot 16\sqrt{\frac{10-4}{10}} = 24{,}78\ldots\,\text{m},$$

$$2x_2 = 16\sqrt{\frac{10-4}{10}} = 12{,}39\ldots\,\text{m}.$$ ◀

Beispiel 2

Ein Stab mit den Drehlagern O und M und der Länge $|\overline{OM}| = r$ ist um O drehbar. Um M dreht sich ein zweiter Stab der Länge $|\overline{MP}| = s$ In der Ausgangsstellung liegen M und P als M_0 und rechts davon P_0 auf der Grundrichtungsachse OP_0. Während sich nun $\overline{OM}$ um den Winkel φ nach links dreht, dreht sich $\overline{MP}$ relativ zu $\overline{OM}$ um 2φ nach rechts (Abb. 7.30).

Auf welcher Kurve bewegt sich P bei fortgesetzter Drehung?

Die Lösung der Aufgabe erfolgt mit Hilfe der Parameterdarstellung der Ellipse: $x = a\cos\varphi$, $y = b\sin\varphi$

Man bestätigt: Aus $\cos\varphi = \frac{x}{a}$ und $\sin\varphi = \frac{y}{b}$ folgt durch Quadrieren und Addieren wegen $\sin^2\varphi + \cos^2\varphi = 1$ die Normalform $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ der Ellipsengleichung.

In der vorliegenden Aufgabe ist $a = r + s$ (halbe Länge der Hauptachse) und $b = r - s$ (halbe Länge der Nebenachse). Der Punkt P bildet mit A und B, den Schnittpunkten der Geraden OM mit dem Haupt- und dem Nebenscheitelkreis (Kreise um O mit den Radien a und b), ein rechtwinkliges Dreieck. Deshalb folgt $x = (r+s)\cos\varphi$, $y = (r-s)\sin\varphi$, die Parameterdarstellung der Ellipse, die als Bahnkurve von P gesucht ist. Ihre Normalform lautet $\frac{x^2}{(r+s)^2} + \frac{y^2}{(r-s)^2} = 1$. ◀

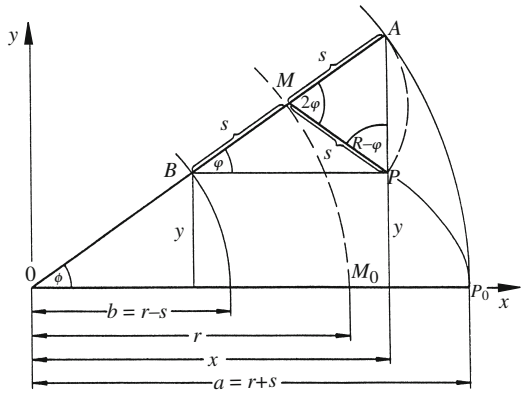


Abb. 7.30 Zu Beispiel 2

7.6 Vektoren

7.6.1 Definitionen

Eine gerichtete und orientierte Strecke bezeichnet man als Vektor. Ein Vektor ist durch drei Größen bestimmt: Richtung, Orientierung und Länge. Vektoren, die in diesen drei Größen übereinstimmen, sind gleich, unabhängig von ihrer Lage in der Ebene oder im Raum (vgl. Abb. 7.31).

Eine Größe, die durch einen einzigen reellen Zahlenwert charakterisiert wird, heißt Skalar. Beispiele für Skalare sind Temperatur, Arbeit, Masse, Energie.

Vektoren dagegen sind Größen, zu deren vollständiger Beschreibung neben einem Zahlenwert, ihrem Betrag (Länge des Vektors), noch die Angabe ihrer Richtung und Orientierung erforderlich sind. Beispiele für Vektoren sind Kraft, Geschwindigkeit, Beschleunigung, magnetische Feldstärke.

Vektoren werden meist mit kleinen lateinischen Buchstaben, die mit einem Pfeil versehen sind, bezeichnet: $\vec{a} = \overrightarrow{PQ}$ (gesprochen: Vektor a, Vektor PQ). Der Punkt P ist der Anfangspunkt und der Punkt Q der Endpunkt des Vektors. Der Betrag $|\vec{a}| = |\overrightarrow{PQ}|$ eines Vektors ist die Länge des Vektors, also die Länge der Verbindungsstrecke $\overline{PQ}$. Der Betrag ist eine nichtnegative reelle Zahl.

Zwei Vektoren $\vec{a}$ und $\vec{b}$ sind gleich, in Zeichen $\vec{a} = \vec{b}$, wenn sie den gleichen Betrag und gleiche Richtung und gleiche Orientierung haben. Vektoren dürfen daher parallel verschoben werden. Gleiche Vektoren gehen durch Parallelverschiebung ineinander über.

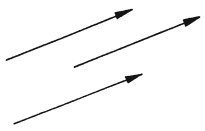


Abb. 7.31 Gleiche Vektoren

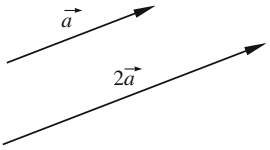


Abb. 7.32 Vektoren $\vec{a}$ und $2\vec{a}$

Im Unterschied zu diesen sogenannten freien Vektoren haben Ortsvektoren $\overrightarrow{OP}$ einen festen Anfangspunkt O. Ortsvektoren können also nicht verschoben werden.

Spezielle Vektoren

Der Nullvektor $\vec{0}$ hat den Betrag 0 und unbestimmte Richtung.

Ein Vektor $\vec{e}$ mit dem Betrag $|\vec{e}| = 1$ heißt Einheitsvektor. Man bezeichnet Einheitsvektoren auch als normierte Vektoren.

7.6.2 Multiplikation eines Vektors mit einem Skalar

Multipliziert man einen Vektor $\vec{a}$ mit einem Skalar (also einer reellen Zahl) $\lambda \in \mathbb{R}$, dann erhält man einen Vektor $\lambda\vec{a}$ mit dem Betrag $|\lambda\vec{a}| = |\lambda| \cdot |\vec{a}|$ ($|\lambda|$-facher Betrag des Vektors $\vec{a}$). Für $\lambda > 0$ haben $\lambda\vec{a}$ und $\vec{a}$ gleiche Richtung und Orientierung, für $\lambda < 0$ haben $\lambda\vec{a}$ und $\vec{a}$ gleiche Richtung und entgegengesetzte Orientierung (Abb. 7.32).

Multiplikation mit $\lambda = -1$ ergibt den Vektor $-\vec{a}$. Dieser Vektor hat den gleichen Betrag und die gleiche Richtung wie der Vektor $\vec{a}$, jedoch die entgegengesetzte Orientierung.

7.6.3 Addition und Subtraktion zweier Vektoren

Sollen zwei Vektoren $\vec{a}$ und $\vec{b}$ addiert werden, so bringt man durch Parallelverschiebung den Anfangspunkt des Vektors $\vec{b}$ in den Endpunkt des Vektors $\vec{a}$. Die Summe $\vec{a}+\vec{b}$ ist dann derjenige Vektor, der vom Anfangspunkt von $\vec{a}$ zum Endpunkt von $\vec{b}$ führt (Abb. 7.33).

Die Subtraktion zweier Vektoren $\vec{a}$ und $\vec{b}$ ist definiert als Addition von $\vec{a}$ und $-\vec{b}$ (Abb. 7.34).

$$\vec{a} - \vec{b} = \vec{a} + (-\vec{b})$$

Legt man die Anfangspunkte von $\vec{a}$ und $\vec{b}$ übereinander, dann ist der Vektor $\vec{a}-\vec{b}$ der Vektor vom Endpunkt von $\vec{b}$ zum Endpunkt von $\vec{a}$.

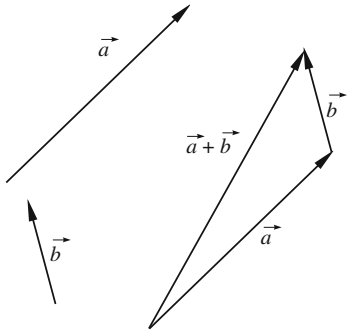


Abb. 7.33 Vektoraddition

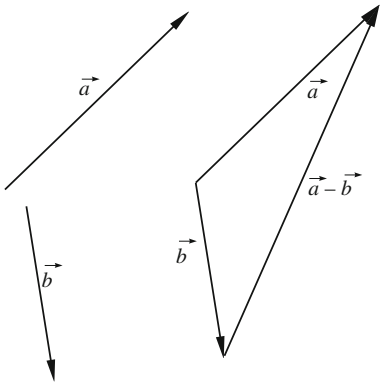


Abb. 7.34 Vektorsubtraktion

Zeichnet man ein Parallelogramm mit den Seiten $\vec{a}$ und $\vec{b}$, so kann man die Diagonale als $\vec{a}+\vec{b}$ oder als $\vec{b}+\vec{a}$ auffassen. Die Addition von Vektoren ist also kommutativ.

Kommutativgesetz

$$\vec{a} + \vec{b} = \vec{b} + \vec{a}$$

Auch das Assoziativgesetz und das Distributivgesetz sind erfüllt.

Assoziativgesetz

$$\vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + \vec{c} = \vec{a} + \vec{b} + \vec{c}$$

Distributivgesetz

$$\lambda \cdot (\vec{a} + \vec{b}) = \lambda \cdot \vec{a} + \lambda \cdot \vec{b} \quad (\lambda \in \mathbb{R})$$

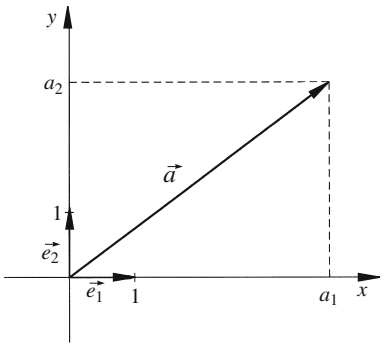


Abb. 7.35 Komponentendarstellung eines Vektors in der Ebene

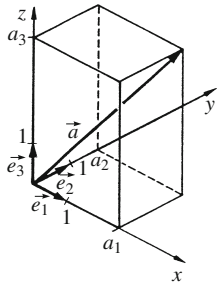


Abb. 7.36 Komponentendarstellung eines Vektors im Raum

7.6.4 Komponentendarstellung von Vektoren in der Ebene

Wählt man in einem kartesischen Koordinatensystem der Ebene einen Einheitsvektor $\vec{e}_1$ mit Richtung und Orientierung wie die positive x-Achse und einen Einheitsvektor $\vec{e}_2$ mit Richtung und Orientierung wie die positive y-Achse, dann lässt sich jeder Vektor $\vec{a}$ in der Ebene in eindeutiger Weise als Linearkombination der beiden sogenannten Basisvektoren $\vec{e}_1$ und $\vec{e}_2$ darstellen (Abb. 7.35).

$$\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2, \quad a_1, a_2 \in \mathbb{R}$$

Die beiden Vektoren $a_1\vec{e}_1$ und $a_2\vec{e}_2$ werden durch Parallelen zu den Basisvektoren $\vec{e}_1$ und $\vec{e}_2$ konstruiert.

Der Vektor $\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2$ wird identifiziert mit dem sogenannten Spaltenvektor

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Dabei heißen a_1 und a_2 die beiden Komponenten oder die kartesischen Koordinaten des Vektors $\vec{a}$. Mit Hilfe der Komponenten lassen sich Addition und Subtraktion von Vektoren sowie die Multiplikation eines Vektors mit einem Skalar folgendermaßen darstellen:

$$\vec{a} + \vec{b} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

$$\vec{a} - \vec{b} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \end{pmatrix}$$

$$\lambda \cdot \vec{a} = \lambda \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \end{pmatrix}$$

Der Betrag $|\vec{a}| = |\overrightarrow{PQ}|$, also die Länge des Vektors $\vec{a} = \overrightarrow{PQ}$, ist die Entfernung zwischen den Punkten P und Q. Nach dem Satz des Pythagoras gilt:

$$|\vec{a}| = \sqrt{a_1^2 + a_2^2}$$

7.6.5 Komponentendarstellung von Vektoren im Raum

Ganz analog wählt man in einem kartesischen Koordinatensystem des Raums drei Einheitsvektoren $\vec{e}_1, \vec{e}_2, \vec{e}_3$ mit Richtung und Orientierung wie die positive x-Achse, die positive y-Achse und die positive z-Achse. Dann lässt sich jeder Vektor $\vec{a}$ im Raum in eindeutiger Weise als Linearkombination der drei Basisvektoren $\vec{e}_1, \vec{e}_2$ und $\vec{e}_3$ darstellen (Abb. 7.36).

$$\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3, \quad a_1, a_2, a_3 \in \mathbb{R}$$

Die drei Vektoren $a_1\vec{e}_1$, $a_2\vec{e}_2$ und $a_3\vec{e}_3$ werden durch Parallelen zu den Basisvektoren $\vec{e}_1$, $\vec{e}_2$ und $\vec{e}_3$ konstruiert.

Der Vektor $\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3$ wird identifiziert mit dem Spaltenvektor

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Dabei heißen a_1, a_2, a_3 die Komponenten oder die kartesischen Koordinaten des Vektors $\vec{a}$.

Mit Hilfe der Komponenten lassen sich auch im Raum Addition und Subtraktion von Vektoren sowie die Multiplikation eines Vektors mit einem Skalar darstellen:

$$\vec{a}+\vec{b}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}+\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}=\begin{bmatrix}a_1+b_1\\a_2+b_2\\a_3+b_3\end{bmatrix}$$

$$\vec{a}-\vec{b}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}-\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}=\begin{bmatrix}a_1-b_1\\a_2-b_2\\a_3-b_3\end{bmatrix}$$

$$\lambda\cdot\vec{a}=\lambda\cdot\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}=\begin{bmatrix}\lambda a_1\\\lambda a_2\\\lambda a_3\end{bmatrix}$$

Der Betrag $|\vec{a}|=|\overrightarrow{PQ}|$, also die Länge des Vektors $\vec{a}=\overrightarrow{PQ}$, ist die Entfernung zwischen den Punkten P und Q. Durch zweimalige Anwendung des Satzes von Pythagoras errechnet man:

$$|\vec{a}|=\sqrt{a_1^2+a_2^2+a_3^2}$$

7.6.6 Skalarprodukt

Für die beiden Vektoren $\vec{a}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}$ und $\vec{b}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$ heißt

$$\vec{a}\cdot\vec{b}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}\cdot\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}=a_1b_1+a_2b_2+a_3b_3$$

Skalarprodukt oder inneres Produkt. Das Skalarprodukt zweier Vektoren ist kein Vektor, sondern eine reelle Zahl, also ein Skalar. Geometrisch ist das Skalarprodukt das Produkt der Länge des Vektors $\vec{a}$ und der Länge der senkrechten Projektion des Vektors $\vec{b}$ auf $\vec{a}$, also, falls $\varphi=\angle(\vec{a},\vec{b})$ den Winkel zwischen $\vec{a}$ und $\vec{b}$ bezeichnet (Abb. 7.37),

$$\vec{a}\cdot\vec{b}=|\vec{a}|\cdot|\vec{b}|\cdot\cos\varphi$$

Für den Winkel $\varphi=\angle(\vec{a},\vec{b})$ gilt somit:

$$\cos\varphi=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|}=\frac{a_1b_1+a_2b_2+a_3b_3}{\sqrt{a_1^2+a_2^2+a_3^2}\cdot\sqrt{b_1^2+b_2^2+b_3^2}}$$

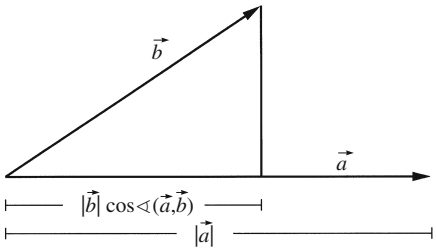


Abb. 7.37 Skalarprodukt: $\vec{a}\cdot\vec{b}=|\vec{a}|\cdot|\vec{b}|\cdot\cos\varphi$

Die folgenden Rechenregeln lassen sich aus der Definition ableiten:

1. $\vec{a}\cdot\vec{b}=\vec{b}\cdot\vec{a}$
2. $(\lambda\cdot\vec{a})\cdot\vec{b}=\vec{a}\cdot(\lambda\cdot\vec{b})=\lambda\cdot(\vec{a}\cdot\vec{b})$
3. $(\vec{a}+\vec{b})\cdot\vec{c}=\vec{a}\cdot\vec{c}+\vec{b}\cdot\vec{c}$
4. $\vec{a}\cdot\vec{b}=0\Leftrightarrow\vec{a}\perp\vec{b}$ ($\vec{a}$ und $\vec{b}$ stehen senkrecht aufeinander)
5. $|\vec{a}|=\sqrt{\vec{a}\cdot\vec{a}}$.

So folgt zum Beispiel aus 4., nämlich dass $\vec{a}\cdot\vec{b}=0$ genau für zwei senkrecht aufeinanderstehende (man sagt auch orthogonale) Vektoren $\vec{a}$ und $\vec{b}$ gilt, dass genau dann der Winkel φ gleich 90° ist ($\Rightarrow\cos\varphi=0$).

Das Skalarprodukt lässt sich entsprechend auch in der Ebene, also für Vektoren mit zwei Komponenten, definieren.

Beispiele

1. Das Skalarprodukt der Vektoren
$\vec{a}=\begin{bmatrix}2\\3\\-1\end{bmatrix}$ und $\vec{b}=\begin{bmatrix}4\\-5\\2\end{bmatrix}$ ist

$$\vec{a}\cdot\vec{b}=\begin{bmatrix}2\\3\\-1\end{bmatrix}\cdot\begin{bmatrix}4\\-5\\2\end{bmatrix}=2\cdot4+3\cdot(-5)+(-1)\cdot2$$
$$=8-15-2=-9.$$

2. Gesucht ist der Winkel φ, den die beiden Vektoren
$\vec{a}=\begin{bmatrix}1\\1\\0\end{bmatrix}$ und $\vec{b}=\begin{bmatrix}0\\1\\1\end{bmatrix}$ miteinander einschließen.
Es gilt $\vec{a}\cdot\vec{b}=1\cdot0+1\cdot1+0\cdot1=1$ und $|\vec{a}|=|\vec{b}|=\sqrt{1+1}=\sqrt{2}$. Somit folgt $\cos\varphi=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|}=\frac{1}{\sqrt{2}\cdot\sqrt{2}}=\frac{1}{2}\Rightarrow\varphi=60°$.
3. Für welchen Wert von c sind die beiden Vektoren
$\vec{a}=\begin{bmatrix}2\\3\\-4\end{bmatrix}$ und $\vec{b}=\begin{bmatrix}1\\2\\c\end{bmatrix}$ orthogonal?

Für das Skalarprodukt gilt

$$\vec{a}\cdot\vec{b} = \begin{bmatrix}2\\3\\-4\end{bmatrix}\cdot\begin{bmatrix}1\\2\\c\end{bmatrix}$$
$$= 2+6-4c = 8-4c = 0 \quad \Rightarrow \quad c = 2.$$

Für $c = 2$ stehen die Vektoren $\vec{a}$ und $\vec{b}$ also senkrecht aufeinander. ◀

7.6.7 Vektorprodukt

Sind $\vec{a} = \begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}$ und $\vec{b} = \begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$ zwei Vektoren im Raum, so heißt der Vektor

$$\vec{a}\times\vec{b} = \begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}\times\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} = \begin{bmatrix}a_2b_3-a_3b_2\\a_3b_1-a_1b_3\\a_1b_2-a_2b_1\end{bmatrix}$$

Vektorprodukt oder Kreuzprodukt oder äußeres Produkt der Vektoren $\vec{a}$ und $\vec{b}$. Das Vektorprodukt ist im Unterschied zum Skalarprodukt nur im Raum definiert (Abb. 7.38).

Das Vektorprodukt besitzt folgende Eigenschaften:

1. $\vec{b}\times\vec{a} = -\vec{a}\times\vec{b}$
2. $\vec{a}\times\vec{b} = \vec{0}$, falls $\vec{a} = \vec{0}$ oder $\vec{b} = \vec{0}$ oder $\vec{a}$ parallel zu $\vec{b}$
3. $(\lambda\vec{a})\times\vec{b} = \vec{a}\times(\lambda\vec{b}) = \lambda(\vec{a}\times\vec{b})$
4. $(\vec{a}+\vec{b})\times\vec{c} = \vec{a}\times\vec{c}+\vec{b}\times\vec{c}$
5. $\vec{a}\times\vec{b}$ steht senkrecht auf den Vektoren $\vec{a}$ und $\vec{b}$
6. $|\vec{a}\times\vec{b}| = |\vec{a}|\cdot|\vec{b}|\sin\varphi = |\vec{a}|\cdot|\vec{b}|\sin\angle(\vec{a}\vec{b})$
7. $\vec{a}, \vec{b}, \vec{a}\times\vec{b}$ bilden in dieser Reihenfolge ein Rechtssystem

Der Vektor $\vec{a}\times\vec{b}$ steht also senkrecht auf $\vec{a}$ und auf $\vec{b}$. Sein Betrag (seine Länge) ist gleich dem Flächeninhalt des von den beiden Vektoren $\vec{a}$ und $\vec{b}$ aufgespannten Parallelogramms. Falls $\vec{a}$ auf dem kürzesten Weg nach $\vec{b}$ gedreht wird, zeigt $\vec{a}\times\vec{b}$ in Richtung der Bewegung einer Schraube mit Rechtsgewinde.

Beispiel

Für die Vektoren $\vec{a} = \begin{bmatrix}-1\\2\\-3\end{bmatrix}$ und $\vec{b} = \begin{bmatrix}2\\1\\-2\end{bmatrix}$ ergibt sich für das Vektorprodukt $\vec{a}\times\vec{b} = \begin{bmatrix}-1\\-8\\-5\end{bmatrix}$.

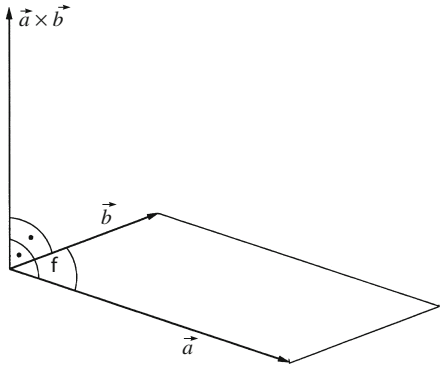


Abb. 7.38 Vektorprodukt $\vec{a}\times\vec{b}$ der Vektoren $\vec{a}$ und $\vec{b}$

Zur Probe kann man etwa Eigenschaft 5. benutzen: Es muss $(\vec{a}\times\vec{b})\cdot\vec{a} = 0$ (und auch $(\vec{a}\times\vec{b})\cdot\vec{b} = 0$) gelten:

$$(\vec{a}\times\vec{b})\cdot\vec{a} = \begin{bmatrix}-1\\-8\\-5\end{bmatrix}\cdot\begin{bmatrix}-1\\2\\-3\end{bmatrix} = 1-16+15 = 0.$$ ◀

7.6.8 Spatprodukt

Sind $\vec{a}, \vec{b}$ und $\vec{c}$ drei Vektoren im Raum, so heißt der Skalar

$$(\vec{a}\times\vec{b})\cdot\vec{c}$$

Spatprodukt. Aus der geometrischen Interpretation des Skalarprodukts folgt, dass $(\vec{a}\times\vec{b})\cdot\vec{c}$ gleich dem Produkt aus der Länge von $\vec{a}\times\vec{b}$ und der Länge der Projektion von $\vec{c}$ auf $\vec{a}\times\vec{b}$ ist. Da $|\vec{a}\times\vec{b}|$ gleich dem Flächeninhalt des von $\vec{a}$ und $\vec{b}$ aufgespannten Parallelogramms ist, stellt $(\vec{a}\times\vec{b})\cdot\vec{c}$ das Volumen des von den Vektoren $\vec{a}, \vec{b}, \vec{c}$ aufgespannten Spates dar, falls die Vektoren eine Lage wie in Abb. 7.39 haben. Spat ist ein anderer Name für Parallelepiped oder Parallelflach. Zeigt $\vec{c}$ nach unten, so ist das Spatprodukt negativ, und es ist dem Betrage nach das Volumen des Spates.

Mit der abkürzenden Schreibweise

$$[\vec{a},\vec{b},\vec{c}] = (\vec{a}\times\vec{b})\cdot\vec{c}$$

für das Spatprodukt können einige Eigenschaften des Spatprodukts formuliert werden:

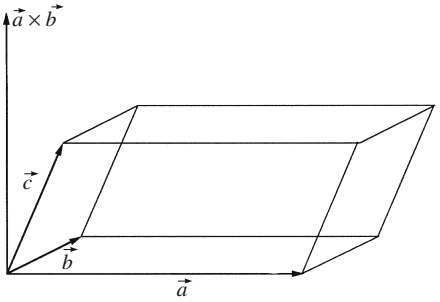


Abb. 7.39 Geometrische Veranschaulichung des Spatprodukts

1.
$$[\vec{a},\vec{b},\vec{c}] = (\vec{a}\times\vec{b})\cdot\vec{c} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - b_1a_2 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$
$$= c_1(a_2b_3 - a_3b_2) - c_2(a_3b_1 - a_1b_3) + c_3(a_1b_2 - b_1a_2)$$

2. Eine zyklische (kreisförmige) Vertauschung der Vektoren ändert das Spatprodukt nicht:
$$[\vec{a},\vec{b},\vec{c}] = [\vec{b},\vec{c},\vec{a}] = [\vec{c},\vec{a},\vec{b}]$$

3. Das Spatprodukt ändert das Vorzeichen (bei gleichem Betrag), falls zwei Vektoren miteinander vertauscht werden:
$$[\vec{b},\vec{a},\vec{c}] = [\vec{c},\vec{b},\vec{a}] = [\vec{a},\vec{c},\vec{b}] = -[\vec{a},\vec{b},\vec{c}]$$

4. $[\vec{a},\vec{b},\vec{c}] = 0 \Leftrightarrow \vec{a},\vec{b},\vec{c}$ liegen in einer Ebene (man sagt dann: $:\vec{a},\vec{b},\vec{c}$ sind linear abhängig)
5. $[\vec{a},\vec{b},\vec{c}] > 0 \Leftrightarrow \vec{a},\vec{b},\vec{c}$ bilden ein Rechtssystem
6. Das Volumen V des von den Vektoren $\vec{a},\vec{b},\vec{c}$ gebildeten Tetraeders ist
$$V = \frac{1}{6}[\vec{a},\vec{b},\vec{c}]$$

Beispiel

Das Volumen V des von den Vektoren
$$\vec{a} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \vec{c} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$
aufgespannten Tetraeders beträgt
$$V = \frac{1}{6}|[\vec{a},\vec{b},\vec{c}]| = \frac{1}{6}|(\vec{a}\times\vec{b})\cdot\vec{c}| = \frac{1}{6}\left|\begin{bmatrix} -1 \\ -3 \\ -1 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}\right| = \frac{1}{6}|-6| = \frac{1}{6}6 = 1.$$
◀

8 Differential- und Integralrechnung

Arnfried Kemnitz



8.1 Folgen 144
8.2 Reihen 146
8.3 Grenzwerte von Funktionen 150
8.4 Ableitung einer Funktion 153
8.5 Integralrechnung 165
8.6 Funktionenreihen 174

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_8

8.1 Folgen

8.1.1 Grundbegriffe

Eine Folge besteht aus Zahlen einer Menge, die in einer bestimmten Reihenfolge angeordnet sind:

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

Sind alle diese Zahlen reelle Zahlen, dann nennt man die Folge auch reelle Zahlenfolge. Die Zahlen der Folge heißen Glieder der Folge.

Handelt es sich um endlich viele Zahlen, so heißt die Folge endlich, andernfalls unendliche Folge.

Eine unendliche Folge lässt sich auch als Funktion (Abbildung) definieren:

$$f : \mathbb{N}^* \to \mathbb{R}, \quad n \mapsto f(n) = a_n$$

Unter den Gliedern einer Folge können auch gleiche Zahlen auftreten.

Eine Folge kann durch direkte Angabe ihrer Glieder oder auch durch einen arithmetischen Ausdruck gegeben sein. Ein solcher arithmetischer Ausdruck kann entweder eine explizite Formel für das Folgenglied a_n oder eine rekursive Definition sein. Bei einer Rekursion wird a_n durch Folgenglieder mit kleineren Indizes definiert.

Schreibweise von Folgen

$$(a_n) = (a_1, a_2, a_3, \ldots)$$

Eine konstante Folge (a_n) ist eine Folge mit $(a_n) = (a, a, a, \ldots)$.

Beispiele

1. $(a_n) = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$
2. $(a_n) = (n) = (1, 2, 3, 4, \ldots)$
3. $(a_n) = (3 - \frac{1}{2^{n-2}}) = (1, 2, \frac{5}{2}, \frac{11}{4}, \frac{23}{8}, \ldots)$
4. $(a_n) = (4 + 3(n-1)) = (4, 7, 10, 13, \ldots)$
5. $(a_n) = ((-1)^{n+1}) = (1, -1, 1, -1, 1, \ldots)$
6. $(a_n) = (\frac{1}{n}) = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$
7. (a_n) mit $a_1 = 0$, $a_{n+1} = \frac{1}{2}(1 + a_n)$ für $n \in \mathbb{N}^*$, also $(a_n) = (0, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots)$ ◀

Die erste Folge ist endlich, alle anderen sind unendlich. Die erste Folge ist durch Angabe ihrer Glieder definiert, die letzte Folge ist rekursiv definiert und alle anderen durch eine explizite Formel.

Monotone Folgen

Eine Folge (a_n) heißt

- monoton wachsend, wenn $a_1 \le a_2 \le a_3 \le \ldots \le a_n \le \ldots$ gilt,
- streng monoton wachsend, wenn $a_1 < a_2 < a_3 < \ldots < a_n < \ldots$ gilt,
- monoton fallend, wenn $a_1 \ge a_2 \ge a_3 \ge \ldots \ge a_n \ge \ldots$ gilt,
- streng monoton fallend, wenn $a_1 > a_2 > a_3 > \ldots > a_n > \ldots$ gilt.

Die Folgen der Beispiele 1 bis 4 und 7 sind streng monoton wachsend, die Folge aus Beispiel 6 ist streng monoton fallend.

Alternierende Folgen

Eine alternierende Folge ist eine Folge, deren Glieder abwechselnd unterschiedliche Vorzeichen haben. Von zwei aufeinander folgenden Gliedern a_k und a_{k+1} einer solchen Folge (a_n) ist also genau ein Glied positiv und eins negativ.

Die Folge aus Beispiel 5 ist alternierend.

Beschränkte Folgen

Eine Folge (a_n) heißt

- nach oben beschränkt, wenn es eine konstante Zahl K_o gibt, so dass für alle Glieder $a_n \le K_o$ gilt,
- nach unten beschränkt, wenn es eine konstante Zahl K_u gibt, so dass für alle Glieder $a_n \ge K_u$ gilt,
- beschränkt, wenn die Folge sowohl nach oben als auch nach unten beschränkt ist, wenn es also zwei Zahlen K_u, K_o gibt mit $K_u \le a_n \le K_o$ für alle $n \in \mathbb{N}^*$.
  Gleichwertig damit ist, dass es eine Konstante $K > 0$ mit $|a_n| \le K$ für alle n gibt.

Monoton wachsende und streng monoton wachsende Folgen sind nach unten beschränkt, monoton fallende und streng monoton fallende Folgen sind nach oben beschränkt.

Die Folgen der Beispiele 1, 3, 5, 6 und 7 sind beschränkt, die Folgen der Beispiele 2 und 4 sind nach unten beschränkt.

8.1.2 Arithmetische Folgen

Bei einer arithmetischen Folge ist die Differenz je zweier aufeinanderfolgender Glieder konstant. Durch das Anfangsglied $a_1 = a$ und diese Differenz d ist die Folge dann eindeutig bestimmt.

$$(a_n) = (a, a + d, a + 2d, a + 3d, \ldots, a + (n-1)d, \ldots)$$

Das n-te Glied einer arithmetischen Folge lautet $a_n = a + (n-1)d$, $n \in \mathbb{N}^*$. Das Glied $a_1 = a$ nennt man Anfangsglied

der Folge und $d = a_{n+1} - a_n$ (für $n = 1, 2, 3, \ldots$) die (konstante) Differenz der Folge. In einer arithmetischen Folge ist jedes Folgenglied a_n ($n \geq 2$) das arithmetische Mittel seiner Nachbarglieder.

Beispiele

1. $(a_n) = (n) = (1, 2, 3, 4, \ldots)$ (arithmetische Folge mit $a = 1$ und $d = 1$)
2. $(a_n) = (4 + 3(n-1)) = (4, 7, 10, 13, \ldots)$ (arithmetische Folge mit $a = 4$ und $d = 3$) ◀

8.1.3 Geometrische Folgen

Bei einer geometrischen Folge ist der Quotient je zweier aufeinanderfolgender Glieder konstant. Durch das Anfangsglied $a_1 = a$ und diesen Quotienten q ist die Folge dann eindeutig bestimmt.

$$(a_n) = (a, aq, aq^2, aq^3, \ldots, aq^{n-1}, \ldots)$$

Das n-te Glied einer geometrischen Folge lautet $a_n = aq^{n-1}$, $n \in \mathbb{N}^*$. Das Glied $a_1 = a$ nennt man Anfangsglied der Folge und $q = \frac{a_{n+1}}{a_n}$ (für $n = 1, 2, 3, \ldots$) den (konstanten) Quotienten der Folge. In einer geometrischen Folge ist jedes Folgenglied a_n ($n \geq 2$) das geometrische Mittel seiner Nachbarglieder (bis eventuell auf das Vorzeichen).

Beispiele

1. $(a_n) = (3 \cdot 2^{n-1}) = (3, 6, 12, 24, \ldots)$ (geometrische Folge mit $a = 3$ und $q = 2$)
2. $(a_n) = (2^{-n}) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots)$ (geometrische Folge mit $a = \frac{1}{2}$ und $q = \frac{1}{2}$) ◀

8.1.4 Grenzwert einer Folge

Man sagt, die Folge (a_n) besitzt den Grenzwert (oder auch Limes genannt) $\lim_{n\to\infty} a_n = a$ oder $(a_n) \to a$ (gesprochen: Limes a_n gleich a), wenn die Abweichung $|a - a_n|$ der Folgenglieder a_n von diesem Wert a für genügend große n beliebig klein wird.

Grenzwert (Limes)

$$\lim_{n\to\infty} a_n = a \quad \text{oder} (a_n) \to a$$

Exakte Definition:

Die Folge (a_n) besitzt den Grenzwert $a_n = a$, wenn sich nach Vorgabe einer beliebig kleinen positiven Zahl ε ein $n_0 \in \mathbb{N}$ so finden lässt, dass für alle $n \geq n_0$ gilt

$$|a - a_n| < \varepsilon$$

Das n_0 hängt offensichtlich von der Wahl von ε ab, also $n_0 = n_0(\varepsilon)$.

Besitzt (a_n) den Grenzwert a, so sagt man, dass (a_n) gegen a konvergiert. Eine Folge, die einen Grenzwert besitzt, heißt konvergent. Eine Folge, die keinen Grenzwert besitzt, heißt dagegen divergent.

Eine Folge besitzt höchstens einen Grenzwert.

Eine Nullfolge ist eine Folge, die den Grenzwert 0 besitzt.

Beispiel

1. Die Folge (a_n) mit $a_n = \frac{1}{10^n}$ hat den Grenzwert $a = 0$, denn die Differenz $|a - a_n| = |0 - \frac{1}{10^n}| = \frac{1}{10^n}$ wird für große n beliebig klein. Wählt man etwa $\varepsilon = \frac{1}{10^{10}}$, so gilt $|a - a_n| < \varepsilon$ für $n \geq 11$. Es gilt also $\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{10^n} = 0$.
   Die Folge (a_n) ist somit eine Nullfolge. ◀

Konvergente Folgen sind beschränkt. Eine beliebige Folge kann also nur konvergent sein, wenn sie beschränkt ist.

Es gilt folgendes Konvergenzkriterium:

Eine monotone und beschränkte Folge ist stets konvergent.

Für konvergente Folgen gelten verschiedene Rechenregeln:

$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$$

$$\lim_{n\to\infty} (a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n$$

$$\lim_{n\to\infty} (a_n \cdot b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$$

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n},$$

falls $b_n \neq 0$ und $\lim_{n\to\infty} b_n \neq 0$

Beispiele

2. $\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n} = 0$
3. $\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{n}{n+1} = 1$
4. $\lim_{n\to\infty} a_n = \lim_{n\to\infty} (3 - \frac{1}{2^{n-2}}) = 3$
5. $\lim_{n\to\infty} a_n = \lim_{n\to\infty} (\frac{1}{2})^n = 0$ ◀

Die Folge aus Beispiel 5 ist eine geometrische Folge. Es gilt:

Jede geometrische Folge mit $a_n = aq^{n-1}$ konvergiert gegen Null, wenn $|q|$, der Betrag von q, kleiner als 1 ist.

8.1.5 Einige Grenzwerte

Für einige wichtige konvergente Zahlenfolgen sind im Folgenden ihre Grenzwerte angegeben.

$$\lim_{n\to\infty} \sqrt[n]{q} = 1 \quad (q > 0)$$

$$\lim_{n\to\infty} \sqrt[n]{n} = 1$$

$$\lim_{n\to\infty} \frac{c_r n^r + c_{r-1} n^{r-1} + \ldots + c_1 n + c_0}{d_s n^s + d_{s-1} n^{s-1} + \ldots + d_1 n + d_0} = \begin{cases} \frac{c_r}{d_r} & \text{für } r = s \\ 0 & \text{für } r < s \end{cases}$$

$$(c_0, c_1, \ldots, c_r, d_0, d_1, \ldots, d_s \in \mathbb{R}, c_r \neq 0, d_s \neq 0)$$

$$\lim_{n\to\infty} \frac{\log_a n}{n} = 0 \quad (a > 0, a \neq 1)$$

$$\lim_{n\to\infty} q^n = 0 \quad (|q| < 1)$$

$$\lim_{n\to\infty} nq^n = 0 \quad (|q| < 1)$$

$$\lim_{n\to\infty} \frac{a^n}{n!} = 0 \quad (a \in \mathbb{R})$$

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e = 2{,}7182818284\ldots$$

8.1.6 Divergente Folgen

Eine Folge, die keinen Grenzwert besitzt, heißt divergent. Bei divergenten Folgen unterscheidet man zwischen bestimmter und unbestimmter Divergenz.

Eine Folge (a_n) heißt bestimmt divergent gegen $+\infty$, wenn zu jeder beliebig großen vorgegebenen Zahl K ein Index n_0 existiert, so dass $a_n > K$ für alle Indizes $n \geq n_0$ gilt. Eine solche bestimmt divergente Folge wächst für $n \to \infty$ über alle Grenzen. Man schreibt dann

$$\lim_{n\to\infty} a_n = \infty$$

Eine Folge (a_n) heißt dagegen bestimmt divergent gegen $-\infty$, wenn zu jeder noch so kleinen vorgegebenen Zahl $-K$ $(K > 0)$ ein Index n_0 existiert, so dass $a_n < -K$ für alle Indizes $n \geq n_0$ gilt. Eine solche bestimmt divergente Folge fällt für $n \to \infty$ unter alle Grenzen. Man schreibt dann

$$\lim_{n\to\infty} a_n = -\infty$$

Eine Folge, die nicht konvergent und nicht bestimmt divergent ist, heißt unbestimmt divergent.

Monoton (streng monoton) wachsende und nicht beschränkte Folgen (a_n) sind bestimmt divergent mit $\lim_{n\to\infty} a_n = \infty$.

Monoton (streng monoton) fallende und nicht beschränkte Folgen (a_n) sind bestimmt divergent mit $\lim_{n\to\infty} a_n = -\infty$.

Beispiele

1. $\lim_{n\to\infty} n = \infty$
   (denn $(a_n) = (n)$ ist streng monoton wachsend und nicht beschränkt)
2. $\lim_{n\to\infty} (-n^3) = -\infty$
   (denn $(a_n) = (-n^3)$ ist streng monoton fallend und nicht beschränkt)
3. $\lim_{n\to\infty} 2^{n+2} = \infty$
   (denn $(a_n) = (2^{n+2})$ ist streng monoton wachsend und nicht beschränkt)
4. $(a_n) = ((-3)^n) = ((-1)^n \cdot 3^n)$ ist unbestimmt divergent ◀

Die Folge aus Beispiel 1 ist eine arithmetische Folge. Es gilt:

Jede arithmetische Folge ist divergent, denn die Differenz zweier aufeinanderfolgender Glieder ist stets d.

Für positive Werte von d werden die Glieder a_n der Folge ab einer Stelle größer als jede beliebig große Zahl. Für negative Werte von d werden die Glieder a_n dagegen ab einer Stelle kleiner als jede vorgegebene beliebig kleine Zahl. Jede arithmetische Folge ist also bestimmt divergent.

Die Folgen aus den Beispielen 3 und 4 sind geometrische Folgen. Es gilt:

Jede geometrische Folge mit $a_n = aq^{n-1}$ ist divergent, wenn der Betrag $|q|$ größer als 1 ist, und zwar für $q > 1$ bestimmt divergent und für $q < -1$ unbestimmt divergent.

8.2 Reihen

8.2.1 Definitionen

Eine Reihe ist die Summe der Glieder einer Folge (Zahlenfolge) (a_n).

$$a_1 + a_2 + \ldots + a_n + \ldots$$

Ist die Folge endlich, so nennt man auch die Reihe endlich. Für unendliche Folgen ergeben sich unendliche Reihen.

$$a_1 + a_2 + \ldots + a_n + \ldots = \sum_{k=1}^{\infty} a_k$$

Das Zeichen ∞ bedeutet dabei, dass die Reihe nicht abbricht. Sie besteht aus unendlich vielen Summanden. Die Zahlen a_n, also die Summanden, heißen auch Glieder der Reihe.

Beispiele

1. $\sum_{k=1}^{10} 2^k = 2^1 + 2^2 + 2^3 + \ldots + 2^{10} = 2 + 4 + 8 + \ldots + 1024$ (endliche Reihe)
2. $\sum_{k=1}^{\infty} \frac{3^k}{k} = 3 + \frac{3^2}{2} + \frac{3^3}{3} + \ldots + \frac{3^n}{n} + \ldots$ (unendliche Reihe) ◀

Folgende Summen heißen Teilsummen oder Partialsummen der Reihe:

$$
\begin{aligned}
s_1 &= a_1, \\
s_2 &= a_1 + a_2, \quad \ldots, \\
s_n &= a_1 + a_2 + a_3 + \ldots + a_n = \sum_{k=1}^{n} a_k, \quad \ldots
\end{aligned}
$$

Man spricht von einer konvergenten unendlichen Reihe, wenn die Folge (s_n) der Partialsummen konvergiert, also einen Grenzwert s besitzt.

$$
s = \lim_{n\to\infty} s_n = \sum_{k=1}^{\infty} a_k
$$

Dieser Grenzwert s heißt die Summe der Reihe. Eine unendliche Reihe ist also genau dann konvergent, wenn die Folge der Partialsummen konvergiert.

Besitzt die Folge der Partialsummen keinen Grenzwert, dann heißt die unendliche Reihe divergent. In diesem Fall können die Partialsummen unbegrenzt wachsen oder oszillieren (die Folge der Partialsummen ist alternierend).

Die unendliche Reihe heißt bestimmt divergent, wenn die Folge (s_n) der Partialsummen bestimmt divergent ist. Ist die Folge der Partialsummen unbestimmt divergent, so heißt auch die unendliche Reihe unbestimmt divergent.

Die Frage nach der Konvergenz einer unendlichen Reihe wird somit auf die Frage nach der Existenz eines Grenzwertes der Folge (s_n) der Partialsummen zurückgeführt.

Die Folge der Glieder (a_n) einer konvergenten Reihe muss gegen Null konvergieren, also eine Nullfolge sein. Diese Bedingung ist notwendig, sie reicht jedoch für die Konvergenz einer unendlichen Reihe nicht aus (vgl. Abschn. 8.2.4).

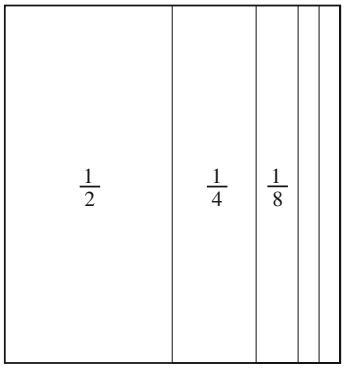


Abb. 8.1 Zur Konvergenz der Reihe $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$

Für konvergente Reihen gelten verschiedene Rechenregeln:

Konvergieren die Reihen $\sum_{k=1}^{\infty} a_k$ und $\sum_{k=1}^{\infty} b_k$, so konvergieren auch die Reihen $\sum_{k=1}^{\infty} (a_k + b_k)$ und $\sum_{k=1}^{\infty} c \cdot a_k$, $c \in \mathbb{R}$, und es gilt

$$
\sum_{k=1}^{\infty} (a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k
$$

$$
\sum_{k=1}^{\infty} c \cdot a_k = c \sum_{k=1}^{\infty} a_k
$$

Beispiele

3. $$
\begin{aligned}
\sum_{k=1}^{6} k \cdot 2^k &= 2 + 2 \cdot 2^2 + 3 \cdot 2^3 + 4 \cdot 2^4 + 5 \cdot 2^5 + 6 \cdot 2^6 \\
&= 2 + 8 + 24 + 64 + 160 + 384 = 642
\end{aligned}
$$

4. $$
\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = 1
$$

   Dass diese unendliche Reihe die Summe 1 hat, kann man sich dadurch klarmachen, dass man ein Quadrat mit der Fläche 1 fortgesetzt halbiert (Abb. 8.1). Die entstehenden Rechtecke haben die Flächeninhalte $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots$, und ihre Summe ist offensichtlich 1, der Flächeninhalt des Quadrats (vgl. auch Abschn. 8.2.3).

5. $$
\sum_{k=1}^{\infty} k = 1 + 2 + 3 + \ldots + n + \ldots
$$

   Diese unendliche Reihe ist bestimmt divergent, denn die Folge $(a_n) = (n)$ ist bestimmt divergent (Abschn. 8.1.6).

6. $$\sum_{k=1}^{\infty}(-1)^k = -1+1-1+1-1+1-\ldots$$

   Für die Partialsummen gilt

   $$s_n = \begin{cases} 0 & \text{falls } n \text{ gerade ist} \\ -1 & \text{falls } n \text{ ungerade ist} \end{cases}$$

   Die unendliche Reihe ist unbestimmt divergent.

7. $$\sum_{k=1}^{\infty}\frac{1}{k(k+1)} = \frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\frac{1}{4\cdot 5}+\ldots$$

   Aus $a_k = \frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$ folgt

   $$\begin{aligned} s_n = \sum_{k=1}^{n} a_k &= \left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right) \\ &\quad +\left(\frac{1}{3}-\frac{1}{4}\right)+\ldots+\left(\frac{1}{n}-\frac{1}{n+1}\right) \\ &= 1-\frac{1}{n+1} \end{aligned}$$

   Wegen $s_n = (1-\frac{1}{n+1}) = 1-\frac{1}{n+1} = 1$ ist die gegebene Reihe konvergent mit dem Grenzwert 1:

   $$\sum_{k=1}^{\infty}\frac{1}{k(k+1)} = 1$$

8. $\sum_{k=1}^{\infty}\frac{k+1}{3k+2}$ ist nicht konvergent, denn die Glieder $a_k = \frac{k+1}{3k+2}$ bilden wegen $\lim_{n\to\infty} a_n = \lim_{n\to\infty}\frac{n+1}{3n+2} = \frac{1}{3}$ keine Nullfolge. ◀

8.2.2 Arithmetische Reihen

Eine arithmetische Reihe entsteht aus den Gliedern einer arithmetischen Folge. Da schon jede unendliche arithmetische Folge divergiert, ist auch jede unendliche arithmetische Reihe divergent. Da unendliche arithmetische Folgen bestimmt divergent sind (Abschn. 8.1.6), sind auch unendliche arithmetische Reihen bestimmt divergent.

Die Summe s_n einer endlichen arithmetischen Reihe $\sum_{k=1}^{n}(a+(k-1)d)$ lässt sich jedoch allgemein berechnen. Wegen $a_1 = a$ folgt $s_n = a_1 + (a_1+d) + (a_1+2d) + \ldots + (a_1+(n-1)d)$. Dreht man die Reihenfolge der Summanden um und beachtet, dass die Differenz zweier aufeinanderfolgender Glieder gleich d ist, so folgt andererseits $s_n = a_n + (a_n-d) + (a_n-2d) + \ldots + (a_n-(n-1)d)$. Schreibt man diese beiden Ausdrücke für s_n untereinander und addiert jeweils die beiden übereinanderstehenden Terme, so folgt

$$\begin{aligned} s_n &= a_1 + (a_1+d) + (a_1+2d) + \ldots + (a_1+(n-1)d) \\ s_n &= a_n + (a_n-d) + (a_n-2d) + \ldots + (a_n-(n-1)d) \end{aligned}$$

und daraus $2s_n = n(a_1+a_n)$, denn jede dieser Summen ist a_1+a_n, und es gibt insgesamt n solcher Summen. Es folgt:

$$s_n = \sum_{k=1}^{n}(a+(k-1)d) = \frac{n}{2}(a_1+a_n)$$

Die Summe einer endlichen arithmetischen Reihe mit n Summanden ist also die Summe des ersten und des letzten Glieds multipliziert mit der halben Anzahl der Summanden.

Beispiele

1. $\sum_{k=1}^{10}(3+(k-1)5) = \frac{10}{2}(3+48) = 255$
2. $\sum_{k=1}^{100} k = 50(1+100) = 5050$
3. $\sum_{k=1}^{100}(3+4k) = \frac{100}{2}(7+403) = 50\cdot 410 = 20.500$ ◀

Die Summe der ersten n natürlichen Zahlen zum Beispiel lässt sich hiermit für beliebiges (beliebig großes) n sehr einfach ausrechnen (vgl. Beispiel 2).

8.2.3 Geometrische Reihen

Eine geometrische Reihe entsteht aus den Gliedern einer geometrischen Folge. Die Summe s_n einer endlichen geometrischen Reihe $\sum_{k=1}^{n} aq^{k-1}$ ergibt sich für $q \neq 1$ aus folgender Rechnung:

$$\begin{aligned} s_n &= a + aq + aq^2 + \ldots + aq^{n-1} \\ qs_n &= aq + aq^2 + \ldots + aq^{n-1} + aq^n \end{aligned}$$

Zieht man die zweite Gleichung von der ersten ab, so folgt $s_n - qs_n = a - aq^n$ und somit für die Summe s_n einer endlichen geometrischen Reihe mit $q \neq 1$:

$$s_n = \sum_{k=1}^{n} aq^{k-1} = a\frac{1-q^n}{1-q} \quad (q \neq 1)$$

Für $q = 1$ gilt $s_n = n \cdot a$.

Beispiele

1. $\sum_{k=1}^{5} 2^{k-1} = \frac{1-2^5}{1-2} = 31$
2. $\sum_{k=1}^{10} 3 \cdot 5^{k-1} = 3\frac{1-5^{10}}{1-5} = 3\frac{9.765.624}{4} = 7.324.218$
3. $\sum_{k=1}^{100} 3^k = 3\frac{1-3^{100}}{1-3} = \frac{3}{2}(3^{100}1)$ ◀

Die Summe $s_n = a\frac{1-q^n}{1-q}$ ist für $q \neq 1$ das n-te Glied der Folge der Partialsummen. Die Größen a und q sind Konstanten, die Konvergenz der Folge hängt nur von der Größe $1 - q^n$ ab. Für $q > 1$ und $q \leq -1$ divergiert die Folge (q^n), die geometrische Reihe ist dann also ebenfalls divergent. Für $q \geq 1$ ist die unendliche geometrische Reihe bestimmt divergent, für $q \leq -1$ ist sie unbestimmt divergent.

Für $|q| < 1$ wird $|q|^n = |q^n|$ beliebig klein, wenn n nur groß genug gewählt wird, das heißt, es gilt $q^n = 0$. Für $|q| < 1$ konvergiert deshalb die Folge (q^n), es gilt dann $(1 - q^n) = 1 - q^n = 1$. In diesem Fall konvergiert die unendliche geometrische Reihe und hat den Grenzwert

$$\begin{aligned} s = \lim_{n\to\infty} s_n &= \sum_{k=1}^{\infty} aq^{k-1} \\ &= \lim_{n\to\infty} a\frac{q^n - 1}{q - 1} \\ &= \frac{a}{1-q} \quad (|q| < 1) \end{aligned}$$

Beispiele

4. $\sum_{k=1}^{\infty} 5 \cdot (-\frac{11}{12})^{k-1} = \frac{5}{1+\frac{11}{12}} = \frac{60}{23}$ $(a = 5, q = -\frac{11}{12})$
5. $\sum_{k=1}^{\infty} (\frac{1}{2})^k = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$ $(a = q = \frac{1}{2})$
6. $\sum_{k=1}^{\infty} 3(\frac{4}{5})^{k-1} = \frac{3}{1-\frac{4}{5}} = 15$ $(a = 3, q = \frac{4}{5})$ ◀

8.2.4 Harmonische Reihen

Ist $(a_n) = (\frac{1}{n})$, so nennt man $\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} \frac{1}{k}$ endliche harmonische Reihe und $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} \frac{1}{k}$ unendliche harmonische Reihe.

Ist $(a_n) = ((-1)^{n+1}\frac{1}{n})$, dann heißt die Reihe alternierende harmonische Reihe.

Die unendliche harmonische Reihe ist bestimmt divergent, wie folgende Rechnung zeigt:

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{1}{k} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \\ &\quad + \frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16} \\ &\quad + \frac{1}{17} + \ldots \\ &= \left(1 + \frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ &\quad + \left(\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}\right) \\ &\quad + \frac{1}{17} + \ldots \\ &> \left(\frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) \\ &\quad + \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}\right) \\ &\quad + \frac{1}{32} + \ldots \\ &= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots \end{aligned}$$

Die unendliche Reihe $\sum_{k=1}^{\infty} \frac{1}{2}$ ist eine arithmetische Reihe (mit $d = 0$) und deshalb bestimmt divergent. Somit folgt

$$\sum_{k=1}^{\infty} \frac{1}{k} = +\infty$$

Die harmonische Reihe ist bestimmt divergent, obwohl die Glieder der Reihe eine Nullfolge bilden.

Die unendliche alternierende harmonische Reihe ist dagegen konvergent.

Beispiele

1. $\sum_{k=1}^{6} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}$
   (endliche harmonische Reihe)
2. $\sum_{k=1}^{\infty} (-1)^{k+1}\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + (-1)^{n+1}\frac{1}{n} + \ldots = \ln 2$
   (unendliche alternierende harmonische Reihe) ◀

8.2.5 Alternierende Reihen

Ist (a_n) eine alternierende Folge, also eine Folge, deren Glieder abwechselnd unterschiedliches Vorzeichen haben, dann nennt

man $\sum_{k=1}^{n} a_k$ eine endliche alternierende Reihe und $\sum_{k=1}^{\infty} a_k$ eine unendliche alternierende Reihe.

Beispiele

1. $\sum_{k=1}^{10} (-1)^k k = -1+2-3+4-5+6-7+8-9+10$
2. $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$
3. $\sum_{k=1}^{n} (-1)^k = \begin{cases} 0 & \text{für gerades } n \\ -1 & \text{für ungerades } n \end{cases}$ ◀

Für alternierende Reihen gibt es ein einfaches Kriterium, mit dem sich die Konvergenz der Reihe untersuchen lässt:

Eine alternierende Reihe $\sum_{k=1}^{\infty} a_k$, bei der $(|a_n|)$, also die Folge der Beträge der Glieder, eine monoton fallende Nullfolge bildet, ist stets konvergent (Leibniz'sches Konvergenzkriterium).

Beispiel

4. $\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k} = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots+(-1)^{n+1}\frac{1}{n}+\ldots$
   Die alternierende harmonische Reihe ist konvergent nach dem Leibniz'schen Konvergenzkriterium, denn die Folge der Beträge der Glieder, also $(|(-1)^{n+1}\frac{1}{n}|) = (\frac{1}{n})$, ist monoton fallend und eine Nullfolge. ◀

8.3 Grenzwerte von Funktionen

8.3.1 Grenzwert an einer endlichen Stelle

Die Funktion $y = f(x)$ besitzt an der Stelle $x = a$ den Grenzwert $\lim_{x\to a} f(x) = A$ oder $f(x) \to A$ für $x \to a$ (gesprochen: Limes $f(x)$ gleich A für x gegen a), wenn sich die Funktion $f(x)$ bei unbegrenzter Annäherung von x an a unbegrenzt an A nähert. Die Variable x nähert sich a unbegrenzt an, es gilt jedoch stets $x \neq a$. Die Funktion $f(x)$ muss an der Stelle $x = a$ den Wert A nicht annehmen und braucht an dieser Stelle auch nicht definiert zu sein (Abb. 8.2).

Grenzwert

$$\lim_{x\to a} f(x) = A \quad \text{oder} \quad f(x) \to A \quad \text{für } x \to a$$

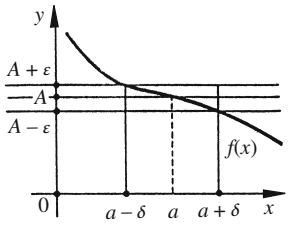


Abb. 8.2 Veranschaulichung des Grenzwertbegriffes

Exakte Definition: Die Funktion $y = f(x)$ besitzt an der Stelle $x = a$ den Grenzwert $\lim_{x\to a} f(x) = A$, wenn sich nach Vorgabe einer beliebig kleinen positiven Zahl ε eine zweite positive Zahl $\delta = \delta(\varepsilon)$ so finden lässt, dass für alle x mit $|x - a| < \delta(\varepsilon)$ gilt $|f(x) - A| < \varepsilon$ eventuell mit Ausnahme der Stelle a.

Der Unterschied $|f(x) - A|$ zwischen den Funktionswerten und dem Grenzwert wird kleiner als jede beliebig vorgegebene positive Zahl ε, wenn die x-Werte sich um weniger als eine passend gewählte, von ε abhängige Zahl $\delta = \delta(\varepsilon)$ vom Wert a unterscheiden, wenn also $0 < |x - a| < \delta(\varepsilon)$ gilt.

Besitzt die Funktion $y = f(x)$ an der Stelle $x = a$ den Grenzwert $\lim_{x\to a} f(x) = A$, so sagt man auch, der Grenzwert $\lim_{x\to a} f(x)$ existiert und ist gleich A.

Beispiele

1. Die Funktion $y = f(x) = x^3$ hat für $x \to 0$ den Grenzwert $A = 0$: $\lim_{x\to 0} x^3 = 0$. Soll etwa $|x^3 - 0|$, der Unterschied zwischen $y = x^3$ und $A = 0$, kleiner als $\varepsilon = 0{,}000001$ sein, so ist dies erfüllt, wenn man für $\delta = \delta(\varepsilon) < 0{,}01$ wählt, denn $(10^{-2})^3 = 10^{-6}$.
   Für ein beliebiges positives ε erfüllt $\delta(\varepsilon) < \sqrt[3]{\varepsilon}$ die geforderte Bedingung.
2. Die Funktion $y = f(x) = \frac{2x^2+5x}{3x}$ ist an der Stelle $x = 0$ nicht definiert, da für $x = 0$ der Nenner Null ist. Es gilt

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \frac{2x^2 + 5x}{3x} = \lim_{x\to 0} \frac{2x + 5}{3}$$

   (Kürzen durch $x \neq 0$) und weiter wegen der Rechenregeln für Grenzwerte (Abschn. 8.3.4)

$$\lim_{x\to 0} f(x) = \frac{2}{3} \lim_{x\to 0} x + \frac{5}{3} = \frac{5}{3}.$$

   Die Funktion $y = f(x)$ besitzt an der Stelle $x = 0$ den Grenzwert $\frac{5}{3}$. ◀

8.3.2 Einseitige Grenzwerte

Die Funktion $y = f(x)$ besitzt an der Stelle $x = a$ den linksseitigen Grenzwert A, wenn sich die Funktion $f(x)$ bei unbegrenzter Annäherung von x von links an a unbegrenzt an A nähert.

Linksseitiger Grenzwert

$$\lim_{\substack{x\to a \\ x<a}} f(x) = \lim_{x\to a-0} f(x) = A$$

Die Funktion $y = f(x)$ besitzt an der Stelle $x = a$ den rechtsseitigen Grenzwert A, wenn sich die Funktion $f(x)$ bei un-

begrenzter Annäherung von x von rechts an a unbegrenzt an A nähert.

Rechtsseitiger Grenzwert

$$\lim_{\substack{x\to a\\ x>a}} f(x) = \lim_{x\to a+0} f(x) = A$$

Die Variable x nähert sich a unbegrenzt an, es gilt jedoch stets $x \neq a$. Die Funktion $f(x)$ muss an der Stelle $x = a$ den Wert A nicht annehmen und braucht an dieser Stelle auch nicht definiert zu sein.

Die Funktion $y = f(x)$ besitzt an der Stelle $x = a$ den Grenzwert A, wenn an dieser Stelle sowohl der linksseitige als auch der rechtsseitige Grenzwert existieren und gleich sind.

Grenzwert

$$\lim_{x\to a} f(x) = A$$

Beispiel

$$f(x) = \begin{cases} 1 & \text{für } x > 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Linksseitiger Grenzwert:

$\lim_{\substack{x\to 0\\ x<0}} f(x) = \lim_{x\to 0-0} f(x) = 0$

Rechtsseitiger Grenzwert:

$\lim_{\substack{x\to 0\\ x>0}} f(x) = \lim_{x\to 0+0} f(x) = 1$

Die Funktion $y = f(x)$ besitzt an der Stelle $x = 0$ sowohl den linksseitigen als auch den rechtsseitigen Grenzwert. Da diese jedoch verschieden sind, existiert der Grenzwert an der Stelle $x = 0$ nicht. ◀

8.3.3 Grenzwert im Unendlichen

Die Funktion $y = f(x)$ besitzt für $x \to \infty$ den Grenzwert A, wenn es zu jedem beliebigen $\varepsilon > 0$ ein hinreichend großes $\omega = \omega(\varepsilon)$ gibt, so dass $|f(x) - A| < \varepsilon$ für alle $x > \omega(\varepsilon)$ gilt. Man schreibt dafür

$$\lim_{x\to\infty} f(x) = A$$

Analog besitzt die Funktion $y = f(x)$ für $x \to -\infty$ den Grenzwert A, wenn es zu jedem beliebigen $\varepsilon > 0$ ein hinreichend großes $\omega = \omega(\varepsilon)$ gilt, so dass $|f(x) - A| < \varepsilon$ für alle $x < -\omega(\varepsilon)$ gilt. Man schreibt dann

$$\lim_{x\to-\infty} f(x) = A$$

Die Grenzwerte $\lim_{x\to\infty} f(x)$ und $\lim_{x\to-\infty} f(x)$ der Funktion $y = f(x)$ beschreiben, falls sie existieren, den Verlauf der Funktion im Unendlichen, das heißt, das Verhalten der Funktion für sehr großes positives und für sehr kleines negatives Argument x.

Beispiele

1. Es ist $\lim_{x\to\infty} \frac{1}{x} = 0$, denn es gilt $|\frac{1}{x} - 0| = |\frac{1}{x}| < \varepsilon$ für alle x, die der Bedingung $x > \omega(\varepsilon) = \frac{1}{\varepsilon}$ genügen. Ebenso gilt $\lim_{x\to-\infty} \frac{1}{x} = 0$.
2. Die Funktion $y = f(x) = \frac{5x+3}{2x+7}$ hat für $x \to \pm\infty$ den Grenzwert $\frac{5}{2}$, also $\lim_{x\to\infty} \frac{5x+3}{2x+7} = \lim_{x\to-\infty} \frac{5x+3}{2x+7} = \frac{5}{2}$, wie folgende Rechnung unter Anwendung der Rechenregeln für Grenzwerte (Abschn. 8.3.4) zeigt:

$$\lim_{x\to\infty} \frac{5x+3}{2x+7} = \lim_{x\to\infty} \frac{5+\frac{3}{x}}{2+\frac{7}{x}} = \lim_{x\to\infty} \frac{5+3\frac{1}{x}}{2+7\frac{1}{x}} = \frac{5}{2}$$

   Die Rechnung für $\frac{5x+3}{2x+7} = \frac{5}{2}$ verläuft ganz analog.
3. Der Grenzwert $\lim_{x\to\infty} \sin x$ existiert nicht. Wie groß man x auch wählt, es lassen sich wegen der Periodizität der Sinusfunktion unendlich viele größere x-Werte angeben, für die die Funktion einen vorgegebenen Wert zwischen -1 und 1 hat. ◀

8.3.4 Rechenregeln für Grenzwerte

Die für Folgen aufgestellten Regeln für das Rechnen mit Grenzwerten (Abschn. 8.1.4) lassen sich auf das Rechnen mit Grenzwerten von Funktionen übertragen.

Gilt $\lim_{x\to a} f(x) = F$ und $\lim_{x\to a} g(x) = G$ für zwei Funktionen $f(x)$ und $g(x)$, so existieren auch die folgenden Grenzwerte:

$$\lim_{x\to a}[f(x) + g(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = F + G$$

$$\lim_{x\to a}[f(x) - g(x)] = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = F - G$$

$$\lim_{x\to a}[c \cdot f(x)] = c \cdot \lim_{x\to a} f(x) = c \cdot F \quad (c \in \mathbb{R})$$

$$\lim_{x\to a}[f(x) \cdot g(x)] = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = F \cdot G$$

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{F}{G} \quad (g(x) \neq 0, G \neq 0)$$

Diese Regeln sagen aus, dass man die Operation der Grenzwertbildung mit der Addition, Subtraktion, Multiplikation und Division (falls $G \neq 0$) vertauschen darf.

Die Regeln wurden schon bei den Beispielen der vorangegangenen Abschnitte angewandt.

8.3.5 Unbestimmte Ausdrücke

Unbestimmte Ausdrücke sind symbolische Ausdrücke der Form

$$\frac{0}{0}, \frac{\infty}{\infty}, 0 \cdot \infty, \infty \cdot 0, \infty - \infty, -\infty + \infty, 0^0, \infty^0, 1^\infty$$

Solche Ausdrücke ergeben sich bei bestimmten Grenzwertaufgaben. Sind zum Beispiel $f(x)$ und $g(x)$ zwei Funktionen mit $f(a) = g(a) = 0$, so ist ihr Quotient $\frac{f(x)}{g(x)}$ an der Stelle $x = a$ nicht definiert. Formales Einsetzen von $x = a$ führt auf den unbestimmten Ausdruck „$\frac{0}{0}$“. Damit soll ausgedrückt werden, dass der Grenzwert $\lim_{x\to a} \frac{f(x)}{g(x)}$ zu ermitteln ist, der das Verhalten des Quotienten in der Nähe der kritischen Stelle a beschreibt, falls er existiert.

Beispiele

1. $$\lim_{x\to\infty} \frac{3x^2 - x - 1}{4x^2 + 3} = ?$$

   Unbestimmter Ausdruck der Form „$\frac{\infty}{\infty}$“.
   Durch Kürzen des Bruches durch x^2 $(x \neq 0)$ ergibt sich

   $$\lim_{x\to\infty} \frac{3x^2 - x - 1}{4x^2 + 3} = \lim_{x\to\infty} \frac{3 - \frac{1}{x} - \frac{1}{x^2}}{4 + \frac{3}{x^2}} = \frac{3}{4}$$

   Der Grenzwert existiert also und ist gleich $\frac{3}{4}$.
2. $\lim_{x\to\infty}(x^2 - x) = ?$
   Unbestimmter Ausdruck der Form „$\infty - \infty$“.
   Durch Umformen ergibt sich

   $$\lim_{x\to\infty} (x^2 - x) = x(x - 1) = \infty \cdot \infty = \infty$$

   Es existiert kein Grenzwert.
3. $$\lim_{x\to\infty} \frac{x^2 - 2}{x^3} = ?$$

   Unbestimmter Ausdruck der Form „$\frac{\infty}{\infty}$“.
   Durch Kürzen des Bruches durch x^3 $(x \neq 0)$ erhält man

   $$\lim_{x\to\infty} \frac{x^2 - 2}{x^3} = \lim_{x\to\infty} \frac{\frac{1}{x} - \frac{2}{x^3}}{1} = \frac{0 - 0}{1} = 0$$

   Der Grenzwert existiert und ist gleich 0.
4. $$\lim_{x\to\infty} \frac{x^3}{x^2 - 2} = ?$$

   Unbestimmter Ausdruck der Form „$\frac{\infty}{\infty}$“.
   Durch Kürzen des Bruches durch x^2 $(x \neq 0)$ ergibt sich

   $$\lim_{x\to\infty} \frac{x^3}{x^2 - 2} = \lim_{x\to\infty} \frac{x}{1 - \frac{2}{x^2}} = \frac{\infty}{1 - 0} = \infty$$

   Es existiert kein Grenzwert. ◄

8.3.6 Stetigkeit einer Funktion

Die Stetigkeit einer Funktion $y = f(x)$ an einer Stelle $x = a$ wird mit Hilfe des Grenzwertes der Funktion an dieser Stelle definiert.

Eine Funktion $y = f(x)$ heißt an der Stelle $x = a$ stetig, wenn $f(x)$ an der Stelle a definiert ist und der Grenzwert $f(x)$ existiert und gleich $f(a)$ ist.

Das ist genau dann der Fall, wenn es zu jedem vorgegebenen $\varepsilon > 0$ ein $\delta = \delta(\varepsilon) > 0$ gibt, so dass $|f(x) - f(a)| < \varepsilon$ für alle x mit $|x - a| < \delta$ gilt.

Ist eine Funktion $y = f(x)$ stetig, dann ändert sich bei kleinen Änderungen der Variablen x auch der Funktionswert $f(x)$ nur geringfügig. Die meisten Funktionen, die in den Anwendungen vorkommen, sind stetig.

Der Graph einer stetigen Funktion ist eine zusammenhängende Kurve. Ist dagegen die Kurve an verschiedenen Stellen (mindestens an einer) unterbrochen, dann heißt die zugehörige Funktion unstetig, und die Werte der unabhängigen Variablen x, an denen die Unterbrechung auftritt, heißen Unstetigkeitsstellen.

Eine an jeder Stelle ihres Definitionsbereichs stetige Funktion $y = f(x)$ heißt stetig.

Sind $f(x)$ und $g(x)$ zwei Funktionen mit dem Definitionsbereich D und dem Wertebereich $W = \mathbb{R}$, und ist c eine reelle Zahl, so gilt:

Sind $f(x)$ und $g(x)$ stetig an der Stelle $x = a$ des Definitionsbereichs D, so sind auch $f(x) + g(x)$, $c \cdot f(x)$, $f(x) \cdot g(x)$, $\frac{f(x)}{g(x)}$ (falls $g(x) \neq 0$ für $x \in D$) und $|f(x)|$ stetig an der Stelle $x = a$. Da die Sinusfunktion $y = \sin x$ eine stetige Funktion ist, folgt hieraus zum Beispiel, dass eine so kompliziert gebaute Funktion wie etwa $f : \mathbb{R} \to \mathbb{R}, f(x) = \frac{x \cdot \sin(x^2+1)}{1+|\sin x|}$ ebenfalls stetig ist.

Beispiele

1. Die Funktion $f(x) = 5x + 2$ ist an jeder Stelle $x = a$ des Definitionsbereichs stetig, denn es gilt $\lim_{x\to a}(5x + 2) = 5a + 2 = f(a)$. Die Funktion ist also eine stetige Funktion.

2. Die Funktion $f(x) = 3x^2$ ist für jedes reelle x stetig, die Funktion ist eine stetige Funktion.
3. Die Funktion $f(x) = \begin{cases} 1 & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$ besitzt für $x = 0$ eine Unstetigkeitsstelle, also ist $y = f(x)$ eine unstetige Funktion. ◀

8.3.7 Unstetigkeitsstellen

Eine Unstetigkeitsstelle ist eine Stelle $x = a$ einer Funktion $y = f(x)$, an der die Funktion nicht stetig ist.

Die Kurve einer Funktion ist an einer Unstetigkeitsstelle unterbrochen. Eine Funktion, die mindestens eine Unstetigkeitsstelle besitzt, heißt unstetig.

Die häufigsten Unstetigkeitsstellen sind Sprungstellen und Pole.

An einer Sprungstelle $x = a$ sind der rechtsseitige Grenzwert $\lim_{x \to a+0} f(x)$ und der linksseitige Grenzwert $\lim_{x \to a-0} f(x)$ voneinander verschieden. Die Funktion $f(x)$ springt beim Durchlaufen des Punktes $x = a$ von einem auf einen anderen endlichen Wert. Die Funktion $f(x)$ braucht für $x = a$ nicht definiert zu sein.

Ein Pol oder eine Unendlichkeitsstelle $x = a$ einer Funktion $y = f(x) = \frac{g(x)}{h(x)}$ ist eine Stelle, für die der Nenner von $f(x)$ den Wert 0 hat und der Zähler von 0 verschieden ist, also $h(a) = 0$ und $g(a) \neq 0$ (Abschn. 5.5). An einer solchen Stelle ist die Funktion also nicht definiert. Die Funktion strebt bei Annäherung an einen Pol nach (plus oder minus) Unendlich. Die Kurve der Funktion läuft an einer solchen Stelle ins Unendliche.

Beispiele

1. $f(x) = \begin{cases} 1 & \text{für } x > 0 \\ 0 & \text{für } x < 0 \end{cases}$
   Linksseitiger Grenzwert:
   $\lim_{\substack{x \to 0 \\ x<0}} f(x) = \lim_{x \to 0-0}(x) = 0$
   Rechtsseitiger Grenzwert:
   $\lim_{\substack{x \to 0 \\ x>0}} f(x) = \lim_{x \to 0+0} f(x) = 1$
   Der linksseitige und der rechtsseitige Grenzwert der Funktion $y = f(x)$ sind verschieden, also besitzt die Funktion bei $x = 0$ eine Sprungstelle. Die Funktion springt beim Durchlaufen des Punktes $x = 0$ von 0 auf 1 (Abb. 8.3).
2. $f(x) = \frac{1}{x}$, $D = \mathbb{R}, x \neq 0$
   Einseitige Grenzwerte:
   $\lim_{\substack{x \to 0 \\ x<0}} \frac{1}{x} = -\infty$, $\lim_{\substack{x \to 0 \\ x>0}} \frac{1}{x} = +\infty$
   Die Funktion $y = \frac{1}{x}$ besitzt bei $x = 0$ einen Pol. Bei Annäherung von links an den Pol strebt die Funktion nach minus Unendlich, bei Annäherung von rechts nach plus Unendlich (Abb. 8.4). ◀

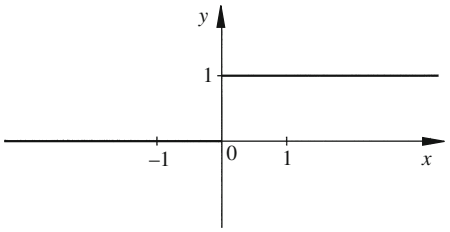


Abb. 8.3 Graph der Funktion von Beispiel 1

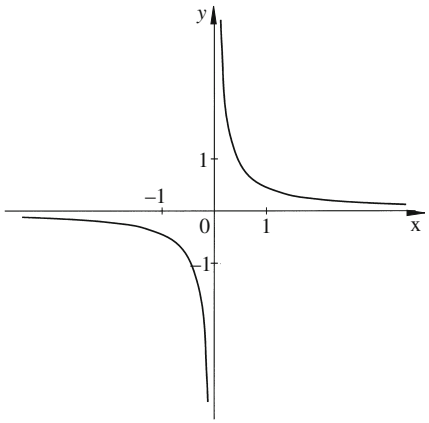


Abb. 8.4 Graph der Funktion von Beispiel 2

8.4 Ableitung einer Funktion

8.4.1 Definitionen

Existiert für eine Funktion $y = f(x)$ mit dem Definitionsbereich D der Grenzwert

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad (x_0 \in D)$$

dann nennt man $f'(x_0)$ die Ableitung der Funktion $f(x)$ an der Stelle $x = x_0$ (gesprochen: f Strich von x_0). Die Funktion $f(x)$ heißt dann differenzierbar in x_0.

Statt $f'(x_0)$ schreibt man auch $y'(x_0)$ oder $\frac{dy}{dx}(x_0)$ oder $\frac{df}{dx}(x_0)$ (gesprochen: y Strich von x_0 bzw. dy nach dx an der Stelle x_0 bzw. df nach dx an der Stelle x_0).

Der Bruch $\frac{f(x)-f(x_0)}{x-x_0}$ heißt auch Differenzenquotient, da im Zähler die Differenz zweier Funktionswerte und im Nenner die Differenz zweier x-Werte steht. Deshalb nennt man den Grenzwert $f'(x_0) = \lim_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ statt Ableitung auch Differentialquotient.

Geometrische Deutung: Ist die Funktion $y = f(x)$ als Kurve in einem kartesischen Koordinatensystem dargestellt, dann ist der Differenzenquotient gleich der Steigung (also dem Tangens des Steigungswinkels β) der Sekante durch die Punkte $P_0(x_0|f(x_0))$

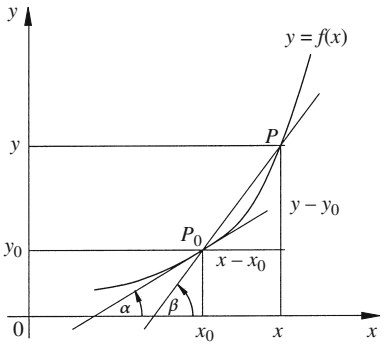


Abb. 8.5 Geometrische Deutungen des Differenzen- und des Differentialquotienten

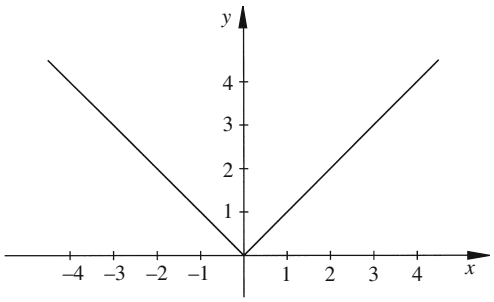


Abb. 8.6 Graph der Funktion von Beispiel 4

und $P(x|f(x))$. Der Grenzwert $f'(x_0)$ ist die Steigung der Tangente in x_0 an den Graphen von $f(x)$, also $f'(x_0) = \tan\alpha$. Dabei ist α der Winkel zwischen der x-Achse und der Tangente an den Graphen in x_0, wobei der Winkel von der positiven x-Achse zur Tangente im entgegengesetzten Drehsinn des Uhrzeigers gemessen wird (Abb. 8.5).

Anschaulich bedeutet die Existenz der Ableitung an der Stelle $x = x_0$, dass der Kurvenverlauf in x_0 glatt ist (keine „Knickstelle“ hat).

Eine Funktion $y = f(x)$ heißt (generell) differenzierbar, wenn sie an jeder Stelle ihres Definitionsbereichs differenzierbar ist. Dann heißt die durch $g(x) = f'(x)$ definierte Funktion $y' = f'(x)$ die Ableitung oder die Ableitungsfunktion von $f(x)$.

Eine an der Stelle x_0 differenzierbare Funktion $y = f(x)$ ist dort auch stetig. Falls $f(x)$ an der Stelle x_0 nicht stetig ist, kann $f(x)$ dort auch nicht differenzierbar sein. Aus der Stetigkeit an der Stelle x_0 folgt jedoch noch nicht die Differenzierbarkeit an dieser Stelle.

Eine Funktion $y = f(x)$ heißt stetig differenzierbar, wenn $f(x)$ differenzierbar ist und die Ableitung $f'(x)$ eine stetige Funktion ist.

Beispiele

1. Für die konstante Funktion $f(x) = c$ $(c \in \mathbb{R})$ gilt

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\to x_0} \frac{c-c}{x-x_0} = \lim_{x\to x_0} 0 = 0$$

   Die Ableitungsfunktion einer konstanten Funktion ist somit $f'(x) = 0$.
2. Für die Funktion $f(x) = ax + b$, $a, b \in \mathbb{R}$, $D = \mathbb{R}$ gilt

$$\begin{aligned} f'(x_0) &= \lim_{x\to x_0} \frac{ax+b-(ax_0+b)}{x-x_0} \\ &= \lim_{x\to x_0} \frac{a\cdot(x-x_0)}{x-x_0} = \lim_{x\to x_0} a = a \end{aligned}$$

   Die Funktion $f(x) = ax + b$ ist ebenfalls eine (überall) differenzierbare Funktion, und es gilt $f'(x) = a$.
3. Die Funktion $f(x) = x^2$ mit $D = \mathbb{R}$ ist in jedem Punkt $x_0 \in D$ differenzierbar. Es ist

$$\frac{f(x)-f(x_0)}{x-x_0} = \frac{x^2-x_0^2}{x-x_0} = \frac{(x+x_0)(x-x_0)}{x-x_0} = x + x_0$$

   und somit

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\to x_0} (x+x_0) = 2x_0$$

   Die Funktion $f(x) = x^2$ ist eine (überall) differenzierbare Funktion, und es gilt $f'(x) = 2x$.
4. Die Funktion $f(x) = |x| = \begin{cases} x & \text{für } x \geq 0 \\ -x & \text{für } x < 0 \end{cases}$ ist (überall) stetig.
   Für $x_0 > 0$ ist $f'(x_0) = 1$, für $x_0 < 0$ ist dagegen $f'(x_0) = -1$.
   Für $x_0 = 0$ gilt wegen $f(0) = 0$

$$\lim_{\substack{x\to 0\\ x>0}} |x| = \lim_{\substack{x\to 0\\ x>0}} \frac{x-0}{x} = \lim_{\substack{x\to 0\\ x>0}} 1 = 1 \quad \text{und}$$

$$\lim_{\substack{x\to 0\\ x<0}} |x| = \lim_{\substack{x\to 0\\ x<0}} \frac{-x-0}{x} = \lim_{\substack{x\to 0\\ x<0}} (-1) = -1$$

   Der Differentialquotient existiert nicht, da der rechtsseitige Grenzwert und der linksseitige Grenzwert nicht gleich sind.
   Die Funktion $f(x) = |x|$ ist an der Stelle $x_0 = 0$ stetig, aber nicht differenzierbar (Abb. 8.6). ◀

8.4.2 Differentiationsregeln

Die folgenden Regeln gelten sowohl für die Ableitungen einer Funktion $y = f(x)$ an einer bestimmten Stelle $x = x_0$ als auch für die Ableitungsfunktionen $y' = f'(x)$.

1. *Konstante Funktion*
   Die Ableitung einer konstanten Funktion ist Null (vgl. Beispiel 1 im Abschn. 8.4.1).

$$y = f(x) = c \ (c \in \mathbb{R}, \text{konstant}) \quad \Rightarrow \quad y' = 0$$

Beispiel

$y = 3 \Rightarrow y' = 0.$ ◀

2. *Faktorregel*
Die Ableitung einer Funktion mal konstantem Faktor ist gleich konstanter Faktor mal abgeleitete Funktion.

$$y = c \cdot f(x) \ (c \in \mathbb{R}, \text{konstant}) \quad \Rightarrow \quad y' = c \cdot f'(x)$$

Herleitung:

$$\begin{aligned}(c \cdot f)'(x_0) &= \lim_{x \to x_0} \frac{c \cdot f(x) - c \cdot f(x_0)}{x - x_0} \\ &= \lim_{x \to x_0} c \cdot \frac{f(x) - f(x_0)}{x - x_0} \\ &= c \cdot \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = c \cdot f'(x_0)\end{aligned}$$

Beispiel

$y = 3x^2 \Rightarrow y' = 3 \cdot 2x = 6x$, denn $\frac{d}{dx} x^2 = 2x.$ ◀

3. *Summenregel*
Die Ableitung der Summe (Differenz) zweier Funktionen ist gleich der Summe (Differenz) der Ableitungen der Funktionen.

$$\begin{aligned} y = f(x) + g(x) \quad &\Rightarrow \quad y' = f'(x) + g'(x) \\ y = f(x) - g(x) \quad &\Rightarrow \quad y' = f'(x) - g'(x)\end{aligned}$$

Herleitung:

$$\begin{aligned}(f + g)'(x_0) &= \lim_{x \to x_0} \frac{f(x) + g(x) - (f(x_0) + g(x_0))}{x - x_0} \\ &= \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} + \frac{g(x) - g(x_0)}{x - x_0} \right) \\ &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} + \lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0} \\ &= f'(x_0) + g'(x_0)\end{aligned}$$

Die Herleitung für die Ableitung der Differenz zweier Funktionen verläuft ganz analog.

Beispiel

$y = x^2 + 3x \Rightarrow y' = 2x + 3$, denn $\frac{d}{dx} x^2 = 2x$ und $\frac{d}{dx} 3x = 3.$ ◀

4. *Produktregel*
Die Ableitung des Produkts zweier Funktionen ist gleich der Summe aus der ersten Funktion multipliziert mit der Ableitung der zweiten Funktion und der zweiten Funktion multipliziert mit der Ableitung der ersten Funktion.

$$y = f(x) \cdot g(x) \quad \Rightarrow \quad y' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Herleitung:

$$\begin{aligned}(f \cdot g)'(x_0) &= \lim_{x \to x_0} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} \\ &= \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} g(x) + f(x_0) \frac{g(x) - g(x_0)}{x - x_0} \right) \\ &= f'(x_0) \cdot g(x_0) + f(x_0) \cdot g'(x_0)\end{aligned}$$

Für die Ableitung des Produkts von drei Funktionen gilt

$$\begin{aligned} & y = f(x) \cdot g(x) \cdot h(x) \\ \Rightarrow \quad & y' = f(x) \cdot g(x) \cdot h'(x) + f(x) \cdot g'(x) \cdot h(x) \\ & \quad + f'(x) \cdot g(x) \cdot h(x)\end{aligned}$$

Mehrfache Anwendung der Produktregel ergibt die Ableitung der Potenzfunktion.

Potenzregel

$$y = x^n \ (n \in \mathbb{N}) \quad \Rightarrow \quad y' = nx^{n-1}$$

Mit Hilfe von Polynomdivision lässt sich dieses Ergebnis auch direkt herleiten:

$$\begin{aligned}\frac{d}{dx} x^n(x_0) &= \lim_{x \to x_0} \frac{x^n - x_0^n}{x - x_0} \\ &= \lim_{x \to x_0} (x^{n-1} + x_0 x^{n-2} + x_0^2 x^{n-3} + \ldots + x_0^{n-1}) \\ &= nx^{n-1}\end{aligned}$$

Durch Anwendung von Quotienten- und Kettenregel (siehe unten) kann man dieses Ergebnis auf reelle Exponenten ausweiten.

$$y = x^r \ (r \in \mathbb{R}) \quad \Rightarrow \quad y' = rx^{r-1}$$

Summen- und Potenzregel zusammen ergeben die Ableitung eines Polynoms.

$$y = \sum_{k=0}^{n} c_k x^k = c_0 + c_1 x + \ldots + c_n x^n$$

$$\Rightarrow \quad y' = \sum_{k=1}^{n} k \cdot c_k x^{k-1} = c_1 + 2c_2 x + \ldots + n c_n x^{n-1}$$

Beispiele

1. $y = 3x^2 \cdot \sin x \Rightarrow y' = 3x^2 \cdot \cos x + 6x \cdot \sin x$
2. $y = x^7 \Rightarrow y' = 7x^6$
3. $y = x^{\frac{7}{3}} \Rightarrow y' = \frac{7}{3} x^{\frac{4}{3}}$
4. $y = 3x^7 - 5x^4 + x^2 + 3 \Rightarrow y' = 21x^6 - 20x^3 + 2x$ ◀

5. *Quotientenregel*
   Die Ableitung des Quotienten zweier Funktionen ist gleich der Differenz der Ableitung der Zählerfunktion multipliziert mit der Nennerfunktion und der Zählerfunktion multipliziert mit der Ableitung der Nennerfunktion dividiert durch das Quadrat der Nennerfunktion.

$$y = \frac{f(x)}{g(x)} \qquad (g(x) \neq 0)$$

$$\Rightarrow \quad y' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}$$

Der Zähler von y' beginnt also mit der Ableitung der Zählerfunktion $f(x)$.

Herleitung:

$$\begin{aligned}
&\left(\frac{f}{g}\right)'(x_0) \\
&= \lim_{x \to x_0} \frac{\frac{f(x)}{g(x)} - \frac{f(x_0)}{g(x_0)}}{x - x_0} \\
&= \lim_{x \to x_0} \frac{1}{g(x)g(x_0)} \left(\frac{f(x) - f(x_0)}{x - x_0} g(x_0) - f(x_0) \frac{g(x) - g(x_0)}{x - x_0} \right) \\
&= \frac{f'(x_0) \cdot g(x_0) - f(x_0) \cdot g'(x_0)}{g^2(x_0)}
\end{aligned}$$

Im Spezialfall, dass $f(x)$ eine konstante Funktion mit $f(x) = 1$ ist, gilt

$$y = \frac{1}{g(x)} \quad \Rightarrow \quad y' = -\frac{g'(x)}{g^2(x)}$$

Beispiele

1. $y = \frac{5x-1}{2x+3} \Rightarrow y' = \frac{(2x+3)\cdot 5 - (5x-1)\cdot 2}{(2x+3)^2} = \frac{17}{(2x+3)^2}$
2. $y = \frac{x^3}{x^2-1} \Rightarrow y' = \frac{3x^2(x^2-1) - 2x \cdot x^3}{(x^2-1)^2} = \frac{x^2(x^2-3)}{(x^2-1)^2}$
3. $y = \frac{1}{x^2+3x} \Rightarrow y' = -\frac{2x+3}{(x^2+3x)^2}$ ◀

6. *Kettenregel*
   Die Kettenregel ist eine Regel zur Differentiation zusammengesetzter Funktionen.
   Ist $y = F(x)$ eine zusammengesetzte Funktion, also $F(x) = f(h(x))$, und setzt man $z = h(x)$, dann ist $y = F(x)$ differenzierbar, wenn die Funktionen $y = f(z)$ und $z = h(x)$ differenzierbar sind, und es gilt

$$\begin{aligned}
y' = F'(x) &= \frac{df}{dz} \cdot \frac{dh}{dx} = f'(z) \cdot h'(x) \\
&= f'(h(x)) \cdot h'(x)
\end{aligned}$$

Man nennt $f'(h(x))$ die äußere Ableitung und $h'(x)$ die innere Ableitung der Funktion $y = f(h(x))$.

Beispiele

1. $y = F(x) = (x^3 - 2x + 1)^3$, also $z = h(x) = x^3 - 2x + 1$ und $y = f(z) = z^3$

$$\begin{aligned}
\Rightarrow \quad y' = F'(x) &= f'(z) \cdot h'(x) = 3z^2 \cdot (3x^2 - 2) \\
&= 3(x^3 - 2x + 1)^2 \cdot (3x^2 - 2)
\end{aligned}$$

2. $y = F(x) = \sqrt{5x^2 - 7x + 8}$, also $z = h(x) = 5x^2 - 7x + 8$ und $y = f(z) = \sqrt{z}$

$$\begin{aligned}
\Rightarrow \quad y' = F'(x) &= f'(z) \cdot h'(x) = \frac{1}{2} z^{-\frac{1}{2}} \cdot (10x - 7) \\
&= \frac{10x - 7}{2\sqrt{5x^2 - 7x + 8}}
\end{aligned}$$

3. $y = F(x) = \sin(4x^2 + 1)$, also $z = h(x) = 4x^2 + 1$ und $y = f(z) = \sin z$

$$\begin{aligned}
\Rightarrow \quad y' = F'(x) &= f'(z) \cdot h'(x) = \cos z \cdot 8x \\
&= 8x \cdot \cos(4x^2 + 1)
\end{aligned}$$

   Die Ableitung von $y = \sin x$ ist $y' = \cos x$ (Abschn. 8.4.5.1). ◀

7. *Ableitung der Umkehrfunktion*
   Ist $y = f(x)$ eine differenzierbare Funktion mit $f'(x) \neq 0$, die eine Umkehrfunktion $y = f^{-1}(x)$ besitzt, so ist auch die Umkehrfunktion differenzierbar, und es gilt

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

Herleitung:

Aus $f(f^{-1}(x)) = x$ (Abschn. 5.2.8) folgt durch Differentiation mit Anwendung der Kettenregel $f'(f^{-1}(x)) \cdot (f^{-1})'(x) = 1$ und daraus wegen $f'(x) \neq 0$ durch Auflösen nach $(f^{-1})'(x)$ die behauptete Gleichung für die Ableitung der Umkehrfunktion.

8.4.3 Höhere Ableitungen

Ist die Funktion $y = f(x)$ differenzierbar oder zumindest in einem ganzen Intervall ihres Definitionsbereichs differenzierbar, so kann dort also an jeder Stelle die Ableitung $f'(x)$ gebildet werden. Dann ist $y = f'(x)$ wieder eine Funktion von x. Ist diese Funktion wieder differenzierbar, so nennt man diese Ableitung der (ersten) Ableitung die zweite Ableitung der Ausgangsfunktion $y = f(x)$, geschrieben $f''(x)$ oder $y''(x)$ oder $\frac{d^2y}{dx^2}(x)$ oder $\frac{d^2f}{dx^2}(x)$ (gesprochen: f zwei Strich von x bzw. y zwei Strich von x bzw. d zwei y nach dx Quadrat an der Stelle x bzw. d zwei f nach dx Quadrat an der Stelle x).

Entsprechend kann es auch eine dritte, vierte, ... Ableitung von $f(x)$ geben. Die n-te Ableitung von $f(x)$ schreibt man

$$f^{(n)}(x) = y^{(n)}(x) = \frac{d^n y}{dx^n}(x) = \frac{d^n f}{dx^n}(x)$$

Beispiele

1. Die Funktion $f(x) = x^5$ hat als (erste) Ableitung $f'(x) = 5x^4$, als zweite Ableitung $f''(x) = 20x^3$, als dritte Ableitung $f'''(x) = 60x^2$, als vierte Ableitung $f^{(4)}(x) = 120x$ und als fünfte Ableitung $f^{(5)}(x) = 120$. Alle höheren Ableitungen sind Null, also $f^{(k)}(x) = 0$ für $k = 6, 7, \ldots$
2. $f(x) = 4x^4 - 12x^3 + 5x - 2 \Rightarrow f'(x) = 16x^3 - 36x^2 + 5$, $f''(x) = 48x^2 - 72x$, $f'''(x) = 96x - 72$, $f^{(4)}(x) = 96$, $f^{(5)}(x) = f^{(6)}(x) = \ldots = 0$
3. $f(x) = \frac{x^2}{(x-1)^2} \Rightarrow f'(x) = -\frac{2x}{(x-1)^3}$, $f''(x) = \frac{2(2x+1)}{(x-1)^4}$, $f'''(x) = -\frac{12(x+1)}{(x-1)^5}, \ldots$ ◀

8.4.4 Ableitungen einiger algebraischer Funktionen

Mit den Differentiationsregeln aus Abschn. 8.4.2 lassen sich die Ableitungen von algebraischen Funktionen berechnen.

Rationale Funktionen

$$\begin{aligned}
&y = c \ (c \text{ konstant}) && \Rightarrow \quad y' = 0\\
&y = x && \Rightarrow \quad y' = 1\\
&y = x^n && \Rightarrow \quad y' = nx^{n-1}\\
&y = c_n x^n + c_{n-1}x^{n-1} + \ldots + c_2 x^2 + c_1 x + c_0 &&\\
\Rightarrow \quad &y' = nc_n x^{n-1} + (n-1)c_{n-1}x^{n-2} + \ldots + 2c_2 x + c_1 &&\\
&y = \frac{1}{x} && \Rightarrow \quad y' = -\frac{1}{x^2}\\
&y = \frac{1}{x^n} && \Rightarrow \quad y' = -\frac{n}{x^{n+1}}\\
&y = \frac{x^m}{x^n} && \Rightarrow \quad y' = \frac{(m-n)x^m}{x^{n+1}}
\end{aligned}$$

Irrationale Funktionen

$$\begin{aligned}
&y = \sqrt{x} && \Rightarrow \quad y' = \frac{1}{2\sqrt{x}}\\
&y = \sqrt[n]{x} && \Rightarrow \quad y' = \frac{1}{n\sqrt[n]{x^{n-1}}}\\
&y = \frac{\sqrt[m]{x}}{\sqrt[n]{x}} && \Rightarrow \quad y' = \frac{n-m}{mn}\,\frac{\sqrt[m]{x}}{\sqrt[n]{x^{n+1}}}
\end{aligned}$$

8.4.5 Ableitungen einiger transzendenter Funktionen

8.4.5.1 Trigonometrische Funktionen

Zur Herleitung der Ableitung der Sinusfunktion wird der Grenzwert von $\frac{\sin x}{x}$ für $x \to 0$ benötigt. Es gilt

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

Beweis (Abb. 8.7):

Zeichnet man einen Einheitskreis in einem kartesischen Koordinatensystem mit dem Mittelpunkt im Koordinatenursprung O, so ist die Länge des Lotes eines Kreispunktes P auf die Abszisse $\sin x$, wenn die Strecke $\overline{OP}$ mit der Abszisse den Winkel x (im Bogenmaß) bildet. Der zum Winkel gehörige Bogen hat die Länge x, da der Radius des Kreises 1 ist (Abschn. 6.2). Die Projektion von $\overline{OP}$ auf die Abszisse hat die Länge $\cos x$. Der Teil der zur Abszisse senkrechten Tangente an den Einheitskreis, der zwischen der Abszisse und der Verlängerung der Strecke $\overline{OP}$ liegt, hat die Länge $\tan x$.

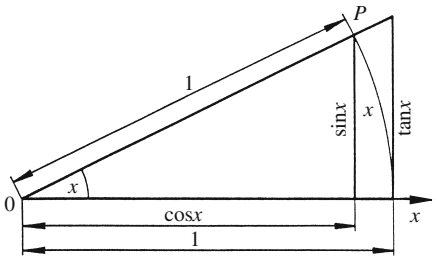


Abb. 8.7 Zur Herleitung von $\lim_{x\to 0}\frac{\sin x}{x}=1$

Die Fläche A_1 des rechtwinkligen Dreiecks mit den Kathetenlängen $\sin x$ und $\cos x$ ist $A_1 = \frac{\sin x\cdot\cos x}{2}$, die Fläche A_2 des Einheitskreissektors mit dem Winkel x ist $A_2 = \frac{x}{2}$, und die Fläche A_3 des rechtwinkligen Dreiecks mit den Kathetenlängen 1 und $\tan x$ ist $A_3 = \frac{1\cdot\tan x}{2}$. Wegen $A_1 < A_2 < A_3$ folgt $\frac{\sin x\cdot\cos x}{2} < \frac{x}{2} < \frac{1\cdot\tan x}{2}$.

Durch Multiplikation mit $\frac{2}{\sin x}$ folgt

$$\cos x < \frac{x}{\sin x} < \frac{1}{\cos x} \quad\Rightarrow\quad \cos x < \frac{\sin x}{x} < \frac{1}{\cos x}$$

Die Grenzwerte der beiden äußeren Funktionen an der Stelle $x = 0$ sind 1, also muss auch der Grenzwert der mittleren Funktion an der Stelle $x = 0$ gleich 1 sein:

$$\lim_{x\to 0}\frac{\sin x}{x} = 1.$$

Die Ableitung der Funktion $y = f(x) = \sin x$ ist die Funktion $y' = f'(x) = \cos x$.

$$y = \sin x \quad\Rightarrow\quad y' = \cos x$$

Beweis:

Für die Differenz $\sin\alpha - \sin\beta$ zweier trigonometrischer Funktionen gilt $\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$. Damit folgt

$$\begin{aligned}(\sin x)'(x_0) &= \lim_{x\to x_0}\frac{\sin x - \sin x_0}{x - x_0} = \lim_{x\to x_0}\frac{2\cos\frac{x+x_0}{2}\sin\frac{x-x_0}{2}}{x - x_0}\\ &= \lim_{x\to x_0}\cos\frac{x+x_0}{2}\cdot\lim_{x\to x_0}\frac{\sin\frac{x-x_0}{2}}{\frac{x-x_0}{2}} = \cos x_0,\end{aligned}$$

denn $\lim_{x\to x_0}\frac{\sin\frac{x-x_0}{2}}{\frac{x-x_0}{2}} = 1$, wie oben bewiesen wurde.

$$y = \cos x \quad\Rightarrow\quad y' = -\sin x$$

Beweis:

Aus $\sin(90° - \alpha) = \cos\alpha$ (Abschn. 6.1) folgt $y = \cos x = \sin(\frac{\pi}{2} - x)$. Anwendung der Kettenregel mit $z = h(x) = \frac{\pi}{2} - x$ und $y = f(z) = \sin z$ ergibt $y' = f'(z)\cdot h'(x) = \cos z\cdot(-1) = -\cos(\frac{\pi}{2} - x)$. Wegen $\cos(90^{circ} - \alpha) = \sin\alpha$ (Abschn. 6.1) folgt die Behauptung $y' = -\sin x$.

$$y = \tan x \quad\Rightarrow\quad y' = \frac{1}{\cos^2 x}$$

Beweis:

Wegen $\tan x = \frac{\sin x}{\cos x}$ folgt mit der Quotientenregel und den Ableitungen von $\sin x$ und $\cos x$:

$$y' = \frac{\cos x\cdot\cos x - \sin x\cdot(-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

$$y = \cot x \quad\Rightarrow\quad y' = -\frac{1}{\sin^2 x}$$

Beweis:

Wegen $\cot x = \frac{\cos x}{\sin x}$ folgt

$$\begin{aligned}y' &= \frac{\sin x\cdot(-\sin x) - \cos x\cdot\cos x}{\sin^2 x} = \frac{-(\sin^2 x + \cos^2 x)}{\sin^2 x}\\ &= \frac{-1}{\sin^2 x}\end{aligned}$$

8.4.5.2 Logarithmusfunktionen

Der Grenzwert der Zahlenfolge $(a_n) = ((1 + \frac{1}{n})^n)$ heißt Euler'sche Zahl (Name nach dem schweizerischen Mathematiker Leonhard Euler, 1707–1783) und wird mit dem Buchstaben e bezeichnet.

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2{,}7182818284\ldots$$

Die Euler'sche Zahl ist eine irrationale Zahl; sie ist Basis der natürlichen Logarithmen (Abschn. 1.8.2).

Mit Hilfe dieses Grenzwertes lassen sich die Ableitungen der Logarithmusfunktionen $y = \log_a x$ berechnen. Es ergibt sich mit den Regeln der Logarithmenrechnung (Abschn. 1.8.3)

$$\begin{aligned}(\log_a x)'(x_0) &= \lim_{x\to x_0}\frac{\log_a x - \log_a x_0}{x - x_0} = \lim_{x\to x_0}\frac{1}{x - x_0}\log_a\frac{x}{x_0}\\ &= \lim_{x\to x_0}\frac{1}{x_0}\frac{x_0}{x - x_0}\log_a\left(1 + \frac{x - x_0}{x_0}\right)\\ &= \lim_{x\to x_0}\frac{1}{x_0}\log_a\left(1 + \frac{1}{\frac{x_0}{x - x_0}}\right)^{\frac{x_0}{x-x_0}}\\ &= \frac{1}{x_0}\log_a\lim_{x\to x_0}\left(1 + \frac{1}{\frac{x_0}{x - x_0}}\right)^{\frac{x_0}{x-x_0}} = \frac{1}{x_0}\log_a e\end{aligned}$$

Bei der vorletzten Umformung wurde (der hier nicht hergeleitete) Satz benutzt, dass der Grenzwert des Logarithmus gleich dem Logarithmus des Grenzwertes ist. Wegen $\log_a e = \frac{\ln e}{\ln a} = \frac{1}{\ln a}$ (Abschn. 1.8.4) ergibt sich

$$y = \log_a x \quad \Rightarrow \quad y' = \log_a e \cdot \frac{1}{x} = \frac{1}{\ln a} \cdot \frac{1}{x}$$

Für die natürliche Logarithmusfunktion mit der Basis $a = e$ folgt

$$y = \ln x \quad \Rightarrow \quad y' = \frac{1}{x}$$

8.4.5.3 Exponentialfunktionen

Die e-Funktion ist die Umkehrfunktion der natürlichen Logarithmusfunktion. Nach der Regel zur Berechnung der Ableitung der Umkehrfunktion (Abschn. 8.4.2) folgt $\frac{d}{dx}e^x = \frac{1}{\frac{1}{e^x}} = e^x$ wegen $\frac{d}{dx}\ln x = \frac{1}{x}$.

$$y = e^x \quad \Rightarrow \quad y' = e^x$$

Daraus folgt unmittelbar $f(x) = f'(x) = f''(x) = f'''(x) = \ldots = f^{(n)}(x) = \ldots = e^x$ für alle $n \in \mathbb{N}$. Die e-Funktion hat also die bemerkenswerte Eigenschaft, dass sie mit allen ihren Ableitungen übereinstimmt.

Für die allgemeine Exponentialfunktion gilt

$$y = a^x \quad \Rightarrow \quad y' = a^x \cdot \ln a$$

Beweis:

Wegen $a^x = e^{x \cdot \ln a}$ (Abschn. 5.7.1) ergibt sich aus der Ableitung der e-Funktion zusammen mit der Kettenregel $y' = (a^x)' = (e^{x \cdot \ln a})' = e^{x \cdot \ln a} \cdot \ln a = a^x \cdot \ln a$.

8.4.5.4 Zusammenfassende Übersicht

Trigonometrische Funktionen

$y = \sin x \quad \Rightarrow \quad y' = \cos x$

$y = \cos x \quad \Rightarrow \quad y' = -\sin x$

$y = \tan x \quad \Rightarrow \quad y' = \frac{1}{\cos^2 x} \quad (x \neq (2k+1)\frac{\pi}{2}, k \in \mathbb{Z})$

$y = \cot x \quad \Rightarrow \quad y' = -\frac{1}{\sin^2 x} \quad (x \neq k\pi, k \in \mathbb{Z})$

Exponentialfunktionen

$y = e^x \quad \Rightarrow \quad y' = e^x = y$

$y = a^x \quad \Rightarrow \quad y' = a^x \ln a \quad (a \in \mathbb{R}, a > 0 \text{ konstant})$

Logarithmusfunktionen

$y = \ln x \quad \Rightarrow \quad y' = \frac{1}{x} (x > 0)$

$y = \log_a x \quad \Rightarrow \quad y' = \frac{1}{x}\log_a e = \frac{1}{lna} \cdot \frac{1}{x}$

$(a \in \mathbb{R}, a > 0, a \neq 1 \text{ konstant}, x > 0)$

8.4.6 Sekanten und Tangenten

Eine Sekante ist eine Gerade, die eine Kurve, also den Graph einer Funktion $y = f(x)$, in (mindestens) zwei Punkten schneidet (Sekante = Schneidende). Der Teil zwischen den Schnittpunkten heißt Sehne.

Die Gleichung der Sekante durch die Punkte $P_1(x_1|f(x_1))$ und $P_2(x_2|f(x_2))$ lautet

$$y = \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1) + f(x_1)$$

Beispiele

1. $f(x) = x^2, P_1(0|0), P_2(1|1)$
   Die Gleichung der Sekante durch die Punkte P_1 und P_2 lautet
   $$y = \frac{1-0}{1-0}(x-0) + 0, \quad \text{also } y = x.$$
2. $f(x) = x^3 - 2x + 1, P_1(-1|2), P_2(2|5)$
   Die Gleichung der Sekante durch die Punkte P_1 und P_2 lautet
   $$y = \frac{5-2}{2-(-1)}(x-(-1)) + 2 = x + 3. \quad \blacktriangleleft$$

Eine Tangente ist eine Gerade, die den Graph einer Funktion $y = f(x)$ in einem Punkt berührt, aber nicht schneidet (Tangente = Berührende).

Die Funktion $f(x)$ hat in dem Punkt $P(a|f(a))$ genau dann eine Tangente, wenn die Funktion in a differenzierbar ist. Die Ableitung der Funktion an der Stelle, also $f'(a)$, ist die Steigung der

Tangente. Die Gleichung der Tangente an die Kurve im Punkt $P(a|f(a))$ lautet

$$y = f'(a)(x-a) + f(a)$$

Beispiele

3. $f(x) = x^2$, $P(1|1)$
   $f'(x) = 2x \Rightarrow f'(1) = 2$
   Die Gleichung der Tangente an die Kurve im Punkt $P(1|1)$ lautet somit $y = 2(x-1) + 1 = 2x - 1$.
4. $f(x) = x^3 - 2x + 1$, $P(1|0)$
   $f'(x) = 3x^2 - 2 \Rightarrow f'(1) = 1$
   Die Gleichung der Tangente an die Kurve im Punkt $P(1|0)$ lautet somit $y = 1 \cdot (x-1) + 0 = x - 1$. ◀

8.4.7 Extremwerte von Funktionen

Eine Funktion $y = f(x)$ besitzt an der Stelle $x = a$ ein relatives Maximum, wenn es eine Umgebung von a gibt, in der alle Funktionswerte kleiner als an der Stelle $x = a$ sind. Dieser Funktionswert $f(a)$ heißt relatives Maximum. Es gilt dann $f(x) < f(a)$ für alle $x \neq a$ aus einer passenden Umgebung von a. Alle benachbarten Funktionswerte sind also kleiner als $f(a)$.

Relatives Maximum $f(a)$

$$f(x) < f(a) \quad \text{für } x \neq a$$

Entsprechend besitzt eine Funktion $y = f(x)$ an der Stelle $x = a$ ein relatives Minimum, wenn es eine Umgebung von a gibt, in der alle Funktionswerte größer als an der Stelle $x = a$ sind. Der Funktionswert $f(a)$ heißt dann relatives Minimum. Für ein relatives Minimum gilt analog $f(x) > f(a)$ für alle $x \neq a$ aus einer geeigneten Umgebung von a. Alle benachbarten Funktionswerte sind also größer als $f(a)$.

Relatives Minimum $f(a)$

$$f(x) > f(a) \quad \text{für } x \neq a$$

Es handelt sich bei einem relativen Maximum oder einem relativen Minimum um eine lokale Eigenschaft, denn es wird nur eine Umgebung von $x = a$ betrachtet.

Das absolute oder globale Maximum einer Funktion $y = f(x)$, die in einem abgeschlossenen Intervall $[c, d]$ differenzierbar ist, ist entweder ein relatives Maximum, oder es wird am Rand, also für $x = c$ oder $x = d$, angenommen. Entsprechend ist das absolute oder globale Minimum ein relatives Minimum, oder es wird an einem der Intervallränder $x = c$ oder $x = d$ angenommen.

Ein Extremwert einer Funktion ist ein Funktionswert $f(a)$, der ein relatives Minimum oder ein relatives Maximum ist. Statt Extremwert sagt man auch Extremum oder relatives Extremum.

Eine notwendige Bedingung dafür, dass die Funktion $y = f(x)$ an der Stelle $x = a$ ein relatives Extremum besitzt, ist das Verschwinden der Ableitung an dieser Stelle, also $f'(a) = 0$ (falls sie existiert). Zur Bestimmung der relativen Extrema müssen alle x berechnet werden, die die Gleichung $f'(x) = 0$ erfüllen.

Eine hinreichende Bedingung für ein relatives Extremum (das heißt, ist die Bedingung erfüllt, dann liegt ein relatives Extremum vor) ist, dass die zweite Ableitung von Null verschieden ist, also $f''(a) \neq 0$. Gilt jedoch auch $f''(a) = 0$, so ist $f(a)$ ein relatives Extremum, wenn es ein gerades n gibt, so dass $f'(a) = f''(a) = \ldots = f^{(n-1)}(a) = 0, f^{(n)}(a) \neq 0$ (n gerade). Ein Extremum liegt vor, wenn die erste an der Stelle a nicht verschwindende Ableitung von gerader Ordnung ist.

Dieses relative Extremum ist ein relatives Minimum, wenn im ersten Fall $f''(a) > 0$ und im zweiten Fall $f^{(n)}(a) > 0$ gilt. Das relative Extremum ist ein relatives Maximum, wenn im ersten Fall $f''(a) < 0$ und im zweiten Fall $f^{(n)}(a) < 0$ gilt.

Geometrisch bedeutet $f'(a) = 0$, dass die Tangente an die Kurve der Funktion im Punkt $P(a|f(a))$ waagerecht, also parallel zur x-Achse, verläuft.

Beispiele

1. $f(x) = x^2$
   $f'(x) = 2x, f''(x) = 2$
   $f'(x) = 0 \Rightarrow x = 0$
   $f''(0) = 2 > 0 \Rightarrow f(0) = 0$ ist ein relatives Minimum von $y = f(x)$
2. $f(x) = -x^4 + 1$
   $f'(x) = -4x^3$, $f''(x) = -12x^2$, $f'''(x) = -24x$, $f^{(4)}(x) = -24$
   $f'(x) = 0 \Rightarrow x = 0$
   $f''(0) = f'''(0) = 0, f^{(4)}(0) = -24 < 0 \Rightarrow f(0) = 0$ ist ein relatives Maximum von $y = f(x)$
3. $f(x) = x^3 - 4x^2 + 4x = x(x-2)^2$
   $f'(x) = 3x^2 - 8x + 4, f''(x) = 6x - 8$
   $f'(x) = 0 \Rightarrow 3x^2 - 8x + 4 = 0 \Rightarrow x_1 = 2, x_2 = \frac{2}{3}$
   $f''(2) = 4 > 0, f''(\frac{2}{3}) = -4 < 0 \Rightarrow f(x_1) = f(2) = 0$ ist ein relatives Minimum und $f(x_2) = f(\frac{2}{3}) = \frac{32}{27}$ ist ein relatives Maximum von $y = f(x)$
4. $f(x) = x^3 - 3x^2 + 3x$
   $f'(x) = 3x^2 - 6x + 3 = 3(x-1)^2, f''(x) = 6x - 6, f''(x) = 6$
   $f'(x) = 0 \Rightarrow x = 1$
   $f''(1) = 0, f'''(1) = 6 \Rightarrow f(x)$ besitzt kein relatives Extremum, bei $x = 1$ liegt der Sattelpunkt $P = (1|1)$, also ein Wendepunkt mit waagerechter Tangente (Abschn. 8.4.8). ◀

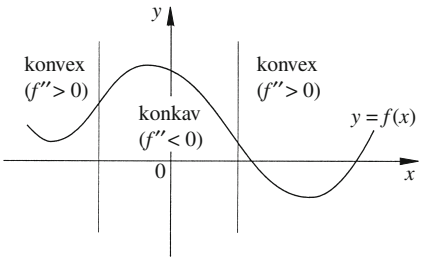


Abb. 8.8 Konkave und konvexe Bereiche der Funktion $y = f(x)$

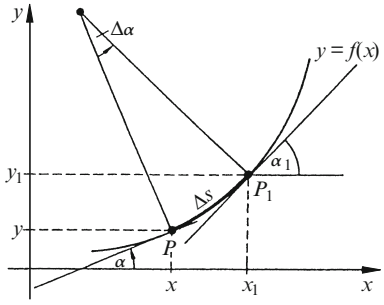


Abb. 8.9 Zur Definition der Krümmung einer Kurve

8.4.8 Krümmungsverhalten von Funktionen

Das Krümmungsverhalten einer Funktion ist die Verteilung von konvexen und konkaven Bereichen der Kurve der Funktion (Abb. 8.8).

Eine Funktion $y = f(x)$ heißt an der Stelle $x = a$ von unten konvex, wenn alle Punkte der Kurve der Funktion in einer Umgebung von a oberhalb der Tangente im Punkt $P(a|f(a))$ liegen.

In einem von unten konvexen Bereich ist die Ableitungsfunktion $y' = f'(x)$ monoton wachsend. Die Funktion $y = f(x)$ hat dort eine Linkskrümmung (der Graph macht in x-Richtung eine Linkskurve). Existiert in dem Bereich auch die zweite Ableitung $f''(x)$, so ist die Kurve konvex, wenn $f''(x) \geq 0$ gilt.

Entsprechend heißt die Funktion an der Stelle $x = a$ von unten konkav (oder von oben konvex), wenn alle Punkte der Kurve der Funktion in einer Umgebung von a unterhalb der Tangente im Punkt $P(a|f(a))$ liegen.

In einem von unten konkaven Bereich ist die Ableitungsfunktion $y' = f'(x)$ monoton fallend. Die Funktion $y = f(x)$ hat dort eine Rechtskrümmung (der Graph macht in x-Richtung eine Rechtskurve). Existiert in dem Bereich auch die zweite Ableitung $f''(x)$, so ist die Kurve konkav, wenn $f''(x) \leq 0$ gilt.

Die Krümmung einer Funktion ist die Abweichung der Kurve der Funktion von der Geraden.

Die Krümmung der Kurve der Funktion $y = f(x)$ im Punkt $P(x|y)$ ist definiert als der Grenzwert κ des Quotienten aus der Differenz der Steigungswinkel α_1, α der Tangenten durch einen Punkt P_1 und durch P an die Kurve und der Länge Δs des Kurvenbogens zwischen den Punkten (falls der Grenzwert existiert), (Abb. 8.9):

$$\kappa = \lim_{P_1 \to P} \frac{\alpha_1 - \alpha}{\Delta s} = \lim_{P_1 \to P} \frac{\Delta\alpha}{\Delta s} = \frac{d\alpha}{ds}$$

Die Krümmung einer Funktion ist in einem konvexen Bereich (Linkskurve) positiv, in einem konkaven Bereich (Rechtskurve) negativ. Für eine Gerade gilt $\kappa = 0$.

Mit Hilfe der Kettenregel berechnet man für die Krümmung in einen Punkt $P(x|y)$ der Funktion $y = f(x)$:

$$\kappa = \frac{f''(x)}{[1 + f'^2(x)]^{\frac{3}{2}}} = \frac{f''(x)}{[\sqrt{1 + f'^2(x)}]^3}$$

Für $\kappa \neq 0$ heißt $\rho = \frac{1}{|\kappa|}$ Krümmungsradius und der Kreis mit diesem Radius Krümmungskreis der Kurve im Punkt $P(x|y)$.

Beispiel

$f(x) = 3x^3 - 1$

$f'(x) = 9x^2, f''(x) = 18x$

Es folgt: $\kappa = \frac{18x}{(1+81x^4)^{\frac{3}{2}}}$.

Krümmung im Punkt $P(1|2)$ zum Beispiel:

$\kappa = \frac{18}{82^{\frac{3}{2}}} \approx 0{,}0242$ ◀

8.4.9 Wendepunkte von Funktionen

Ein Wendepunkt einer Funktion $y = f(x)$ ist ein Punkt $P(a|f(a))$, in dem sich das Krümmungsverhalten der Kurve ändert. In einem Wendepunkt findet der Übergang von einem konvexen zu einem konkaven Bereich oder umgekehrt statt. Die Kurve liegt in der unmittelbaren Nähe eines Wendepunktes nicht auf einer Seite der Tangente, sondern wird von dieser durchsetzt.

Eine notwendige Bedingung für die Existenz eines Wendepunktes $P(a|f(a))$ einer Funktion $y = f(x)$ ist das Verschwinden der zweiten Ableitung im Wendepunkt, also $f''(a) = 0$ (falls sie existiert). Zur Bestimmung der Wendepunkte müssen alle x berechnet werden, die die Gleichung $f''(x) = 0$ erfüllen.

Eine hinreichende Bedingung für einen Wendepunkt ist, dass die dritte Ableitung von Null verschieden ist, also $f'''(a) \neq 0$. Gilt jedoch auch $f'''(a) = 0$, so hat $f(x)$ an der Stelle a einen Wendepunkt, wenn es ein ungerades n gibt, so dass $f''(a) = f'''(a) = \ldots = f^{(n-1)}(a) = 0, f^{(n)}(a) \neq 0$ (n ungerade). Ein Wendepunkt liegt vor, wenn die erste an der Stelle a nicht verschwindende Ableitung von ungerader Ordnung ist.

Falls in einem Wendepunkt $P(a|f(a))$ auch noch die erste Ableitung verschwindet, wenn also zusätzlich $f'(a) = 0$ gilt, dann ist dort die Tangente waagerecht. Ein solcher Wendepunkt heißt Sattelpunkt.

Beispiele

1. $f(x) = x^3 - 4x^2 + 4x = x(x-2)^2$
   $f'(x) = 3x^2 - 8x + 4, f''(x) = 6x - 8, f'''(x) = 6$
   $f''(x) = 0 \Rightarrow 6x - 8 = 0 \Rightarrow x = \frac{4}{3}$
   $f'''(\frac{4}{3}) = 6 \neq 0 \Rightarrow$ bei $x = \frac{4}{3}$ liegt der Wendepunkt
   $P = (\frac{4}{3}|f(\frac{4}{3})) = (\frac{4}{3}|\frac{16}{27})$
2. $f(x) = x^3 - 3x^2 + 3x$
   $f'(x) = 3x^2 - 6x + 3 = 3(x-1)^2, f''(x) = 6x - 6,$
   $f'''(x) = 6$
   $f''(x) = 0 \Rightarrow x = 1$
   $f'''(1) = 6 \neq 0 \Rightarrow f(x)$ besitzt bei $x = 1$ einen Wendepunkt
   Da auch $f'(1) = 0$ gilt, ist dort die Tangente waagerecht, und somit ist $P = (1|1)$ ein Sattelpunkt.
3. $f(x) = -x^5$
   $f'(x) = -5x^4, f''(x) = -20x^3, f'''(x) = -60x^2,$
   $f^{(4)}(x) = -120x, f^{(5)}(x) = -120$
   $f''(x) = 0 \Rightarrow x = 0$
   $f'''(0) = f^{(4)}(0) = 0, f^{(5)}(0) = -120 \neq 0 \Rightarrow f(x)$ besitzt bei $x = 0$ einen Wendepunkt
   Da auch $f'(0) = 0$ gilt, ist dort die Tangente waagerecht, und somit ist $P = (0|0)$ ein Sattelpunkt. ◀

8.4.10 Kurvendiskussion

Eine Kurvendiskussion ist die Untersuchung einer Funktion $y = f(x)$ bzw. des Graphen der Funktion auf typische Eigenschaften. Dazu gehören die Untersuchung auf Symmetrie und Monotonie sowie die Bestimmung von Definitionsbereich, Nullstellen, relativen Extrema, Wendepunkten, Unstetigkeitsstellen und Asymptoten.

Beispiel

$$f(x) = \frac{1}{2}x(x-2)^3$$

Ableitungen:

$$\begin{aligned} f'(x) &= \frac{1}{2}(x-2)^3 + \frac{3}{2}x(x-2)^2 \\ &= \frac{1}{2}(x-2)^2(x-2+3x) = (x-2)^2(2x-1) \\ f''(x) &= 2(x-2)(2x-1) + 2(x-2)^2 \\ &= (x-2)(4x-2+2x-4) = 6(x-1)(x-2) \\ f'''(x) &= 6(x-1) + 6(x-2) = 6(2x-3) \end{aligned}$$

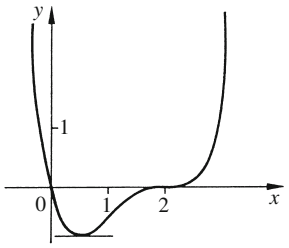


Abb. 8.10 Graph der Funktion $f(x) = \frac{1}{2}x(x-2)^3$

Definitionsbereich:

$$D = \mathbb{R}$$

Nullstellen:

$$f(x) = \frac{1}{2}x(x-2)^3 = 0 \text{ (Abb. 8.10)}$$
$$\Rightarrow \quad x_1 = 0, x_2 = 2$$

Relative Extremwerte:

$$f'(x) = (x-2)^2(2x-1) = 0$$
$$\Rightarrow \quad x_3 = 2, x_4 = \frac{1}{2}$$

$f''(x_3) = f''(2) = 0, f'''(2) = 6 > 0$ (n ungerade) $\Rightarrow$ bei $x_3 = 2$ Wendepunkt; wegen $f'(2) = 0$ ist $P(2|0)$ ein Sattelpunkt

$$f''(x_4) = f''\left(\frac{1}{2}\right) = 6 \cdot \frac{1}{2} \cdot \frac{3}{2} > 0$$

$\Rightarrow$ Minimum bei $x_4 = \frac{1}{2}$

Wendepunkte:

$$f''(x) = 6(x-1)(x-2) = 0$$
$$\Rightarrow \quad x_5 = 1, x_6 = x_3 = 2$$

$f''(x_5) = f''(1) \neq 0 \Rightarrow$ Wendepunkt bei $x_5 = 1$

Sattelpunkt bei $x_6 = x_3 = 2$ (siehe oben).

Zusammenfassung:

Die Funktion $f(x) = \frac{1}{2}x(x-2)^3$ hat die Nullstellen $x_1 = 0$ und $x_2 = 2$, das relative Minimum $f(\frac{1}{2}) = \frac{1}{2} \cdot \frac{1}{2} \cdot (-\frac{3}{2})^3 = -\frac{27}{32}$, den Wendepunkt $P(1|-\frac{1}{2})$ (denn $f(1) = \frac{1}{2} \cdot 1 \cdot (-1)^3 = -\frac{1}{2}$) und den Sattelpunkt $P(2|0)$. Die Funktion besitzt keine Unstetigkeitsstellen und Asymptoten, sie ist weder zur y-Achse noch zum Koordinatenursprung symmetrisch. Die Funktion ist streng monoton fallend im Intervall $(-\infty, \frac{1}{2}]$ und streng monoton wachsend im Intervall $[\frac{1}{2}, \infty)$. ◀

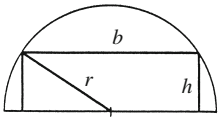


Abb. 8.11 Zu Anwendungsbeispiel 1

8.4.11 Anwendungsbeispiele

1. Ein halbrunder Balken soll so besäumt werden, dass ein rechtwinkliger Balken mit maximalem Widerstandsmoment W entsteht (Abb. 8.11).
   Die Gleichung für das Widerstandsmoment lautet:

$$W = \frac{hb^2}{6} \tag{1}$$

   Nach dem Satz des Pythagoras gilt für die Beziehung zwischen b und h:

$$\left(\frac{b}{2}\right)^2 + h^2 = r^2 \tag{2}$$

   Auflösen von (2) nach b^2:

$$b^2 = 4(r^2 - h^2)$$

   Einsetzen in (1):

$$W = \frac{h}{6} \cdot 4(r^2 - h^2) = \frac{2}{3}(r^2 h - h^3)$$

   Da r eine feste Größe ist, hängt W nur von h ab, das heißt, W ist eine Funktion von h: $W = W(h)$. Notwendige Voraussetzung für ein Maximum von W ist das Verschwinden der Ableitung: $W' = 0$.
   Berechnung der Ableitung:

$$W'(h) = \frac{2}{3}(r^2 - 3h^2)$$

$$W'(h) = 0 \quad \Rightarrow \quad \frac{2}{3}(r^2 - 3h^2) = 0$$

$$\Rightarrow \quad r^2 - 3h^2 = 0 \quad \Rightarrow \quad h = \frac{1}{3}r\sqrt{3}$$

   (Da die Höhe h nicht negativ sein kann, kommt für das Maximum nur das positive Vorzeichen in Frage.) Wegen $W''(h) = -4h$ ist für $h = \frac{1}{3}r\sqrt{3}$ die zweite Ableitung negativ, es liegt also ein Maximum vor.
   Ergebnis:
   $h = \frac{1}{3}r\sqrt{3}$ und $b = \frac{2}{3}r\sqrt{6}$ sind die Abmessungen für das maximale Widerstandsmoment, es beträgt $W = \frac{2}{9}r\sqrt{3}(r^2 - \frac{1}{3}r^2) = \frac{4}{27}r^3\sqrt{3}$.
2. Aus einem kreiskegelförmigen Stück Holz soll ein Zylinder größtmöglichen Rauminhalts (Gewichts) gedreht werden. Welchen Radius x und welche Höhe y hat dieser Zylinder, wenn r der Radius und h die Höhe des Kegels sind (Abb. 8.12)?

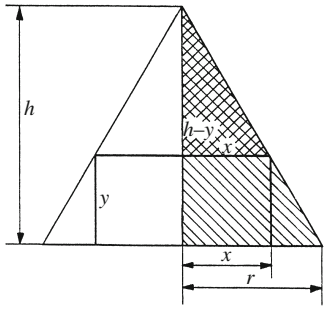


Abb. 8.12 Zu Anwendungsbeispiel 2

$$V = \pi x^2 y \qquad \text{Zylindervolumen} \tag{1}$$

$$\frac{h-y}{x} = \frac{h}{r} \qquad \text{Beziehung zwischen } x \text{ und } y \tag{2}$$

   Die Beziehung zwischen x und y folgt aus der Ähnlichkeit der schraffierten Dreiecke.
   Auflösen von (2) nach y:

$$y = h\left(1 - \frac{x}{r}\right)$$

   Einsetzen in (1):

$$V = \pi x^2 h\left(1 - \frac{x}{r}\right) = \pi h\left(x^2 - \frac{1}{r}x^3\right)$$

   h ist eine feste Größe, V ist also eine Funktion der Variablen x: $V = V(x)$.
   Berechnung der Ableitung:

$$V'(x) = \pi h(2x - \frac{3}{r}x^2)$$

$$V'(x) = 0 \quad \Rightarrow \quad \pi h\left(2x - \frac{3}{r}x^2\right) = \pi h x\left(2 - \frac{3}{r}x\right) = 0$$

$$\Rightarrow \quad x_1 = 0 \quad \text{und} \quad x_2 = \frac{2}{3}r$$

   Wegen $V''(x) = \pi h(2 - \frac{6}{r}x)$ gilt $V''(x_1) > 0$ und $V''(x_2) < 0$, das heißt, bei x_1 liegt ein Minimum und bei x_2 ein Maximum vor.
   Ergebnis:
   $x = \frac{2}{3}r$ und $y = \frac{1}{3}h$ sind Radius und Höhe des gesuchten Zylinders, das maximale Zylindervolumen beträgt $V = \frac{4}{27}\pi r^2 h$.

8.4.12 Näherungsverfahren zur Nullstellenbestimmung

In vielen Fällen ist es nicht möglich oder nicht notwendig, die Nullstellen von Funktionen exakt zu berechnen. Gerade in vielen praktischen Anwendungen genügen oftmals angenäherte Werte. Zur Bestimmung solcher sogenannten Näherungslösungen gibt es eine Reihe von Näherungsverfahren.

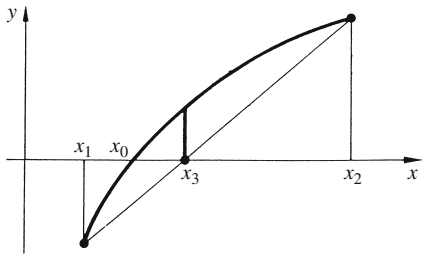


Abb. 8.13 Regula falsi

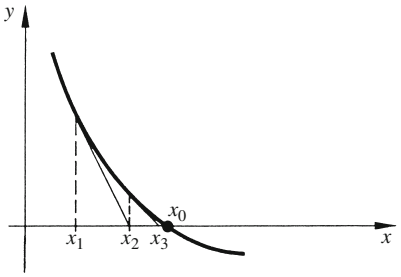


Abb. 8.14 Newton'sches Verfahren

8.4.12.1 Regula falsi

Regula falsi ist ein Verfahren zur näherungsweisen Bestimmung einer Nullstelle einer stetigen Funktion (Abb. 8.13).

Für eine stetige Funktion $y = f(x)$ wird eine Nullstelle, also eine Stelle x_0 mit $f(x_0) = 0$, gesucht. Sind x_1 und x_2 zwei Stellen in der Nähe der Nullstelle x_0, deren Funktionswerte unterschiedliche Vorzeichen haben (also $f(x_1) \cdot f(x_2) < 0$), dann erhält man eine bessere Näherung, indem man durch die Punkte $P_1(x_1|f(x_1))$ und $P_2(x_2|f(x_2))$ die Verbindungsgerade (Sekante) legt. Der Schnittpunkt x_3 der Verbindungsgeraden mit der x-Achse liefert einen verbesserten Näherungswert für die Nullstelle x_0.

$$x_3 = x_2 - \frac{x_2 - x_1}{f(x_2) - f(x_1)} f(x_2)$$

Das folgt aus $\frac{x_2 - x_3}{x_2 - x_1} = \frac{f(x_2)}{f(x_2) - f(x_1)}$, was sich mit dem zweiten Strahlensatz ergibt.

Dieses Verfahren lässt sich zur Bestimmung immer besserer Näherungslösungen für die Nullstelle x_0 beliebig oft wiederholen. Im nächsten Schritt wendet man das Verfahren auf x_3 und den Wert x_1 oder x_2 an, dessen Funktionswert ein von $f(x_3)$ verschiedenes Vorzeichen hat.

Diese Methode zur Bestimmung von Näherungswerten einer Nullstelle einer stetigen Funktion heißt auch Sekantenverfahren.

Beispiel

$f(x) = x^3 + 2x^2 + 10x - 20$

Für $x_1 = 1$ und $x_2 = 1{,}5$ gilt $f(x_1) = f(1) = -7$ und $f(x_2) = f(1{,}5) = 2{,}875$. Eine bessere Näherungslösung für die Nullstelle von $f(x)$, die zwischen 1 und 1,5 liegt, erhält man mit Regula falsi:

$$x_3 = 1{,}5 - \frac{1{,}5 - 1}{2{,}875 - (-7)} 2{,}875$$
$$= 1{,}5 - \frac{0{,}5}{9{,}875} 2{,}875 = 1{,}3544\ldots$$

Wegen $f(x_3) = -0{,}3020\ldots < 0$ lässt sich im nächsten Schritt das Verfahren auf x_3 und x_2 anwenden:

$$x_4 = x_3 - \frac{x_3 - x_2}{f(x_3) - f(x_2)} f(x_3)$$
$$= 1{,}3544\ldots - \frac{1{,}3544\ldots - 1{,}5}{-0{,}3020\ldots - 2{,}875}(-0{,}3020\ldots)$$
$$= 1{,}3682\ldots$$

Es gilt $f(x_4) = -0{,}0113\ldots$, das heißt, x_4 ist schon eine gute Näherung für die Nullstelle x_0.

Will man die Näherung weiter verbessern, so wendet man Regula falsi im nächsten Schritt auf x_4 und x_2 an ($f(x_4)$ und $f(x_3)$ haben dasselbe Vorzeichen, deshalb kann das Verfahren nicht auf x_4 und x_3 angewandt werden). ◀

8.4.12.2 Newton'sches Verfahren

Das Newton'sche Verfahren ist eine Methode zur näherungsweisen Bestimmung einer Nullstelle einer stetig differenzierbaren Funktion (Abb. 8.14).

Bei diesem Verfahren wird die Funktion in der Nähe einer Nullstelle nicht durch eine Sekante wie bei Regula falsi, sondern durch eine Tangente ersetzt.

Für eine stetig differenzierbare Funktion $y = f(x)$ wird eine Nullstelle, also eine Stelle x_0 mit $f(x_0) = 0$ gesucht. Ist x_1 eine Stelle in der Nähe der Nullstelle x_0, dann ersetzt man die Funktion durch die Tangente in dem Punkt $P(x_1|f(x_1))$. Der Schnittpunkt x_2 dieser Tangente mit der x-Achse ergibt einen neuen Näherungswert für die Nullstelle x_0.

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

Damit x_2 tatsächlich ein besserer Näherungswert als x_1 für die Nullstelle x_0 ist, muss in der Umgebung von x_0 die Bedingung $\left|\frac{f(x) \cdot f''(x)}{[f'(x)]^2}\right| < 1$ erfüllt sein.

Dasselbe Verfahren lässt sich auch auf x_2 anwenden. Man erhält als weitere Verbesserung den Wert $x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$. Allgemein findet man durch folgende Iterationsvorschrift aus x_1 eine Folge von verbesserten Näherungswerten $x_2, x_3, x_4, \ldots$ für die Nullstelle x_0.

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 1, 2, 3, \ldots$$

Diese Methode zur Bestimmung von Näherungswerten einer Nullstelle einer stetig differenzierbaren Funktion heißt Newton'sches Verfahren (nach dem englischen Mathematiker Isaac Newton, 1642–1727) oder auch Tangentenverfahren.

Beispiel

$f(x) = x^3 + 2x^2 + 10x - 20$

Wegen $f'(x) = 3x^2 + 4x + 10$ erhält man die Iterationsvorschrift

$$x_{k+1} = x_k - \frac{x_k^3 + 2x_k^2 + 10x_k - 20}{3x_k^2 + 4x_k + 10}$$

Für die Anfangsnäherung $x_1 = 1$ gilt $f(x_1) = f(1) = -7$, und man berechnet

$$x_2 = 1{,}4117\ldots \quad \text{mit } f(x_2) = 0{,}9175\ldots,$$
$$x_3 = 1{,}3693\ldots \quad \text{mit } f(x_3) = 0{,}0111\ldots,$$
$$x_4 = 1{,}3688\ldots \quad \text{mit } f(x_4) = 0{,}000001\ldots$$

Die Zahl x_4 ist also schon eine sehr gute Näherung für die Nullstelle x_0. ◀

8.5 Integralrechnung

8.5.1 Unbestimmtes Integral

Ist $y = f(x)$ eine Funktion mit einem Intervall I als Definitionsbereich, dann heißt eine differenzierbare Funktion $F(x)$ mit demselben Intervall I als Definitionsbereich eine Stammfunktion von $f(x)$, wenn für alle $x \in I$ gilt

$$F'(x) = f(x)$$

Die Funktion $f(x)$ heißt dann integrierbar.

Ist $F(x)$ eine Stammfunktion von $f(x)$, so ist auch $F(x) + c$ für eine beliebige Konstante c eine Stammfunktion, denn eine additive Konstante verschwindet bei der Differentiation. Somit ist $\{F(x) + C \mid C \in \mathbb{R}\}$ die Menge aller Stammfunktionen von $f(x)$. Stammfunktionen sind also bis auf eine additive Konstante eindeutig bestimmt.

Beispiele

1. Funktion: $f(x) = x^2 - 2x - 3$
   Stammfunktion: $F(x) = \frac{1}{3}x^3 - x^2 - 3x$, aber zum Beispiel auch $F_1(x) = \frac{1}{3}x^3 - x^2 - 3x + 5$
2. Funktion: $f(x) = \sin x$
   Stammfunktion: $F(x) = -\cos x$ oder etwa $F_1(x) = -\cos x + 3$
3. Funktionen: $f(x) = x^k$ $(k \in \mathbb{R}, k \neq -1)$
   Stammfunktionen: $F(x) = \frac{x^{k+1}}{k+1} + C$ $(C \in \mathbb{R})$
4. Funktion: $f(x) = x^{-1} = \frac{1}{x}$
   Stammfunktionen: $F(x) = \ln x + C$ $(C \in \mathbb{R})$
5. Funktion: $f(x) = e^x$
   Stammfunktionen: $f(x) = e^x + C$ $(k \in \mathbb{R})$ ◀

Die Gesamtheit aller Stammfunktionen $F(x) + C$ heißt unbestimmtes Integral der Funktion $y = f(x)$, gesprochen: Integral über $f(x)\,dx$ und geschrieben

$$\int f(x)dx = F(x) + C$$

Das Zeichen $\int$ heißt Integralzeichen, und $f(x)$ heißt Integrand. Die Variable x nennt man Integrationsvariable und C Integrationskonstante.

Die Konstante C soll andeuten, dass $F(x)$ durch die Funktion $f(x)$ bis auf eine additive Konstante bestimmt ist.

Beispiele

1. $\int x^3 dx = \frac{1}{4}x^4 + C$
2. $\int \cos x dx = \sin x + C$
3. $\int (x^4 - 3x^2 + 1)dx = \frac{1}{5}x^5 - 3 \cdot \frac{1}{3}x^3 + x + C = \frac{1}{5}x^5 - x^3 + x + C$ ◀

8.5.2 Integrationsregeln

Die folgenden Integrationsregeln zur Berechnung der unbestimmten Integrale von Funktionen lassen sich durch Differentiation der entsprechenden Gleichung beweisen.

1. *Faktorregel*
   Ein konstanter Faktor im Integranden kann vor das Integralzeichen gezogen werden.

$$\int cf(x)dx = c\int f(x)dx \quad (c \in \mathbb{R})$$

Beispiel

1.
$$\int 3x\,dx = 3\int x\,dx = 3\cdot\frac{1}{2}x^2 + C = \frac{3}{2}x^2 + C \quad \blacktriangleleft$$

2. *Potenzregel*

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + C$$

Beweis:

$$\frac{d}{dx}\left(\frac{1}{n+1}x^{n+1} + C\right) = x^n$$

Beispiel

2.
$$\int x^5 dx = \frac{1}{6}x^6 + C \quad \blacktriangleleft$$

3. *Summenregel*
Das unbestimmte Integral einer Summe ist gleich der Summe der unbestimmten Integrale (falls Stammfunktionen existieren).

$$\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx$$

Beispiel

3.
$$\begin{aligned}\int (4x^3 - 3x^2 + 5)dx &= \int 4x^3 dx - \int 3x^2 dx + \int 5dx \\ &= 4\int x^3 dx - 3\int x^2 dx + \int 5dx \\ &= 4\cdot\frac{1}{4}x^4 - 3\cdot\frac{1}{3}x^3 + 5x + C \\ &= x^4 - x^3 + 5x + C \quad \blacktriangleleft\end{aligned}$$

4. Ist der Integrand ein Bruch, in dem der Zähler die Ableitung des Nenners ist, dann ist das unbestimmte Integral gleich dem natürlichen Logarithmus des Absolutbetrages der Nennerfunktion.

$$\int \frac{f'(x)}{f(x)}dx = \ln|f(x)| + C$$

Beweis:

Nach der Kettenregel zur Differentiation zusammengesetzter Funktionen gilt:

$$\frac{d}{dx}(lnf(x) + C) = \frac{f'(x)}{f(x)}$$

Beispiel

4.
$$\int \frac{2x+3}{x^2+3x-5}dx = \ln|x^2 + 3x - 5| + C \quad \blacktriangleleft$$

5. *Partielle Integration*
Lässt sich die Funktion $f(x)$ als Produkt zweier Funktionen $g(x) = u(x)$ und $h(x) = v'(x)$ darstellen, also $f(x) = g(x)\cdot h(x) = u(x)\cdot v'(x)$, dann gilt

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx$$

Beweis:

Mit der Produktregel der Differentialrechnung ergibt sich

$$\frac{d}{dx}(u(x)\cdot v(x)) = u'(x)\cdot v(x) + u(x)\cdot v'(x),$$

woraus mit der Summenregel der Integralrechnung die Behauptung folgt.

Mit dieser Methode wird ein Integral der Form $\int u(x)v'(x)dx$ auf das oft leichter berechenbare Integral $\int u'(x)v(x)dx$ zurückgeführt.

Beispiele

5. $\int \ln x dx$
Setzt man $u(x) = \ln x$ und $v'(x) = 1$, dann ist $u'(x) = \frac{1}{x}$ und $v(x) = x$, und es ergibt sich

$$\begin{aligned}\int \ln x dx &= \int 1\cdot\ln x dx = x\cdot\ln x - \int x\cdot\frac{1}{x}dx \\ &= x\cdot\ln x - \int dx = x\cdot\ln x - x + C\end{aligned}$$

6. $\int xe^x dx$
Setzt man $u(x) = x$ und $v'(x) = e^x$, dann ist $u'(x) = 1$ und $v(x) = e^x$, und es folgt

$$\begin{aligned}\int xe^x dx &= xe^x - \int 1\cdot e^x dx = xe^x - e^x + C \\ &= (x-1)e^x + C\end{aligned}$$

7. $\int x \cdot \cos x dx$
Setzt man $u(x) = x$ und $v'(x) = \cos x$, dann ist $u'(x) = 1$ und $v(x) = \sin x$, und es ergibt sich

$$\int x \cdot \cos x dx = x \cdot \sin x - \int 1 \cdot \sin x dx$$
$$= x \cdot \sin x + \cos x + C$$

8. $\int x^n \cdot \ln x dx \; (n \in \mathbb{N})$
Setzt man $u(x) = \ln x$ und $v'(x) = x^n$, dann ist $u'(x) = \frac{1}{x}$ und $v(x) = \frac{x^{n+1}}{n+1}$, und es ergibt sich

$$\int x^n \cdot \ln x dx = \frac{x^{n+1}}{n+1} \cdot \ln x - \int \frac{1}{x} \cdot \frac{x^{n+1}}{n+1} dx$$
$$= \frac{x^{n+1}}{n+1} \cdot \ln x - \frac{1}{n+1} \int x^n dx$$
$$= \frac{x^{n+1}}{n+1} \cdot \ln x - \frac{1}{(n+1)^2} x^{n+1} + C \quad \blacktriangleleft$$

6. *Substitutionsmethode*
Durch Substitution $x = \varphi(t)$ der unabhängigen Variablen einer Funktion $y = f(x)$, also Einführung einer neuen Variablen t, ergibt sich für das unbestimmte Integral

$$\int f(x) dx = \int f(\varphi(t)) \varphi'(t) dt$$

Durch geeignete Substitution kann das Integral auf der rechten Seite der Gleichung einfacher zu berechnen sein als das Ausgangsintegral $\int f(x) dx$. Die Substitution muss so gewählt sein, dass $x = \varphi(t)$ nach t differenzierbar ist.

Beispiele

9. $\int \frac{dx}{(2+3x)^2}$
Substituiert man $2 + 3x = t$, also $x = \varphi(t) = \frac{t-2}{3}$, dann ist $\varphi'(t) = \frac{dx}{dt} = \frac{1}{3}$ oder $dx = \frac{dt}{3}$, und es ergibt sich

$$\int \frac{dx}{(2+3x)^2} = \int \frac{1}{t^2} \cdot \frac{dt}{3} = -\frac{1}{3t} + C$$
$$= -\frac{1}{3} \cdot \frac{1}{2+3x} + C$$

10. $\int (1+x)^n dx \; (n \in \mathbb{N})$
Substituiert man $x = \varphi(t) = t - 1$, dann ist $\varphi'(t) = \frac{dx}{dt} = 1$, also $dx = dt$, und es ergibt sich

$$\int (1+x)^n dx = \int t^n dt = \frac{t^{n+1}}{n+1} + C$$
$$= \frac{(x+1)^{n+1}}{n+1} + C$$

11. $\int (x^2+7)^8 \cdot x dx$
Substituiert man $x^2 + 7 = t$, also $x = \varphi(t) = \sqrt{t-7}$, dann ist nach der Kettenregel $\varphi'(t) = \frac{dx}{dt} = \frac{1}{2} \cdot \frac{1}{\sqrt{t-7}} \cdot 1$ oder $dx = \frac{dt}{2} \cdot \frac{1}{\sqrt{t-7}}$, und es ergibt sich

$$\int (x^2+7)^8 \cdot x dx = \int t^8 \cdot \sqrt{t-7} \cdot \frac{dt}{2} \cdot \frac{1}{\sqrt{t-7}}$$
$$= \frac{1}{2} \int t^8 dt = \frac{1}{2} \cdot \frac{1}{9} \cdot t^9 + C$$
$$= \frac{1}{18} (x^2+7)^9 + C \quad \blacktriangleleft$$

Das letzte Integral lässt sich noch einfacher berechnen, wenn man die obige Substitutionsgleichung von rechts nach links liest (mit der Substitution $u = \varphi(x)$).

$$\int f(\varphi(x)) \varphi'(x) dx = \int f(u) du$$

Beispiel

12.
$$\int (x^2+7)^8 \cdot x dx = \frac{1}{2} \int u^8 du = \frac{1}{18} u^9 + C$$
$$= \frac{1}{18} (x^2+7)^9 + C$$

mit der Substitution $u = x^2 + 7$, woraus $du = 2x dx$ folgt. $\blacktriangleleft$

Spezialfall

$$\int [f(x)]^n \cdot f'(x) dx = \frac{[f(x)]^{n+1}}{n+1} + C \quad (n \neq -1)$$

Beispiel

13.
$$\int \cos^5 x \cdot \sin x dx = -\int \cos^5 x \cdot (-\sin x) dx$$
$$= -\frac{1}{6} \cos^6 x + C \quad \blacktriangleleft$$

7. *Partialbruchzerlegung*
Die Integration gebrochener rationaler Funktionen $y = f(x)$ mit

$$f(x) = \frac{a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0}{b_m x^m + b_{m-1} x^{m-1} + \ldots + b_2 x^2 + b_1 x + b_0}$$

wird oftmals durch eine Partialbruchzerlegung von $f(x)$ (Abschn. 5.5.2) einfacher oder überhaupt erst möglich.

Beispiele

14. $\int \frac{6x^2-x+1}{x^3-x}dx$
Partialbruchzerlegung der Funktion liefert (vgl. Abschn. 5.5.2, Beispiel 3):

$$\frac{6x^2-x+1}{x^3-x} = \frac{1}{x} + \frac{3}{x-1} + \frac{4}{x+1}$$

Mit der Summenregel folgt:

$$\begin{aligned}\int \frac{6x^2-x+1}{x^3-x}dx &= \int \frac{1}{x}dx + \int \frac{3}{x-1}dx + \int \frac{4}{x+1}dx \\ &= \ln|x| + 3\ln|x-1| + 4\ln|x+1| + C\end{aligned}$$

15. $\int \frac{3x^2-x+1}{x^3-2x^2+x}dx$
Durch Partialbruchzerlegung erhält man (vgl. Abschn. 5.5.2, Beispiel 2):

$$\frac{3x^2-x+1}{x^3-2x^2+x} = \frac{1}{x} + \frac{2}{x-1} + \frac{3}{(x-1)^2}$$

Daraus errechnet sich das unbestimmte Integral der Funktion:

$$\begin{aligned}\int \frac{3x^2-x+1}{x^3-2x^2+x}dx &= \int \frac{1}{x}dx + \int \frac{2}{x-1}dx + \int \frac{3}{(x-1)^2}dx \\ &= \ln|x| + 2\ln|x-1| - \frac{3}{x-1} + C\end{aligned}$$

16. $\int \frac{12x^3-27x^2-8x+37}{3x^2-3x-6}dx$
Polynomdivision ergibt für den Integranden:

$$\frac{12x^3-27x^2-8x+37}{3x^2-3x-6} = 4x-5+\frac{x+7}{3(x^2-x-2)}$$

Nullstellen des Nenners: $x_1 = 2, x_2 = -1$
Zerlegung des Nennerpolynoms:

$$x^2-x-2 = (x-2)(x+1)$$

Ansatz zur Zerlegung des Bruches in Partialbrüche:

$$\frac{x+7}{3(x^2-x-2)} = \frac{A}{x-2} + \frac{B}{x+1}$$

Multiplikation mit dem Hauptnenner:

$$\begin{aligned}x+7 &= 3A(x+1) + 3B(x-2) \\ &= (3A+3B)x + (3A-6B)\end{aligned}$$

Vergleich der Koeffizienten von x und der Absolutglieder:

$$1 = 3A+3B, \quad 7 = 3A-6B$$

Lösung dieses Gleichungssystems mit zwei Gleichungen und zwei Variablen A und B:

$$A = 1, \quad B = -\frac{2}{3}$$

Damit ergibt sich für das unbestimmte Integral:

$$\begin{aligned}&\int \frac{12x^3-27x^2-8x+37}{3x^2-3x-6}dx \\ &= \int 4xdx - \int 5dx + \int \frac{1}{x-2}dx - \int \frac{2}{3(x+1)}dx \\ &= 2x^2-5x+\ln(x-2) - \frac{2}{3}\ln(x+1) + C\end{aligned}$$ ◀

8.5.3 Unbestimmte Integrale einiger algebraischer Funktionen

Mit den Integrationsregeln aus Abschn. 8.5.2 lassen sich die unbestimmten Integrale von algebraischen Funktionen berechnen.

Rationale Funktionen

$$\int a dx = ax + C$$

$$\int x dx = \frac{1}{2}x^2 + C$$

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C$$

$$\int (a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0)dx$$

$$= \frac{a_n}{n+1}x^{n+1} + \frac{a_{n-1}}{n}x^n + \ldots + \frac{a_1}{2}x^2 + a_0 x + C$$

$$\int \frac{1}{x}dx = \ln|x| + C$$

$$\int \frac{1}{x^n}dx = -\frac{1}{n-1}\frac{1}{x^{n-1}} + C \quad (n \neq 1)$$

$$\int \frac{x^m}{x^n}dx = \frac{1}{m-n+1}\frac{x^{m+1}}{x^n} + C \quad (n \neq m+1)$$

Irrationale Funktionen

$$\int \sqrt{x}dx = \frac{2}{3}x^{\frac{3}{2}} + C$$

$$\int \sqrt[n]{x}dx = \frac{n}{n+1}\sqrt[n]{x^{n+1}} + C$$

$$\int \frac{\sqrt[m]{x}}{\sqrt[n]{x}}dx = \frac{mn}{n-m+mn}\frac{\sqrt[m]{x^{m+1}}}{\sqrt[n]{x}} + C$$

8.5.4 Unbestimmte Integrale einiger transzendenter Funktionen

Auch für einige transzendente Funktionen lassen sich die unbestimmten Integrale mit den Integrationsregeln aus Abschn. 8.5.2 berechnen.

Trigonometrische Funktionen

$$\int \sin x dx = -\cos x + C$$

$$\int \cos x dx = \sin x + C$$

$$\int \tan x dx = -\ln|\cos x| + C$$

$$\int \cot x dx = \ln|\sin x| + C$$

$$\int \frac{1}{\cos^2 x}dx = \tan x + C \quad \left(x \neq (2k+1)\frac{\pi}{2}, k \in \mathbb{Z}\right)$$

$$\int \frac{1}{\sin^2 x}dx = -\cot x + C \quad (x \neq k\pi, k \in \mathbb{Z})$$

Exponentialfunktionen

$$\int e^x dx = e^x + C$$

$$\int a^x dx = \frac{1}{\ln a} \cdot a^x + C \quad (a \in \mathbb{R}, a > 0 \text{ konstant})$$

Logarithmusfunktionen

$$\int \ln x dx = x \cdot (\ln x - 1) + C \quad (x > 0)$$

$$\int \log_a x dx = \frac{1}{\ln a} \cdot x \cdot (\ln x - 1) + C$$

$$(a \in \mathbb{R}, a > 0 \text{ konstant}, x > 0)$$

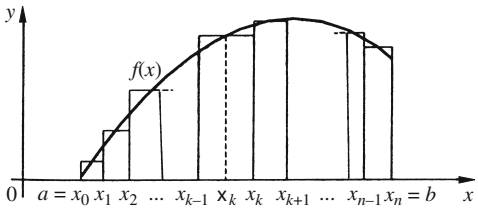


Abb. 8.15 Zur Definition des bestimmten Integrals

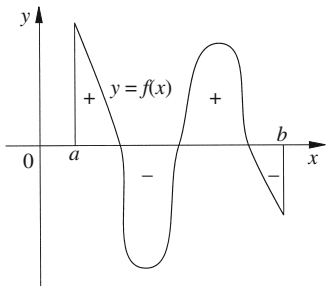


Abb. 8.16 Bestimmtes Integral

8.5.5 Bestimmtes Integral

Ist $y = f(x)$ eine beschränkte Funktion mit einem abgeschlossenen Intervall als Definitionsbereich, also $D = [a, b]$, dann ist das bestimmte Integral von $f(x)$ definiert durch $\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k)\Delta x_k$, falls dieser Grenzwert existiert und unabhängig von der Wahl der Zahlen x_k und ξ_k ist (gesprochen: Integral von a bis b über $f(x)$ dx). Dabei ist $a = x_0 < x_1 < \ldots < x_n = b$ eine Einteilung (Zerlegung) des Intervalls $[a, b]$ mit $\Delta x_k = x_k - x_{k-1}$ und ξ_k, $k = 1, 2, \ldots, n$, ein beliebiger Zwischenpunkt mit $x_{k-1} \leq \xi_k \leq x_k$ (Abb. 8.15).

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k)\Delta x_k$$

Die Funktion $f(x)$ heißt dann im Intervall $[a, b]$ integrierbar. Das Zeichen $\int$ heißt Integralzeichen. Man nennt a die untere Integrationsgrenze, b die obere Integrationsgrenze, $f(x)$ den Integranden und x die Integrationsvariable.

Diese Integraldefinition geht auf Bernhard Riemann zurück (deutscher Mathematiker, 1826–1866).

Gilt $f(x) \geq 0$ für alle $x \in [a, b]$, dann ist $\int_a^b f(x)dx$ gleich dem Inhalt des von der Kurve (Graph der Funktion $y = f(x)$) und der x-Achse zwischen $x = a$ und $x = b$ berandeten Fläche. Für $f(x) \leq 0$ für alle $x \in [a, b]$ ist $\int_a^b f(x)dx$ der negative Flächeninhalt. Besitzt $y = f(x)$ in $[a, b]$ Nullstellen, so ist $\int_a^b f(x)dx$ die Differenz der Flächeninhalte oberhalb („+“) und unterhalb („−“) der x-Achse (Abb. 8.16).

Existenz des bestimmten Integrals:

Jede in einem Intervall $[a, b]$ stetige Funktion ist dort auch integrierbar. Auch jede im Intervall $[a, b]$ beschränkte Funktion,

die in $[a,b]$ nur endlich viele Unstetigkeitsstellen besitzt, ist in diesem Intervall integrierbar.

Beispiele

1. Für die Funktion $f(x)=c$, $c\in\mathbb{R}$, $D=[a,b]$ und eine beliebige Einteilung $a=x_0<x_1<\ldots<x_n=b$ des Intervalls $[a,b]$ gilt

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{n}f(\xi_k)\Delta x_k &= \lim_{n\to\infty}\sum_{k=1}^{n}c\cdot\Delta x_k\\ &= \lim_{n\to\infty}c\sum_{k=1}^{n}(x_k-x_k-1)\\ &= \lim_{n\to\infty}c\cdot(b-a)\end{aligned}$$

Also ist die Funktion $f(x)$ im Intervall $[a,b]$ integrierbar, und es gilt

$$\int_a^b c\,dx = c\cdot(b-a)$$

2. Es sei $f(x)=x$, $D=[a,b]$. Da $f(x)$ stetig ist, ist $f(x)$ in $[a,b]$ integrierbar. Wählt man für eine Intervalleinteilung $x_k=a+k\Delta$ mit $\Delta=\frac{b-a}{n}$, also $a=x_0<x_1=a+\Delta<x_2=a+2\Delta<\ldots<x_n=a+n\Delta=b$, dann folgt $\Delta x_k=x_k-x_{k-1}=a+k\Delta-(a+(k-1)\Delta)=\Delta$. Wählt man außerdem $\xi_k=x_k$, also $\xi_k=a+k\Delta$, dann gilt

$$\begin{aligned}\sum_{k=1}^{n}f(\xi_k)\Delta x_k &= \sum_{k=1}^{n}x_k\Delta=\sum_{k=1}^{n}(a+k\Delta)\Delta\\ &= \sum_{k=1}^{n}a\cdot\Delta+\sum_{k=1}^{n}k\cdot\Delta^2\\ &= a\cdot\Delta\sum_{k=1}^{n}1+\Delta^2\sum_{k=1}^{n}k\\ &= a\cdot\frac{b-a}{n}\cdot n+\left(\frac{b-a}{n}\right)^2\cdot\frac{(n+1)n}{2}\\ &= a(b-a)+\frac{(b-a)^2}{2}\cdot\left(1+\frac{1}{n}\right),\end{aligned}$$

denn $\sum_{k=1}^{n}1=n$ und $\sum_{k=1}^{n}k=\frac{(n+1)n}{2}$. Für das bestimmte Integral ergibt sich dann

$$\begin{aligned}\int_a^b x\,dx &= \lim_{n\to\infty}\left(a(b-a)+\frac{(b-a)^2}{2}\cdot\left(1+\frac{1}{n}\right)\right)\\ &= a(b-a)+\frac{(b-a)^2}{2}\lim_{n\to\infty}\left(1+\frac{1}{n}\right)\\ &= a(b-a)+\frac{(b-a)^2}{2}=\frac{b^2}{2}-\frac{a^2}{2},\end{aligned}$$

denn es gilt $\lim_{n\to\infty}(1+\frac{1}{n})=1+\lim_{n\to\infty}\frac{1}{n}=1$. ◀

8.5.6 Hauptsatz der Differential- und Integralrechnung

Der Hauptsatz der Differential- und Integralrechnung liefert den Zusammenhang zwischen bestimmtem und unbestimmtem Integral einer Funktion $y=f(x)$.

Ist die Funktion $y=f(x)$ mit $D=[a,b]$ im Intervall $[a,b]$ integrierbar, und besitzt $f(x)$ eine Stammfunktion $F(x)$, so gilt

$$\int_a^b f(x)dx = F(b)-F(a)$$

Das bestimmte Integral ist also Funktionswert von F an der oberen Intervallgrenze minus Funktionswert von F an der unteren Intervallgrenze. Dabei ist $F(x)$ eine beliebige Stammfunktion von $f(x)$.

Statt $F(b)-F(a)$ schreibt man auch

$$F(x)\Big|_{x=a}^{x=b}=F(x)\Big|_a^b.$$

Mit diesem Satz wird die Berechnung des bestimmten Integrals einer Funktion auf die Berechnung einer Stammfunktion der Funktion zurückgeführt. Der Satz stellt somit den Zusammenhang zwischen dem bestimmten und dem unbestimmten Integral einer Funktion $y=f(x)$ her. Er wurde von Gottfried Wilhelm Leibniz (deutscher Mathematiker, 1646–1716) und Isaac Newton (englischer Mathematiker, 1642–1727) entdeckt.

Beispiele

1. $\int_a^b x\,dx=\frac{1}{2}x^2\Big|_a^b=\frac{1}{2}(b^2-a^2)$
2. $\int_1^3\frac{dx}{x}=\ln x\Big|_1^3=\ln 3$
3. $\int_1^5 x^3dx=\frac{1}{4}x^4\Big|_1^5=\frac{1}{4}5^4-\frac{1}{4}1^4=\frac{5^4-1}{4}=156$
4. $\int_0^\pi\sin x\,dx=-\cos x\Big|_0^\pi=-\cos\pi-(-\cos 0)$ $=1+1=2$
5. $\int_0^{2\pi}\cos x\,dx=\sin x\Big|_0^{2\pi}=\sin 2\pi-\sin 0=0$ ◀

8.5.7 Eigenschaften des bestimmten Integrals

Die folgenden Eigenschaften zur Berechnung des bestimmten Integrals einer Funktion lassen sich mit Hilfe der Definition beweisen.

1. *Vertauschung der Integrationsgrenzen*

$$\int_b^a f(x)dx=-\int_a^b f(x)dx$$

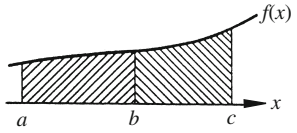


Abb. 8.17 Zusammenfassen der Integrationsintervalle

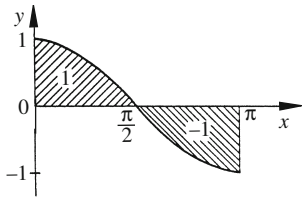


Abb. 8.18 $\int_0^{\pi} \cos x\,dx = 0$

Beispiel

$$\int_6^2 x\,dx = \frac{1}{2}x^2\Big|_6^2 = \frac{1}{2}2^2 - \frac{1}{2}6^2 = 2 - 18 = -16$$

$$-\int_2^6 x\,dx = -\frac{1}{2}x^2\Big|_2^6 = -\left(\frac{1}{2}6^2 - \frac{1}{2}2^2\right)$$

$$= -(18 - 2) = -16 \quad \blacktriangleleft$$

2. *Zusammenfassen der Integrationsintervalle* (Abb. 8.17)

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$$

Beispiel (Abb. 8.18)

$$\int_0^{\pi} \cos x\,dx = \int_0^{\frac{\pi}{2}} \cos x\,dx + \int_{\frac{\pi}{2}}^{\pi} \cos x\,dx$$

Einzelberechnung der Integrale:

$$\int_0^{\pi} \cos x\,dx = \sin x\big|_0^{\pi} = \sin\pi - \sin 0 = 0 - 0 = 0$$

$$\int_0^{\frac{\pi}{2}} \cos x\,dx = \sin x\big|_0^{\pi/2} = \sin\frac{\pi}{2} - \sin 0 = 1 - 0 = 1$$

$$\int_{\pi/2}^{\pi} \cos x\,dx = \sin x\big|_{\pi/2}^{\pi} = \sin\pi - \sin\frac{\pi}{2} = 0 - 1 = -1$$

$\blacktriangleleft$

3. *Gleiche untere und obere Integrationsgrenze*

$$\int_a^a f(x)dx = 0$$

Beispiel

$$\int_3^3 x^3 dx = \frac{1}{4}x^4\Big|_3^3 = \frac{1}{4}3^4 - \frac{1}{4}3^4 = 0 \quad \blacktriangleleft$$

4. Existieren die bestimmten Integrale $\int_a^b f(x)dx$ und $\int_a^b g(x)dx$, so gilt für beliebige $c_1, c_2 \in \mathbb{R}$

$$\int_a^b (c_1 \cdot f(x) + c_2 \cdot g(x))dx = c_1 \int_a^b f(x)dx + c_2 \int_a^b g(x)dx$$

Beispiel

$$\int_1^4 (2x - 4x^3)dx = 2\int_1^4 x\,dx - 4\int_1^4 x^3 dx$$

Einzelberechnung der Integrale:

$$\int_1^4 (2x - 4x^3)dx = (x^2 - x^4)\big|_1^4$$
$$= (4^2 - 4^4) - (1^2 - 1^4)$$
$$= -240 - 0 = -240$$

$$2\int_1^4 x\,dx = 2\left(\frac{1}{2}x^2\Big|_1^4\right)$$
$$= 2\left(\frac{1}{2}\cdot 4^2 - \frac{1}{2}\cdot 1^2\right)$$
$$= 2\left(8 - \frac{1}{2}\right) = 15$$

$$4\int_1^4 x^3 dx = -4\left(\frac{1}{4}x^4\Big|_1^4\right)$$
$$= -4\left(\frac{1}{4}\cdot 4^4 - \frac{1}{4}\cdot 1^4\right)$$
$$= -4\left(64 - \frac{1}{4}\right) = -255 \quad \blacktriangleleft$$

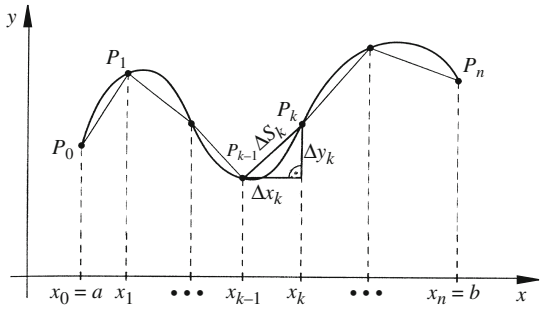


Abb. 8.19 Streckenzug zu einer Zerlegung Z

8.5.8 Einige Anwendungen der Integralrechnung

Es gibt sehr viele Anwendungen der Integralrechnung in der Technik und in den Ingenieurwissenschaften. Im Folgenden sind exemplarisch einige davon genannt.

Bogenlänge

Die Länge eines Kurvenstücks bezeichnet man als Bogenlänge.

Lässt sich der Bogen durch eine stetig differenzierbare Funktion $y = f(x), f : [a,b] \to W$ beschreiben, dann gilt für die Bogenlänge s

$$s = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx$$

Beweis:

Ist $a = x_0 < x_1 < \ldots < x_n = b$ eine Zerlegung Z des Intervalls $[a,b]$ mit $\Delta x_k = x_k - x_{k-1}$, $k = 1,2,\ldots,n$, dann sind $P_k = P_k(x_k|y_k)$ mit $y_k = f(x_k)$ Punkte des Kurvenstücks (Abb. 8.19).

Die geradlinige Verbindung dieser Punkte ergibt einen Streckenzug, dessen Länge s_Z die Summe der Längen der Teilstrecken Δs_k ist:

$$s_Z = \sum_{k=1}^{n} \Delta S_k = \sum_{k=1}^{n} \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}$$

$$= \sum_{k=1}^{n} \sqrt{\Delta x_k^2 + \Delta y_k^2} = \sum_{k=1}^{n} \sqrt{1 + \left(\frac{\Delta y_k}{\Delta x_k}\right)^2}\,\Delta x_k$$

Der Mittelwertsatz der Differentialrechnung sagt aus, dass es für eine stetig differenzierbare Funktion $y = f(x), f : [a,b] \to W$ mindestens eine Stelle ξ mit $a < \xi < b$ gibt, so dass $\frac{f(b)-f(a)}{b-a} = f'(\xi)$. Es folgt, dass es in allen Intervallen (x_{k-1}, x_k) Zwischenstellen ξ_k gibt mit $\frac{\Delta y_k}{\Delta x_k} = \frac{f(x_k)-f(x_{k-1})}{x_k - x_{k-1}} = f'(\xi_k)$. Damit folgt für die Länge s_Z des Streckenzugs Z:

$$s_z = \sum_{k=1}^{n} \sqrt{1 + [f'(\xi_k)]^2}\,\Delta x_k$$

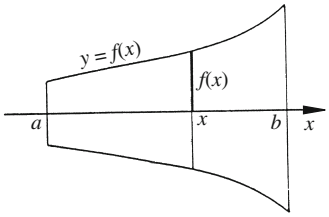


Abb. 8.20 Rotationskörper

Mit f' ist auch $\sqrt{1+f'^2}$ in $[a,b]$ stetig, und somit folgt für die Bogenlänge s mit der Definition des bestimmten Integrals:

$$s = \lim_{n\to\infty} \sum_{k=1}^{n} \sqrt{1 + [f'(\xi_k)]^2}\,\Delta x_k = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx$$

Beispiel

Bogen: $y = \sqrt{1-x^2}$, $D = [a,b] = [-1,1]$ (Halbkreis)

Bogenlänge:

$$s = \int_{-1}^{1} \sqrt{1 + \left(\frac{-x}{\sqrt{1-x^2}}\right)^2}\,dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx$$

$$= \arcsin x\Big|_{-1}^{1} = \pi$$ ◀

Volumen und Mantelfläche von Rotationskörpern

Ein Rotationskörper ist ein Körper, der entsteht, wenn die Kurve einer Funktion $y = f(x)$ mit $f(x) \geq 0$ um die x-Achse (Rotationsachse) zwischen $x = a$ und $x = b$ rotiert (oder die inverse Funktion um die y-Achse). Rotationskörper sind aus dem Alltag bekannt: Vasen, Gläser oder gedrechselte Figuren zum Beispiel.

Ein Rotationskörper ist durch zwei Schnitte senkrecht zur Rotationsachse begrenzt. Die von der Kurve, der x-Achse und den Geraden $x = a$ und $x = b$ begrenzte Fläche heißt die erzeugende Fläche des Rotationskörpers.

Die Kugel ist zum Beispiel ein Rotationskörper. Sie entsteht durch Rotation eines Kreises mit dem Mittelpunkt im Koordinatenursprung um eine der beiden Achsen. Auch gerade Kreiskegel und gerade Kreiszylinder sind Rotationskörper (Abb. 8.20).

Für das Volumen V und für den Inhalt A_M der Mantelfläche eines Rotationskörpers gilt

$$\text{Volumen} \qquad V = \pi \int_a^b f^2(x)\,dx$$

$$\text{Mantelfläche} \qquad A_M = 2\pi \int_a^b f(x)\sqrt{1 + [f'(x)]^2}\,dx$$

Beispiel

Die Gleichung des oberen Halbkreises mit dem Radius r lautet (explizite Form in kartesischen Koordinaten)

$$y = \sqrt{r^2 - x^2}, \quad D = [-r, r].$$

Die Ableitung dieser Funktion ist

$$y'(x) = \frac{1}{2}(-2x)\frac{1}{\sqrt{r^2 - x^2}} = \frac{-x}{\sqrt{r^2 - x^2}}$$

Somit berechnet man nach den obigen Formeln für das Volumen V und die Oberfläche A_O (hier: Mantelfläche = Oberfläche) einer Kugel mit dem Radius r

$$\begin{aligned} V &= \pi \int_{a=-r}^{b=r} (r^2 - x^2)dx = 2\pi \int_0^r (r^2 - x^2)dx \\ &= 2\pi\left(r^2 x - \frac{1}{3}x^3\right)\Big|_0^r = 2\pi \cdot \frac{2}{3}r^3 = \frac{4}{3}\pi r^3 \\ A_O &= 2\pi \int_{a=-r}^{b=r} \sqrt{r^2 - x^2}\sqrt{1 + \frac{x^2}{r^2 - x^2}}dx \\ &= 4\pi \int_0^r \sqrt{r^2 - x^2}\sqrt{\frac{r^2}{r^2 - x^2}}dx = 4\pi \int_0^r r dx \\ &= 4\pi \cdot rx|_0^r = 4\pi r^2 \end{aligned}$$

◀

Volumen und Mantelfläche von Rotationskörpern lassen sich auch mit Hilfe der Guldin'schen Regeln berechnen (benannt nach dem Mathematiker Paul Guldin, 1577–1643):

1. Guldin'sche Regel

Das Volumen eines Rotationskörpers ist gleich dem Produkt aus dem Inhalt der auf einer Seite der Rotationsachse liegenden erzeugenden Fläche und der Länge des Weges, den der Flächenschwerpunkt bei der Rotation zurücklegt.

2. Guldin'sche Regel

Der Inhalt der Mantelfläche eines Rotationskörpers ist gleich dem Produkt aus der Länge des auf einer Seite der Rotationsachse liegenden erzeugenden Kurvenstücks und der Länge des Weges, den der Schwerpunkt des erzeugenden Kurvenstücks bei der Rotation zurücklegt.

Massenträgheitsmoment von Zylindern

1. *Definition des Massenträgheitsmoments*
   Dreht sich ein Massenpunkt der Masse m im Abstand r um eine Rotationsachse, so gilt für die Geschwindigkeit $v = \omega r$, wobei ω die Winkelgeschwindigkeit ist. Für die kinetische Energie gilt:

$$W = \frac{m}{2}v^2 = \frac{m}{2}(\omega r)^2$$

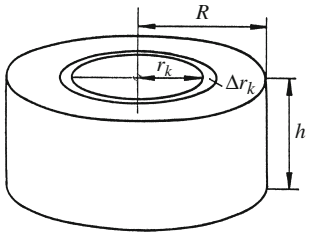


Abb. 8.21 Zum Massenträgheitsmoment eines Vollzylinders

Betrachtet man ein System von n Massenpunkten der Massen Δm_k, $k = 1, 2, \ldots, n$, die sich mit der gleichen Winkelgeschwindigkeit ω um eine Achse drehen, so gilt für die kinetische Energie:

$$W = \sum_{k=1}^{n} \frac{\Delta m_k(\omega r_k)^2}{2} = \frac{\omega^2}{2}\sum_{k=1}^{n} r_k^2 \cdot \Delta m_k$$

Der Vergleich mit einer geradlinig bewegten Masse zeigt, dass der Geschwindigkeit v die Winkelgeschwindigkeit ω entspricht und der Masse m bei der Rotation die Summe $\sum_{k=1}^{n} r_k^2 \cdot \Delta m_k$. Diese Summe wird Massenträgheitsmoment J genannt:

$$J = \sum_{k=1}^{n} r_k^2 \cdot \Delta m_k$$

2. *Massenträgheitsmoment eines Vollzylinders* (Abb. 8.21)
   Der betrachtete Vollzylinder hat die Höhe h und ist homogen, das heißt, er hat konstante Dichte ρ, und er rotiert um die Zylinderachse.
   Denkt man sich den Zylinder in n Hohlzylinder der Dicke Δr_k zerlegt, dann gilt für deren Volumen $\Delta V_k \approx 2\pi r_k \cdot \Delta r_k \cdot h$. Ist Δr_k sehr klein, dann haben alle Massenpunkte eines Hohlzylinders nahezu den gleichen Abstand r_k von der Drehachse, und es gilt für das Massenträgheitsmoment J_v des Vollzylinders

$$\begin{aligned} J_V &\approx \sum_{k=1}^{n} r_k^2 \cdot \Delta m_k = \sum_{k=1}^{n} r_k^2 \cdot \rho \cdot \Delta V_k \\ &\approx \rho \sum_{k=1}^{n} r_k^2 \cdot 2\pi r_k \cdot \Delta r_k \cdot h = 2\pi\rho h \sum_{k=1}^{n} r_k^3 \Delta r_k \end{aligned}$$

Wählt man die Unterteilung immer feiner, also $\Delta r_k \to 0$, so folgt mit der Definition des bestimmten Integrals

$$\begin{aligned} J_V &= 2\pi\rho h \lim_{n\to\infty} \sum_{k=1}^{n} r_k^3 \Delta r k = 2\pi\rho h \int_0^R r^3 dr \\ &= 2\pi\rho h \cdot \frac{R^4}{4} = \frac{1}{2}\pi\rho h R^4 = \frac{R^2}{2} \cdot \pi\rho h R^2 = \frac{R^2}{2} \cdot m_V \end{aligned}$$

$m_v = \pi\rho h R^2$ ist die Masse des Vollzylinders

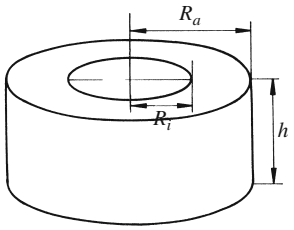


Abb. 8.22 Zum Massenträgheitsmoment eines Hohlzylinders

3. *Massenträgheitsmoment eines Hohlzylinders* (Abb. 8.22)
Das Massenträgheitsmoment J_H eines Hohlzylinders mit dem Außendurchmesser R_a, dem Innendurchmesser R_i, der Höhe h und der konstanten Dichte ρ in Bezug auf die Zylinderachse erhält man als Differenz der Trägheitsmomente zweier Vollzylinder.

$$J_H = 2\pi\rho h \int_0^{R_a} r^3 dr - 2\pi\rho h \int_0^{R_i} r^3 dr = 2\pi\rho h \int_{R_i}^{R_a} r^3 dr$$

$$= 2\pi\rho h \left(\frac{r^4}{4} \Big|_{R_i}^{R_a} \right) = \frac{1}{2}\pi\rho h (R_a^4 - R_i^4)$$

$$= \frac{R_a^2 + R_i^2}{2} \cdot \pi\rho h (R_a^2 - R_i^2) = \frac{R_a^2 + R_i^2}{2} \cdot m_H$$

$mH = \pi\rho h(R_a^2 - R_i^2)$ ist die Masse des Hohlzylinders

8.6 Funktionenreihen

Zur Untersuchung von Eigenschaften gegebener Funktionen ist es oftmals sinnvoll, eine Funktion näherungsweise durch eine unendliche Reihe darzustellen. Eine solche Darstellung heißt Entwicklung der Funktion in eine unendliche Reihe. Häufig wählt man, gerade bei praktischen Anwendungen, als Darstellung eine Potenzreihe oder, besonders bei Funktionen, die periodische Vorgänge beschreiben, Fourier-Reihen.

8.6.1 Definitionen

Für jedes $n = 1, 2, 3, \ldots$ sei $f_n(x)$ eine Funktion mit dem Definitionsbereich $D \subseteq \mathbb{R}$. Die Folge dieser Funktionen heißt Funktionenfolge.

$$(f_n(x)) = (f_1(x), f_2(x), f_3(x), \ldots)$$

Wählt man $x = x_0 \in D$ fest, so erhält man eine gewöhnliche Zahlenfolge (a_n) mit $a_n = f_n(x_0)$ (Abschn. 8.1).

Beispiel

$f_n(x) = x^n \Rightarrow$ Funktionenfolge:

$(f_n(x)) = (x^n) = (x, x^2, x^3, x^4, \ldots)$

$x = x_0 = 2 \Rightarrow$ Zahlenfolge:

$(2^n) = (2, 2^2, 2^3, 2^4, \ldots) = (2, 4, 8, 16, \ldots)$ ◀

Die Konvergenz von Funktionenfolgen wird analog zur Konvergenz von Zahlenfolgen definiert:

Die Funktionenfolge $(f_n(x))$ konvergiert in einem Bereich $E \subseteq D \subseteq \mathbb{R}$ gegen die Funktion $f(x)$, wenn für jedes $x \in E$ die Folge $(f_n(x))$ gegen den Funktionswert $f(x)$ konvergiert. Das heißt, zu jedem $x \in E$ lässt sich nach Vorgabe einer positiven Zahl ε ein $n_0 \in \mathbb{N}$ (im Allgemeinen sowohl von x als auch von ε abhängig, also $n_0 = n_0(\varepsilon, x)$) so finden, dass für alle $n \geq n_0$ gilt $|f(x) - f_n(x)| < \varepsilon$. Man nennt dieses punktweise Konvergenz der Funktionenfolge und schreibt dafür

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{für } x \in E$$

Hängt dagegen die Größe von n_0 nur von ε ab und nicht auch von x, also $n_0 = n_0(\varepsilon)$, dann heißt die Funktionenfolge gleichmäßig konvergent in E.

Die Funktion $f(x)$ heißt Grenzfunktion der Funktionenfolge $(f_n(x))$.

Beispiel

$$f_n(x) = x^n, \quad D = \left[0, \frac{1}{2}\right]$$

Für ein beliebiges $\varepsilon > 0$ wählt man n_0 so, dass $\frac{1}{2^{n_0}} < \varepsilon$ ist. Für alle $n \geq n_0$ und für alle $x \in D$ gilt dann:

$$|f_n(x) - 0| = x^n \leq \frac{1}{2^n} \leq \frac{1}{2^{n_0}} < \varepsilon.$$

Die Funktionenfolge ist also gleichmäßig konvergent im Definitionsbereich D mit der Grenzfunktion $f(x) = 0$. ◀

Eine Funktionenreihe ist die Summe der Glieder einer Funktionenfolge $(f_n(x))$.

Funktionenreihe

$$f_1(x) + f_2(x) + \ldots + f_n(x) + \ldots = \sum_{k=1}^{\infty} f_k(x)$$

Die folgenden Summen heißen Partialsummen der Funktionenreihe.

$$
\begin{aligned}
F_1(x) &= f_1(x),\\
F_2(x) &= f_1(x) + f_2(x),\\
F_3(x) &= f_1(x) + f_2(x) + f_3(x),\\
&\vdots\\
F_n(x) &= f_1(x) + f_2(x) + f_3(x) + \ldots + f_n(x)\\
&= \sum_{k=1}^{n} f_k(x)
\end{aligned}
$$

In Analogie zu den gewöhnlichen Reihen heißt eine Funktionenreihe in einem Bereich $E \subseteq D \subseteq \mathbb{R}$ konvergent, wenn die Folge der Partialsummen $F_n(x) = \sum_{k=1}^{n} f_k(x)$ der Funktionenreihe in E gegen eine Funktion $F(x)$ konvergiert. Man schreibt dann

$$F(x) = \lim_{n\to\infty} \sum_{k=1}^{n} f_k(x) = \sum_{k=1}^{\infty} f_k(x)$$

Ist die Konvergenz der Folge der Partialsummen in E punktweise, dann heißt auch die Funktionenreihe in E punktweise konvergent. Ist dagegen die Konvergenz der Folge der Partialsummen gleichmäßig, so heißt auch die Funktionenreihe gleichmäßig konvergent.

Die Funktion $F(x)$ heißt Grenzfunktion der Funktionenreihe in E. Man nennt $F(x)$ auch die durch die Funktionenreihe dargestellte Funktion (im Konvergenzbereich E).

Die Funktionenreihe $\sum_{k=1}^{\infty} f_k(x)$ heißt in E absolut konvergent, wenn $\sum_{k=1}^{\infty} |f_k(x)| < \infty$ für jedes $x \in E$ gilt. Bei absolut konvergenten Funktionenreihen darf die Summationsreihenfolge der Ausgangsreihe beliebig geändert werden, ohne dass sich dadurch der Wert der Reihe ändert.

Beispiel

Funktionenreihe:

$$\sum_{k=1}^{\infty} \frac{x^2}{(1+x^2)^k}, \quad x \in \mathbb{R}$$

Für $x = 0$ gilt $f_k(x) = \frac{x^2}{(1+x^2)^k} = 0$.

Für festes $x \neq 0$ ist $\frac{1}{1+x^2} < 1$, die Reihe $\sum_{k=1}^{\infty} f_k(x) = \sum_{k=1}^{\infty} x^2 \left(\frac{1}{1+x^2}\right)^k$ ist also eine konvergente geometrische Reihe (Abschn. 8.2.3) mit der Grenzfunktion

$$
\begin{aligned}
F(x) &= \sum_{k=1}^{\infty} x^2 \left(\frac{1}{1+x^2}\right)^k = x^2 \sum_{k=1}^{\infty} \left(\frac{1}{1+x^2}\right)^k\\
&= x^2 \cdot \frac{1}{1-\frac{1}{1+x^2}} = x^2 \cdot \frac{1}{\frac{x^2}{1+x^2}} = 1 + x^2.
\end{aligned}
$$

Die Funktionenreihe $\sum_{k=1}^{\infty} \frac{x^2}{(1+x^2)^k}$ ist für alle $x \in \mathbb{R}$ konvergent mit der Grenzfunktion

$$F(x) = \begin{cases} 1 + x^2 & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases}$$ ◀

8.6.2 Potenzreihen

Funktionenreihen $\sum_{k=0}^{\infty} f_k(x)$ mit $f_k(x) = a_k(x - x_0)^k$, wobei a_k, $k = 0, 1, 2, \ldots$, und x_0 reelle Zahlen sind, heißen Potenzreihen.

Potenzreihe

$$\sum_{k=0}^{\infty} a_k(x - x_0)^k = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots$$

Die reellen Zahlen $a_0, a_1, a_2, \ldots$ heißen Koeffizienten der Potenzreihe, x_0 heißt ihr Entwicklungspunkt.

Der Konvergenzbereich einer Potenzreihe ist die Menge aller Zahlen, die man für x einsetzen kann, so dass die entstehende numerische Reihe konvergiert.

Beispiel

Zum Konvergenzbereich der Potenzreihe $\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \ldots$ gehört die Zahl $\frac{1}{2}$, denn die numerische Reihe $\sum_{k=0}^{\infty} (\frac{1}{2})^k$ konvergiert, die Zahl -2 aber nicht, da die Reihe $\sum_{k=0}^{\infty} (-2)^k$ divergiert. ◀

Für den Konvergenzbereich einer Potenzreihe gibt es drei Möglichkeiten:

1. Die Reihe konvergiert nur im Entwicklungspunkt x_0.
2. Die Reihe konvergiert für alle $x \in \mathbb{R}$.
3. Die Reihe konvergiert für alle x aus einem endlichen offenen Intervall $(x_0 - r, x_0 + r)$. In den beiden Randpunkten $x_0 - r$ und $x_0 + r$ kann die Reihe konvergent oder divergent sein.

Die Zahl r heißt Konvergenzradius der Potenzreihe. Sie wird bestimmt durch $r = \lim_{n\to\infty} |\frac{a_n}{a_{n+1}}|$ oder $r = \frac{1}{\lim_{n\to\infty} \sqrt[n]{|a_n|}}$, falls diese Grenzwerte existieren.

Die erste Möglichkeit für den Konvergenzbereich einer Potenzreihe folgt aus $\lim_{n\to\infty} \sqrt[n]{|a_n|} = \infty$, es ist dann $r = 0$. Die zweite Möglichkeit folgt aus $\lim_{n\to\infty} \sqrt[n]{|a_n|} = 0$, es ist dann $r = \infty$.

Beispiele

1. $\sum_{k=1}^{\infty} k^k x^k$

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{n^n} = n = \infty$$

Somit ist $r = 0$, der Konvergenzbereich besteht also nur aus dem Entwicklungspunkt x_0.

2. $\sum_{k=0}^{\infty} \frac{2^{k-1}}{k^{k+1}} x^k$

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{\frac{2^n}{2n \cdot n^n}} = \lim_{n\to\infty} \frac{1}{\sqrt[n]{2} \cdot \sqrt[n]{n}} \cdot \frac{2}{n} = 0,$$

denn $\sqrt[n]{2} = \sqrt[n]{n} = 1$.
Es ist also $r = \infty$, die Reihe konvergiert für alle $x \in \mathbb{R}$.

3. $\sum_{k=0}^{\infty} k^2 5^k (x-2)^k$

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{n^2 5^n} = \lim_{n\to\infty} 5 \cdot \sqrt[n]{n^2} = 5 \lim_{n\to\infty} \left(\sqrt[n]{n}\right)^2 = 5$$

Der Konvergenzradius ist folglich $r = \frac{1}{5}$. Um den genauen Konvergenzbereich zu bestimmen, müssen die Randpunkte $x_1 = 2 - \frac{1}{5} = \frac{9}{5}$ und $x_2 = 2 + \frac{1}{5} = \frac{11}{5}$ untersucht werden. Für x_1 erhält man die numerische Reihe $\sum_{k=0}^{\infty} k^2 5^k (-\frac{1}{5})^k = \sum_{k=0}^{\infty} (-1)^k k^2$ und für x_2 die Reihe $\sum_{k=0}^{\infty} k^2 5^k (\frac{1}{5})^k = \sum_{k=0}^{\infty} k^2$.
Beide numerischen Reihen divergieren, die Potenzreihe konvergiert somit in keinem der Randpunkte. Der Konvergenzbereich der Potenzreihe ist das offene Intervall $(\frac{9}{5}, \frac{11}{5})$. ◀

Im Konvergenzbereich der Potenzreihe wird die Grenzfunktion $F(x)$ durch die Potenzreihe dargestellt.

$$F(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

Eine durch eine Potenzreihe dargestellte Funktion $F(x)$ ist im Innern des Konvergenzbereichs stetig. Eine Potenzreihe konvergiert in jedem abgeschlossenen Teilintervall des Konvergenzbereichs gleichmäßig. Jede Potenzreihe darf im Innern des Konvergenzbereichs gliedweise differenziert und integriert werden.

Die so entstehenden Potenzreihen haben den gleichen Konvergenzradius r wie die Ausgangsreihe. Für $x \in (x_0 - r, x_0 + r)$ gilt somit

$$F(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

$$F'(x) = \sum_{k=0}^{\infty} k \cdot a_k (x - x_0)^{k-1}$$

$$\int F(x)dx = \sum_{k=0}^{\infty} \frac{a_k}{k+1} (x - x_0)^{k+1} + C$$

Die Integrationskonstante ist durch Einsetzen (zum Beispiel $x = x_0$) zu bestimmen.

Beispiel

Potenzreihe:

$$F(x) = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

(Abschn. 8.2.3)

Konvergenzradius:

$$r = \left| \frac{a_n}{a_{n+1}} \right| = \frac{1}{1} = 1$$

In den Randpunkten $x_1 = -1$ und $x_2 = 1$ divergiert die numerische Reihe, der Konvergenzbereich der Potenzreihe ist somit das offene Intervall $(-1, 1)$.

Differentiation:

$$F'(x) = \sum_{k=0}^{\infty} k x^{k-1} = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2} \quad \text{für } x \in (-1, 1)$$

Integration:

$$\int F(x)dx = \sum_{k=0}^{\infty} \frac{1}{k+1} x^{k+1} + C = \int \frac{dx}{1-x} = -\ln(1-x)$$

Aus $x = 0$ folgt $C = 0$, es gilt daher

$$\sum_{k=0}^{\infty} \frac{1}{k+1} x^{k+1} = -\ln(1-x) \quad \text{für } x \in (-1, 1)$$ ◀

Für zwei Potenzreihen $F(x) = \sum_{k=0}^{\infty} a_k(x-x_0)^k$ und $G(x) = \sum_{k=0}^{\infty} b_k(x-x_0)^k$ mit dem gleichen Entwicklungspunkt x_0 gelten folgende Rechenregeln (x liegt dabei im Innern des Konvergenzbereichs beider Reihen):

Summe

$$F(x)+G(x) = \sum_{k=0}^{\infty}(a_k+b_k)(x-x_0)^k$$

Differenz

$$F(x)-G(x) = \sum_{k=0}^{\infty}(a_k-b_k)(x-x_0)^k$$

Produkt

$$\begin{aligned} F(x)\cdot G(x) &= a_0b_0 + (a_0b_1+a_1b_0)(x-x_0) \\ &\quad + (a_0b_2+a_1b_1+a_2b_0)(x-x_0)^2+\ldots \\ &= \sum_{k=0}^{\infty}(a_0b_k+a_1b_{k-1}+\ldots+a_kb_0)(x-x_0)^k \end{aligned}$$

Quotient

$$\frac{F(x)}{G(x)} = \sum_{k=0}^{\infty} c_k(x-x_0)^k$$

Die Produktdarstellung nennt man Cauchy-Produkt (nach dem französischen Mathematiker Augustin Louis Cauchy, 1789–1857). Den Quotienten berechnet man durch Multiplikation mit dem Nenner

$$\sum_{k=0}^{\infty} a_k(x-x_0)^k = \Big(\sum_{k=0}^{\infty} b_k(x-x_0)^k\Big)\Big(\sum_{k=0}^{\infty} c_k(x-x_0)^k\Big)$$

und anschließendem Koeffizientenvergleich der Potenzen von $(x-x_0)^k$ auf beiden Seiten des Gleichheitszeichens.

8.6.3 Fourier-Reihen

In der Praxis treten häufig periodische Vorgänge auf (zum Beispiel Schwingungen in der Elektrotechnik, Akustik oder Optik), die sich am besten durch periodische Funktionen beschreiben lassen. Oftmals ist es dann notwendig oder vorteilhaft, eine solche periodische Funktion exakt oder angenähert durch eine Funktionenreihe, deren Summanden bestimmte Summen aus trigonometrischen Funktionen sind, darzustellen.

Eine Funktion, deren Funktionsgleichung die Bedingung $f(x+T) = f(x)$ für alle x aus dem Definitionsbereich D erfüllt, wobei $T \neq 0$ eine Konstante ist, heißt periodische Funktion (Abschn. 5.2.7). Die kleinste positive Zahl T mit dieser Eigenschaft heißt Periode der Funktion, man nennt die Funktion $y = f(x)$ auch T-periodisch.

Eine T-periodische Funktion $y = f(x)$ ist eindeutig bestimmt durch ihr Verhalten auf einem beliebigen Intervall der Länge T. Diese Tatsache führte zur Einführung von Fourier-Reihen.

Funktionenreihen $\sum_{k=0}^{\infty} f_k(x)$ mit $f_k(x) = a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x$ für $k = 1,2,3,\ldots$ (a_k und b_k sind reelle Zahlen) und $f_0(x) = \frac{a_0}{2}$ heißen trigonometrische Reihen.

Trigonometrische Reihe

$$\sum_{k=0}^{\infty} f_k(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x\right)$$

Dabei sind (a_k), $k = 0,1,2,\ldots$ und (b_k), $k = 1,2,3,\ldots$ Zahlenfolgen.

Ist eine trigonometrische Reihe für alle $x \in \mathbb{R}$ konvergent, so stellt sie in $\mathbb{R}$ eine Funktion $F(x)$ dar:

$$F(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x\right)$$

In diesem Fall ist $F(x)$ eine $2p$-periodische Funktion, denn es gilt für alle $x \in \mathbb{R}$:

$$\begin{aligned} &F(x+2p) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}(x+2p) + b_k\sin\frac{\pi k}{p}(x+2p)\right) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\left(\frac{\pi k}{p}x+2\pi k\right) + b_k\sin\left(\frac{\pi k}{p}x+2\pi k\right)\right) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x\right) = F(x) \end{aligned}$$

Sind die numerischen Reihen $\sum_{k=1}^{\infty} a_k$ und $\sum_{k=1}^{\infty} b_k$ absolut konvergent (Abschn. 8.6.1), dann gilt:

1. Die trigonometrische Reihe

$$\frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x\right)$$

   ist für alle $x \in \mathbb{R}$ konvergent.
2. Die Funktion $F(x)$ mit

$$F(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x + b_k\sin\frac{\pi k}{p}x\right)$$

   ist in $\mathbb{R}$ stetig.

Es sei $y = f(x)$ eine $2p$-periodische Funktion, die über dem Intervall $[-p,p]$ integrierbar ist. Dann heißen die Zahlen a_k und b_k die Fourier-Koeffizienten der Funktion $f(x)$.

Fourier-Koeffizienten von $f(x)$

$$a_k = \frac{1}{p}\int_{-p}^{p} f(x)\cdot\cos\frac{\pi k}{p}x\,dx \quad (k\in\mathbb{N})$$

$$b_k = \frac{1}{p}\int_{-p}^{p} f(x)\cdot\sin\frac{\pi k}{p}x\,dx \quad (k\in\mathbb{N}^*)$$

Die mit Hilfe dieser Fourier-Koeffizienten gebildete trigonometrische Reihe $\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos\frac{\pi k}{p}x+b_k sin\frac{\pi k}{p}x)$ heißt die zur Funktion $f(x)$ gehörende oder die durch $f(x)$ erzeugte oder die $f(x)$ formal zugeordnete Fourier-Reihe (Name nach dem französischen Mathematiker Jean-Baptiste-Joseph Fourier, 1768–1830). Man schreibt:

$f(x)$ formal zugeordnete Fourier-Reihe

$$f(x)\sim\frac{a_0}{2}+\sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi k}{p}x+b_k\sin\frac{\pi k}{p}x\right) \qquad (*)$$

Das Zeichen $\sim$ soll andeuten, dass die Zuordnung von $f(x)$ zu ihrer Fourier-Reihe eine formale Zuordnung ist. Es ist nichts ausgesagt über die Konvergenz dieser Fourier-Reihe und nichts darüber, ob im Falle der Konvergenz die Fourier-Reihe die Funktion $f(x)$ auch darstellt, das heißt, ob in $(*)$ die Gleichheit gilt.

Da die Integranden der Fourier-Koeffizienten $2p$-periodisch sind, kann dort jedes Intervall der Länge $2p$ als Integrationsintervall verwendet werden, zum Beispiel $[0, 2p]$.

Ist $f(x)$ eine gerade Funktion, das heißt, gilt $f(x)=f(-x)$ (Abschn. 5.2.2), so folgt

$f(x)$ gerade Funktion

$$a_k = \frac{2}{p}\int_{0}^{p} f(x)\cdot\cos\frac{\pi k}{p}x\,dx \quad (k\in\mathbb{N})$$

$$b_k = 0 \quad (k\in\mathbb{N}^*)$$

Für eine ungerade Funktion $f(x)$, also $f(x)=-f(-x)$, folgt

$f(x)$ ungerade Funktion

$$a_k = 0 \quad (k\in\mathbb{N})$$

$$b_k = \frac{2}{p}\int_{0}^{p} f(x)\cdot\sin\frac{\pi k}{p}x\,dx \quad (k\in\mathbb{N}^*)$$

Die Bestimmung der Fourier-Koeffizienten einer Funktion heißt harmonische Analyse.

Der Satz von Dirichlet (nach dem Mathematiker Peter Gustav Lejeune Dirichlet, 1805–1859) gibt hinreichende Bedingungen dafür an, dass eine Funktion $f(x)$ durch ihre (konvergente) Fourier-Reihe dargestellt wird, das heißt dafür, dass in $(*)$ Gleichheit für alle x gilt.

Satz von Dirichlet:

(1) Die Funktion $f(x)$ mit der Periode $2p$ ist im Intervall $[-p,p]$ bis auf endlich viele Sprungstellen (Abschn. 8.3.7) stetig.
(2) Das Intervall $[-p,p]$ lässt sich so in endlich viele Teilintervalle zerlegen, dass $f(x)$ in den einzelnen Teilintervallen monoton (monoton fallend oder monoton wachsend) (Abschn. 5.2.1) ist.

Gelten (1) und (2) für eine Funktion $f(x)$, so konvergiert die Fourier-Reihe von $f(x)$ und stellt die Funktion auch dar.

An den Sprungstellen der Funktion nimmt die Fourier-Reihe den Mittelwert an, das heißt, für eine Sprungstelle x_1 gilt

$$f(x_1)=\frac{1}{2}\left(\lim_{x\to x_1+0}f(x)+\lim_{x\to x_1-0}f(x)\right)$$

(Abschn. 8.3.2).

Häufig kommt es vor, dass die Funktion $f(x)$ nur in dem Intervall $[-p,p]$ definiert, nicht aber periodisch ist. In diesem Fall setzt man die Funktion $f(x)$ einfach nach links und nach rechts periodisch fort und bestimmt dann die Fourier-Reihe zu der fortgesetzten Funktion (siehe Beispiele 2–4, Abb. 8.23, 8.24, 8.25 und 8.26).

In den folgenden Beispielen werden die Fourier-Reihen von Funktionen berechnet, die häufig in der Praxis auftreten.

Beispiele

1. Rechteckpuls
Es sei $f(x)$ die 2π-periodische Funktion mit

$$f(x)=\begin{cases} A & \text{für } |x|<\frac{\pi}{2}\\ \frac{A}{2} & \text{für } |x|=\frac{\pi}{2}\\ 0 & \text{für } \frac{\pi}{2}<|x|\le\pi\end{cases}$$

(periodisch fortsetzen).
Bestimmung der formalen Fourier-Reihe von $f(x)$:
Da $f(x)$ eine gerade Funktion ist, gilt $b_k=0$ für die Fourier-Koeffizienten b_k.
Berechnung der a_k:
$k=0$:

$$a_0=\frac{2}{\pi}\int_0^{\pi}f(x)dx=\frac{2}{\pi}\int_0^{\frac{\pi}{2}}A\,dx=A$$

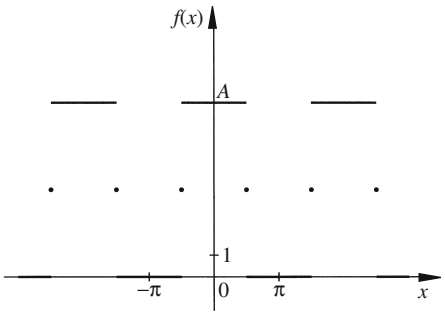


Abb. 8.23 Rechteckpuls

$k \in \mathbb{N}^*$:

$$a_k = \frac{2}{\pi}\int_0^{\pi} f(x)\cos kx dx$$

$$= \frac{2A}{\pi}\int_0^{\frac{\pi}{2}} \cos kx dx = \frac{2A}{k\pi}\sin\frac{k\pi}{2}$$

$$= \begin{cases} (-1)^{n+1}\cdot\frac{2A}{(2n-1)\pi} & \text{für } k \text{ ungerade}, k = 2n-1 \\ 0 & \text{für } k \text{ gerade}, k = 2n \end{cases}$$

Fourier-Reihe von $f(x)$:

$$f(x) \sim \frac{A}{2} + \frac{2A}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-1}\cdot\cos(2n-1)x$$

$$= \frac{A}{2} + \frac{2A}{\pi}\left(\cos x - \frac{\cos 3x}{3} + \frac{cos5x}{5} \mp \ldots\right)$$

Darstellbarkeit:
Die Funktion ist in den Teilintervallen $(-\pi, -\frac{\pi}{2}), (-\frac{\pi}{2}, \frac{\pi}{2})$ und $(\frac{\pi}{2}, \pi)$ stetig und monoton. An den Sprungstellen ist der Funktionswert gleich dem arithmetischen Mittel der einseitigen Grenzwerte. Nach dem Satz von Dirichlet wird die Funktion $f(x)$ somit durch ihre Fourier-Reihe dargestellt:

$$f(x) = \frac{A}{2} + \frac{2A}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-1}\cdot\cos(2n-1)x$$

2. Dreieckpuls
Man setze die Funktion $f(x) = 1 - |x|$, $-1 \le x \le 1$ periodisch fort und entwickle sie in eine Fourier-Reihe. Die periodisch fortgesetzte Funktion hat die Periode 2, in den Formeln muss deshalb $2p = 2$, also $p = 1$ gesetzt werden. Da $f(x)$ eine gerade Funktion ist, folgt $b_k = 0$ für $k = 1, 2, 3, \ldots$
Berechnung der a_k:
$k = 0$:

$$a_0 = 2\int_0^1 (1-x)dx = -(1-x)^2\Big|_0^1 = 1,$$

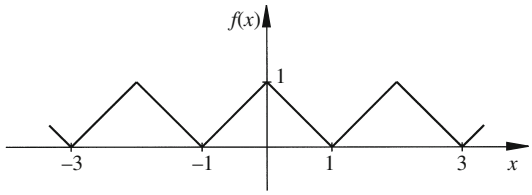


Abb. 8.24 Graph der periodisch fortgesetzten Funktion von Beispiel 2

denn im Intervall $[0, 1]$ gilt $1 - |x| = 1 - x$.
$k \neq 0$:

$$a_k = 2\int_0^1 (1-x)\cos\pi kx dx$$

$$= 2\left(\int_0^1 \cos\pi kx dx - \int_0^1 x\cos\pi kx dx\right)$$

$$= 2\left(\frac{1}{\pi k}\sin\pi kx\Big|_0^1 \left(\frac{x}{\pi k}\sin\pi kx + \frac{1}{(\pi k)^2}\cos\pi kx\right)\Big|_0^1\right)$$

$$= 2\left(-\frac{1}{(\pi k)^2}(\cos\pi k - 1)\right)$$

$$= \frac{2}{(\pi k)^2}(1-(-1)^k)$$

$$= \begin{cases} \frac{4}{(\pi k)^2} & \text{für } k \text{ ungerade}, k = 2n-1 \\ 0 & \text{für } k \text{ gerade}, k = 2n \end{cases}$$

Das zweite Integral berechnet man mit zweimaliger partieller Integration.
Fourier-Reihe von $f(x)$:

$$f(x) \sim \frac{1}{2} + \sum_{n=1}^{\infty} a_{2n-1}\cos\pi(2n-1)x$$

$$= \frac{1}{2} + \sum_{n=1}^{\infty}\frac{4}{\pi^2(2n-1)^2}\cos(2n-1)\pi x$$

Darstellbarkeit:
(1) Die Funktion $f(x) = 1 - |x|$ ist im Intervall $[-1, 1]$ überall stetig.
(2) Zerlegt man das Periodenintervall $[-1, 1]$ in die zwei Teilintervalle $[-1, 0]$ und $[0, 1]$, so ist $f(x)$ in $[-1, 0]$ monoton wachsend und in $[0, 1]$ monoton fallend.
Nach dem Satz von Dirichlet wird die Funktion $f(x)$ somit durch ihre Fourier-Reihe dargestellt:

$$f(x) = \frac{1}{2} + \sum_{n=1}^{\infty}\frac{4}{\pi^2(2n-1)^2}\cos(2n-1)\pi x$$

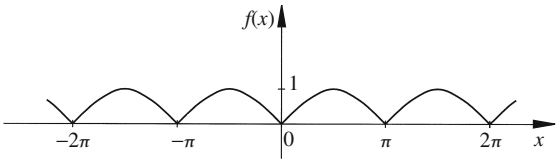


Abb. 8.25 Graph der periodisch fortgesetzten Funktion von Beispiel 3

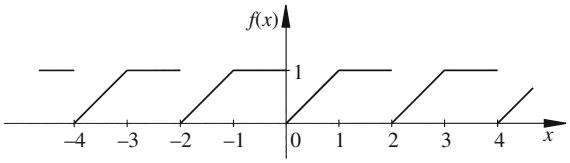


Abb. 8.26 Graph der periodisch fortgesetzten Funktion von Beispiel 4

3. Man setze die Funktion $y = \sin x$, $0 \le x \le \pi$ periodisch fort und entwickle sie in eine Fourier-Reihe.
Die fortgesetzte Funktion ist die Funktion $f(x) = |\sin x|$, sie hat die Periode π, es folgt also $p = \frac{\pi}{2}$.
Wegen $|\sin(-x)| = |-\sin x| = |\sin x|$ ist $f(x)$ gerade, es folgt $b_k = 0$ für alle k.
Berechnung der a_k:
$k = 0$:

$$a_0 = \frac{2}{\pi}\int_0^{\pi} \sin x dx = \frac{2}{\pi}(-\cos x)\Big|_0^{\pi} = \frac{2}{\pi}(1+1) = \frac{4}{\pi}$$

denn im Intervall $[0, \pi]$ gilt $|\sin x| = \sin x$.
$k \neq 0$:

$$a_k = \frac{2}{\pi}\int_0^{\pi} \sin x \cos 2kx dx$$

$$= \frac{2}{\pi}\left(\frac{1}{2k}\sin x \sin 2kx\Big|_0^{\pi} - \frac{1}{2k}\int_0^{\pi} \cos x \sin 2kx dx\right)$$

$$= \frac{2}{\pi}\left(\frac{1}{4k^2}\cos x \cos 2kx\Big|_0^{\pi} + \frac{1}{4k^2}\int_0^{\pi} \sin x \cos 2kx dx\right)$$

Bei beiden Umformungen wurde die Methode der partiellen Integration angewandt. Es folgt

$$a_k = \frac{2}{\pi}\cdot\frac{4k^2}{4k^2-1}\left(\frac{1}{4k^2}\cos x \cos 2kx\right)\Big|_0^{\pi}$$
$$= \frac{2}{\pi(4k^2-1)}(-1-1) = -\frac{4}{\pi(4k^2-1)}$$

Fourier-Reihe von $f(x)$:

$$f(x) \sim \frac{2}{\pi} - \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{cos2kx}{4k^2-1}$$

Darstellbarkeit:
(1) Die Funktion $f(x) = |\sin x|$ ist im Intervall $[0, \pi]$ stetig.
(2) Die Funktion $f(x)$ ist in $[0, \frac{\pi}{2}]$ monoton steigend und in $[\frac{\pi}{2}, \pi]$ monoton fallend.
Nach dem Satz von Dirichlet gilt somit Gleichheit:

$$f(x) = \frac{2}{\pi} - \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{\cos 2kx}{4k^2-1}$$

4. Man setze die Funktion

$$f(x) = \begin{cases} x & \text{für } 0 < x \le 1 \\ 1 & \text{für } 1 \le x < 2 \end{cases}$$

periodisch fort und berechne die zugehörige Fourier-Reihe.
Die fortgesetzte Funktion hat die Periode 2, es folgt $p = 1$.
Die Funktion ist weder gerade noch ungerade, es müssen also sowohl die Fourier-Koeffizienten a_k als auch b_k berechnet werden.
Berechnung der a_k:
$k = 0$:

$$a_0 = \int_0^2 f(x)dx = \int_0^1 x dx + \int_1^2 1 dx = \frac{1}{2}x^2\Big|_0^1 + x\Big|_1^2 = \frac{3}{2}$$

$k \neq 0$:

$$a_k = \int_0^2 f(x)\cos \pi kx dx$$

$$= \int_0^1 x\cos \pi kx dx + \int_1^2 \cos \pi kx dx$$

$$= \left(\frac{\cos \pi kx}{k^2\pi^2} + x\frac{\sin \pi kx}{k\pi}\right)\Big|_0^1 + \frac{\sin \pi kx}{k\pi}\Big|_1^2$$

$$= \frac{\cos \pi k}{k^2\pi^2}\frac{1}{k^2\pi^2}$$

$$= \frac{1}{k^2\pi^2}((-1)^k - 1) = \begin{cases} 0 & \text{für } k \text{ gerade} \\ -\frac{2}{k^2\pi^2} & \text{für } k \text{ ungerade} \end{cases}$$

Berechnung der b_k:

$$b_k = \int_0^2 f(x)\sin \pi kx dx$$

$$= \int_0^1 x\sin \pi kx dx + \int_1^2 \sin \pi kx dx$$

$$= \left(\frac{\sin \pi kx}{k^2\pi^2} - \frac{x\cos \pi kx}{k\pi}\right)\Big|_0^1 + \frac{(-\cos \pi kx)}{k\pi}\Big|_1^2$$

$$= -\frac{\cos \pi k}{k\pi} - \frac{\cos \pi 2k}{k\pi} + \frac{cos\pi k}{k\pi} = -\frac{1}{k\pi}$$

Sowohl bei der Berechnung von a_k als auch von b_k wurde zur Lösung des ersten Integrals zweimalige partielle Integration durchgeführt.
Fourier-Reihe von $f(x)$:

$$f(x) \sim \frac{3}{4} + \sum_{k=1}^{\infty} \left(\frac{(-1)^k - 1}{k^2 \pi^2} \cos \pi k x - \frac{1}{k\pi} \sin \pi k x \right)$$

Darstellbarkeit:
Die Funktion ist im Intervall $(0, 2)$ stetig und monoton. Mit der Definition $f(0) = f(2) = \frac{1}{2}(1 + 0) = \frac{1}{2}$ sind die Bedingungen des Satzes von Dirichlet erfüllt, die Funktion $f(x)$ wird somit durch ihre Fourier-Reihe dargestellt:

$$f(x) = \frac{3}{4} + \sum_{k=1}^{\infty} \left(\frac{(-1)^k - 1}{k^2 \pi^2} \cos \pi k x - \frac{1}{k\pi} \sin \pi k x \right)$$

◀

9 Anhang

Arnfried Kemnitz



9.1 Symbole und Bezeichnungsweisen 184
9.2 Das griechische Alphabet 185

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_9

9.1 Symbole und Bezeichnungsweisen

Die folgenden mathematischen Zeichen sind nach DIN 1302 (April 1994) und DIN 5473 (Juli 1992) zusammengestellt.

$=$	gleich
$\neq$	ungleich
$\approx$	ungefähr gleich
$<$	kleiner als
$\leq$	kleiner oder gleich
$>$	größer als
$\geq$	größer oder gleich
$\ll$	sehr viel kleiner als
$\gg$	sehr viel größer als
$\sim$	proportional
$\pm$	plus oder minus
$\mp$	minus oder plus
$\sum_{k=1}^{n} a_k$	$= a_1 + a_2 + a_3 + \ldots + a_n$; Summe über a_k von $k = 1$ bis $k = n$
$\prod_{k=1}^{n} a_k$	$= a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n$; Produkt über a_k von $k = 1$ bis $k = n$
$\{a, b, c\}$	Menge aus den Elementen a, b, c
$\{x \mid E(x)\}$	Menge aller x, die die Eigenschaft $E(x)$ haben
$\in$	Element von
$\notin$	nicht Element von
$\subseteq$	Teilmenge
$\varnothing$	leere Menge
$\bigcup$	Vereinigung von Mengen
$\bigcap$	Durchschnitt von Mengen
$\lvert M\rvert$	Mächtigkeit der Menge M
$A \wedge B$	A und B
$A \vee B$	A oder B
$\neg A$	nicht A (Negation von A)
$A \Rightarrow B$	aus A folgt B
$A \Leftrightarrow B$	A und B sind äquivalent (gleichwertig)
(a, b)	geordnetes Paar
(a, b, c)	geordnetes Tripel
$\parallel$	parallel
$\overline{AB}$	Strecke AB
$\lvert\overline{AB}\rvert$	Länge (Betrag) der Strecke AB
$\vec{a}$	Vektor a
$\overrightarrow{PQ}$	Vektor PQ
$\lvert\vec{a}\rvert$, $\lvert\overrightarrow{PQ}\rvert$	Länge des Vektors
$\sim$	ähnlich
$\cong$	kongruent
$\mathbb{N}$	$= \{0, 1, 2, 3, \ldots\}$; Menge der natürlichen Zahlen
$\mathbb{Z}$	$= \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$; Menge der ganzen Zahlen
$\mathbb{Q}$	$= \{\frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\}$; Menge der rationalen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{C}$	$= \{z = a + bj \mid a, b \in \mathbb{R}, j = \sqrt{-1}\}$; Menge der komplexen Zahlen
$\mathbb{N}^*$	$= \{1, 2, 3, \ldots\}$; Menge der natürlichen Zahlen ohne die Null
$\mathbb{Z}^*$	$= \{\ldots, -3, -2, -1, 1, 2, 3, \ldots\} = \{x \mid x \in \mathbb{Z}, x \neq 0\}$; Menge der ganzen Zahlen ohne die Null
$\mathbb{Q}^*$	$= \{\frac{m}{n} \mid m, n \in \mathbb{Z}^*\} = \{x \mid x \in \mathbb{Q}, x \neq 0\}$; Menge der rationalen Zahlen ohne die Null
$\mathbb{R}^*$	$= \{x \mid x \in \mathbb{R}, x \neq 0\}$; Menge der reellen Zahlen ohne die Null
$\mathbb{Z}^+$	$= \mathbb{N}^* = \{1, 2, 3, \ldots\} = \{x \mid x \in \mathbb{Z}, x > 0\}$; Menge der positiven ganzen Zahlen
$\mathbb{Q}^+$	$= \{\frac{m}{n} \mid m, n \in \mathbb{N}^*\} = \{x \mid x \in Q, x > 0\}$; Menge der positiven rationalen Zahlen
$\mathbb{R}^+$	$= \{x \mid x \in \mathbb{R}, x > 0\}$; Menge der positiven reellen Zahlen
$\mathbb{P}$	$= \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \ldots\}$; Menge der Primzahlen
j	$= \sqrt{-1}$; imaginäre Einheit
∞	unendlich (größer als jede reelle Zahl)
$-\infty$	minus unendlich (kleiner als jede reelle Zahl)
$n!$	$= 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$; n Fakultät
$\binom{n}{k}$	$= \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}{1\cdot 2\cdot 3\cdot\ldots\cdot k}$; Binomialkoeffizient n über k
$\lvert a\rvert$	Betrag oder Absolutbetrag einer Zahl a
a^n	a hoch n, n-te Potenz von a
$\sqrt{a}$	Wurzel aus a
$\sqrt[n]{a}$	n-te Wurzel aus a
$\log_a b$	Logarithmus b zur Basis a
$\lg b$	dekadischer Logarithmus (Zehnerlogarithmus), Logarithmus zur Basis $a = 10$
$\ln b$	natürlicher Logarithmus, Logarithmus zur Basis $a = e = 2{,}71828182\ldots$
$\operatorname{ld} b$	binärer Logarithmus (Zweierlogarithmus), Logarithmus zur Basis $a = 2$
$[a, b]$	$= \{x \mid x \in \mathbb{R} \text{ und } a \leq x \leq b\}$; abgeschlossenes beschränktes Intervall
(a, b)	$= \{x \mid x \in \mathbb{R} \text{ und } a < x < b\}$; offenes beschränktes Intervall
$[a, b)$	$= \{x \mid x \in \mathbb{R} \text{ und } a \leq x < b\}$; halboffenes beschränktes Intervall
$(a, b]$	$= \{x \mid x \in \mathbb{R} \text{ und } a < x \leq b\}$; halboffenes beschränktes Intervall
$[a, \infty)$	$= \{x \mid x \in \mathbb{R} \text{ und } x \geq a\}$; halboffenes Intervall, nach rechts unbeschränkt
(a, ∞)	$= \{x \mid x \in \mathbb{R} \text{ und } x > a\}$; offenes Intervall, nach rechts unbeschränkt
$(-\infty, a]$	$= \{x \mid x \in \mathbb{R} \text{ und } x \leq a\}$; halboffenes Intervall, nach links unbeschränkt
$(-\infty, a)$	$= \{x \mid x \in \mathbb{R} \text{ und } x < a\}$; offenes Intervall, nach links unbeschränkt
$(-\infty, \infty)$	$= \{x \mid x \in \mathbb{R}\}$; offenes Intervall, nach links und nach rechts unbeschränkt
(a_n)	$= (a_1, a_2, a_3, \ldots)$; Folge, Zahlenfolge
$\lim_{n\to\infty} a_n$	Limes, Grenzwert der Folge (a_n)
$\lim_{x\to a} f(x)$	Grenzwert (Limes) der Funktion $f(x)$ für x gegen a
$\lim_{x\to a-0} f(x)$	linksseitiger Grenzwert der Funktion $y = f(x)$ an der Stelle $x = a$
$\lim_{x\to a+0} f(x)$	rechtsseitiger Grenzwert der Funktion $y = f(x)$ an der Stelle $x = a$

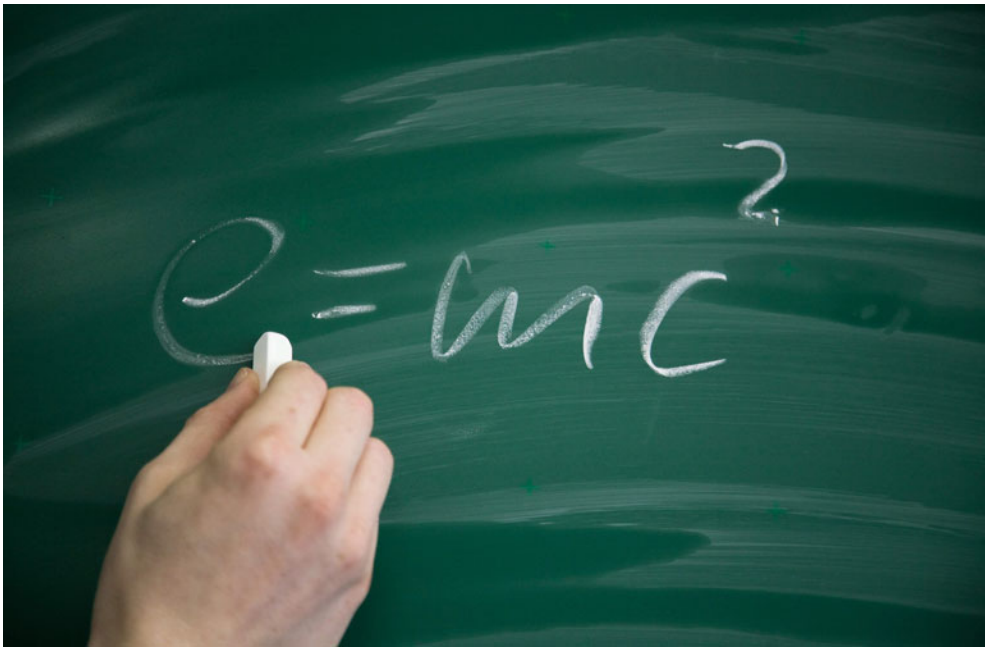
$f'(x_0)$	Ableitung von $f(x)$ an der Stelle $x = x_0$
$\frac{df}{dx}(x_0)$	Ableitung von $f(x)$ an der Stelle $x = x_0$
$f'(x)$	Ableitung der Funktion $f(x)$
$f''(x)$	zweite Ableitung der Funktion $f(x)$
$f'''(x)$	dritte Ableitung der Funktion $f(x)$
$f^{(n)}(x)$	n-te Ableitung der Funktion $f(x)$
$\int f(x)dx$	unbestimmtes Integral der Funktion $y = f(x)$
$\int_a^b f(x)dx$	bestimmtes Integral der Funktion $y = f(x)$ von $x = a$ bis $x = b$
$(f_n(x))$	$= (f_1(x), f_2(x), f_3(x), \ldots)$; Funktionenfolge

9.2 Das griechische Alphabet

Alpha	A	α	Jota	I	ι	Rho	P	ρ
Beta	B	β	Kappa	K	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	T	τ
Delta	Δ	δ	My	M	μ	Ypsilon	Υ	υ
Epsilon	E	ε	Ny	N	ν	Phi	Φ	φ
Zeta	Z	ζ	Xi	Ξ	ξ	Chi	X	χ
Eta	H	η	Omikron	O	o	Psi	Ψ	ψ
Theta	Θ	ϑ	Pi	Π	π	Omega	Ω	ω

Physik

II

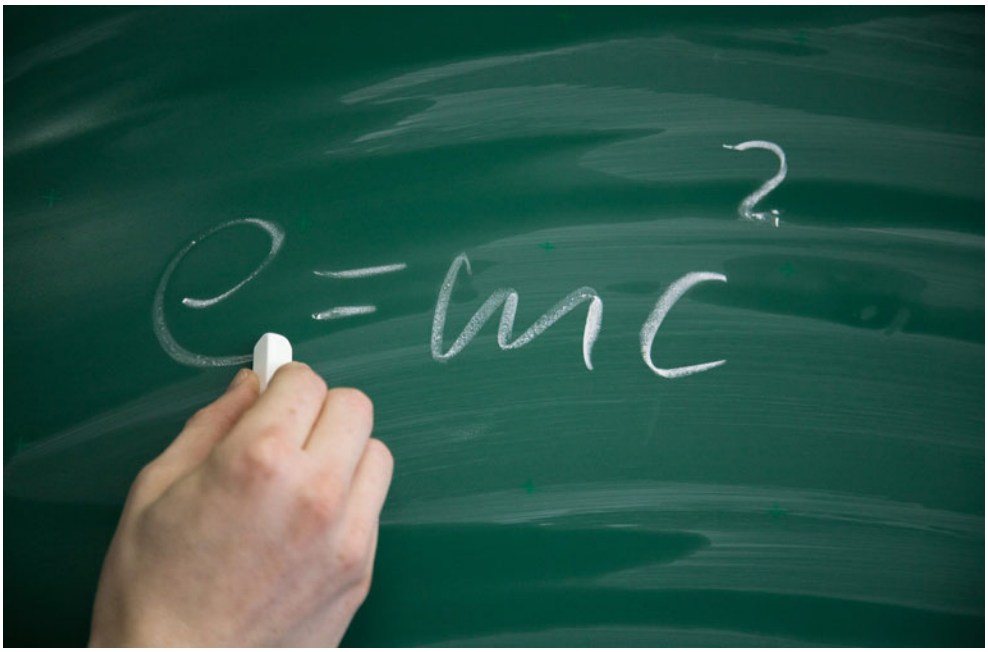


10 Einführung 189
11 Mechanik 193
12 Thermodynamik 213
13 Schwingungen 221
14 Wellen 227
15 Akustik 233
16 Optik 237
17 Anhang 253

Einführung

10

Horst Steffen



10.1 Physikalische Größen 190
10.2 SI-System 190

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_10

Die Physik ist ein Teilgebiet der Naturwissenschaften und beschäftigt sich mit der leblosen Umwelt. In der Physik wird versucht, die Gesetzmäßigkeiten der unbelebten Materie durch Beobachtungen und Messungen zu erfassen und in einer mathematischen Gleichung darzustellen. Ist diese bekannt, so kann man die physikalischen Gesetze für technische Zwecke ausnutzen. Die Physik wird in folgende Gebiete unterteilt: Mechanik, Thermodynamik (Wärmelehre), Elektrizität und Magnetismus, Wellenlehre, Akustik, Optik, Atom- und Kernphysik, Festkörperphysik, Relativitätstheorie.

10.1 Physikalische Größen

Eine *physikalische Größe* kennzeichnet Eigenschaften, Zustände oder Größen von messbaren Objekten. Sie ist das Produkt einer *Maßzahl* und einer *Einheit*.

$$Größe = Maßzahl \cdot Einheit$$

Um Größen und ihre Einheiten deutlich zu unterscheiden, werden für sie unterschiedliche Symbole verwendet.

$$\text{Spannung} = 100 \cdot 1\ \text{Volt}; \quad U = 100 \cdot 1\,\text{V}; \quad U = 100\,\text{V}$$

In Gleichungen werden immer physikalische Größen miteinander verbunden, das heißt, dass sowohl die Maßzahlen, aber auch die Einheiten auf beiden Seiten der Gleichung miteinander übereinstimmen müssen.

10.1.1 Skalare

Viele Größen sind neben ihrer Einheit allein durch ihre Maßzahlen eindeutig bestimmt, dazu gehören z. B. Temperatur, Masse, Energie, Leistung, Widerstand. Solche Größen werden *skalare Größen* oder *Skalare* genannt.

10.1.2 Vektoren

Bei anderen Größen reichen diese Angaben alleine nicht aus, sondern es muss zur vollständigen Beschreibung noch eine Richtung angegeben werden. Zum Beispiel ist es nicht ausreichend zu sagen, ein Auto habe eine Strecke von 100 km zurückgelegt, wenn nicht auch die Richtung der Bewegung mit angegeben wurde.

Solche gerichteten Größen werden *Vektoren* genannt. Zur vollständigen Angabe gehört ein *Betrag* (Maßzahl, Einheit) und eine *Richtung*.

Beispiele für Vektoren: Kraft, Geschwindigkeit, elektrische und magnetische Feldstärke. Wenn die Vektoreigenschaft einer Größe hervorgehoben werden soll, so wird dies entweder durch einen Pfeil über dem Größenzeichen $\vec{F}$ oder durch Fettdruck $\boldsymbol{F}$ des Zeichens kenntlich gemacht. Für die mathematische Behandlung von Gleichungen mit Vektoren wird die *Vektorrechnung* benötigt.

10.2 SI-System

Die Einheiten der physikalischen Größen sind seit 1960 im „*Système International d'Unités*“, kurz *SI-System*, festgelegt und in der Bundesrepublik Deutschland gesetzlich vorgeschrieben. Das System besteht aus *Basisgrößen* und *abgeleiteten Größen*. Die Basisgrößen sind in Tab. 10.1 angegeben.

Definitionen der Basisgrößen

1 Sekunde ist das 9.192.631.770fache der Periodendauer der dem Übergang zwischen den beiden Hyperfeinstrukturniveaus des Grundzustandes von Atomen des Nuklids ^{133}Cs entsprechenden Strahlung. (1967)

1 Meter ist die Länge der Strecke, die Licht im Vakuum während der Dauer von 1/299.792.458 Sekunden durchläuft. (1983)

1 Kilogramm ist die Masse des internationalen Kilogrammprototyps. (1889)

1 Ampere ist die Stärke eines zeitlich unveränderlichen Stroms, der, durch zwei im Vakuum parallel im Abstand von 1 Meter voneinander angeordnete, geradlinige, unendlich lange Leiter von vernachlässigbar kleinem kreisförmigen Querschnitt fließend, zwischen diesen Leitern je 1 Meter Leiterlänge die Kraft $2 \cdot 10^{-7}$ Newton hervorruft. (1948)

1 Kelvin ist der 273,16te Teil der thermodynamischen Temperatur des Tripelpunktes des Wassers. (1967)

1 Candela ist die Lichtstärke in einer bestimmten Richtung, einer Strahlungsquelle, die monochromatische Strahlung der Frequenz 540 THz aussendet und deren Strahlstärke in dieser Richtung $\frac{1}{683}\,\frac{\text{W}}{\text{sr}}$ beträgt.

Tab. 10.1 Basisgrößen und Basiseinheiten

Gebiet	Basisgröße	Formelzeichen	Basiseinheit	Einheitenzeichen
Mechanik	Zeit	t	Sekunde	s
	Länge	l	Meter	m
	Masse	m	Kilogramm	kg
Elektrotechnik	Stromstärke	I	Ampere	A
Thermodynamik	Temperatur	T	Kelvin	K
Optik	Lichtstärke	I_L	Candela	cd
Chemie	Stoffmenge	n	Mol	mol

Tab. 10.2 Vorsätze für dezimale Vielfache

Wert	Vorsatz	Zeichen	Wert	Vorsatz	Zeichen
10^{18}	Exa	E	10^{-1}	Dezi	d
10^{15}	Peta	P	10^{-2}	Zenti	c
10^{12}	Tera	T	10^{-3}	Milli	m
10^{9}	Giga	G	10^{-6}	Mikro	m
10^{6}	Mega	M	10^{-9}	Nano	n
10^{3}	Kilo	k	10^{-12}	Piko	p
10^{2}	Hekto	h	10^{-15}	Femto	f
10^{1}	Deka	da	10^{-18}	Atto	a

1 Mol ist die Stoffmenge eines Systems bestimmter Zusammensetzung, das aus ebenso vielen Teilchen besteht, wie Atome in (12/1000) kg des Nuklids ^{12}C enthalten sind. (1971)

Abgeleitete Größen

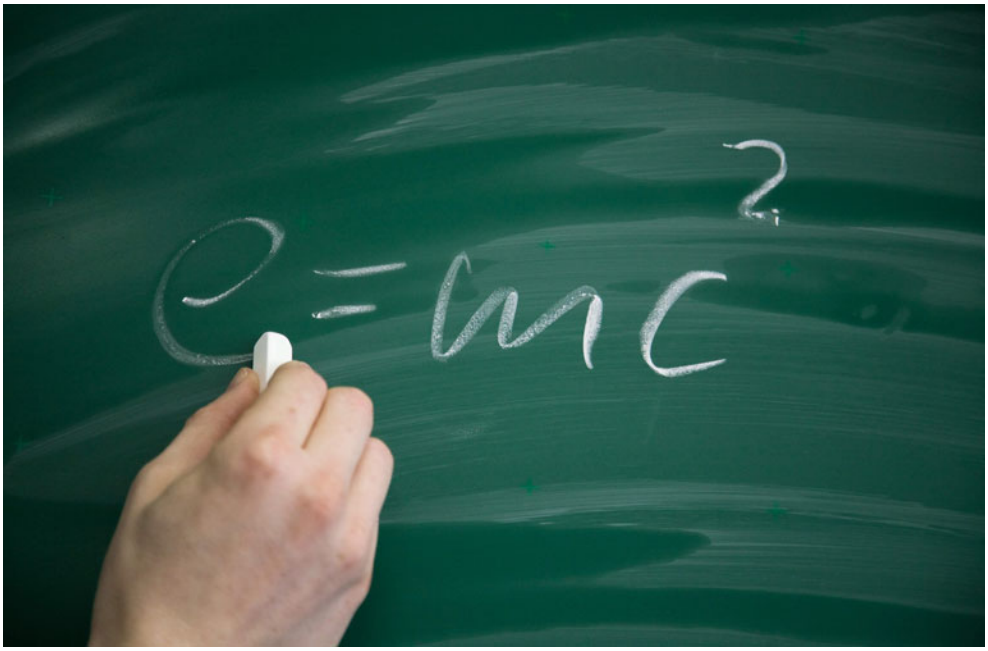
Aus den Basisgrößen lassen sich die SI-Einheiten aller anderen Größen ableiten. Eine Zusammenfassung der wichtigsten Größen finden Sie in Kap. 17.

Durch Vorsätze können dezimale Vielfache oder Teile der Maßeinheiten gebildet und damit umständlicher zu benutzende Zehnerpotenzen vermieden werden. In Tab. 10.2 sind die Vorsilben und Kurzzeichen für die Vorsätze zusammengestellt. Doppelvorsätze, wie z. B. nmm sind nicht zulässig.

Mechanik

11

Horst Steffen



11.1 Kinematik des Massenpunktes 194
11.2 Dynamik 199
11.3 Elastische Verformung fester Körper 210
11.4 Mechanik der ruhenden Flüssigkeiten und Gase 210
11.5 Hydrodynamik 212

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_11

11.1 Kinematik des Massenpunktes

Die Kinematik beschreibt die Bewegung von Körpern im Raum. Ein Punkt im Raum wird durch seine *Ortskoordinaten* festgelegt. Diese ändern sich während der Bewegung des Körpers mit der Zeit. Bei größeren Systemen können einzelne Teile des Systems völlig unterschiedliche Bewegungen durchführen, so bewegt sich bei einem fahrenden Auto ein Punkt auf der Karosserie anders als ein Punkt auf dem Reifen. Da sich aber jeder Körper aus einzelnen Massenpunkten zusammensetzt, ist die Beschreibung der Bewegung eines einzelnen Massenpunktes von grundlegender Bedeutung.

11.1.1 Eindimensionale Bewegungen

Eine Bewegung wird dann eindimensional genannt, wenn sie nur auf einer vorgeschriebenen Bahn erfolgen kann, wie es z. B. bei Schienenfahrzeugen oder auch Werkzeugschlitten der Fall ist. Zu ihrer Beschreibung ist dann neben der Zeitabhängigkeit nur eine *Ortskoordinate* ausreichend.

11.1.1.1 Geschwindigkeit

Eine wichtige Grundgröße der Kinematik ist die *Geschwindigkeit*. Sie gibt an, welcher Weg Δs in der Zeit Δt zurückgelegt wird. Die Geschwindigkeit ist ein Vektor, denn der Endzustand einer Bewegung hängt von der Richtung der Geschwindigkeit ab.

$$v = \frac{\Delta s}{\Delta t} \qquad \begin{array}{c|c|c} v & \Delta s & \Delta t \\ \hline \mathrm{m/s} & \mathrm{m} & \mathrm{s} \end{array} \qquad (11.1)$$

$\Delta s = s_2 - s_1$, Differenz der Ortskoordinaten.

$\Delta t = t_2 - t_1$, Differenz der entsprechenden Zeiten.

Ist der Betrag der Geschwindigkeit überall gleich, so spricht man von einer *gleichförmigen Bewegung*. Die Geschwindigkeit ist dann unabhängig von der Größe des Zeitabschnittes. Ändert sich dagegen die Geschwindigkeit während der Beobachtungszeit (Beispiel: anfahrendes Auto), so kann man die *Momentangeschwindigkeit* oder *Augenblicksgeschwindigkeit* nur bestimmen, wenn der Zeitabschnitt Δt, in dem der zurückgelegte Weg Δs gemessen wird, beliebig klein gemacht wird, im Grenzfall gegen 0. Ist dies nicht möglich, erhält man die *Durchschnittsgeschwindigkeit* oder auch *mittlere Geschwindigkeit*. Die Geschwindigkeit, mit der sich ein Körper geradlinig bewegt, nennt man auch *Translationsgeschwindigkeit*.

Eine Masse m befindet sich zum Zeitpunkt $t = 0$ an einem Ort mit der Ortskoordinate s_0. Sie hat eine konstante Geschwindigkeit v_0. Der nach Ablauf einer Zeit t zurückgelegte Weg s errechnet sich nach:

$$s = s_0 + v_0 \cdot t \qquad (11.2)$$

mit s_0 und v_0 als Anfangswerte der Ortskoordinate und der Geschwindigkeit.

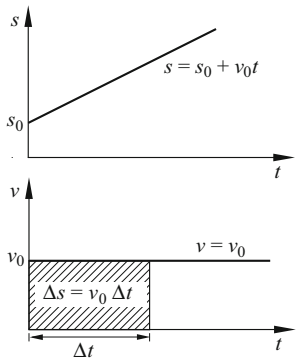


Abb. 11.1 $s(t)$ und $v(t)$-Diagramm einer gleichförmigen Bewegung

Beispiel

Ein Auto fährt mit einer konstanten Geschwindigkeit von 50 km/h. Um 9 Uhr ist es 30 km von seinem Startpunkt entfernt.

a) Welche Zeit braucht es für einen Weg von 20 km?
b) Wo befindet es sich um 11 Uhr?

Lösung:

a) $v = \frac{\Delta s}{\Delta t} \Rightarrow \Delta t = \frac{\Delta s}{v} = \frac{20\,\mathrm{km}}{50\,\mathrm{km\,h^{-1}}} = 0{,}4\,\mathrm{h} = 24\,\mathrm{min}$
b) $s = 30\,\mathrm{km} + 2\,\mathrm{h} \cdot 50\,\frac{\mathrm{km}}{\mathrm{h}} = 130\,\mathrm{km}$ ◀

11.1.1.2 Beschleunigung

Wenn sich die Geschwindigkeit im Lauf der Zeit ändert, liegt eine *beschleunigte Bewegung* vor. Die Beschleunigung a ist der Quotient aus der Änderung der Geschwindigkeit Δv in der Zeit Δt. Wie bei der Geschwindigkeit sind auch hier *Momentanbeschleunigung* und *Durchschnittsbeschleunigung* zu unterscheiden.

$$a = \frac{\Delta v}{\Delta t} \qquad \begin{array}{c|c|c} a & \Delta v & \Delta t \\ \hline \mathrm{m/s^2} & \mathrm{m/s} & \mathrm{s} \end{array} \qquad (11.3)$$

$\Delta v = v_2 - v_1$, Differenz der Geschwindigkeiten.

$\Delta t = t_2 - t_1$, Differenz der entsprechenden Zeiten.

Ist die Beschleunigung konstant, so liegt eine *gleichmäßig beschleunigte Bewegung* vor. Hat ein Körper zum Zeitpunkt $t = 0$ die Anfangsgeschwindigkeit v_0 und befindet er sich am Ort s_0, so ändern sich seine Geschwindigkeit v und die Ortskoordinate s mit der Zeit entsprechend der folgenden Gleichungen:

$$v = v_0 + at \qquad (11.4)$$

$$s = s_0 + v_0 t + \frac{1}{2} a t^2 \qquad (11.5)$$

Ist $a = 0$, so liegt eine gleichförmige Bewegung vor. Mit diesen Gleichungen können auch verzögerte Bewegungen berechnet werden; a ist dann negativ einzusetzen, der Körper wird langsamer und somit abgebremst (Abb. 11.1).

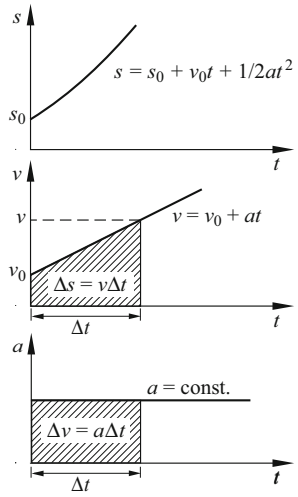


Abb. 11.2 $s(t)$-, $v(t)$- und $a(t)$-Diagramm einer gleichmäßig beschleunigten Bewegung

Beispiel

Ein Eisenbahnzug, der sich 20 km von seinem Startbahnhof befindet, fährt 30 min lang mit konstanter Geschwindigkeit $v_0 = 60\,\text{km/h}$. Dann wird er mit einer Beschleunigung $a = -0{,}222\,\text{m/s}^2$ abgebremst. Wie lang ist sein Bremsweg, und wie weit ist er dann von seinem Startbahnhof entfernt?

Lösung: Um die Bremszeit zu berechnen, wird in (11.4) die Endgeschwindigkeit $v = 0$ eingesetzt:

Bremszeit ($v = 0$):

$$t_B = -\frac{v_0}{a} = \frac{60.000\,\text{m}}{3600\,\text{s} \cdot 0{,}222\,\text{m s}^{-2}} = 75\,\text{s}$$

Entfernung von Bahnhof zu Beginn des Bremsvorganges nach (11.2):

$$s_0 = 20\,\text{km} + 60\,\frac{\text{km}}{\text{h}} \cdot 0{,}5\,\text{h} = 50\,\text{km}$$

Entfernung von Bahnhof zu Ende des Bremsvorganges nach (11.5):

$$s = 50\,\text{km} + \frac{60\,\text{km} \cdot 75\,\text{s}}{3600\,\text{s}} - \frac{0{,}222\,\text{m} \cdot 75^2\,\text{s}^2}{2\,\text{s}^2}$$
$$s = 50{,}626\,\text{km}$$ ◀

Trägt man in einer Grafik den zurückgelegten Weg als Funktion der Zeit auf, so erhält man das $s(t)$-Diagramm. Bei Auftragung der momentanen Geschwindigkeit als Funktion der Zeit das $v(t)$-Diagramm. Im $s(t)$-Diagramm ist die Momentangeschwindigkeit anschaulich als Steigung der Weg-Zeit-Kurve abzulesen, während der in einer Zeit Δt zurückgelegte Weg Δs aus dem $v(t)$-Diagramm als Fläche unter der Kurve bestimmt werden kann ($\Delta s = v \cdot \Delta t$, Abb. 11.2).

11.1.1.3 Freier Fall

Ein Beispiel für eine Bewegung unter dem Einfluss einer konstanten Beschleunigung ist die Bewegung an der Erdoberfläche allein unter dem Einfluss der Erdanziehungskraft, d. h. ohne Luftreibung und andere Kräfte. Die *Erdbeschleunigung* hat für alle Körper den mittleren Wert von $g = 9{,}81\,\text{m/s}^2$. Startet ein Körper aus der Ruhe, so gelten (11.4) und (11.5) mit $v_0 = 0$ und $a = -g$ Die Höhe h wird von der Erdoberfläche aus in positiver Richtung nach oben gemessen und entspricht der Anfangskoordinate s_0 Die Fallzeit bestimmt sich aus der Bedingung $s(t_F) = 0$.

Fallzeit:

$$0 = h - \frac{1}{2}gt_F^2 \tag{11.6}$$

$$t_F = \sqrt{\frac{2h}{g}} \tag{11.7}$$

Die Fallgeschwindigkeit zu einem beliebigen Zeitpunkt t berechnet sich nach

$$v_F(t) = g \cdot t \tag{11.8}$$

daraus folgt für die Aufprallgeschwindigkeit oder Endgeschwindigkeit v_e, die ja nach der Fallzeit t_F erreicht ist:

$$v_e = g \cdot t_F \tag{11.9}$$

$$v_e = g \cdot \sqrt{\frac{2h}{g}} = \sqrt{2hg} \tag{11.10}$$

Beispiel

Ein Körper fällt von einem Turm der Höhe $h = 20\,\text{m}$ im freien Fall.

a) Wann kommt er unten an?
b) Wie groß ist dann seine Geschwindigkeit?

Lösung:

a)

$$t_F = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2 \cdot 20\,\text{m}}{9{,}81\,\text{m/s}^2}}$$
$$t_F = \sqrt{\frac{40\,\text{m}}{9{,}81\,\text{m/s}^2}} = 2{,}02\,\text{s}$$

b)

$$v_e = \sqrt{2hg} = \sqrt{2 \cdot 20\,\text{m} \cdot 9{,}81\,\frac{\text{m}}{\text{s}^2}}$$
$$v_e = 19{,}8\,\frac{\text{m}}{\text{s}}$$ ◀

11.1.1.4 Senkrechter Wurf

Beim Wurf nach oben ist v_0 positiv, beim Wurf nach unten negativ einzusetzen. Hierbei kann ebenfalls eine Anfangshöhe

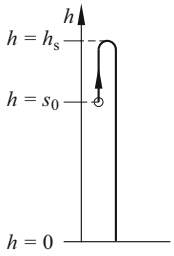


Abb. 11.3 Senkrechter Wurf

$s_0 = h$ angenommen werden. Die maximale Steighöhe und die Steigzeit beim Wurf nach oben folgen aus der Bedingung, dass am Umkehrpunkt die momentane Geschwindigkeit $v = 0$ und am Ende des Fluges die Höhe $h = 0$ sein muss. Aus diesen Bedingungen folgen aus (11.4) und (11.5) die nachfolgenden Formeln. Bei negativen Werten der Geschwindigkeit ist die Bewegung abwärts gerichtet (Abb. 11.3).

momentane Geschwindigkeit	$v(t) = v_0 - gt$	(11.11)
momentane Höhe	$h(t) = s_0 + v_0 t - \frac{1}{2}gt^2$	(11.12)
Flugzeit $(h(t) = 0)$	$t_F = \frac{v_0 + \sqrt{v_0^2 + 2s_0 g}}{g}$	(11.13)
Endgeschwindigkeit	$v_e = -\sqrt{v_0^2 + 2s_0 g}$	(11.14)
Steigzeit $(v(t) = 0)$	$t_s = \frac{v_0}{g}$	(11.15)
maximale Höhe $(t = t_s)$	$h_s = s_0 + \frac{v_0^2}{2g}$	(11.16)

Die Formeln (11.11)–(11.15) gelten auch für den senkrechten Fall nach unten, während die Formel (11.16) hier dann keinen Sinn ergibt.

Beispiel

Von einem 10 m hohen Turm wird ein Stein mit einer Anfangsgeschwindigkeit von 20 m/s senkrecht nach oben geworfen. Wie groß sind die maximale Höhe und die gesamte Flugzeit?

Lösung:

$$h_s = 10\,\text{m} + \frac{(20\,\text{m/s})^2}{2 \cdot 9{,}81\,\text{m s}^{-2}} = 30{,}39\,\text{m}$$

$$t_f = \frac{20\,\text{m/s} + \sqrt{(20\,\text{m/s})^2 + 2 \cdot 10\,\text{m} \cdot 9{,}81\,\text{m s}^{-2}}}{9{,}81\,\text{m s}^{-2}}$$

$$= 4{,}528\,\text{s}$$ ◀

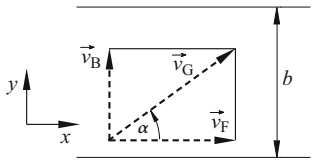


Abb. 11.4 Geschwindigkeiten im Fluss

11.1.2 Zusammengesetzte Bewegungen

Im Gegensatz zu eindimensionalen Bewegungen sind hier bei allen vektoriellen Größen zwei Angaben notwendig. Als Richtungen sollen x- und y-Richtung festgelegt sein, die entsprechenden Größen werden durch die Indizes x und y unterschieden.

Führt ein Körper gleichzeitig Bewegungen in x- und y-Richtung aus, sogenannte *zusammengesetzte Bewegungen*, so überlagern sich diese Bewegungen unabhängig voneinander und der Endzustand ist derselbe, als wenn die einzelnen Bewegungen nacheinander ausgeführt worden wären.

Als Beispiel soll eine Bewegung in einem strömenden Fluss der Breite b betrachtet werden. Der Fluss fließt in x-Richtung mit einer Geschwindigkeit v_F. Der Geschwindigkeitsvektor kann dann in eine Komponente parallel zum Ufer und eine senkrecht zum Ufer aufgeteilt werden, dies sollen die x- und y-Richtung sein. Eine Schreibweise hierfür ist die *Komponentenschreibweise* vektorieller Größen:

$$\vec{v} = (v_x, v_y) \vec{v} = (v_x, v_y) \tag{11.17}$$

und somit

$$\vec{v}_F = (v_{Fx}, 0) \vec{v}_F = (v_{Fx}, 0).$$

In dem Wasser bewegt sich ein Boot mit einer Eigengeschwindigkeit $\vec{v}_B$ *relativ* zum Wasser. Ist das Wasser in Ruhe, so ist dies auch die Geschwindigkeit *relativ* zum Grund $\vec{v}_G$.

$$\vec{v}_F = (v_{Fx}, 0)$$

In einem strömenden Fluss ist nun $\vec{v}_B$ unterschiedlich von $\vec{v}_G$.

Die Geschwindigkeit des Bootes kann nun entweder *relativ* zum Grund oder *relativ* zum Wasser angegeben werden. Der Zusammenhang ist (Abb. 11.4):

$$\vec{v}_G = \vec{v}_B + \vec{v}_F = (v_{Fx}, v_{By}) \tag{11.18}$$

Es gilt also, dass sich die einzelnen Komponenten unabhängig addieren.

Hieraus lassen sich die folgenden Größen berechnen:

Zeit zum Überqueren	$t = \frac{b}{v_{By}}$	(11.19)
Weg in x-Richtung	$s_{Gx} = v_{Fx} \cdot t$	(11.20)
Weg in y-Richtung	$s_{Gy} = v_{By} \cdot t$	(11.21)
Gesamter Weg	$s_G = \sqrt{s_{Gx}^2 + s_{Gy}^2}$	(11.22)
Geschwindigkeit	$v_G = \sqrt{v_{Fx}^2 + v_{By}^2}$	(11.23)
Richtung	$\alpha = \arctan\left(\frac{v_{By}}{v_{Fx}}\right)$	(11.24)

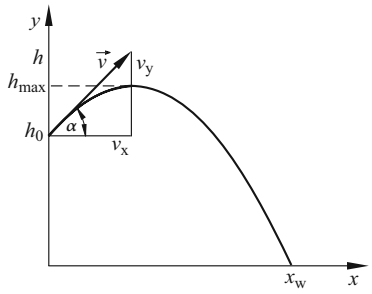


Abb. 11.5 Schiefer Wurf

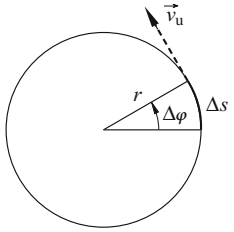


Abb. 11.6 Kreisbewegung

Beispiel

In einem Fluss der Breite 500 m fließt Wasser mit einer Strömungsgeschwindigkeit von 2 m/s. Senkrecht zum Ufer startet ein Boot mit einer Eigengeschwindigkeit von 10 m/s. Wie lange braucht das Boot für die Überquerung, und wie groß ist die seitliche Abdrift d?

Lösung:

$$\textit{Zeit} \qquad t = \frac{500\,\text{m}}{10\,\text{m/s}} = 50\,\text{s}$$

$$\textit{Abdrift} \qquad d = 2\,\frac{\text{m}}{\text{s}} \cdot 50\,\text{s} = 100\,\text{m} \quad \blacktriangleleft$$

11.1.2.1 Schiefer Wurf

Ein weiteres Beispiel für eine *zusammengesetzte Bewegung* ist die Bewegung eines Steines, der unter einem bestimmten Winkel α mit einer Anfangsgeschwindigkeit v_0 geworfen wird (*schiefer Wurf*). In diesem Fall handelt es sich um die Überlagerung einer gleichmäßigen Bewegung in x-Richtung mit einer gleichförmig beschleunigten Bewegung (freier Fall) in y-Richtung. Die Flugbahn unter dem Einfluss einer konstanten Kraft (Erdanziehung) ist in diesem Fall eine Parabel (*Wurfparabel*, Abb. 11.5).

Es soll auch zugelassen werden, dass der Stein in einer Höhe h_0 abgeworfen wird. Aus den Bedingungen, dass am Ende der Bewegung der Wert für $y = 0$ sein muss und am höchsten Punkt der Wert für $v_y = 0$ sein muss, folgen die Gleichungen für den *Schiefen Wurf* aus einer Anfangshöhe h_0 und mit einer Anfangsgeschwindigkeit v_0

$$v_{0x} = v_0 \cdot \cos\alpha \tag{11.25}$$

$$v_{0y} = v_0 \cdot \sin\alpha \tag{11.26}$$

$$v_x(t) = v_0 \cdot \cos\alpha = \text{const} \tag{11.27}$$

$$v_y(t) = v_0 \cdot \sin\alpha - gt \tag{11.28}$$

$$y(t) = h_0 + v_0 \cdot \sin\alpha - \frac{1}{2}gt^2 \tag{11.29}$$

$$x(t) = v_0 \cdot t \cdot \cos\alpha \tag{11.30}$$

$$\textit{Flugzeit} \qquad t_F = \frac{v_0 \cdot \sin\alpha + \sqrt{(v_0 \cdot \sin\alpha)^2 + 2gh_0}}{g} \tag{11.31}$$

$$\textit{Flughöhe} \quad h_{\max} = h_0 + \frac{v_0^2 \cdot \sin^2\alpha}{2g} \tag{11.32}$$

$$\textit{Flugweite} \quad x_w = v_0 \cdot t_F \cdot \cos\alpha \tag{11.33}$$

Wenn in (11.29) bis (11.32) der Wert für h_0 auf 0 gesetzt wird, so ergeben sich die Gleichungen für den Fall eines *Schiefen Wurfes* mit Anfangshöhe 0.

$$\textit{Flugzeit} \qquad t_F = \frac{2\,v_0 \sin\alpha}{g} \tag{11.34}$$

$$\textit{Flughöhe} \qquad h_{\max} = \frac{v_0^2 \sin^2\alpha}{2g} \tag{11.35}$$

$$\textit{Flugweite} \qquad x_w = v_0 \cdot t_F \cdot \cos\alpha = \frac{2 \cdot v_0^2 \cdot \sin\alpha \cdot \cos\alpha}{g}$$

$$x_w = \frac{v_0^2 \cdot \sin 2\alpha}{g} \tag{11.36}$$

Beispiel

Ein Stein wird unter einem Winkel von 30° mit einer Anfangsgeschwindigkeit von $20\,\text{m}\,\text{s}^{-1}$ geworfen. Wie weit fliegt er, und wann trifft er auf den Boden?

Lösung:

$$\textit{Flugweite} \qquad x_w = \frac{(20\,\text{m}\,\text{s}^{-1})^2 \cdot \sin 60°}{9{,}81\,\text{m}\,\text{s}^{-2}} = 35{,}31\,\text{m}$$

$$\textit{Flugzeit} \qquad t_F = \frac{2 \cdot 20\,\text{m}\,\text{s}^{-1} \cdot \sin 30°}{9{,}81\,\text{m}\,\text{s}^{-2}} = 2{,}04\,\text{s} \quad \blacktriangleleft$$

11.1.3 Kreisbewegung

Bei einer Kreisbewegung bewegt sich eine punktförmige Masse auf einer Kreisbahn mit dem Radius r. Wenn in gleichen Zeiten Δt gleiche Strecken Δs auf dem Umfang zurückgelegt werden, so überstreicht auch die Verbindungslinie zum Zentrum des Kreises in gleichen Zeiten Δt gleiche Winkel $\Delta\varphi$ (Abb. 11.6).

11.1.3.1 Bahngeschwindigkeit

Unter *Bahngeschwindigkeit* oder auch *Umfangsgeschwindigkeit* versteht man die Geschwindigkeit, mit der sich eine Masse m auf dem Umfang eines Kreises mit dem Radius r bewegt. Wenn sich die Masse in der Zeit Δt um die Strecke Δs weiterbewegt hat, wird von der Verbindungslinie zwischen der Masse und dem Zentrum des Kreises der Winkel $\Delta\varphi$ überstrichen. Zwischen den Größen r, Δs und $\Delta\varphi$ gilt die Gleichung:

$$\Delta\varphi = \frac{\Delta s}{r} \qquad \begin{array}{c|c|c} \varphi & s & r \\ \hline \text{rad} & \text{m} & \text{m} \end{array} \tag{11.37}$$

$$\Delta s = r \cdot \Delta\varphi \tag{11.38}$$

Umfangsgeschwindigkeit

$$v_u = \frac{\Delta s}{\Delta t} = r\frac{\Delta\varphi}{\Delta t} \tag{11.39}$$

oder

Bahngeschwindigkeit

$$v_u = r\omega \tag{11.40}$$

mit der Abkürzung $\omega = \frac{\Delta\varphi}{\Delta t}$.

Die Größe ω wird *Winkelgeschwindigkeit* genannt.

11.1.3.2 Winkelgeschwindigkeit

Die *Winkelgeschwindigkeit* ω ist durch

$$\omega = \frac{\Delta\varphi}{\Delta t} \qquad \begin{array}{c|c|c} \omega & \Delta\phi & \Delta t \\ \hline 1/\text{s} & \text{rad} & \text{s} \end{array} \tag{11.41}$$

definiert, wobei $\Delta\varphi$ der in der Zeiteinheit Δt überstrichene Winkel ist.

Bei einer Kreisbewegung mit konstanter *Winkelgeschwindigkeit* ω (*gleichförmige Kreisbewegung*) ist die Umfangsgeschwindigkeit vom Betrag her konstant, allerdings ändert sie laufend die Richtung.

11.1.3.3 Kreisfrequenz

Oft werden die *Drehzahl* n oder auch die *Frequenz* f einer kreisförmigen Bewegung angegeben. Im Gegensatz zur Frequenz f wird die Größe w auch *Kreisfrequenz* genannt.

$$\omega = 2\pi f \qquad \omega = \frac{2\pi n}{60} \qquad \begin{array}{c|c|c} \omega & f & n \\ \hline 1/\text{s} & \text{Hz} & 1/\text{min} \end{array} \tag{11.42}$$

11.1.3.4 Winkelbeschleunigung

Werden in gleichen Zeiten ungleiche Wegstrecken auf dem Umfang zurückgelegt, ändert sich also die Umfangsgeschwindigkeit, muss der Körper eine *Tangentialbeschleunigung* erfahren. Jetzt werden in gleichen Zeiten ungleiche Winkel überstrichen, daher ändert sich die Winkelgeschwindigkeit ebenfalls. In diesem Fall liegt eine *Winkelbeschleunigung* vor. Analog zur linearen Beschleunigung wird die Winkelbeschleunigung definiert:

$$\alpha = \frac{\Delta\omega}{\Delta t} \qquad \begin{array}{c|c|c} \alpha & \Delta\omega & \Delta t \\ \hline 1/\text{s}^2 & 1/\text{s} & \text{s} \end{array} \tag{11.43}$$

In diesem Fall gelten analoge Gleichungen zu (11.4) und (11.5).

$$\omega = \omega_0 + \alpha t \tag{11.44}$$

$$\varphi = \varphi_0 + \omega_0 t + \frac{1}{2}\alpha t^2 \tag{11.45}$$

Hierbei sind w_0 und j_0 die Anfangswerte zur Zeit $t = 0$ der Winkelgeschwindigkeit und des Winkels.

Beispiel

Ein Elektromotor läuft mit einer Drehzahl $n = 600\,\text{min}^{-1}$. Nach dem Abschalten wird er mit konstanter Winkelbeschleunigung a abgebremst, bis er nach 50 Umdrehungen zum Stillstand kommt.

a) Wie groß ist die Winkelbeschleunigung?
b) Wie lange ist die Bremszeit t_B?

Lösung:

a) Anfangswerte:

$$\varphi_0 = 0, \quad \omega_0 = 2\pi\frac{n}{60}\,\text{s}^{-1} = 62{,}83\,\text{s}^{-1}$$

50 Umdrehungen ergeben:

$$\varphi = 50 \cdot 2\pi = 314{,}16\,\text{rad}$$

aus (11.44) folgt:

$$\alpha = -\frac{\omega_0}{t_B}$$

Winkelbeschleunigung:

$$\alpha = -\frac{62{,}83\,\text{s}}{10\,\text{s}} = -6{,}28\,\text{s}^{-2}$$

b) In (11.45) eingesetzt:

$$\varphi = \omega_0 t_B - \frac{\omega_0 t_B^2}{2t_B} = \frac{1}{2}\omega_0 t_B$$

Bremszeit:

$$t_B = \frac{2\varphi}{\omega_0} = \frac{2 \cdot 314{,}16\,\text{rad}}{62{,}83\,\text{s}^{-1}} = 10\,\text{s}$$ ◀

11.2 Dynamik

In der *Kinematik* wird die Bewegung von Massen durch geeignete Formeln beschrieben, ohne dass nach der Ursache für eine Bewegung oder eine Änderung eines Bewegungszustandes gefragt wird. In der *Dynamik* werden diese Ursachen untersucht.

11.2.1 Newton'sche Axiome

Isaac Newton (1643 bis 1727) hat drei grundlegende Axiome formuliert, die das Verhalten von Körpern unter dem Einfluss äußerer Kräfte und das Zusammenspiel von Kräften untereinander beschreiben. Diese *Newton'schen Axiome* sind die Grundlagen der klassischen Mechanik und werden in der Tab. 11.1 aufgeführt:

Das Aktionsgesetz lässt eine zeitliche Änderung sowohl der Masse als auch der Geschwindigkeit zu. In der allgemeinen Form gilt:

$$\vec{F} = \frac{\Delta \vec{p}}{\Delta t} = \frac{\Delta(m \cdot \vec{v})}{\Delta t} = m \cdot \frac{\Delta \vec{v}}{\Delta t} + \vec{v} \cdot \frac{\Delta m}{\Delta t} \tag{11.46}$$

Ist die Masse konstant, so ist $\Delta m = 0$, und es gilt mit $\frac{\Delta \vec{v}}{\Delta t} = \vec{a}$:

$$\vec{F} = m \cdot \vec{a} \tag{11.47}$$

Dieses Gesetz wird auch als *Newton'sches Grundgesetz* bezeichnet, gilt in dieser Form aber nur für konstante Massen.

11.2.2 Kraft

Nach dem *Newton'schen Grundgesetz* ist die Kraft $\vec{F}$ bei konstanter Masse proportional zur Beschleunigung $\vec{a}$. Die Kraft ist eine vektorielle Größe. Kraft und Beschleunigung haben dieselbe Richtung.

$$\vec{F} = m \cdot \vec{a} \qquad \begin{array}{c|c|c} F & m & a \\ \hline \mathrm{kg\,m/s^2} & \mathrm{kg} & \mathrm{m/s^2} \end{array} \tag{11.48}$$

Die Einheit der Kraft im SI-System ist $1\ \frac{\mathrm{kg\,m}}{\mathrm{s^2}}$, hierfür wird die Abkürzung 1 N (1 Newton) verwendet.

$$1\,\mathrm{N} = 1\,\frac{\mathrm{kg\,m}}{\mathrm{s^2}} \tag{11.49}$$

An der Erdoberfläche wirkt auf alle Körper die *Gewichtskraft* oder *Schwerkraft*

$$F_G = m \cdot g \tag{11.50}$$

mit der Erdbeschleunigung g

$$g = 9{,}81\,\frac{\mathrm{m}}{\mathrm{s^2}} \tag{11.51}$$

Eine nicht mehr zulässige Einheit der Kraft ist 1 kp (1 Kilopond). Dies ist die Gewichtskraft auf die Masse von 1 kg. Somit gilt:

$$1\,\mathrm{kp} = 9{,}81\,\mathrm{N} \tag{11.52}$$

Hängt eine Masse m an einer Feder, so wird die Feder um eine Strecke x gedehnt und zwar solange, bis die Rückstellkraft der Feder und die Gewichtskraft auf die Masse m entgegengesetzt gleich groß sind. Die Rückstellkraft der Feder wird durch die Federeigenschaften beeinflusst und in der Federkonstanten c festgelegt. Da die Rückstellkraft entgegengesetzt zur Auslenkung gerichtet ist, gilt

$$F_{\mathrm{Rück}} = -cx \tag{11.53}$$

$$c = -\frac{F_{\mathrm{Rück}}}{x} \qquad \begin{array}{c|c|c|c} c & F & x \\ \hline \mathrm{N/m} & \mathrm{N} & \mathrm{m} \end{array} \tag{11.54}$$

11.2.2.1 Zerlegung und Zusammensetzung von Kräften

Kräfte sind Vektoren und müssen somit vektoriell addiert werden. Die grafische Lösung für die Addition zweier Kräfte und die sich daraus ergebende *resultierende Kraft* erfolgt so, dass der Anfangspunkt des zweiten Vektors in den Endpunkt des ersten Vektors verschoben wird. Es entsteht ein Parallelogramm, das *Kräfteparallelogramm*.

Mit Hilfe der trigonometrischen Gleichungen lassen sich die folgenden Beziehungen für die Addition von Kräften ableiten (Abb. 11.7):

$$x\text{-Komponente} \qquad F_x = F_1 \cos\alpha + F_2 \cos\beta \tag{11.55}$$

$$y\text{-Komponente} \qquad F_y = F_1 \sin\alpha + F_2 \sin\beta \tag{11.56}$$

daraus folgt für die *Resultierende Kraft*:

$$F = \sqrt{F_x^2 + F_y^2} \tag{11.57}$$

$$F = \sqrt{F_1^2 + F_2^2 + 2F_1F_2\cos(\alpha - \beta)} \tag{11.58}$$

Tab. 11.1 Newton'sche Axiome

Newton'sche Axiome	Formulierung	Gleichung
1. Axiom: *Trägheitsgesetz*	Jeder Körper beharrt im Zustand der Ruhe oder der gleichförmig geradlinigen Bewegung, solange er nicht durch äußere Kräfte gezwungen wird, diesen Zustand zu ändern	
2. Axiom: *Aktionsgesetz*	Die zeitliche Änderung der Bewegungsgröße (Impuls) $\vec{p} = m\vec{v}$ ist gleich der resultierenden Kraft $\vec{F}$	$\vec{F} = \frac{\Delta \vec{p}}{\Delta t}$
3. Axiom: *Wechselwirkungsgesetz actio = reactio*	Wirkt ein Körper 1 auf einen Körper 2 mit der Kraft F_{12}, so wirkt der Körper 2 auf den Körper 1 mit einer gleich großen, entgegengesetzten Kraft F_{21}	$\vec{F}_{12} = -\vec{F}_{21}$

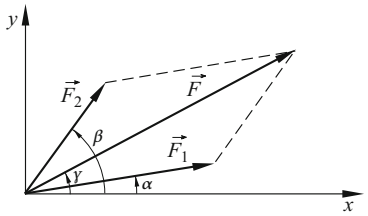


Abb. 11.7 Kräfteparallelogramm

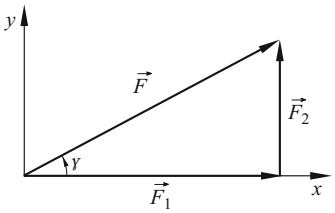


Abb. 11.8 Zerlegung von Kräften

Nach Umkehrung dieser Gleichungen lassen sich die Teilkräfte aus der Resultierenden und den Winkeln berechnen:

$$F_1 = F\frac{\sin(\beta-\gamma)}{\sin(\beta-\alpha)} \tag{11.59}$$

$$F_2 = F\frac{\sin(\gamma-\alpha)}{\sin(\beta-\alpha)} \tag{11.60}$$

Oft ist es notwendig, eine Kraft $\vec{F}$ in zwei senkrecht zueinanderstehende Komponenten zu zerlegen. In (11.59) und (11.60) ist dann $a = 0°$ und $b = 90°$ zu setzen (Abb. 11.8):

$$F_1 = F\cos\gamma \tag{11.61}$$

$$F_2 = F\sin\gamma \tag{11.62}$$

Der Winkel der Resultierenden mit der Kraft F_1 errechnet sich zu:

$$\gamma = \arctan(F_2/F_1) \tag{11.63}$$

11.2.2.2 Schiefe Ebene

Befindet sich ein Körper auf einer *Schiefen Ebene*, so wirkt auf ihn die Schwerkraft in der in Abb. 11.9 angezeigten Richtung. Diese Schwerkraft kann in eine Komponente senkrecht zur Unterlage, der *Normalkraft* F_N, und eine Komponente parallel zur Unterlage, der *Hangabtriebskraft*, zerlegt werden.

Die Beträge dieser Kräfte sind:

$$F_H = mg\sin\alpha \tag{11.64}$$

$$F_N = mg\cos\alpha \tag{11.65}$$

Die Hangabtriebskraft beschleunigt den Körper, während die Normalkraft den Druck auf die Unterlage bewirkt. Durch die Normalkraft wird wegen der immer vorhandenen Reibung eine der Hangabtriebskraft entgegengesetzte Reibungskraft F_R verursacht.

$$F_R = \mu F_N = \mu mg \tag{11.66}$$

Die Größe μ wird als Reibungszahl bezeichnet und hängt von der Beschaffenheit der Oberfläche der beiden an der Reibung beteiligten Körper ab und von der Art der Bewegung. Sie ist eine dimensionslose Zahl. Bei der Bewegung eines Fahrzeuges ist die *Rollreibung* zu berücksichtigen. Einige typische Werte für die Reibungszahl sind in Tab. 11.2 zusammengestellt. Hierbei ist zwischen *Haftreibung* und *Gleitreibung* zu unterscheiden.

Tab. 11.2 Werte für die Reibungszahl μ

Stoffpaar	Haftreibung	Gleitreibung
Stahl/Stahl	0,15	0,12
Stahl/Holz	0,5–0,6	0,2–0,5
Stahl/Eis	0,027	0,014
Holz/Holz	0,65	0,2–0,4
Gummi/Asphalt	0,9	0,85
Gummi/Beton	0,65	0,5
Gummi/Eis	0,2	0,15

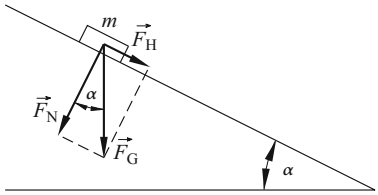


Abb. 11.9 Schiefe Ebene

Beispiel

Ein Fahrzeug der Masse $m = 1000\,\text{kg}$ befindet sich auf einer Schiefen Ebene mit dem Neigungswinkel $a = 10°$. Die Reibungszahl sei $\mu = 0{,}1$. Wie groß ist die Beschleunigung des Wagens?

Lösung:

Normalkraft:

$$F_N = mg\cos\alpha = 1000\,\text{kg}\cdot 9{,}81\,\text{m s}^2\cdot 0{,}984 = 9660\,\text{N}$$

Reibungskraft:

$$F_R = \mu F_N = 0{,}1\cdot 9660\,\text{N} = 966\,\text{N}$$

Hangabtriebskraft:

$$F_H = mg\sin\alpha = 1000\,\text{kg}\cdot 9{,}81\,\text{m s}^2\cdot\sin 10° = 1703\,\text{N}$$

Resultierende Kraft:

$$F = F_H - F_R = 1703\,\text{N} - 966\,\text{N} = 737\,\text{N}$$

Beschleunigung:

$$a = F/m = 737\,\text{N}/1000\,\text{kg} = 0{,}737\,\text{m/s}^2$$ ◀

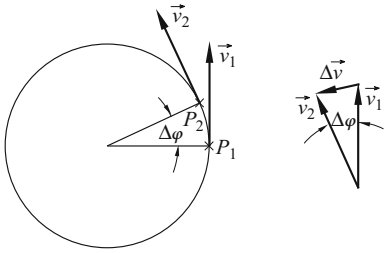


Abb. 11.10 Zentripetal-Beschleunigung

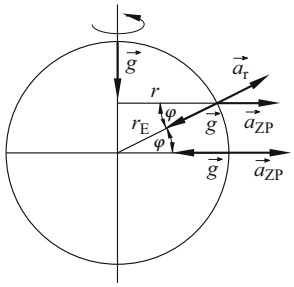


Abb. 11.11 Fallbeschleunigung und geographische Breite

11.2.2.3 Kräfte bei Kreisbewegungen

Soll sich eine Masse m auf einer Kreisbahn mit konstanter Winkelgeschwindigkeit bewegen, so bleibt zwar der Betrag der Umfangsgeschwindigkeit konstant, nicht aber die Richtung (siehe Abb. 11.10). Die Masse m hat sich in der Zeit Δt vom Punkt P_1 nach P_2 bewegt. Dabei hat sich der Vektor der Geschwindigkeit von $\vec{v}_1$ nach $\vec{v}_2$ verändert, der Betrag ist aber geblieben. Da sich die Richtung geändert hat, gilt

$$\Delta\vec{v} = \vec{v}_2 - \vec{v}_1 \tag{11.67}$$

Die Geschwindigkeitsänderung lässt sich auch durch die Winkeländerung ausdrücken:

$$\Delta v = v \cdot \Delta\varphi \tag{11.68}$$

Die Richtung von $\Delta\vec{v}$ ist zum Zentrum der Kreisbahn gerichtet, die Geschwindigkeit hat sich daher in Richtung auf das Zentrum geändert, somit muss auch eine Beschleunigung in Richtung auf das Zentrum der Kreisbahn erfolgen

$$a_{Zp} = \frac{\Delta v}{\Delta t} = \frac{v \cdot \Delta\varphi}{\Delta t} = v\omega \tag{11.69}$$

und mit (11.40)

$$a_{Zp} = \omega^2 r \tag{11.70}$$

$$a_{Zp} = \omega^2 r \quad \begin{array}{c|c|c} \alpha_{Zp} & \omega & r \\ \hline \mathrm{m/s^2} & 1/\mathrm{s} & \mathrm{m} \end{array} \tag{11.71}$$

Diese Beschleunigung ist die *Zentripetalbeschleunigung*.

Wegen des 2. Newton'schen Axioms wirkt daher eine Kraft, die *Zentripetalkraft*

$$F_{Zp} = m\omega^2 r \tag{11.72}$$

Sie muss aufgewendet werden, um eine Masse m auf einer Kreisbahn zu halten. Die entgegengesetzt gerichtete gleich große Kraft ist die *Zentrifugalkraft*. Die dazugehörende Beschleunigung heißt *Zentrifugalbeschleunigung*. Diese Kraft tritt bei allen Rotationsbewegungen auf, dabei ist mit dem Radius r in (11.72) der Abstand von der Drehachse gemeint. Als Beispiel soll die Abhängigkeit der Fallbeschleunigung auf der Erdoberfläche von der geographischen Breite untersucht werden. Dabei wird angenommen, dass die Erde eine homogene Kugel sei mit dem Radius $r_E = 6378\,\mathrm{km}$. Die Fallbeschleunigung ist die Resultierende aus Erdbeschleunigung und Zentrifugalbeschleunigung. Am Pol ist der Abstand von der Drehachse $= 0$, somit tritt hier keine Zentrifugalbeschleunigung auf, und für die Fallbeschleunigung gilt $a_F = g$. Am Äquator ist die Zentrifugalbeschleunigung maximal, nämlich $a_{Zp} = \omega^2 r_E$. Da sich die Erde in 24 Stunden einmal um sich selbst dreht, gilt am Äquator:

$$a_{Zp} = \left(\frac{2\pi}{24 \cdot 60 \cdot 60\,\mathrm{s}}\right)^2 \cdot 6378 \cdot 10^3\,\mathrm{m}$$

$$a_{Zp} = 0{,}034\,\frac{\mathrm{m}}{\mathrm{s^2}}$$

Fallbeschleunigung

$$\vec{a}_F = \vec{g} + \vec{a}_{Zp} \tag{11.73}$$

und, da Zentrifugalbeschleunigung und Erdbeschleunigung entgegengesetzt wirken

$$a_F = (9{,}81 - 0{,}034)\,\frac{\mathrm{m}}{\mathrm{s^2}} = 9{,}776\,\frac{\mathrm{m}}{\mathrm{s^2}}$$

Für andere geographische Breiten muss entsprechend Abb. 11.11 der Abstand von der Drehachse r und die Komponente der Zentrifugalbeschleunigung a_r in Richtung auf den Erdmittelpunkt bestimmt werden, denn nur diese Komponente wirkt der Erdbeschleunigung, die ja zum Erdmittelpunkt gerichtet ist, entgegen. Es gelten die folgenden Gleichungen:

$$r = r_E \cdot \cos\varphi \tag{11.74}$$

$$a_{Zp}^{\varphi} = \omega^2 \cdot r = \omega^2 \cdot r_E \cdot \cos\varphi \tag{11.75}$$

$$a_r = a_{Zp}^{\varphi} \cdot \cos\varphi \tag{11.76}$$

$$a_r = \omega^2 \cdot r_E \cdot \cos^2\varphi \tag{11.77}$$

und damit

$$a_F = g - \omega^2 \cdot r_E \cdot \cos^2\varphi \tag{11.78}$$

11.2.3 Impuls

Im 2. Newton'schen Axiom wird die zeitliche Änderung der Bewegungsgröße p gleich der resultierenden Kraft gesetzt. Die Bewegungsgröße ist der *Impuls*. Bei konstanter Kraft F gilt:

$$\vec{p} = m\vec{v} \quad \begin{array}{c|c|c} p & m & v \\ \hline \mathrm{N\,s} & \mathrm{kg} & \mathrm{m/s} \end{array} \tag{11.79}$$

$$\vec{F} = \frac{\Delta\vec{p}}{\Delta t} \Rightarrow \Delta\vec{p} = \vec{F} \cdot \Delta t \tag{11.80}$$

Teil II

Die Kraft ist also gleich der zeitlichen Änderung des Impulses. Die Größe

$$\int_{t_1}^{t_2} F\,dt = p_2 - p_1 = \Delta p \tag{11.81}$$

wird als Kraftstoß bezeichnet. Ist die Kraft zeitlich konstant, so vereinfacht sich (11.81) zu

$$\Delta\vec{p} = \vec{F}\cdot\Delta t = \vec{F}(t_2 - t_1) \tag{11.82}$$

Die Impulsänderung $\Delta\vec{p}$, die ja auch ein Vektor ist, hat die Richtung der angreifenden resultierenden Kraft.

Beispiel

Eine konstante Kraft von 2 kN wirkt 10 s lang auf ein ruhendes Fahrzeug der Masse $m = 800\,\text{kg}$. Wie groß sind

a) der Kraftstoß,
b) der Impuls,
c) die Geschwindigkeit nach 10 s?

Lösung:

a) $F\cdot\Delta t = 2000\,\text{N}\cdot 10\,\text{s} = 20.000\,\text{N s}$

b) $\Delta p = p_2 - p_1 = F\cdot\Delta t \Rightarrow p = 20.000\,\text{N s}$

c) $v = \frac{\Delta p}{m} = \frac{20.000\,\text{N s}}{800\,\text{kg}} = 25\,\frac{\text{m}}{\text{s}} = 90\,\frac{\text{km}}{\text{h}}$ ◀

11.2.3.1 Impulserhaltungssatz

Wirkt nun auf ein System keine äußere resultierende Kraft, so ist entsprechend (11.82) die Änderung des Impulses $\Delta\vec{p} = 0$ Daraus folgt, dass in diesem Fall der Impuls konstant sein muss. Dieses gilt für den *Gesamtimpuls* des betrachteten Systems. Besteht das System aus mehreren Massen mit *Einzelimpulsen* $\vec{p}_i$, so gilt der *Impulserhaltungssatz* für den *Gesamtimpuls*

$$\vec{p}_{\text{ges}} = \vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \ldots + \vec{p}_n = \text{const} \tag{11.83}$$

Dabei können sich die Einzelimpulse durchaus ändern, wenn nur der Gesamtimpuls konstant bleibt. Eine weitergehende Betrachtung wird im Abschn. 11.2.5 „Stoßprozesse“ durchgeführt.

Beispiel

Aus einem Boot der Masse 200 kg, welches sich in ruhendem Wasser (äußere Kräfte = 0) mit einer Geschwindigkeit von 2 m/s bewegt, springt in Fahrtrichtung ein 80 kg schwerer Mann. Dabei stößt er sich 0,2 s lang mit einer Kraft von 300 N ab.

Wie groß sind die Geschwindigkeiten des Bootes und des Mannes direkt nach dem Sprung?

Lösung:

Gesamtimpuls zu Beginn:

$$p_g = (200\,\text{kg} + 80\,\text{kg})\cdot 2\,\text{m/s} = 560\,\text{N s}$$

Impulsänderung durch den Sprung:

$$\Delta p = F\Delta t = 300\,\text{N}\cdot 0{,}2\,\text{s} = 60\,\text{N s}$$

Impuls des Bootes nach dem Sprung:

$$p'_B = 200\,\text{kg}\cdot 2\,\text{m/s} - 60\,\text{N s} = 340\,\text{N s}$$

Geschwindigkeit des Bootes:

$$v'_B = 340\,\text{N s}/200\,\text{kg} = 1{,}7\,\text{m/s}$$

Impuls des Mannes nach dem Sprung:

$$p'_M = 80\,\text{kg}\cdot 2\,\text{m/s} + 60\,\text{N s} = 220\,\text{N s}$$

Geschwindigkeit des Mannes:

$$v'_M = 220\,\text{N s}/80\,\text{kg} = 2{,}75\,\text{m/s}$$

Gesamtimpuls am Ende:

$$p'_g = (340 + 220)\,\text{N s} = 560\,\text{N s}$$

Die positiven Vorzeichen bei beiden Geschwindigkeiten zeigen an, dass sich sowohl der Mann wie auch das Boot weiter in der ursprünglichen Fahrtrichtung des Bootes bewegen, der Gesamtimpuls hat sich nicht verändert. Da der Impuls ein Vektor ist, muss die Impulserhaltung auch für jede Komponente gelten. ◀

Beispiel

Der Mann aus dem voranstehenden Beispiel vorher springt nicht in Fahrtrichtung, sondern senkrecht zur Fahrtrichtung in positiver y-Richtung vom Boot. Wie groß ist die Geschwindigkeit des Bootes und des Mannes direkt nach dem Sprung?

Lösung:

Die Geschwindigkeit des Bootes und des Mannes in Fahrtrichtung bleiben unverändert.

$$\Delta p_y = F_y\cdot\Delta t = 300\,\text{N}\cdot 0{,}2\,\text{s} = 60\,\text{N s}$$

$$v_{B_y} = \frac{-60\,\text{Ns}}{200\,\text{kg}} = -0{,}3\,\frac{\text{m}}{\text{s}}$$

$$v_{M_y} = \frac{60\,\text{Ns}}{80\,\text{kg}} = -0{,}75\,\frac{\text{m}}{\text{s}}$$

Das Boot bewegt sich also in entgegengesetzter Richtung zum Mann. ◀

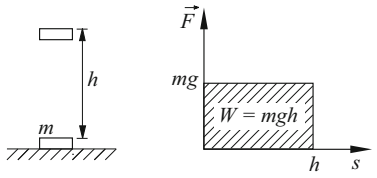


Abb. 11.12 Hubarbeit

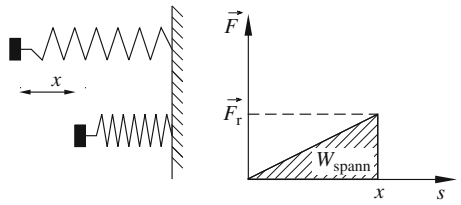


Abb. 11.13 Spannarbeit

11.2.4 Arbeit, Leistung, Wirkungsgrad und Energie

11.2.4.1 Arbeit

Wirkt eine Kraft auf eine Masse m und verschiebt sie dabei die Masse m um den Weg Δs, so hat die Kraft den Zustand des Körpers verändert, es wurde *Arbeit* verrichtet. Schließt die Kraft einen Winkel α mit der Richtung von Δs ein, so gilt für die Teilarbeit ΔW auf dieser Wegstrecke:

$$\Delta W = F \cdot \Delta s \cdot \cos\alpha \tag{11.84}$$

Es wirkt nur die Projektion der Kraft in Richtung des Weges. Die für den gesamten Weg aufzubringende Arbeit ist dann aus der Summe der Teilarbeiten zu berechnen.

$$W = \sum \Delta W \tag{11.85}$$

Ist die Kraft während des gesamten Vorganges konstant und parallel zum Weg, so folgt aus (11.84) und (11.85) für die gesamte Arbeit

$$W = F \cdot s \qquad \begin{array}{c|c|c} W & F & s \\ \hline \mathrm{N\,m} & \mathrm{N} & \mathrm{m} \end{array} \tag{11.86}$$

Die Einheit der Arbeit ist 1 N m, dafür wird die Abkürzung 1 J (Joule) verwendet.

$$1\,\mathrm{J} = 1\,\mathrm{N\,m} = 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^2}} \tag{11.87}$$

Immer dann, wenn eine Masse m gegen eine Kraft F bewegt wird, muss Arbeit verrichtet werden. Wird die Masse in Richtung der Kraft bewegt, wird Arbeit gewonnen. Wird z. B. eine Masse m gegen die Gewichtskraft um eine Höhe $\Delta s = h$ angehoben, so ist dafür die *Hubarbeit* (Abb. 11.12)

$$W_{\mathrm{hub}} = F_G \cdot h = mgh \tag{11.88}$$

gegen die konstante Gewichtskraft zu verrichten. Die Fläche im F-s-Diagramm ist somit gleich der Arbeit.

Wird ein Körper beschleunigt, so ist *Beschleunigungsarbeit* zu verrichten. Nach (11.5) gilt, dass für den zusätzlich während der Beschleunigungszeit Δt zurückgelegten Weg Δs gilt:

$$\Delta s = \frac{1}{2} a (\Delta t)^2$$

Aus (11.3) folgt

$$\Delta t = \frac{\Delta v}{a}$$

und somit

$$\Delta s = \frac{1}{2} a \left(\frac{\Delta v}{a}\right)^2 = \frac{(\Delta v)^2}{2a} = \frac{v_2^2 - v_1^2}{2a}.$$

Damit gilt für die *Beschleunigungsarbeit*

$$W_{\mathrm{beschl}} = ma \cdot \Delta s$$

und somit

$$W_{\mathrm{beschl}} = \frac{1}{2} m (v_2^2 - v_1^2) \tag{11.89}$$

Hierbei ist vorausgesetzt, dass keine weiteren Arbeiten, wie z. B. Reibungsarbeit, zu verrichten sind. Wird ein Körper bei vorhandener *Reibungskraft* auf horizontalem Weg bewegt, so ist jetzt *Reibungsarbeit* gegen die Reibungskraft zu verrichten:

Reibungsarbeit

$$W_R = F_R \cdot (s_2 - s_1) = F_N \cdot \mu \cdot g \cdot (s_2 - s_1) \tag{11.90}$$

Um eine Feder zusammenzudrücken, muss ebenfalls Arbeit gegen die Rückstellkraft der Feder geleistet werden. Wird eine Feder um die Strecke x zusammengedrückt, so ist die Rückstellkraft nicht konstant, sondern gemäß Abb. 11.13 von der Auslenkung abhängig. Auch hier ist die Spannarbeit gleich der gestrichelten Fläche in Abb. 11.13. Es gilt

$$W_{\mathrm{spann}} = \frac{1}{2} F_r x \tag{11.91}$$

$$W_{\mathrm{spann}} = \frac{1}{2} c x^2 \tag{11.92}$$

11.2.4.2 Leistung, Wirkungsgrad

Die *Leistung* P gibt an, wieviel Zeit Δt zur Verrichtung einer Arbeit ΔW benötigt wird.

$$P = \frac{\Delta W}{\Delta t} \qquad \begin{array}{c|c|c|c} P & \Delta W & \Delta t \\ \hline \mathrm{N\,m/s} & \mathrm{N\,m} & \mathrm{s} \end{array} \tag{11.93}$$

Hieraus folgt:

$$P = \frac{\Delta W}{\Delta t} = \frac{F \cdot \Delta s}{\Delta t} \tag{11.94}$$

$$P = F \cdot v \tag{11.95}$$

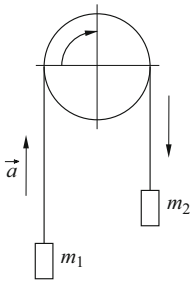


Abb. 11.14 Antriebsleistung

Die SI-Einheit der Leistung ist das *Watt*.

$$1\,\mathrm{W} = 1\,\frac{\mathrm{J}}{\mathrm{s}} = 1\,\frac{\mathrm{N\,m}}{\mathrm{s}} \tag{11.96}$$

Bei allen Maschinen und Antrieben gibt es immer Verluste durch Reibung, Wärme etc. Die von der Maschine geleistete Arbeit ist immer kleiner als die zugeführte Arbeit. Das Verhältnis von abgegebener Arbeit zu zugeführter Arbeit wird *Wirkungsgrad* η genannt.

$$\eta = \frac{W_{\mathrm{ab}}}{W_{\mathrm{zu}}} \leq 1 \tag{11.97}$$

Der Wirkungsgrad ist eine dimensionslose Größe.

Beispiel

Ein beladener Förderkorb mit der zulässigen Gesamtmasse $m_1 = 1000\,\mathrm{kg}$ ist über eine Rolle mit dem Gegengewicht $m_2 = 400\,\mathrm{kg}$ verbunden. Der Korb wird zu Beginn mit der Beschleunigung $a = 1\,\mathrm{m/s^{-2}}$ aufwärts beschleunigt, bis er dann mit einer konstanten Geschwindigkeit von $5\,\mathrm{m/s^{-1}}$ nach oben fährt. Die Reibungskraft $F_R$wird als konstant mit 500 N angenommen. Welche Dauerleistung und welche maximale Leistung muss die Antriebsmaschine bei einem Wirkungsgrad von 0,8 leisten?

Lösung:

Während der Phase der beschleunigten Bewegung ist die Kraft:

$$F_a = m_1(g+a) - m_2(g-a) + F_R = 7786\,\mathrm{N}$$
$$P_a = \frac{F_a v_2}{\eta} = 48{,}67\,\mathrm{kW}$$

während der Phase der gleichförmigen Bewegung gilt:

$$F_v = (m_1 - m_2)g + F_R = 6386\,\mathrm{N}$$
$$P_v = \frac{F_v v_2}{\eta} = 39{,}91\,\mathrm{kW}$$ ◀

11.2.4.3 Energie

Wird einem Körper Arbeit zugeführt, so erhöht sich seine *Energie E*. Energie und Arbeit werden in der gleichen Einheit gemessen (11.87).

$$\Delta E = E_2 - E_1 = \Delta W \tag{11.98}$$

Wird einem Körper *Hubarbeit* zugeführt, so erhöht sich seine *Lageenergie* oder auch *potentielle Energie* E_{pot}. Durch Beschleunigungsarbeit wird die *kinetische Energie* E_{kin} erhöht. Es gilt für die *potentielle Energie*

$$E_{\mathrm{pot}} = mgh \tag{11.99}$$

und für die *kinetische Energie*

$$E_{\mathrm{kin}} = \frac{1}{2}mv^2 \tag{11.100}$$

Die unterschiedlichen Energieformen können während des Ablaufes eines Vorganges in andere Energieformen umgewandelt werden. Es gilt jedoch der grundlegende Satz von der *Erhaltung der Energie in einem abgeschlossenen System.*

Energieerhaltungssatz

In einem abgeschlossenen System ist die Summe aller Energien, die Gesamtenergie, zu jedem Zeitpunkt konstant. Energie kann sich in verschiedene Formen umwandeln und zwischen verschiedenen Teilen des Systems austauschen.

$$E_1 + E_2 + E_3 + \ldots + E_n = const \tag{11.101}$$

Beispiel

Eine Stahlkugel ($m = 1\,\mathrm{kg}$) fällt aus einer Höhe von 2 m auf eine Stahlplatte.

Wie groß ist ihre Geschwindigkeit kurz vor dem Aufprall?

Lösung:

Energien in Höhe h:

$$E_{\mathrm{pot}} = mgh, \quad E_{\mathrm{kin}} = 0$$

vor dem Aufprall:

$$E_{\mathrm{pot}} = 0, \quad E_{\mathrm{kin}} = \frac{1}{2}mv^2$$
$$mgh = \frac{1}{2}mv^2$$

Aufprallgeschwindigkeit

$$v_e = \sqrt{2gh} \quad \text{siehe auch (11.10)}$$
$$v_e = \sqrt{2 \cdot 9{,}81\,\frac{\mathrm{m}}{\mathrm{s^2}} \cdot 2\,\mathrm{m}} = 6{,}26\,\frac{\mathrm{m}}{\mathrm{s}}$$ ◀

Beispiel

Eine Stahlkugel ($m = 1\,\text{kg}$) fällt aus einer Höhe von $h_0 = 2\,\text{m}$ auf eine Stahlplatte und springt dann bis zu einer Höhe von $h_1 = 1{,}90\,\text{m}$ zurück.

Welche Energie geht durch Verformungsarbeit während des Aufpralles verloren?

Lösung:

Umwandlungsenergie = Differenz der potentiellen Energien.

$$\Delta E = mg(h_0 - h_1) \tag{11.102}$$

$$\Delta E = 0{,}981\,\text{J}$$ ◀

11.2.5 Stoßprozesse

Stoßen zwei Körper zusammen, so berühren sie sich kurzzeitig und ändern ihren Bewegungszustand. Je nach Art der Energieübertragung unterteilt man mechanische Stöße in *elastische* und *inelastische Stöße*, nach den geometrischen Verhältnissen in *gerade* und *schiefe Stöße*, *zentrale* und *exzentrische Stöße*. Im Rahmen diese Buches werden nur gerade, zentrale Stöße behandelt. Die Stoßpartner bewegen sich also auf einer Geraden und treffen sich auf der Verbindungslinie ihrer Mittelpunkte.

11.2.5.1 Elastischer Stoß

Beim elastischen Stoß ist die Bewegungsenergie E vor dem Stoß gleich der Bewegungsenergie E' nach dem Stoß. (Größen nach dem Stoß werden durch $'$ gekennzeichnet.) Weiterhin gilt der Impulserhaltungssatz (11.83) $\vec{p} = \vec{p}'$.

Energieerhaltung

$$\frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 = \frac{1}{2}m_1' v_1'^2 + \frac{1}{2}m_2' v_2'^2 \tag{11.103}$$

Impulserhaltung

$$m_1\vec{v}_1 + m_2\vec{v}_2 = m_1\vec{v'}_1 + m_2\vec{v'}_2 \tag{11.104}$$

Geschwindigkeiten nach dem Stoß

$$\vec{v}_1' = \frac{(m_1 - m_2)\vec{v}_1 + 2m_2\vec{v}_2}{m_1 + m_2} \tag{11.105}$$

$$\vec{v}_2' = \frac{(m_2 - m_1)\vec{v}_2 + 2m_1\vec{v}_1}{m_1 + m_2} \tag{11.106}$$

Beispiel

Eine Kugel der Masse $m_1 = 2\,\text{kg}$, $v_1 = 3\,\text{m/s}$ stößt elastisch mit einer ruhenden Kugel $m_2 = 1\,\text{kg}$, $v_2 = 0\,\text{m/s}$ zusammen. Wie groß sind die Geschwindigkeiten nach dem Stoß?

Lösung:

$$v_1' = \frac{(2-1)\,\text{kg}\cdot 3\,\text{m/s} + 2\cdot 1\,\text{kg}\cdot 0\,\text{m/s}}{2\,\text{kg} + 1\,\text{kg}} = 1\,\text{m/s}$$

$$v_2' = \frac{(1-2)\,\text{kg}\cdot 0\,\text{m/s} + 2\cdot 2\,\text{kg}\cdot 3\,\text{m/s}}{2\,\text{kg} + 1\,\text{kg}} = 4\,\text{m/s}$$

Haben beide Kugeln gleiche Massen, so ergibt sich: $\vec{v}_1' = \vec{v}_2$ und $\vec{v}_2' = \vec{v}_1$. ◀

11.2.5.2 Inelastischer Stoß

Beim inelastischen Stoß wird ein Teil der Bewegungsenergie in andere Energiearten, z. B. Verformungsenergie oder Wärmeenergie ΔE, umgewandelt.

$$E_1 + E_2 + \ldots + E_n = E_1' + E_2' + \ldots + E_n' + \Delta E \tag{11.107}$$

Um ΔE berechnen zu können, sind weitere Angaben notwendig. Haben nach dem Stoß beide Körper dieselbe Geschwindigkeit, so liegt ein *vollkommen inelastischer* oder auch *unelastischer Stoß* vor.

$$\vec{v}_1' = \vec{v}_2' = \vec{v}' \tag{11.108}$$

$$m_1\vec{v}_1 + m_2\vec{v}_2 = (m_1 + m_2)\vec{v}' \tag{11.109}$$

$$\vec{v}' = \frac{m_1\vec{v}_1 + m_2\vec{v}_2}{m_1 + m_2} \tag{11.110}$$

Beispiel

Ein Auto ($m_1 = 1000\,\text{kg}$, $\vec{v}_1 = 100\,\text{km/h}$) stößt frontal vollkommen inelastisch mit einem zweiten entgegenkommenden Wagen ($m_2 = 1500\,\text{kg}$, $\vec{v}_2 = -120\,\text{km/h}$) zusammen. Die Dauer des Stoßes sei 0,5 s, beide Fahrer wiegen 100 kg. Das Minuszeichen bei $\vec{v}_2$ bedeutet, dass $\vec{v}_2$ entgegengesetzt zu $\vec{v}_1$ gerichtet ist.

a) Wie groß ist die Geschwindigkeit beider Fahrzeuge nach dem Aufprall?
b) Welche Kräfte haben beide Fahrer auszuhalten?

Lösung:

a) nach (11.110)

$$\vec{v}' = \frac{1100\,\text{kg}\cdot 100\,\text{km h}^{-1} - 1600\,\text{kg}\cdot 120\,\text{km h}^{-1}}{(1100 + 1600)\,\text{kg}}$$

$$\vec{v}' = -30{,}37\,\text{km h}^{-1}$$

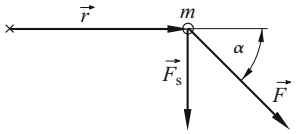


Abb. 11.15 Drehmoment

b) nach (11.82)

$$\vec{F}_1 = \frac{\Delta p_1}{\Delta t} = \frac{100\,\text{kg} \cdot (-30{,}37 - 100)\,\text{km h}^{-1}}{0{,}5\,\text{s}}$$
$$= -2747\,\text{N}$$
$$\vec{F}_2 = \frac{\Delta p_2}{\Delta t} = \frac{100\,\text{kg} \cdot (-30{,}37 + 120)\,\text{km h}^{-1}}{0{,}5\,\text{s}}$$
$$= 1792{,}6\,\text{N}$$ ◀

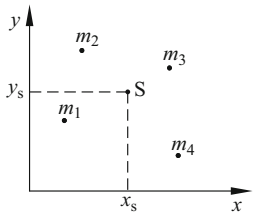


Abb. 11.16 Schwerpunkt

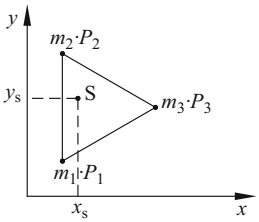


Abb. 11.17 Beispiel Schwerpunkt

11.2.6 Rotation

11.2.6.1 Drehmoment

Um einen Körper in Drehbewegung oder *Rotation* zu versetzen, muss auf diesen Körper eine resultierende Kraft $\vec{F}$ wirken, die in einem Abstand $\vec{r}$ von der Drehachse angreifen muss. Die Richtung der Kraft darf aber nicht auf der Verbindungslinie vom Zentrum der Drehbewegung zum Angriffspunkt der Kraft liegen. Der Winkel α darf also nicht 0° sein. Ist dies der Fall, so bewirkt nur die Komponente der Kraft F_s senkrecht zur Verbindungslinie Angriffspunkt–Drehachse eine Rotation (Abb. 11.15).

Aus $\vec{r}$ und $\vec{F}$ wird durch die Gleichung

$$\vec{M} = \vec{r} \times \vec{F} \qquad \begin{array}{c|c|c} M & r & F \\ \hline \text{N m} & \text{m} & \text{N} \end{array} \qquad (11.111)$$

ein Vektor $\vec{M}$ definiert. Dieser heißt *Drehmoment*. Der Betrag von $\vec{M}$ ist durch

$$M = r \cdot F_s = r \cdot F \cdot \sin\alpha \qquad (11.112)$$

gegeben.

Mit dieser Formel wird allerdings nur der Betrag des Drehmomentes berechnet, die Richtung ergibt sich aus der *Rechte-Hand-Regel*:

Der Daumen der rechten Hand wird in die Richtung von $\vec{r}$ gehalten, der Zeigefinger in Richtung $\vec{F}_s$, der Mittelfinger zeigt die Richtung von $\vec{M}$ an.

11.2.6.2 Massenmittelpunkt, Schwerpunkt

Greifen an einem *starren Körper*, also einem System von fest miteinander verbundenen Massenpunkten, beliebige äußere Kräfte an, so wird das System eine Translation und eine Rotation durchführen. Das System ist nur dann im *statischen Gleichgewicht*, wenn die Summe aller äußeren Kräfte und die Summe aller äußeren Drehmomente = 0 ist.

$$\sum \vec{F}_a = 0 \qquad (11.113)$$
$$\sum \vec{M}_a = 0 \qquad (11.114)$$

Dieses Gleichgewicht stellt sich dann ein, wenn der Körper in einem ganz bestimmten Punkt durch eine Stützkraft $\vec{F}_s$ unterstützt wird. Diese muss entgegengesetzt gleich der Gewichtskraft aller Massenpunkte sein.

$$\vec{F}_S + \sum_{i=1}^{N} m_i \vec{g} = 0 \qquad (11.115)$$

Dadurch ist (11.113) erfüllt. Weiterhin muss bezüglich einer beliebigen Drehachse das Gesamtdrehmoment gleich der Summe der Einzeldrehmomente sein.

$$\vec{M}_{\text{ges}} + \sum_{i=1}^{N} \vec{M}_i = 0 \qquad (11.116)$$

Aus diesen Gleichungen kann die Koordinate dieses Punktes bestimmt werden. Dieser Punkt ist der *Schwerpunkt* oder *Massenmittelpunkt* des starren Körpers. Die Lage hängt von der Massenverteilung ab. Gezeichnet sind vier willkürlich angeordnete Massen und ein noch zu bestimmender Schwerpunkt S. Die Schwerpunkt-Koordinaten berechnen sich bei einer räumlichen Anordnung zu (Abb. 11.16 und 11.17)

$$x_s = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i} \quad y_s = \frac{\sum_{i=1}^{N} m_i y_i}{\sum_{i=1}^{N} m_i} \quad z_s = \frac{\sum_{i=1}^{N} m_i z_i}{\sum_{i=1}^{N} m_i} \qquad (11.117)$$

Beispiel

Berechnung der Schwerpunktkoordinaten folgender Anordnung: $m_1 = 200\,\text{g}$, $m_2 = 300\,\text{g}$, $m_3 = 100\,\text{g}$. Die Massen sitzen auf den Endpunkten eines gleichseitigen Dreiecks. Die Punktkoordinaten sind: $P_1 = (20/20)$, $P_2 = (20/80)$, $P_3 = (71{,}962/50)$.

Lösung:

$$x_s = \frac{200\,\text{g} \cdot 20 + 300\,\text{g} \cdot 20 + 100\,\text{g} \cdot 71{,}9652}{200\,\text{g} + 300\,\text{g} + 100\,\text{g}} = 28{,}66$$

$$y_s = \frac{200\,\text{g} \cdot 20 + 300\,\text{g} \cdot 80 + 100\,\text{g} \cdot 50}{200\,\text{g} + 300\,\text{g} + 100\,\text{g}} = 55$$

Dieser Punkt ist eingezeichnet. ◀

11.2.6.3 Drehimpuls

Bei Bewegungen auf einer Kreisbahn kann in Analogie zur Berechnung des Drehmomentes auch das Produkt rp, mit dem Impuls p, gebildet werden. Die hier vereinfacht wiedergegebene Ableitung gilt allerdings nur für Kreisbewegungen in einer Ebene, damit kann mit den Beträgen der Vektoren gerechnet werden. Mit den Gleichungen für Kreisbewegungen gilt:

$$r \cdot p = r \cdot m \cdot v \tag{11.118}$$

$$r \cdot p = r \cdot m \cdot \omega \cdot r \tag{11.119}$$

$$r \cdot p = m \cdot r^2 \cdot \omega \tag{11.120}$$

Die Größe mr^2 ist das *Trägheitsmoment* J eines Massenpunktes.

$$J = mr^2 \qquad \begin{array}{c|c|c} J & m & r^2 \\ \hline \text{kg m}^2 & \text{kg} & \text{m}^2 \end{array} \tag{11.121}$$

Das Produkt rp wird Drehimpuls L genannt.

$$L = r \cdot p \tag{11.122}$$

$$L = J\omega \tag{11.123}$$

$$L = r \cdot p \qquad \begin{array}{c|c|c} L & r & p \\ \hline \text{kg m}^2/\text{s} & \text{m} & \text{kg m/s} \end{array} \tag{11.124}$$

Weiterhin gilt für das *Drehmoment*

$$M = r \cdot F = r \cdot \frac{\Delta p}{\Delta t} \tag{11.125}$$

$$M = \frac{\Delta L}{\Delta t} \tag{11.126}$$

Wirken keine äußeren Drehmomente, so ist $\Delta L = 0$ und somit $L = \text{const}$ der Gesamtdrehimpuls bleibt also konstant. Dies gilt allgemein, so dass hier die vektorielle Schreibweise verwendet werden kann.

$$\vec{L}_{\text{ges}} = \vec{L}_1 + \vec{L}_2 + \vec{L}_3 + \ldots + \vec{L}_n = \text{const} \tag{11.127}$$

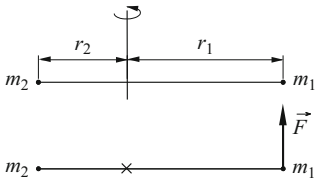


Abb. 11.18 Rotation zweier Punktmassen

Beispiel

Zwei gleiche Massen, $m = 100\,\text{g}$, rotieren im Abstand $r = 30\,\text{cm}$ von der Drehachse mit einer Winkelgeschwindigkeit $\omega = 2/s$.

a) Wie groß ist der Drehimpuls?
b) Wie ändert sich die Winkelgeschwindigkeit, wenn der Abstand der Massen halbiert wird?

Lösung:

a)

$$L_1 = 2 \cdot mr_1^2 \cdot \omega_1$$

$$L_1 = 2 \cdot 0{,}1\,\text{kg} \cdot (0{,}3\,\text{m})^2 \cdot 2\,\text{s} = 0{,}036\,\frac{\text{kg m}^2}{\text{s}}$$

b) Da der Drehimpuls konstant bleibt, gilt:

$$2 \cdot mr_1^2 \cdot \omega_1 = 2 \cdot mr_2^2 \cdot \omega_2$$

$$r_2 = \frac{r_1}{2}$$

$$\omega_2 = \frac{2mr_1^2\omega_1}{2mr_1^2/4} = 4\omega_1$$

$$\omega_2 = 8\,\text{s}^{-1}$$ ◀

11.2.6.4 Trägheitsmoment

Wird eine Punktmasse m, die sich auf einer Kreisbahn bewegt, durch die Kraft F beschleunigt, so gilt nach (11.3), (11.47) und (11.113):

$$M = r \cdot ma = r \cdot m\frac{\Delta v}{\Delta t} = r \cdot m \cdot r\frac{\Delta \omega}{\Delta t}$$

und mit der Winkelbeschleunigung α (11.43).

$$M = mr^2\alpha \tag{11.128}$$

oder auch

$$M = J\alpha \tag{11.129}$$

J ist das bereits oben eingeführte *Massenträgheitsmoment* einer punktförmigen Masse m im Abstand r von der Drehachse.

Sind zwei Massen m_1 und m_2 in unterschiedlichen Abständen r_1 und r_2 auf einer massenlosen Stange gelagert, und greift an

Teil II

Tab. 11.3 Trägheitsmomente spezieller Körper

Punktmasse	$J = mr^2$
Stab Achse durch Stabende, senkrecht zum Stab	$J = \frac{1}{3}ml^2$
Stab Achse durch Stabmitte, senkrecht zum Stab	$J = \frac{1}{12}ml^2$
Vollzylinder Drehachse = Längsachse	$J = \frac{1}{2}mr^2$
Dünne Scheibe Drehachse senkrecht zur Scheibe	$J = \frac{1}{2}mr^2$
Hohlzylinder Drehachse = Längsachse	$J = \frac{1}{2}m(r_a^2 + r_i^2)$
Dünner Ring Drehachse senkrecht zum Ring	$J = mr^2$
Kugel	$J = \frac{2}{5}mr^2$

der Masse m_1 die Kraft F an, so werden beide Massen wegen des angreifenden Drehmomentes in Rotation versetzt. Das Gesamtträgheitsmoment beider Massen kann durch folgende Überlegung bestimmt werden: Die Drehmomenten Gleichung lautet:

$$M = r_1 \cdot F = r_1 m_1 a_1 + r_2 m_2 a_2 \tag{11.130}$$

Da die Winkelbeschleunigung α überall gleich ist, gilt

$$\alpha = \frac{a_2}{r_2} = \frac{a_1}{r_1} \tag{11.131}$$

$$M = r_1^2 m_1 \alpha + r_2^2 m_2 \alpha \tag{11.132}$$

$$M = (r_1^2 m_1 + r_2^2 m_2)\alpha \tag{11.133}$$

$$M = J_{\text{ges}} \cdot \alpha \tag{11.134}$$

mit

$$J_{\text{ges}} = r_1^2 m_1 + r_2^2 m_2 \tag{11.135}$$

$$J_{\text{ges}} = J_1 + J_2 \tag{11.136}$$

Für n einzelne Massenpunkte gilt

$$J_{\text{ges}} = \sum_{i=1}^{n} J_i \tag{11.137}$$

$$J_{\text{ges}} = \sum_{i=1}^{n} r_i^2 m_i \tag{11.138}$$

Bei homogenen Körpern wird diese Summe zum Integral. Das Trägheitsmoment nicht punktförmiger Massen hängt von der Geometrie des Körpers und von der Lage der Drehachse ab

$$J_{\text{ges}} = \int r^2 \, dm = \int_{\text{Vol}} r^2 \rho \, dV \tag{11.139}$$

mit der Dichte

$$\rho = \frac{m}{V} \tag{11.140}$$

In Tab. 11.3 sind einige Beispiele für unterschiedliche Geometrien gegeben, wobei die Drehachse immer parallel zu einer Symmetrieachse liegt und durch den Schwerpunkt des Körpers führt.

11.2.6.5 Rotationsenergie

Rotiert ein Körper um eine beliebige Drehachse A, so hat jeder Massenpunkt *kinetische Energie* nach (11.100). Für einen Massenpunkt gilt mit (11.40) und (11.89):

$$E_{\text{kin}} = \frac{1}{2} m r^2 \omega^2 \tag{11.141}$$

und mit dem *Trägheitsmoment* $J = mr^2$ eines Massenpunktes:

$$E_{\text{rot}} = \frac{1}{2} J_A \omega^2 \tag{11.142}$$

Diese Rotationsenergie ist ein weitere mögliche Bewegungsenergie. Somit gibt es zwei Bewegungsenergien: *Translations-* und *Rotationsenergie*, die unter den Oberbegriff *kinetische Energie* fallen.

$$E_{\text{rot}} = \frac{1}{2} J_A \omega^2 \quad \begin{array}{c|c|c} E_{\text{rot}} & J & \omega^2 \\ \hline \text{J} & \text{kg m}^2 & 1/\text{s}^2 \end{array} \tag{11.143}$$

Beispiel

Eine Kugel mit dem Radius $r = 5\,\text{cm}$ rollt ohne Reibung auf einer schiefen Ebene (Neigungswinkel $a = 20°$) aus einer Höhe $h = 2\,\text{m}$. Wie groß ist ihre Translationsgeschwindigkeit am Ende der schiefen Ebene?

Lösung:

Energien in Höhe h:

$$E_{\text{pot}} = mgh$$
$$E_{\text{kin}} = 0 \tag{11.144}$$

am Ende der Ebene:

$$E_{\text{pot}} = 0$$
$$E_{\text{kin}} = \frac{1}{2} m v^2 + \frac{1}{2} J \omega^2 \tag{11.145}$$

Aus dem Energieerhaltungssatz folgt, dass die Gesamtenergie in Höhe h gleich der Gesamtenergie am Ende der schiefen Ebene sein muss, daher gilt:

$$mgh = \frac{1}{2} m v^2 + \frac{1}{2} \cdot \frac{2}{5} m r^2 \frac{v^2}{r^2} \tag{11.146}$$

$$gh = \frac{v^2}{2} + \frac{v^2}{5} = \frac{7}{10} v^2 \tag{11.147}$$

$$v = \sqrt{\frac{10}{7} gh} \tag{11.148}$$

$$v = \sqrt{\frac{10}{7} \cdot 9{,}81 \frac{\text{m}}{\text{s}^2} \cdot 2\,\text{m}} = 5{,}3 \frac{\text{m}}{\text{s}} \quad \blacktriangleleft$$

Die im vorigen Abschnitt angegebenen Massenträgheitsmomente beziehen sich auf die Rotation um eine Achse durch den Schwerpunkt. Liegt die Drehachse nicht durch den Schwerpunkt, ist aber das Trägheitsmoment J_s bezüglich der Schwerpunktachse bekannt, so lässt sich das Trägheitsmoment J_A

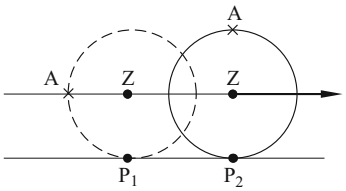


Abb. 11.19 Steiner'scher Satz

bezüglich der beliebigen Achse durch folgende Überlegung bestimmen:

Eine Rotation um eine beliebige Achse kann ersetzt werden durch eine Rotation um die Schwerpunktachse und eine Translation des Schwerpunktes. Die Kugel in Abb. 11.19 soll um eine Viertelumdrehung von links nach rechts rollen. Der momentane Drehpunkt ist der Auflagepunkt P. Während dieser Drehung hat sich der Punkt A auf dem Umfang des Kreises bewegt. Das Zentrum des Kreises Z (Schwerpunkt der Kugel) hat sich linear um die Strecke $\overline{P_1P_2}$ weiterbewegt. Den Endzustand kann man auch erreichen, indem die Kugel zunächst in der Stellung 1 eine Viertelumdrehung macht und dann ohne Drehung um die Strecke $\overline{P_1P_2}$ nach rechts verschoben wird.

Die Rotationsenergie für die Drehung um P ist:

$$E_{\text{kin}}^{\text{rot}} = \frac{1}{2} J_P \omega^2 \tag{11.149}$$

In der zweiten Betrachtungsweise gilt:

$$E_{\text{kin}} = \frac{1}{2} m v_S^2 + \frac{1}{2} J_S \omega^2 \tag{11.150}$$

$$E_{\text{kin}} = \frac{1}{2} (m r^2 + J_S) \omega^2 \tag{11.151}$$

Somit gilt:

$$J_P = m r^2 + J_S \tag{11.152}$$

Diese letzte Gleichung ist der *Steiner'sche Satz*. Die Drehachse durch P hat von der Drehachse durch den Schwerpunkt den Abstand r.

Beispiel

Das vorige Beispiel soll mit Hilfe des Steiner'schen Satzes gelöst werden:

Lösung:

$$J_P = m r^2 + \frac{2}{5} m r^2 \tag{11.153}$$

$$mgh = \frac{1}{2} \cdot \frac{7}{5} m r^2 \omega^2 \tag{11.154}$$

$$mgh = \frac{1}{2} \cdot \frac{7}{5} m v^2 \tag{11.155}$$

$$v = \sqrt{\frac{10}{7} g h} \tag{11.156}$$

$$v = \sqrt{\frac{10}{7} \cdot 9{,}81 \frac{\text{m}}{\text{s}^2} \cdot 2\,\text{m}} = 5{,}3 \frac{\text{m}}{\text{s}}$$ ◀

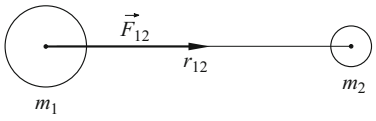


Abb. 11.20 Gravitationsgesetz

11.2.7 Gravitation

Zwischen zwei Massen m_1 und m_2 besteht immer eine Anziehungskraft, die *Gravitationskraft* F_G. Die Richtung dieser Kraft liegt in der Verbindungslinie beider Massen, für den Betrag gilt:

$$F_G = \frac{\gamma m_1 m_2}{r_{12}^2} \tag{11.157}$$

mit der Gravitationskonstanten

$$\gamma = 6{,}67259 \cdot 10^{-11} \frac{\text{N m}^2}{\text{kg}^2} \tag{11.158}$$

Der Abstand r_{12} ist bei einer räumlichen kugelförmigen homogenen Ausdehnung der Massen der Abstand der beiden Kugelzentren.

Die Erdanziehungskraft an der Oberfläche der Erde auf eine Masse m kann ebenfalls durch (11.157) bestimmt werden. Der Abstand r_{12} ist in diesem Fall gleich dem Erdradius $r_E = 6378\,\text{km}$. Die Gewichtskraft auf eine Masse m lässt sich ausdrücken durch:

$$\frac{\gamma m_E m}{r_E^2} = mg \tag{11.159}$$

Mit der bekannten Erdbeschleunigung $g = 9{,}81\,\text{m/s}^2$ kann die Masse der Erde bestimmte werden:

$$m_E = \frac{g r_E^2}{\gamma} \tag{11.160}$$

$$m_E = \frac{9{,}81 \frac{\text{m}}{\text{s}^2} \cdot (6370 \cdot 10^3\,\text{m})^2}{6{,}67 \cdot 10^{-11} \frac{\text{N m}^2}{\text{kg}^2}} \tag{11.161}$$

$$m_E = 5{,}97 \cdot 10^{24}\,\text{kg} \tag{11.162}$$

Auch die Bewegung der Planeten lässt sich durch (11.157) deuten. Sie ist von *Johannes Kepler* in den drei Kepler'schen Gesetzen beschrieben.

1. Die Planeten bewegen sich auf Ellipsen, in deren gemeinsamen Brennpunkt die Sonne steht.
2. Jeder Strahl von der Sonne zu einem Planeten überstreicht in gleichen Zeiten gleiche Flächen.
3. Die Quadrate der Umlaufzeiten der Planeten verhalten sich wie die dritten Potenzen der großen Halbachsen ihrer Bahnen um die Sonne.

11.3 Elastische Verformung fester Körper

Elastische Verformungen sind *reversible* Vorgänge unter dem Einfluss einer äußeren Kraft, nach Beendigung der Einwirkung geht die Verformung vollständig zurück. Entscheidend für die Größe der Verformung ist, abgesehen von den Materialeigenschaften des untersuchten Festkörpers, die *Spannung*. Diese ist definiert:

$$\textit{Spannung}\quad S = \frac{\Delta F}{\Delta A} \qquad \begin{array}{c|c|c} S & \Delta F & \Delta A \\ \hline \mathrm{N\,m^{-2}} & \mathrm{N} & \mathrm{m^2} \end{array} \qquad (11.163)$$

Dabei wird die zur Fläche ΔA senkrechte Komponente von S die *Normalspannung* s und die zur Fläche parallele Komponente Schubspannung t genannt.

$$\textit{Normalspannung} \qquad \sigma = \frac{\Delta F_n}{\Delta A} \qquad (11.164)$$

$$\textit{Schubspannung} \qquad \tau = \frac{\Delta F_t}{\Delta A} \qquad (11.165)$$

Unter Einwirkung einer Normalspannung wird sich ein Körper dehnen. Die *Dehnung* ε ist die relative Längenänderung bezogen auf die Ausgangslänge l_0.

$$\textit{Dehnung}\ \varepsilon = \frac{\Delta l}{l_0} = \frac{l_1 - l_0}{l_0} \qquad (11.166)$$

Zwischen *Normalspannung* und *Dehnung* gilt im Bereich elastischer Verformung das *Hooke'sche Gesetz*

$$\sigma = \varepsilon \cdot E \qquad (11.167)$$

E ist eine materialabhängige Größe und heißt: *Elastizitätsmodul*.

$$E = \frac{\sigma}{\varepsilon} \qquad \begin{array}{c|c|c} E & \varepsilon & \sigma \\ \hline \mathrm{N\,m^{-2}} & 1 & \mathrm{N\,m^{-2}} \end{array} \qquad (11.168)$$

In der Technik ist für den *Elastizitätsmodul* E die Einheit $\mathrm{N/mm^2}$ geläufig.

Unter dem Einfluss einer Spannung verändert sich normalerweise auch das Volumen eines Körpers. Es gilt für einen Stab der Dicke d unter einseitiger Belastung:

$$\frac{\Delta V}{V} = \varepsilon \cdot (1 - 2\mu) \qquad (11.169)$$

mit der *Poissonzahl*

$$\mu = -\frac{\Delta d/d_0}{\varepsilon} \qquad (11.170)$$

Nur wenn $\mu = 0{,}5$ ist, bleibt das Volumen unverändert.

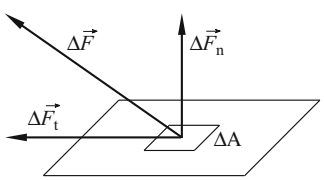


Abb. 11.21 Spannung

11.4 Mechanik der ruhenden Flüssigkeiten und Gase

Im Gegensatz zu Festkörpern, bei denen die Atome auf festen Plätzen sitzen, können in Flüssigkeiten und Gasen die Atome bzw. Moleküle leicht gegeneinander verschoben werden. Die Atomabstände in Flüssigkeiten und Festkörpern sind ähnlich, in Gasen wesentlich größer. Festkörper und Flüssigkeiten haben ähnliche Dichten, Gase eine viel geringere Dichte. Für die Dichte ρ gilt:

$$\rho = \frac{m}{V} \qquad \begin{array}{c|c|c} \rho & m & V \\ \hline \mathrm{kg\,m^{-3}} & \mathrm{kg} & \mathrm{m^3} \end{array} \qquad (11.171)$$

11.4.1 Druck

Wird nun an einer Stelle der Flüssigkeit oder des Gases auf die Oberfläche eine Kraft ausgeübt, so überträgt sich diese auf das ganze Volumen. Das Verhältnis Kraft zu Fläche wird *Druck* p genannt. Wegen der leichten Verschiebbarkeit der Moleküle ist der Druck überall gleich groß und wirkt senkrecht auf die Oberfläche.

$$p = \frac{\Delta F}{\Delta A} \qquad \begin{array}{c|c|c} p & \Delta F & \Delta A \\ \hline \mathrm{N\,m^{-2}} & \mathrm{N} & \mathrm{m^{-2}} \end{array} \qquad (11.172)$$

Die Einheit ist 1 Pascal:

$$1\,\mathrm{Pa} = 1\,\frac{\mathrm{N}}{\mathrm{m^2}} \qquad (11.173)$$

Eine weitere außerhalb des SI-Systems, aber nach DIN 1301 zulässige Bezeichnung ist

$$1\,\mathrm{bar} = 10^5\,\mathrm{Pa} \qquad (11.174)$$

11.4.2 Kompressibilität

Flüssigkeiten und Gase werden durch äußeren Druck komprimiert, ihr Volumen nimmt ab.

$$V = V_a(1 - \chi\Delta p) \qquad (11.175)$$

χ ist die *Kompressibilität*

$$\chi = -\frac{1}{V_a} \cdot \frac{\Delta V}{\Delta p} \qquad (11.176)$$

Da sich das Volumen verkleinert, muss sich die Dichte vergrößern.

$$\rho = \frac{\rho_a}{(1 - \chi\Delta p)} \qquad (11.177)$$

ρ_a und V_a sind die dabei jeweiligen Anfangswerte.

Da χ von Flüssigkeiten sehr viel kleiner ist als bei Gasen, werden Flüssigkeiten oft als *inkompressibel* betrachtet. Technische Anwendung: *hydraulische Presse*.

11.4.3 Volumenausdehnung

Durch Temperaturerhöhung wird normalerweise das Volumen vergrößert. (Ausnahme: Anomalie des Wassers zwischen 0 °C und 4 °C, Wasser hat bei 4 °C sein kleinstes Volumen und damit seine größte Dichte)

$$V = V_0(1 + \gamma\Delta\vartheta) \tag{11.178}$$

γ ist der *Volumenausdehnungskoeffizient* oder *kubische Ausdehnungskoeffizient*, V_0 ist das Volumen bei 0 °C, $\Delta\vartheta$ wird dann ebenfalls auf 0 °C bezogen. Die Dichte wird bei Temperaturerhöhung kleiner (Ausnahme s. o.)

$$\rho = \frac{\rho_0}{1 + \gamma\Delta\vartheta} \tag{11.179}$$

11.4.4 Hydrostatischer Druck in Flüssigkeiten

Auf jedes Flüssigkeitsmolekül wirkt die Schwerkraft. Auf den Boden eines mit Flüssigkeit gefüllten Gefäßes wirkt dann die Gewichtskraft aller in der Flüssigkeit befindlichen Moleküle unter der Annahme, dass sich die Dichte nicht ändert:

$$F_G = mg = \rho Vg \tag{11.180}$$
$$F_G = \rho Ahg$$

Damit wird der Druck

$$p = \frac{F_G}{A} = \rho gh \tag{11.181}$$

Dieser durch die Schwerkraft verursachte Druck wird *Schweredruck* genannt. Herrscht über der Flüssigkeit noch ein weiterer äußerer Druck p_a, z. B. der Luftdruck, so ist der Gesamtdruck am Boden

$$p_h = p_a + \rho gh \tag{11.182}$$

Dieser Gesamtdruck ist der *hydrostatische Druck*. Er ist neben dem äußeren Druck nur von der Höhe der Flüssigkeitssäule über dem Boden abhängig, nicht aber von der gesamten Masse der Flüssigkeitssäule. Er ist damit unabhängig von der Form des Gefäßes und wirkt auch auf alle Seitenwände.

11.4.5 Schweredruck in Gasen

Der Schweredruck von Gasen errechnet sich ebenfalls durch die Gewichtskraft der über einer Fläche stehenden Gassäule. Im Gegensatz zu Flüssigkeiten ändert sich die Dichte in der Gassäule merklich. Wendet man diese Überlegung auf den Luftdruck an, so lässt sich der Luftdruck in einer bestimmten Höhe über der Erdoberfläche berechnen durch die *barometrische Höhenformel*

$$p = p_0 \cdot e^{-\frac{\rho_0 gh}{p_0}} \tag{11.183}$$

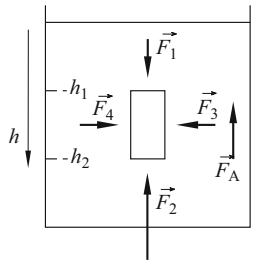


Abb. 11.22 Auftrieb

Nach DIN 5450 gilt: Für eine Lufttemperatur von 0 °C ist $p_0 = 1{,}01325 \cdot 10^5$ Pa und $\rho_0 = 1{,}293\ \frac{\text{kg}}{\text{m}^3}$. Bei der Ableitung von (11.183) ist vorausgesetzt, dass die Temperatur konstant bleibt.

Beispiel

In welcher Höhe ist der Luftdruck auf die Hälfte des Wertes an der Erdoberfläche gesunken?

Lösung:

$$\ln\frac{p}{p_0} = \ln 0{,}5 = -\frac{\rho_0 gh}{p_0}$$

$$h = -\frac{p_0}{\rho_0 g}\ln 0{,}5 = 5537\,\text{m}$$ ◀

11.4.6 Auftrieb

Auf einen Körper wirkt von allen Seiten der hydrostatische Druck nach (11.182), (Abb. 11.22).

Da der Druck in größeren Tiefen zunimmt, entsteht dadurch eine resultierende Kraft nach oben, die *Auftriebskraft* $\vec{F}_A$.

$$\vec{F}_A = \sum \vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 \tag{11.184}$$

Da sich die Seitenkräfte paarweise aufheben, verbleibt

$$\vec{F}_A = \vec{F}_1 + \vec{F}_2$$
$$F_A = A(p_2 - p_1) = \rho_{\text{Fl}} gA(h_2 - h_1)$$
$$F_A = g\rho_{\text{Fl}} V_{\text{Fl}} = gm_{\text{Fl}} \tag{11.185}$$

Die Auftriebskraft ist gleich der Gewichtskraft der verdrängten Flüssigkeit. Sie ist nach oben gerichtet. Der gleiche Sachverhalt gilt auch in Gasen, nur sind dort wegen der geringeren Dichten der Gase auch die Auftriebskräfte wesentlich kleiner.

11.5 Hydrodynamik

In der Hydrodynamik wird das Verhalten von strömenden Flüssigkeiten unter dem Einfluss äußerer Kräfte untersucht.

11.5.1 Kontinuitätsgleichung

Der *Massenstrom* $\dot{m}$ gibt an, welche Masse einer Flüssigkeit pro Zeiteinheit durch einen Querschnitt A strömt.

Massenstrom

$$\dot{m} = \rho v A \qquad \begin{array}{c|c|c|c} \dot{m} & \rho & v & A \\ \hline \mathrm{kg\,s^{-1}} & \mathrm{kg\,m^{-3}} & \mathrm{m\,s^{-1}} & \mathrm{m^2} \end{array} \qquad (11.186)$$

Wenn in der betrachteten Flüssigkeit keine Quellen oder Senken vorhanden sind, ist der Massenstrom konstant.

$$\rho_1 v_1 A_1 = \rho_2 v_2 A_2 = \text{const} \qquad (11.187)$$

Ist die Flüssigkeit *inkompressibel*, so ist ihre Dichte ρ konstant, und es gilt

$$\dot{V} = \frac{\dot{m}}{\rho} = vA = \text{const} \qquad (11.188)$$

mit dem Volumenstrom $\dot{V}$.

Volumenstrom

$$\dot{V} = Av \qquad \begin{array}{c|c|c} \dot{V} & A & v \\ \hline \mathrm{m^3\,s^{-1}} & \mathrm{m^2} & \mathrm{m\,s^{-1}} \end{array} \qquad (11.189)$$

Gleichung (11.187) oder (11.188) ist die *Kontinuitätsgleichung*.

11.5.2 Bernoulli-Gleichung

Muss gegen einen Druck p ein Flüssigkeitsvolumen ΔV am Ort 1 in eine Strömung gebracht werden, so ist hierfür die Arbeit

$$W_1 = F_1 \cdot \Delta s = p_1 A_1 \Delta s = p_1 \Delta V_1 \qquad (11.190)$$

aufzubringen. Tritt diese Flüssigkeit am Ort 2 wieder aus, so wird die Arbeit

$$W_2 = p_2 \Delta V_2$$

frei. Da die Flüssigkeiten inkompressibel sein sollen, gilt $\Delta V_1 = \Delta V_2 = \Delta V$.

Für die Differenz der Arbeiten gilt somit:

$$\Delta W = (p_2 - p_1)\Delta V \qquad (11.191)$$

Bei einer strömenden Flüssigkeit haben die Teilvolumina auch kinetische Energie, befinden sie sich auf unterschiedlicher Höhe über der Erdoberfläche, so kommt noch potentielle Energie hinzu. Der Energieerhaltungssatz gilt dann in der Form:

$$p_1 \Delta V + \frac{1}{2} m v_1^2 + mgh_1 = p_2 \Delta V + \frac{1}{2} m v_2^2 + mgh_2 \qquad (11.192)$$

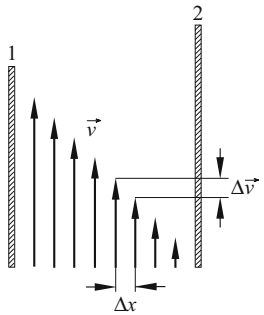


Abb. 11.23 Innere Reibung

oder mit $m = \rho \cdot \Delta V$

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2 \qquad (11.193)$$

oder allgemein

$$p_{\text{ges}} = p + \frac{1}{2}\rho v^2 + \rho g h = \text{const} \qquad (11.194)$$

Gleichung (11.194) ist die *Bernoulli-Gleichung* inkompressibler Flüssigkeiten ohne Reibungsverluste. Der Gesamtdruck setzt sich aus dem *Betriebsdruck* p, dem *Staudruck* oder *dynamischen Druck* $\frac{1}{2}\rho v^2$ und dem *Schweredruck* $\rho g h$ zusammen und ist konstant.

Treten Reibungsverluste auf, wird (11.193) in der Form

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2 + p_V \qquad (11.195)$$

geschrieben, mit dem Druckverlust p_V, der in der Praxis durch eine Verlusthöhe

$$h_V = \frac{p_V}{\rho g} \qquad (11.196)$$

angegeben wird.

11.5.3 Innere Reibung

Zwischen den Molekülen einer Flüssigkeit treten Anziehungskräfte auf. Der Einfluss dieser Kräfte auf die Strömung wird *innere Reibung* genannt. Um eine bewegliche Platte 1 der Fläche A mit konstanter Geschwindigkeit v parallel zu einer festen Wand 2 zu verschieben, ist eine Kraft F_R notwendig (Abb. 11.23). In der Flüssigkeit wird sich im Idealfall ein lineares Geschwindigkeitsgefälle $\Delta v/\Delta x$ einstellen. Die kennzeichnende Größe dieser Reibung ist die *Viskosität*, die durch (11.198) definiert ist.

$$\textit{Reibungskraft} \quad F_R = \eta A \frac{\Delta v}{\Delta x} \qquad (11.197)$$

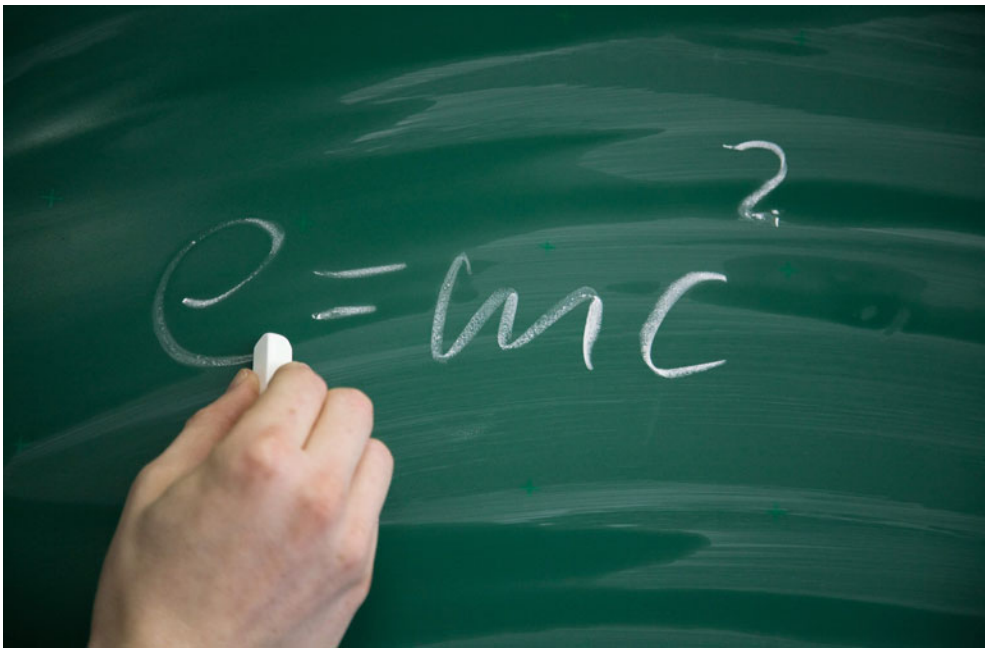
mit der *dynamischen Viskosität*:

$$\eta = \frac{F_R \cdot \Delta x}{A \cdot \Delta v} \qquad \begin{array}{c|c|c|c|c} \eta & F_R & \Delta x & A & \Delta v \\ \hline \mathrm{N\,s/m^2} & \mathrm{N} & \mathrm{m} & \mathrm{m^2} & \mathrm{m\,s^{-1}} \end{array} \qquad (11.198)$$

Thermodynamik

12

Horst Steffen



12.1 Grundbegriffe 214
12.2 Temperatur 214
12.3 Thermische Ausdehnung 214
12.4 Ideale Gase 216
12.5 Wärmeübertragung 219

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_12

12.1 Grundbegriffe

Die Thermodynamik beschreibt das Verhalten auch komplizierter Vielteilchensysteme mit Hilfe einiger wenigen Größen, den *Zustandsgrößen*. Alle beteiligten Teilchen werden zu einem thermodynamischen *System* zusammengefasst. Dieses wird als *abgeschlossenes System* bezeichnet, wenn weder Energie noch Masse mit der Umgebung ausgetauscht werden.

Die *Zustandsgrößen* werden bei der mathematischen Behandlung des thermodynamischen Systems verwendet. Messbare *Zustandsgrößen* sind Druck (siehe Abschn. 11.4.1), Temperatur und Volumen, davon abgeleitet sind z. B. die *Entropie* und die *Innere Energie*. Als Bezugsmenge wird oft das Mol (siehe Tab. 10.1) benutzt, in diesem Fall spricht man von *molaren Größen*.

12.2 Temperatur

Die Temperatur ist ein Maß für die fühlbare Wärme eines Materials. Beim Kontakt zweier Körper mit verschiedenen Temperaturen kann man beobachten, dass sich die Temperaturen ausgleichen, und dass nach einer bestimmten Zeit beide Körper die gleiche Temperatur haben. Im thermodynamischen Gleichgewicht haben alle Stoffe gleiche Temperatur.

12.2.1 Einheiten

Es gibt im praktischen Gebrauch verschiedene Temperaturskalen. Diese werden durch *Fixpunkte* festgelegt. Dies sind Temperaturen, bei denen ein wohldefinierter physikalischer Vorgang abläuft. Fixpunkte können Gefrierpunkt und Siedepunkt des Wassers sein. Die Temperatureinheiten zwischen den Fixpunkten sind dann willkürlich festgelegt. Im europäischen Raum ist die *Celsius-Skala* (ϑ), der Abstand zwischen Gefrier- und Siedepunkt des Wassers ist in 100 Teile unterteilt, im angloamerikanischen Raum die *Fahrenheit-Skala* (ϑ_F) gebräuchlich. Die SI-Skala ist die *Kelvin-Skala* (T).

Skala	Gefrierpunkt	Siedepunkt
Celsius	0 °C	100 °C
Fahrenheit	32 °F	212 °F
Kelvin	273,15 K	373,15 K

Eine Temperaturerhöhung von $\Delta\vartheta = 1\,°\mathrm{C}$ entspricht auch einer Erhöhung von $\Delta T = 1\,\mathrm{K}$. Die Umrechnungsformel lautet:

$$T = (\vartheta + 273{,}15)\,\mathrm{K} \qquad \begin{array}{c|c} T & \vartheta \\ \hline \mathrm{K} & °\mathrm{C} \end{array} \qquad (12.1)$$

Sollen Werte der Celsius-Skala in Fahrenheit umgerechnet werden, so gilt die Beziehung

$$\vartheta = \frac{5}{9}(\delta_F - 32)\,°\mathrm{C} \qquad \begin{array}{c|c} \vartheta & \vartheta_F \\ \hline °\mathrm{C} & °\mathrm{F} \end{array} \qquad (12.2)$$

Beispiel

Ein Körper hat eine Temperatur von 300 K. Welche Celsius- und Fahrenheit-Werte ergeben dies?

Lösung:

$$\vartheta = T - 273{,}15\,\mathrm{K} = 300\,\mathrm{K} - 273{,}15\,\mathrm{K} = 26{,}85\,°\mathrm{C}$$

$$\vartheta_F = \left(\frac{9}{5}\vartheta + 32\right) F = \left(\frac{9}{5}\cdot 26{,}85 + 32\right)\mathrm{F}$$

$$= 80{,}33\,°\mathrm{F}$$ ◀

12.2.2 Temperaturmessung

Die Temperaturmessung ist eine indirekte Messung, es werden physikalische Größen gemessen, die temperaturabhängig sind. Aus der gemessenen Größe wird dann die gesuchte Temperatur berechnet. Zu diesen temperaturabhängigen Effekten gehören z. B.: Ausdehnung fester, flüssiger und gasförmiger Körper, Änderung des elektrischen Widerstandes von Metallen und Halbleitern, Thermospannung und Thermoelemente, optische Strahlung.

12.3 Thermische Ausdehnung

Fast alle Stoffe dehnen sich bei Erwärmung aus. Wenn ein Stoff erwärmt wird, so führen die Atome bzw. Moleküle Schwingungen mit größerer Amplitude aus, die Schwingungsmittelpunkte zweier benachbarter Atome rücken auseinander, makroskopisch hat dies eine Ausdehnung des Körpers zur Folge.

Aufgetragen (Abb. 12.1) ist die potentielle Energie zwischen zwei Atomen als Funktion des Abstandes. Ein Atom ist am Ort $r = 0$, das andere Atom hat hiervon den Abstand r. Bei der Temperatur T_1 ist der mittlere Abstand zwischen beiden Atomen gleich r_1, bei der Temperatur $T_2 > T_1$ ist dieser Abstand r_2. Um die Strecke $\Delta r = r_2 - r_1$ hat sich der mittlere Abstand also vergrößert.

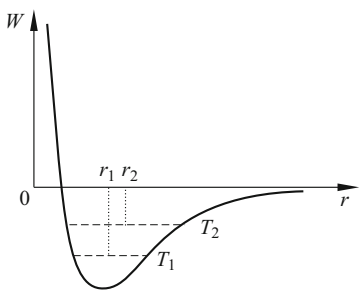


Abb. 12.1 Thermische Ausdehnung

Tab. 12.1 Linearer Ausdehnungskoeffizient

Material	$a_l \cdot 10^6$ in K^{-1}
Aluminium	23,8
Messing	19
Kupfer	16,4
V_2A-Stahl	16
Glas	9

Tab. 12.2 Volumen-Ausdehnungskoeffizient bei 20 °C

Material	$a_v \cdot 10^3/K^{-1}$
Wasser	0,208
Quecksilber	0,182
Dieselkraftstoff	1,0

12.3.1 Feste Stoffe

Betrachtet wird ein Stab der Länge l_1 bei der Temperatur T_1. Wird dieser Stab auf die Temperatur T_2erwärmt, so dehnt er sich um die Strecke $\Delta l = l_2 - l_1$ aus. Für die meisten Festkörper gilt:

$$l_2 = l_1 \cdot (1 + \alpha_l (T_2 - T_1)) \tag{12.3}$$

oder

$$\frac{\Delta l}{l} = \alpha_l \cdot \Delta T \tag{12.4}$$

Die Größe a_1 heißt *linearer Ausdehnungskoeffizient* oder *Längenausdehnungskoeffizient*. Diese Größe ist materialabhängig und kann in bestimmten Temperaturgrenzen als konstant betrachtet werden. Einige typische Zahlenwerte sind in Tab. 12.1 dargestellt. Bei Temperaturerhöhung findet nicht nur eine Verlängerung, sondern auch eine Volumenvergrößerung statt. Für den *Volumenausdehnungskoeffizienten* a_v gilt:

$$\alpha_v = 3\alpha_l \tag{12.5}$$

und für die Volumenänderung:

$$\frac{\Delta V}{V} = \alpha_v \cdot \Delta T \tag{12.6}$$

Da die Masse unverändert bleibt, muss sich bei Volumenvergrößerung die Dichte verkleinern.

$$\rho_2 = \frac{\rho_1}{1 + \alpha_v \Delta T} \tag{12.7}$$

12.3.2 Flüssigkeiten

Bei Flüssigkeiten kann eine Ausdehnung nur in allen Richtungen erfolgen, daher ist hier nur der *Volumenausdehnungskoeffizienten* a_v sinnvoll. Im Allgemeinen gilt auch hier, dass sich bei Temperaturerhöhung das Volumen vergrößert und somit die Dichte verkleinert. Eine Ausnahme bildet das Wasser, dessen Dichte bei Erwärmung ab 0 °C zunächst zunimmt und bei 4 °C seinen größten Wert hat. Anschließend nimmt die Dichte ab (*Anomalie des Wassers*). Aufgrund dieses Effektes frieren Seen im Winter von oben her zu, und das Wasser bleibt am Boden flüssig. Einige typische Zahlenwerte sind in Tab. 12.2 angegeben.

12.3.3 Gase

Bei den tatsächlich vorhandenen realen Gasen haben die Moleküle ein Eigenvolumen und üben untereinander Anziehungskräfte aus. Der Einfluss dieser Größen ist schwer zu erfassen, deshalb geht man zunächst von einer Vereinfachung oder auch Idealisierung aus. Unter solchen idealen Gasen versteht man solche Gase, bei denen gilt:

> Die Gasmoleküle besitzen kein Eigenvolumen. Anziehungskräfte zwischen den Gasmolekülen sind nicht vorhanden.
>
> Es finden nur elastische Stöße statt.

Solche Gase können z. B. nicht verflüssigt werden.

Für diese idealen Gase gilt bei Erwärmung unter konstantem Druck folgendes Gesetz, welches von *Gay-Lussac* zu Beginn des 19. Jahrhunderts durch Experimente festgestellt wurde:

$$V_v = V_0 (1 + \alpha_v \vartheta)_{p=\text{const}} \tag{12.8}$$

Hierbei ist ϑ die Temperatur in °C und V_0 das betrachtete Gasvolumen bei $\vartheta = 0\,°C$. Gleichung (12.8) ergibt eine Gerade, deren Schnittpunkt mit der Temperaturachse bei −273,15 °C (= 0 K) liegt. Daraus folgt für den Ausdehnungskoeffizienten idealer Gase:

$$\alpha_v = \frac{1}{273{,}15\,\text{K}} = 0{,}003661\,\text{K}^{-1} \tag{12.9}$$

Aus (12.8) folgt weiter:

$$V_T = V_0 \left(1 + \frac{1}{273{,}15}(T - 273{,}15)\right)$$

Mit $T_0 = 273{,}15\,\text{K}$ ergibt dies $V_T = T\frac{V_0}{T_0}$

$$\frac{V}{T} = \text{const}_{p=\text{const}} \tag{12.10}$$

Wird das Volumen jedoch konstant gehalten und die Temperatur erhöht, so steigt auch der Druck des Gases. Hierfür gilt:

$$p_\vartheta = p_0 (1 + \alpha_v \vartheta)_{V=\text{const}} \tag{12.11}$$

und analog zu (12.10):

$$\frac{p}{T} = \text{const}_{V=\text{const}} \tag{12.12}$$

Bei konstanter Temperatur folgt aus (12.10) und (12.12) das Gesetz von *Boyle-Mariotte*

$$pV = \text{const}_{T=\text{const}} \tag{12.13}$$

Teil II

12.4 Ideale Gase

12.4.1 Allgemeine Zustandsgleichung idealer Gase

Die Gleichungen (12.10), (12.12) und (12.13) lassen sich zu einer einzigen Gleichung zusammenfassen, sofern die Stoffmenge des betrachteten Gases in der Einheit mol angegeben wird. Für eine Stoffmenge von n Mol kann man diese Gleichungen in der Form

$$\frac{pV}{T} = \frac{p_0 V_0}{T_0} n \tag{12.14}$$

zusammenfassen, wobei die Größen, die mit 0 indiziert sind, die Normgrößen eines idealen Gases bei der Temperatur 0 °C sind:

$$\begin{aligned} p_0 &= 101.325\,\text{Pa} \\ V_0 &= 22{,}41383\,\text{dm}^3/\text{mol} \\ T_0 &= 273{,}15\,\text{K} \end{aligned} \tag{12.15}$$

mit der Abkürzung

$$R = \frac{p_0 V_0}{T_0} \tag{12.16}$$

folgt

$$pV = nRT \tag{12.17}$$

Die Größe R ist die *universelle Gaskonstante* und hat den Zahlenwert:

$$R = \frac{101.325\,\text{Pa} \cdot 22{,}41383\,\text{dm}^3}{273{,}15\,\text{K} \cdot \text{mol}} = 8{,}31441\,\frac{\text{J}}{\text{mol K}} \tag{12.18}$$

Da in einem Mol eines Gases stets die gleiche Anzahl von Molekülen vorhanden ist, nämlich die

Avogadro-Konstante

$$N_A = 6{,}022045 \cdot 10^{23}\,\text{mol}^{-1}, \tag{12.19}$$

kann die universelle Gaskonstante R auch durch molekulare Größen bestimmt werden:

$$R = k \cdot N_A \tag{12.20}$$

mit der Boltzmann-Konstante

$$k = 1{,}38066 \cdot 10^{-23}\,\frac{\text{J}}{\text{K}} \tag{12.21}$$

Wird die Menge des zu betrachtenden Gases nicht in der Einheit mol, sondern in kg angegeben, so muss für jede Gassorte eine *spezielle Gaskonstante* R_s angegeben werden. Zur Umrechnung muss die Dichte ρ_0 des Gases bei 0 °C und die betrachtete Gasmasse m in kg bekannt sein. Es gilt:

$$R_s = \frac{p_0}{\rho_0 T_0} \tag{12.22}$$

und

$$pV = mR_s T \tag{12.23}$$

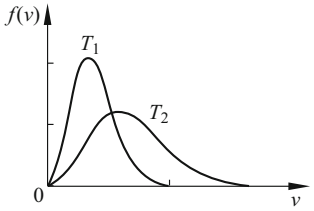


Abb. 12.2 Maxwell'sche Geschwindigkeitsverteilung

12.4.2 Kinetische Gastheorie

Die makroskopischen Zustandsgrößen p und T eines Gases hängen mit der Teilchenbewegung, den Stößen der Moleküle mit der Wand und der kinetischen Energie der Moleküle zusammen. Die Moleküle und ihre Bewegung werden durch ihre Masse m_M, die Dichte ρ und ihre mittlere Geschwindigkeit v_m definiert. Die Zusammenhänge zwischen diesen Größen werden in der *kinetischen Gastheorie* abgeleitet. Die Geschwindigkeitsverteilung wird durch die *Maxwell'sche Geschwindigkeitsverteilung* beschrieben. Diese Verteilungsfunktion gibt die Wahrscheinlichkeit dafür an, dass eine Geschwindigkeit im Intervall zwischen v und $v + \Delta v$ liegt.

$$f(v) \cdot \Delta v = 4\pi v^2 \left(\frac{m_M}{2\pi kT}\right)^{\frac{3}{2}} \cdot e^{-\frac{m_M v^2}{2kT}} \cdot \Delta v \tag{12.24}$$

In Abb. 12.2 mit willkürlichen Einheiten ist $T_2 = 3 \cdot T_1$.

Die Verteilungskurve hat das Maximum bei der wahrscheinlichsten Geschwindigkeit v_w, die mittlere Geschwindigkeit v_m liegt bei

$$v_m = \sqrt{\overline{v^2}} = \sqrt{\frac{3}{2}} v_w \tag{12.25}$$

v_m ist von der Temperatur T abhängig

$$v_M = \sqrt{\frac{3kT}{m_M}} \tag{12.26}$$

Für den Druck gilt:

$$p = \frac{1}{3}\frac{N}{V} m_M \overline{v^2} \tag{12.27}$$

oder

$$p = \frac{1}{3}\rho \overline{v^2} \tag{12.28}$$

und für die mittlere kinetische Energie:

$$\overline{E_{\text{kin}}} = \frac{1}{2} m_M \overline{v^2} \tag{12.29}$$

oder

$$\overline{E_{\text{kin}}} = \frac{3}{2} kT \tag{12.30}$$

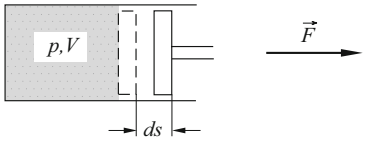


Abb. 12.3 Volumenarbeit

12.4.3 Wärmeenergie

Die Temperatur eines Gases hängt nach (12.30) mit der Bewegungsenergie der Moleküle zusammen. Erhöht man diese Energie z. B. durch Zuführung von Wärmeenergie ΔQ, so muss sich die Temperatur um ΔT erhöhen, solange keine Phasenänderungen ablaufen.

$$\Delta Q = C \cdot \Delta T \qquad \begin{array}{c|c|c} \Delta Q & C & \Delta T \\ \hline \mathrm{J} & \mathrm{J/K} & \mathrm{K} \end{array} \qquad (12.31)$$

Der Faktor C ist die *Wärmekapazität*.

$$C = \frac{\Delta Q}{\Delta T} \qquad (12.32)$$

Je nach Bezugsmenge wird die spezifische (bezogen auf 1 kg) oder molare (bezogen auf 1 mol) Wärmekapazität benutzt.

$$\textit{spezifische Wärmekapazität} \qquad c = \frac{C}{m} \qquad (12.33)$$

$$\textit{molare Wärmekapazität} \qquad C_m = \frac{C}{n} \qquad (12.34)$$

Die spezifische Wärmekapazität von Wasser beträgt z. B. $4{,}182\ \frac{\mathrm{kJ}}{\mathrm{kg\,K}}$.

Wird einem System von außen Wärmeenergie ΔQ zugeführt, so wird nicht die gesamte Energie zur Erhöhung der Temperatur aufgebraucht, ein Teil kann in der Regel in Arbeit ΔW umgewandelt und nach außen abgegeben werden. Wird die abgegebene Arbeit negativ gezählt, so gilt

$$\Delta Q = \Delta U - \Delta W \qquad (12.35)$$

Die Größe ΔU ist die *innere Energie* des Systems und wird durch die innere Bewegung der Teilchen bestimmt. Im folgenden soll als abgegebene Arbeit nur mechanische Arbeit betrachtet werden, die durch Volumenänderung, d. h. Bewegung eines Kolbens, verrichtet werden kann. Dehnt sich ein Gas aus und bewegt dabei einen Kolben um die Strecke Δs, so wird dabei die Arbeit ΔWverrichtet.

$$\Delta W = -F \cdot \Delta s = -\frac{F \cdot A \cdot \Delta s}{A} \qquad (12.36)$$

oder

$$\Delta W = -p \cdot \Delta V \qquad (12.37)$$

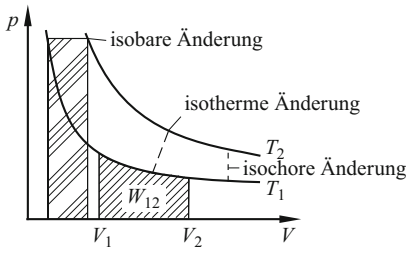


Abb. 12.4 p-V-Diagramm

12.4.4 Zustandsänderungen idealer Gase

Bei der Diskussion von Zustandsänderungen wird eine Zustandsgröße konstant gehalten und dann die Variation der anderen untersucht. Hier sollen nur die Zustandsgrößen T, V und p und deren Variation in den folgenden Prozessen untersucht werden. Die betrachtete Stoffmenge ist 1 mol. Die Ausgangsgleichung ist das allgemeine Gasgesetz (12.17).

Isotherme Zustandsänderungen

Isotherme Zustandsänderungen sind solche bei konstanter Temperatur T. Gleichung (12.17) wird dann zu

$$pV = \text{const} \qquad (12.38)$$

Zur grafischen Darstellung wird das p-V-Diagramm benutzt. In dieser Darstellung sind die *Isotherme* idealer Gase Hyperbeln (Abb. 12.4).

Dargestellt sind zwei Isotherme T_2 und T_1, mit $T_2 > T_1$.

Ändert ein Gas seinen Zustand von 1 nach 2, so folgt für die Volumenarbeit durch Integration von (12.37)

$$W_{12} = -RT \ln \frac{V_2}{V_1} \qquad (12.39)$$

Ist $V_2 > V_1$, expandiert das Gas also, so ist W_{12} negativ und wird vom Gas geleistet. Anschaulich entspricht die Volumenarbeit der schraffierten Fläche im p-V-Diagramm.

Isochore Zustandsänderungen

Bei isochoren Zustandsänderungen wird das Volumen konstant gehalten. Diese Änderung wird im p-V-Diagramm durch eine senkrechte Strecke zwischen Ausgangs- und Endpunkt dargestellt (gestrichelte Linie in Abb. 12.4). Bei einer isochoren Änderung muss dem System Wärme zugeführt werden, da sich die Temperatur erhöht. Wegen des konstanten Volumens wird keine Volumenarbeit abgegeben.

$$W_{12} = 0 \qquad (12.40)$$

Führt man von außen Wärmeenergie zu, wird nach (12.35) die Temperatur erhöht, und die gesamte zugeführte Wärmeenergie

steht zur Vergrößerung der inneren Energie ΔU und somit zur Temperaturerhöhung zur Verfügung. Für 1 mol gilt:

$$\Delta Q = C_{mv}\Delta T \tag{12.41}$$

und somit

$$\Delta U = C_{mv}\Delta T \tag{12.42}$$

Isobare Zustandsänderungen

Bei isobaren Zustandsänderungen wird der Druck konstant gehalten. Diese Änderung wird im p-V-Diagramm durch eine waagerechte Strecke zwischen Ausgangs- und Endpunkt dargestellt (Abb. 12.4). Bei einer isobaren Änderung muss dem System Wärme zugeführt werden, da sich die Temperatur erhöht und außerdem Volumenarbeit, deren Größe dem schraffierten Rechteck entspricht, abgegeben wird.

$$W_{12} = -p \cdot (V_2 - V_1) \tag{12.43}$$

Da die Volumenarbeit $\neq 0$ ist, wird die Temperaturerhöhung bei zugeführter Wärmeenergie anders sein als bei isochoren Prozessen. Auch hier werde 1 mol betrachtet:

$$\Delta Q = C_{mp}\Delta T \tag{12.44}$$

mit (12.35) und (12.37) gilt:

$$\Delta Q = C_{mv}\Delta T + p\Delta V \tag{12.45}$$
$$p\Delta V = R\Delta T \tag{12.46}$$
$$C_{mp}\Delta T = C_{mv}\Delta T + R\Delta T \tag{12.47}$$
$$C_{mp} - C_{mv} = R \tag{12.48}$$

Isentrope Zustandsänderung

Isentrope oder auch adiabatische Zustandsänderungen laufen ohne Wärmeaustausch mit der Umgebung ab. Für diese Zustandsänderungen gelten die *Poisson'schen Gleichungen*:

$$pV^{\chi} = \text{const} \tag{12.49}$$
$$TV^{\chi-1} = \text{const} \tag{12.50}$$
$$\frac{T^{\chi}}{p^{\chi-1}} = \text{const} \tag{12.51}$$

Die Größe χ heißt *Isentropenexponent*, hängt von der Gasart ab, und ist immer > 1

$$\chi = \frac{C_{mp}}{C_{mv}} \tag{12.52}$$

In der Tab. 12.3 ist für einige Gase der Isentropenexponent aufgeführt.

Im p-V-Diagramm verlaufen Isentrope steiler als Isotherme und schneiden daher die Isotherme (siehe Abb. 12.4).

Tab. 12.3 Isentropenexponenten

Gas	χ
Helium	1,67
Argon	1,67
Wasserstoff	1,41
Sauerstoff	1,40
Luft	1,40
Kohlendioxid	1,30
Methan	1,32
Ammoniak	1,31

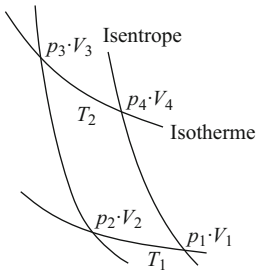


Abb. 12.5 Carnot-Prozess

12.4.5 Kreisprozesse

Unter Kreisprozessen wird eine Abfolge von Zustandsänderungen verstanden, die nach beliebigen Zwischenzuständen wieder zum Ausgangszustand zurückkehrt. Als Beispiel soll hier der Carnot-Prozess vorgestellt werden.

Carnot'scher Kreisprozess

Beim Carnot-Prozess wird Wärme in mechanische Arbeit umgewandelt.

Der Beginn sei im Punkt 1 auf der Isothermen T_1 bei V_1 und p_1. Es werden nacheinander die folgenden Schritte durchgeführt (Abb. 12.5):

1. Schritt isotherme Kompression auf p_2, V_2
2. Schritt isentrope Kompression auf T_2, p_3, V_3
3. Schritt isotherme Expansion auf p_4, V_4
4. Schritt isentrope Expansion auf T_1, p_1, V_1, also zum Ausgangspunkt.

Bei den einzelnen Schritten werden in unterschiedlicher Weise Wärme zu- oder abgeführt und Arbeit geleistet oder ins System gesteckt. Hier soll nur die gesamte Bilanz angegeben werden:

Die Nutzarbeit des Kreisprozesses für einen Umlauf entspricht der im p-V-Diagramm eingeschlossenen Fläche. Um unterschiedliche Kreisprozesse vergleichen zu können, wird der *thermische Wirkungsgrad* η verwendet. Dieser ist definiert durch

$$\eta = \frac{W}{Q_{\text{zu}}} \tag{12.53}$$

Für den Carnot-Prozess gilt:

$$\eta_C = \frac{T_2 - T_1}{T_2} \tag{12.54}$$

Tab. 12.4 Wärmeleitfähigkeit λ ausgewählter Stoffe

Material (feste Stoffe)	λ in W/m K	Material (Flüssigkeiten)	λ in W/m K	Material (Gase)	λ in W/m K
Aluminium	221	Wasser	0,6	Luft	0,026
Eisen	67	Benzin	0,14	Kohlendioxid	0,015
Kupfer	393	Transformatorenöl	0,13	Wasserdampf	0,031
Normalbeton	2,1	Xylol	0,13	Argon	0,016
Eis	2,2	Glyzerin	0,28		
Glas	0,8	Quecksilber	9,2		
Mineralfaser	0,04	Aceton	0,17		

Es lässt sich zeigen, dass der Carnot-Prozess den theoretisch höchsten thermischen Wirkungsgrad aller denkbaren Kreisprozesse besitzt. Er ist immer kleiner als 1 und nur abhängig von den Temperaturen T_2 und T_1.

Beispiel

Welchen Wirkungsgrad besitzt eine Maschine, die einen Carnot-Prozess zwischen den Temperaturen 100 °C und 0 °C durchführt?

Lösung:

obere Temperatur $T_2 = (100 + 273{,}15)\,\mathrm{K} = 373{,}15\,\mathrm{K}$

untere Temperatur $T_1 = (0 + 273{,}15)\,\mathrm{K} = 273{,}15\,\mathrm{K}$

Wirkungsgrad $$\eta = \frac{373{,}15 - 273{,}15}{373{,}15} \cdot 100\,\% = 26{,}8\,\%$$ ◀

12.5 Wärmeübertragung

Wärme kann grundsätzlich nur vom System höherer Temperatur zum System mit niedrigerer Temperatur übertragen werden. Wärme kann durch Wärmeleitung, Wärmeströmung oder Wärmestrahlung transportiert werden.

12.5.1 Wärmeleitung

Wärmeleitung findet durch Energieübertragung ohne Materietransport statt. Damit dieser Energietransport geschehen kann, muss eine Koppelung zwischen benachbarten Atomen vorhanden sein. Beispiel: Wärmetransport durch eine feste Wand. Hier können benachbarte Atome durch Schwingung Energie weitergeben. Als Maß dient die *Wärmestromdichte* q_{th}

$$q_{\mathrm{th}} = \frac{\Delta Q}{A\,\Delta t} \qquad \begin{array}{c|c|c|c} q_{\mathrm{th}} & \Delta Q & A & \Delta t \\ \hline \mathrm{W/m^2} & \mathrm{W\,s} & \mathrm{m^2} & \mathrm{s} \end{array} \qquad (12.55)$$

Wenn angenommen wird, dass der Wärmetransport nur in einer Richtung, die als x-Richtung bezeichnet wird, stattfinden kann, so gilt für die *Wärmestromdichte*:

$$q_{\mathrm{th}} = -\lambda \frac{\Delta \vartheta}{\Delta x} \qquad (12.56)$$

Die Größe l ist die *Wärmeleitfähigkeit* des Kontaktes in der Einheit $\frac{\mathrm{W}}{\mathrm{m\,K}}$, $\Delta\vartheta$ die Temperaturdifferenz. Die Wärmeleitfähigkeit ist temperaturabhängig. In der Tab. 12.4 sind einige Werte bei 20 °C angegeben.

12.5.2 Wärmeströmung

Wärmeströmung oder auch *Konvektion* ist stets mit einem Materietransport verbunden und findet in Flüssigkeiten oder Gasen statt. Beispiel: Zentralheizung. Die mathematische Behandlung ist sehr kompliziert und kann hier nicht behandelt werden.

12.5.3 Wärmestrahlung

Unter Wärmestrahlung wird die Wärmeübertragung durch elektromagnetische Strahlung verstanden. Jeder Körper sendet und empfängt Wärme durch Strahlung, wobei die Absorption und Emission vom Material und der Oberflächenbeschaffenheit abhängt. Dies wird durch den Emissionsgrad ε angegeben. ε ist eine Zahl < 1. Die gesamte Leistungsbilanz eines Körpers der ebenen Fläche A und der Temperatur T_1 in einer Umgebung der Temperatur T_2 wird durch das *Stefan-Boltzmann-Gesetz* beschrieben:

$$S = A\varepsilon\sigma(T_2^4 - T_1^4) \qquad (12.57)$$

mit der Konstanten

$$\sigma = 5{,}67 \cdot 10^{-8}\,\frac{\mathrm{W}}{\mathrm{m^2\,K^4}} \qquad (12.58)$$

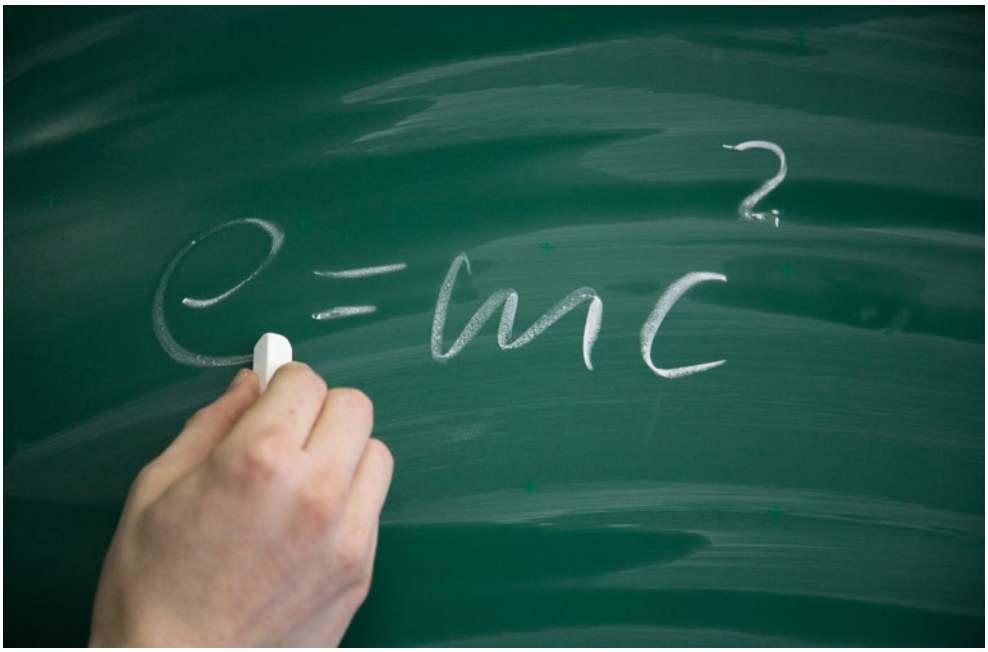
(siehe auch Abschn. 16.4).

Teil II

Schwingungen

13

Horst Steffen



13.1 Freie ungedämpfte harmonische Schwingungen 222
13.2 Gedämpfte Schwingungen 223
13.3 Erzwungene Schwingungen 223
13.4 Überlagerung harmonischer Schwingungen 224

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_13

Schwingungen sind *periodische* Vorgänge sowohl mechanischer (z. B. Masse an einer Feder) als auch elektromagnetischer Systeme (z. B. Schwingkreis). Kennzeichnend ist, dass Energie periodisch umgewandelt wird. Die Zeit für eine Periode wird durch die *Schwingungszeit* T angegeben. Die Anzahl der Perioden pro Zeiteinheit wird *Frequenz* f genannt:

$$f = \frac{1}{T} \tag{13.1}$$

Oft wird auch die *Kreisfrequenz* ω verwendet.

$$\textit{Kreisfrequenz} \quad \omega = 2\pi f = \frac{2\pi}{T} \tag{13.2}$$

Wird die Auslenkung eines schwingenden Systems mit y bezeichnet, so gilt, dass sich nach einer Periode ein bestimmter Schwingungszustand wiederholt. Bei einer ungedämpften Schwingung gilt:

$$y(t) = y(t+T) \tag{13.3}$$

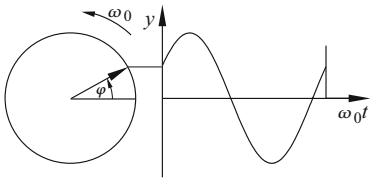


Abb. 13.1 Harmonische Schwingung

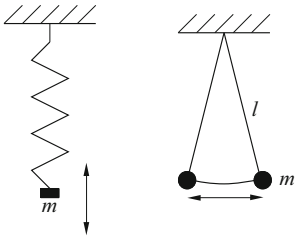


Abb. 13.2 Feder- und Fadenpendel

13.1 Freie ungedämpfte harmonische Schwingungen

Es gibt eine Vielzahl von möglichen Schwingungen, deren Auslenkung sich durch eine Sinus-Funktion beschreiben lässt. Diese Schwingungen werden *harmonische Schwingungen* genannt (Abb. 13.1). Da eine Sinus-Funktion sich als Projektion einer gleichmäßigen Kreisbewegung darstellen lässt (siehe Abschn. 6.4 „Graphen der trigonometrischen Funktionen“), kann der Winkel φ durch $\omega_0 \cdot t$ ersetzt werden:

$$y(t) = \hat{y}\sin(\omega_0 t) \tag{13.4}$$

$y(t)$ ist die *momentane Auslenkung* oder *Augenblickswert*, $\hat{y}$ die *Amplitude* oder *maximale Auslenkung*, ω_0 ist die *Kreisfrequenz* der ungedämpften Schwingung. Diese wird auch *Eigenfrequenz* genannt. Bei (13.4) ist die Auslenkung y zur Zeit $t = 0$ ebenfalls $= 0$

$$y(0) = 0$$

Ist jedoch der Nulldurchgang der Schwingung nicht zur Zeit $t = 0$, so liegt eine Phasenverschiebung um einen Phasenwinkel φ vor, und die Schwingung lässt sich durch

$$y(t) = \hat{y}\sin(\omega_0 t + \varphi) \tag{13.5}$$

darstellen. Der Phasenwinkel kann positiv oder negativ sein. Bei positivem Phasenwinkel wird die Sinuskurve nach links verschoben.

Die Schwingungszeit T bzw. die Frequenz f hängt je nach schwingungsfähigem System von den unterschiedlichsten Systemgrößen ab. Wird z. B. eine Masse m an einer Feder mit der Federkonstanten c angehängt, so gilt bei diesem *Federpendel* für die Schwingungszeit (Abb. 13.2):

$$\textit{Federpendel} \quad T = 2\pi\sqrt{\frac{m}{c}} \tag{13.6}$$

T	m	c
s	kg	N/m

Wird eine Masse m an einem Faden der Länge l angehängt und zu Schwingungen mit kleiner Auslenkung angeregt, so gilt für dieses *Fadenpendel* oder auch *mathematisches Pendel*:

$$\textit{Fadenpendel} \quad T = 2\pi\sqrt{\frac{l}{g}} \tag{13.7}$$

T	l	g
s	m	$\mathrm{m\,s^{-2}}$

g ist dabei die *Erdbeschleunigung*, die Schwingungszeit ist bei diesem Pendel unabhängig von der Masse m.

Beispiel

Eine Pendeluhr geht in 24 Stunden um 1 Stunde nach. Die Pendellänge ist $l_1 = 50\,\mathrm{cm}$. Welche Länge l_2 muss das Pendel haben, damit die Uhr richtig geht? Das Pendel wird als mathematisches Pendel angenommen.

Lösung:

$$\frac{T_1}{T_2} = \frac{2\pi\sqrt{l_1}}{2\pi\sqrt{l_2}} = \sqrt{\frac{l_1}{l_2}}$$

$$l_2 = l_1\frac{T_2^2}{T_1^2} = 50\,\mathrm{cm}\cdot\left(\frac{23\,\mathrm{h}}{24\,\mathrm{h}}\right)^2 = 45{,}9\,\mathrm{cm} \qquad \blacktriangleleft$$

Frequenzen elektromagnetischer Schwingungen werden durch L und C festgelegt. So gilt für einen Schwingkreis:

$$\textit{elektrischer Schwingkreis} \quad T = 2\pi\sqrt{LC} \tag{13.8}$$

Allen diesen Beispielen gemeinsam ist aber, dass die mathematische Beschreibung der Schwingung durch (13.5) gegeben ist. Kennzeichnung jeder ungedämpften Schwingung ist, dass die Amplitude $\hat{y}$ zeitlich konstant bleibt, dem System wird keine Energie entzogen.

Beispiel

Eine Masse $m = 500\,\text{g}$ hängt an einem Faden der Länge $l = 2\,\text{m}$. Mit welcher Frequenz f schwingt die Masse?

Lösung:

$$T = 2\pi\sqrt{\frac{2\,\text{m}}{9{,}81\,\text{m}\,\text{s}^{-2}}} = 2{,}84\,\text{s}$$

$$f = \frac{1}{T} = \frac{1}{2{,}84\,\text{s}} = 0{,}35\,\text{Hz} \quad \blacktriangleleft$$

13.2 Gedämpfte Schwingungen

Wird einem schwingungsfähigen System während der Schwingung Energie entzogen, so verringert sich die Amplitude im Lauf der Zeit. Dieser Energieverlust kann in der Mechanik z. B. durch Reibung, bei elektromagnetischen Schwingungen durch ohmsche Verluste oder Ummagnetisierungsverluste verursacht werden. Wegen dieser Verluste nimmt die Amplitude ab und die Ausschläge werden immer kleiner. Sind die Verluste proportional zur Geschwindigkeit (oder zum elektrischen Strom), so lässt sich die Schwingung durch folgende Gleichung wiedergeben (Abb. 13.3):

$$y = \hat{y}e^{-\frac{t}{\tau}}\sin(\omega t + \varphi) \tag{13.9}$$

oder

$$y = \hat{y}e^{-\delta t}\sin(\omega t + \varphi) \tag{13.10}$$

mit

$$\delta = \frac{1}{\tau} \tag{13.11}$$

Die für die Abnahme der Amplitude maßgebende e-Funktion enthält eine für die Dämpfung charakteristische Größe τ, die auch als *Zeitkonstante* des Systems bezeichnet wird und angibt, zu welchem Zeitpunkt die Amplitude auf den Wert $\frac{1}{e} = 0{,}3678 \approx 37\,\%$ der Ausgangsamplitude gefallen ist.

Die Frequenz ω der gedämpften Schwingung ist kleiner als die Frequenz ω_0 der freien Schwingung.

$$\omega = \sqrt{\omega_0^2 - \delta^2} \tag{13.12}$$

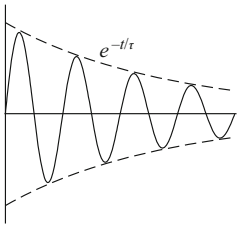


Abb. 13.3 Gedämpfte Schwingung

Diese Abweichungen sind aber für kleine Dämpfungen vernachlässigbar.

Beispiel

Ein Federpendel mit einer Federkonstanten $c = 30\,\text{N/m}$, an dem eine Masse m von $2\,\text{kg}$ hängt, führt gedämpfte Schwingungen mit einer Anfangsamplitude $\hat{y} = 10\,\text{cm}$ aus. Die Zeitkonstante τ ist $2\,\text{s}$, der Phasenwinkel φ ist $0°$. Welche Auslenkung liegt nach einer Zeit von $t = 5\,\text{s}$ vor?

Lösung:

Kreisfrequenz $$\omega = \sqrt{\frac{c}{m}} = \sqrt{\frac{30\,\text{N/m}}{2\,\text{kg}}} = 3{,}87\,\text{s}^{-1} \tag{13.13}$$

Auslenkung $$y = 5\,\text{cm}\cdot e^{-5/2}\cdot\sin(3{,}87\cdot 5) \tag{13.14}$$

$$= 0{,}169\,\text{cm} \quad \blacktriangleleft$$

13.3 Erzwungene Schwingungen

Bei einer erzwungenen Schwingung greift von außen über eine Koppelung eine periodische Kraft mit einer Frequenz Ω an. Das schwingungsfähige System, welches aufgrund seiner Bauart im ungedämpften Fall mit der Eigenfrequenz ω_0 schwingen würde, wird dann zu erzwungenen Schwingungen mit der Frequenz der angreifenden Kraft angeregt. Dabei ist auch mit einer Phasenverschiebung zwischen angreifender Kraft und System zu rechnen. Für das System gilt dann:

$$y = \hat{y}\sin(\Omega t + \varphi) \tag{13.15}$$

Die Amplitude $\hat{y}$ und die Phasenverschiebung hängen von der Frequenz der angreifenden Kraft ab. Für ein Federpendel der Masse m, an der eine Kraft F angreift, gilt für die Amplitude:

$$\hat{y} = \frac{F/m}{\sqrt{(\omega_0^2 - \Omega^2)^2 + (2\delta\Omega)^2}} \tag{13.16}$$

und den Phasenwinkel

$$\tan\varphi = \frac{2\delta\omega_0}{(\omega_0^2 - \Omega^2)} \tag{13.17}$$

Die Amplitudenfunktion hat bei kleiner Dämpfung dann ein Maximum, wenn die Erregerfrequenz und die Eigenfrequenz übereinstimmen ($\Omega/\omega_0 = 1$). Dieses Verhalten wird als *Resonanz* bezeichnet. Der Kurvenverlauf für verschiedene Dämpfungen ist in Abb. 13.4 gezeigt. Aufgetragen ist die Amplitude als Funktion von Ω/ω_0. Allerdings verschiebt sich bei größeren Dämpfungen die Lage des Maximums zu kleineren Erregerfrequenzen.

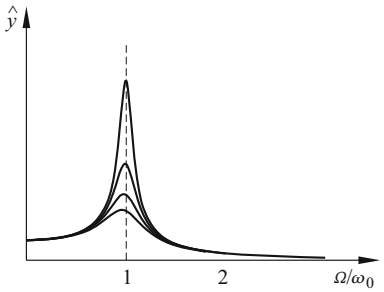


Abb. 13.4 Resonanzkurve

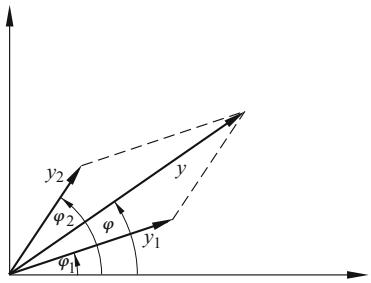


Abb. 13.6 Zeigerdiagramm

13.4 Überlagerung harmonischer Schwingungen

Werden Schwingungen überlagert, so müssen verschiedene Möglichkeiten betrachtet werden: Die Schwingungen können gleiche oder unterschiedliche Frequenz und gleiche oder senkrecht zueinander liegende Schwingungsrichtungen haben. Diese Fälle sollen für sinusförmige (harmonische) Schwingungen untersucht werden.

13.4.1 Schwingungsrichtung parallel zueinander

Gleiche Frequenz

Betrachtet werden zwei Schwingungen, die unterschiedliche Phasenwinkel haben können, aber mit gleicher Frequenz und gleicher Schwingungsrichtung schwingen. Die Schwingungen werden durch die Formeln (Abb. 13.5)

$$y_1(t) = \hat{y}_1 \sin(\omega t + \varphi_1) \tag{13.18}$$

$$y_2(t) = \hat{y}_2 \sin(\omega t + \varphi_2) \tag{13.19}$$

beschrieben.

Die grafische Darstellung kann entweder im $y(t)$-Diagramm oder im *Zeigerdiagramm* erfolgen. Letztere wird in der Elektrotechnik oft verwendet (Abb. 13.6). Unter einem *Zeiger* wird ein zweidimensionaler Pfeil verstanden, dessen Länge der Amplitude und dessen Winkel dem Phasenwinkel entspricht. Dieser Zeiger rotiert mit der Frequenz ω um den Koordinatenursprung. Durch die Überlagerung (Addition) zweier Schwingungen entsteht wieder eine harmonische Schwingung gleicher Frequenz.

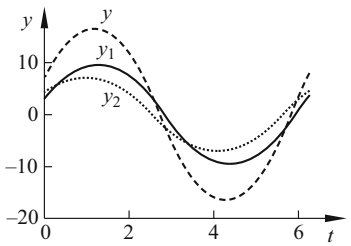


Abb. 13.5 Addition von Sinusschwingungen

Die Berechnung der Amplitude und der Phasenlage erfolgt anhand des Zeigerdiagramms.

Mit dem Kosinussatz (siehe Abschn. 6.7) gilt für die Amplitude:

$$\hat{y} = \sqrt{\hat{y}_1^2 + 2\hat{y}_1\hat{y}_2\cos(\varphi_2 - \varphi_1) + \hat{y}_2^2} \tag{13.20}$$

und für den Phasenwinkel:

$$\tan\varphi = \frac{\hat{y}_1\sin\varphi_1 + \hat{y}_2\sin\varphi_2}{\hat{y}_1\cos\varphi_1 + \hat{y}_2\cos\varphi_2} \tag{13.21}$$

Aus diesen Gleichungen lassen sich einige Spezialfälle ableiten:

1) gleiche Amplitude, Phasendifferenz $= \pi, 3\pi$:

$$\hat{y} = \sqrt{\hat{y}_1^2 - 2\hat{y}_1\hat{y}_2 + y_2^2} \tag{13.22}$$

$$\hat{y} = \sqrt{(\hat{y}_1 - \hat{y}_2)^2} = 0 \tag{13.23}$$

also Auslöschung der resultierenden Schwingung.

2) Phasendifferenz $= 0$:

$$\hat{y} = \sqrt{\hat{y}_1^2 + 2\hat{y}_1\hat{y}_2 + \hat{y}_2^2} \tag{13.24}$$

$$\hat{y} = \sqrt{(\hat{y}_1 + \hat{y}_2)^2} = \hat{y}_1 + \hat{y}_2 \tag{13.25}$$

also Addition der Amplituden.

Geringer Frequenzunterschied

Sind die Frequenzen zweier sich überlagernder Schwingungen fast gleich, treten *Schwebungen* auf. Hierbei verändert sich die Amplitude der resultierenden Schwingung periodisch. Im folgenden wird angenommen, dass beide Ausgangsamplituden gleich sind.

$$y_1(t) = \hat{y}\sin(\omega_1 t) \tag{13.26}$$

$$y_2(t) = \hat{y}\sin(\omega_2 t) \tag{13.27}$$

$$y_{\text{res}}(t) = y_1(t) + y_2(t) \tag{13.28}$$

$$y_{\text{res}} = \hat{y}[\sin(\omega_1 t) + \sin(\omega_2 t)] \tag{13.29}$$

Mit

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cdot\cos\frac{\alpha-\beta}{2} \tag{13.30}$$

folgt aus (13.29)

$$y_{\text{res}}(t) = 2\hat{y}\sin\left(\frac{\omega_1+\omega_2}{2}t\right)\cdot\cos\left(\frac{\omega_1-\omega_2}{2}t\right) \tag{13.31}$$

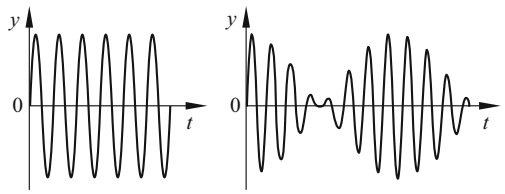


Abb. 13.7 Schwebung

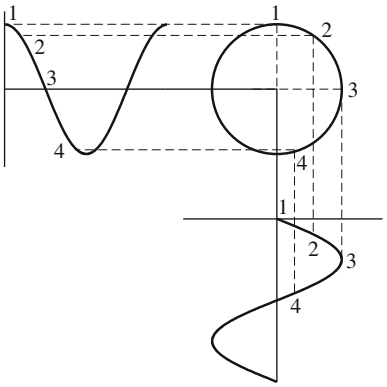


Abb. 13.8 Verfahren zur punktweisen Konstruktion von Lissajous-Figuren ($\varphi = 90°$)

Bei geringen Frequenzdifferenzen können die Argumente der trigonometrischen Funktionen mit $\Delta\omega = \omega_1 - \omega_2$ umgeschrieben werden:

$$y_{\text{res}} = 2\hat{y} \sin \omega_{\text{res}} t \cdot \cos\left(\frac{\Delta\omega}{2} t\right) \tag{13.32}$$

Die resultierende Schwingung ist ebenfalls eine harmonische Schwingung mit der Frequenz

$$\omega_{\text{res}} = \frac{\omega_1 + \omega_2}{2} \tag{13.33}$$

aber die Amplitude ändert sich periodisch zwischen 0 und $2\hat{y}$. Während einer Periode der Kosinus-Funktion ändert sich die Amplitude zweimal von 0 auf $2\hat{y}$, somit ist die *Schwebungsfrequenz*

$$\omega_S = 2\frac{\omega_1 - \omega_2}{2} = \Delta\omega \tag{13.34}$$

Dieses Verhalten ist in Abb. 13.7 wiedergegeben, hierbei sind zwei Schwingungen überlagert, deren Frequenzen sich um 15 % unterscheiden.

13.4.2 Schwingungsrichtung senkrecht zueinander

Werden zwei Schwingungen mit ganzzahligem Frequenzverhältnis, die senkrecht zueinander schwingen, überlagert, so entstehen geschlossene Figuren (s. Abb. 13.8).

Lissajous-Figuren

Die beiden Schwingungsrichtungen seien x- und y-Richtung. Die Schwingungen sind durch

$$x(t) = \hat{x} \cdot \sin(\omega_x t)$$

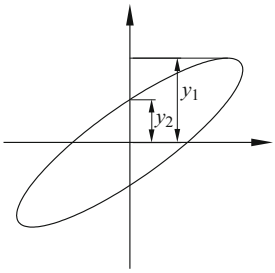


Abb. 13.9 Bestimmung des Phasenwinkels

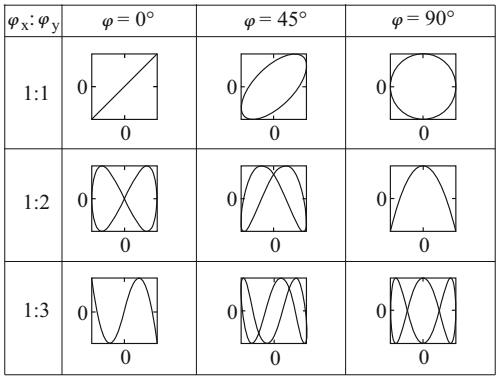


Abb. 13.10 Lissajous-Figuren

und

$$y(t) = \hat{y} \cdot \sin(\omega_y t + \varphi) \tag{13.35}$$

gegeben. Die Lissajous-Figuren können im Prinzip punktweise konstruiert werden. Für den Fall gleicher Frequenzen und Phasendifferenz $\varphi = 90°$ ist das Verfahren in Abb. 13.8 dargestellt.

Auf der x- und auf der y-Kurve werden die Auslenkungen zu gleichen Zeitpunkten, hier mit $1, \ldots, 4$ bezeichnet, abgemessen und in einem rechtwinkligen Koordinatensystem aufgetragen. Für die vorgegebene Phasendifferenz ergibt sich als Lissajous-Figur ein Kreis. Sind die Amplituden nicht gleich, ist aber die Phasendifferenz 90°, so ergeben sich Ellipsen, deren Hauptachsen parallel zu den Koordinatenachsen liegen. Ist die Phasendifferenz nicht 90°, so ergeben sich ebenfalls Ellipsen, allerdings liegen deren Hauptachsen nicht mehr parallel zu den Koordinatenachsen. Diese Lissajous-Figuren lassen sich in der Elektrotechnik mit Hilfe eines Oszilloskops darstellen; sie dienen zur Messung des Phasenwinkels zwischen zwei Spannungen. Werden bei einer gemessenen Ellipse nach Abb. 13.9 die beiden Werte y_1 und y_2, also einmal das Maximum in y-Richtung, zum anderen der Schnittpunkt mit der y-Achse, gemessen, so lässt sich hieraus der Phasenwinkel φ bestimmen zu:

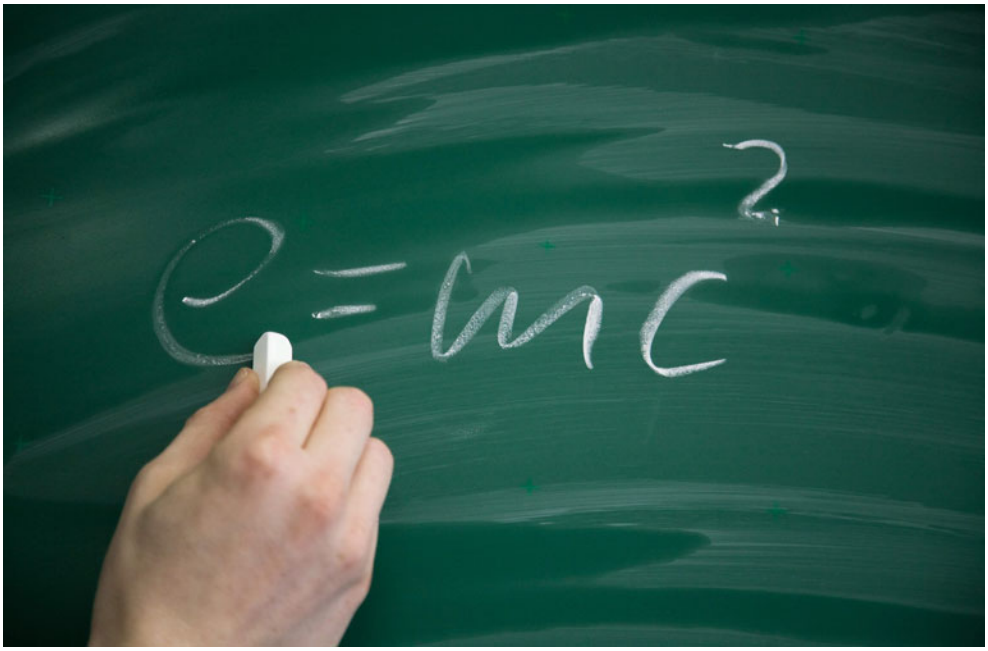
$$\varphi = \arcsin\frac{y_2}{y_1} \tag{13.36}$$

In Abb. 13.10 sind für die Frequenzverhältnisse 1 : 1, 1 : 2 und 1 : 3 für verschiedene Phasenwinkel die *Lissajous-Figuren* angegeben.

Wellen

14

Horst Steffen



14.1 Harmonische Wellen 228
14.2 Huygens'sches Prinzip 230
14.3 Dopplereffekt 231

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_14

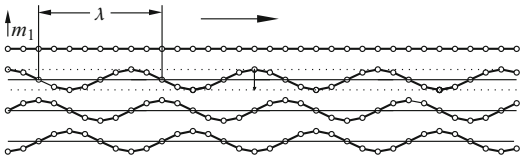


Abb. 14.1 Wellenausbreitung in einer linearen Kette

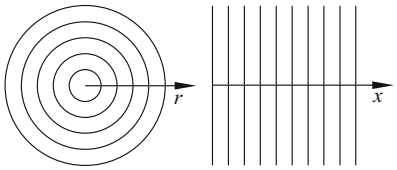


Abb. 14.2 Wellentypen

Wenn schwingungsfähige Systeme miteinander so verbunden sind, dass sich die Schwingung von einem System zum nächsten System übertragen kann, so setzt sich eine an einem System angeregte Schwingung im Raum fort, und man spricht von einer *Welle*. Dies soll im Abb. 14.1 veranschaulicht werden. Dabei ist es zur Ausbreitung von Wellen notwendig, dass die einzelnen Systeme durch gegenseitige Rückstellkräfte gekoppelt sind.

Dargestellt sind zwei Momentaufnahmen einer Reihe von Massen, die durch Federn verbunden sind. Die Masse m_1 wird angestoßen und schwingt mit der Frequenz f um ihre Ruhelage. Wegen der Koppelung durch die Federn setzt sich diese Schwingung fort, und es entsteht eine Welle, die sich hier nach rechts ausbreitet. Die einzelnen Massen schwingen um ihre Ruhelage, bewegen sich also nicht in Ausbreitungsrichtung der Welle. Es findet kein Materietransport, sondern nur Energietransport statt. Wenn die Schwingungsrichtung der Massen senkrecht zur Ausbreitungsrichtung der Welle ist, so wird dieser Wellentyp *Transversalwelle* genannt. Im Abb. 14.1 ist mit der Größe λ der Abstand zwischen zwei gleichen Schwingungszuständen eingetragen. Dieser Abstand ist die *Wellenlänge* λ. Ist dagegen die Schwingungsrichtung der einzelnen Massen parallel zur Ausbreitungsrichtung der Welle, so liegt eine *Longitudinalwelle* vor. Bei einer Longitudinalwelle ändert sich in Ausbreitungsrichtung der Welle die Dichte der schwingungsfähigen Systeme. In Abb. 14.1 ist nur eine einzige Reihe von schwingungsfähigen Massen gezeichnet. Werden allerdings viele Reihen gleichzeitig angeregt, liegt eine räumliche Wellenausbreitung vor.

Werden gleiche Schwingungszustände in benachbarten Ketten miteinander verbunden, so erhält man eine zusammenhängende Fläche, die *Wellenfront* genannt wird. Die Form der Wellenfront hängt von der Form des Wellenerregers ab. In Abb. 14.2 sind zwei verschiedene Wellenfronten gezeichnet, links die Fronten einer *Kugelwelle*, rechts die Fronten einer *ebenen Welle*. Da die Wassermoleküle in beiden Typen nur vertikal schwingen, liegt hier in beiden Fällen eine Transversalwelle vor.

14.1 Harmonische Wellen

14.1.1 Ausbreitung

Wie bei den harmonischen Schwingungen werden die Wellen *harmonisch* genannt, deren mathematische Beschreibung durch eine Sinus-Funktion möglich ist. Im Gegensatz zu den Schwingungen, bei denen ja nur eine zeitliche Änderung der Auslenkung y zu betrachten war, ist bei Wellen auch eine räumliche Abhängigkeit der Auslenkung zu berücksichtigen. Die Welle in Abb. 14.1 hat sich nach der Schwingungszeit $T = \frac{1}{f}$ der einzelnen Systeme um die Wellenlänge λ nach rechts fortbewegt. Hiermit lässt sich die Ausbreitungsgeschwindigkeit c definieren:

Ausbreitungsgeschwindigkeit

$$c = \frac{\lambda}{T} \tag{14.1}$$

$$c = \lambda \cdot f \tag{14.2}$$

Wird die Masse $m_1$zu einer Sinusschwingung $y = \hat{y} \cdot \sin(\omega t)$ angeregt, so werden auch alle anderen Massen sinusförmig schwingen, allerdings zeitversetzt gegen m_1. Eine Masse, die um x von m_1 entfernt ist, wird um die Zeit $t_1 = \frac{x}{c}$ später denselben Schwingungszustand erreichen. Es gilt:

$$y = \hat{y} \cdot \sin[\omega(t - t_1)] \tag{14.3}$$

$$y = \hat{y} \cdot \sin\left[\frac{2\pi}{T}\left(t - \frac{x}{c}\right)\right] \tag{14.4}$$

$$y = \hat{y} \cdot \sin\left[2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)\right] \tag{14.5}$$

Gleichung (14.5) beschreibt die zeitliche und räumliche Ausbreitung einer ebenen harmonischen Welle. Die Amplitude $\hat{y}$ bleibt konstant. Die Schwingungsgleichung kann durch Einführung der *Wellenzahl* k umgeschrieben werden:

$$\textit{Wellenzahl} \quad k = \frac{2\pi}{\lambda} \tag{14.6}$$

$$y = \hat{y} \cdot \sin(\omega t - kx) \tag{14.7}$$

In (14.5) und (14.7) bedeutet ein negatives Zeichen in der Klammer eine Ausbreitung der Welle in positiver x-Richtung.

Bei einer *Kugelwelle* ändert sich die Amplitude mit dem Abstand vom Erregerort. Dieser Wellentyp wird durch folgende Gleichungen beschrieben.

$$y = \frac{\hat{y}_1}{r} \cdot \sin\left[2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)\right] \tag{14.8}$$

$$y = \frac{\hat{y}_1}{r} \cdot \sin(\omega t - kx) \tag{14.9}$$

$\hat{y}_1$ist die Amplitude für den Abstand $r = 1$.

Tab. 14.1 Phasengeschwindigkeit

Wellentyp	Phasengeschwindigkeit
Longitudinalwellen in Gasen, Schallwellen in Gasen	$c = \sqrt{\frac{\chi p}{\rho}}$
Longitudinalwellen in Flüssigkeiten	$c = \sqrt{\frac{K}{\rho}}$
Longitudinalwellen in Festkörpern	$c = \sqrt{\frac{E}{\rho}}$
Elektromagnetische Wellen in Materie	$c = \frac{1}{\sqrt{\varepsilon_r \varepsilon_0 \mu_r \mu_0}}$

χ Isentropenexponent, p Druck, ρ Massendichte, K Kompressionsmodul, E Elastizitätsmodul, ε_0 Elektrische Feldkonstante, ε_r Permittivitätszahl, μ_0 Permeabilität, μ_r relative Permeabilität

Beispiel

Eine harmonische Welle habe eine Phasengeschwindigkeit $c = 300\,\mathrm{m/s}$ und eine Frequenz von 100 Hz. Zur Zeit $t = 0$ wird sie mit einer Amplitude $\hat{y} = 10\,\mathrm{cm}$ angeregt.

a) Wie groß ist die Wellenlänge λ?
b) Wie groß ist die Auslenkung am Ort $x = 9\,\mathrm{m}$ nach einer Zeit von 0,15 s?

Lösung:

a) nach (14.1)

$$\lambda = \frac{c}{f} = \frac{300\,\mathrm{m\,s^{-1}}}{100\,\mathrm{s^{-1}}} = 3\,\mathrm{m} \tag{14.10}$$

b)

$$y = 10\,\mathrm{cm} \cdot \sin\left[2\pi\left(\frac{0{,}15\,\mathrm{s}}{0{,}01\,\mathrm{s}} - \frac{9\,\mathrm{m}}{3\,\mathrm{m}}\right)\right]$$
$$y = 10\,\mathrm{cm} \cdot \sin[2\pi(15-3)]$$
$$= 10\,\mathrm{cm} \cdot \sin(24\pi) = 0\,\mathrm{cm}$$ ◀

Die Ausbreitungsgeschwindigkeit c, besser als *Phasengeschwindigkeit* bezeichnet, da zur Bildung der Wellenfronten Orte gleicher Phase miteinander verbunden werden, hängt vom Wellentyp und vom Medium, in dem sich die Welle bewegt, ab. In Tab. 14.1 sind einige Beispiele angegeben.

14.1.2 Interferenz

Unter Interferenz wird die Überlagerung von Wellen verstanden. Dabei gilt, dass bei nicht zu großen Amplituden der beteiligten Wellen sich jede Welle ungestört ausbreiten kann und die momentanen Auslenkungen addiert werden können. Dies wird das Verfahren der ungestörten *Superposition* genannt.

Interferenz von Wellen gleicher Frequenz

Zwei Wellen gleicher Frequenz, die sich in gleicher Richtung ausbreiten, können sich im allgemeinen durch ihre Amplitude und durch eine Verschiebung gleicher Schwingungszustände unterscheiden. Diese räumliche Verschiebung ist der *Gangunterschied* Δ. Die beiden Wellen sind nach (14.7) durch

$$y_1 = \hat{y}_1 \cdot \sin(\omega t - kx) \tag{14.11}$$

und

$$y_2 = \hat{y}_2 \cdot \sin\left(\omega t - kx + \frac{2\pi\Delta}{\lambda}\right) \tag{14.12}$$

gegeben. Die resultierende Welle lässt sich durch Addition der Teilwellen bestimmen. Sie hat dieselbe Frequenz und dieselbe Wellenlänge, damit auch dieselbe Phasengeschwindigkeit wie die beiden Ursprungswellen, aber eine andere Amplitude. Haben beide Wellen die gleiche Amplitude $\hat{y}_1 = \hat{y}_2 = \hat{y}_0$, so gilt:

resultierende Welle

$$y_{\mathrm{res}} = 2\hat{y}_0 \cos\left(\frac{\pi\Delta}{\lambda}\right) \sin\left(\omega t - kx + \frac{\pi\Delta}{\lambda}\right) \tag{14.13}$$

$$y_{\mathrm{res}} = \hat{y}_{\mathrm{res}} \sin\left(\omega t - kx + \frac{\pi\Delta}{\lambda}\right) \tag{14.14}$$

Amplitude

$$\hat{y}_{\mathrm{res}} = 2\hat{y}_0 \cos\left(\frac{\pi\Delta}{\lambda}\right) \tag{14.15}$$

Sonderfälle:

$$\text{a)} \quad \Delta = n\lambda \quad \text{mit } n = 0, 1, 2, \ldots \tag{14.16}$$

Die Phasenlage der resultierenden Welle ist gegenüber den Ausgangswellen unverändert, die Amplitude ist doppelt so groß wie die Ausgangsamplituden. Für gerade Werte von n hat die Kosinusfunktion den Wert 1, für ungerade Werte den Wert -1. In diesem Fall wird aber die Welle wegen des Summanden $\frac{\pi\Delta}{\lambda}$ in der Sinusfunktion um $\frac{\lambda}{2}$ verschoben, sodass die ursprüngliche Phasenlage beibehalten wird *(konstruktive Interferenz)*.

$$\text{b)} \quad \Delta = (2n+1)\frac{\lambda}{2} \quad \text{mit } n = 0, 1, 2, \ldots \tag{14.17}$$

In diesen Fällen wird $\hat{y}_{\mathrm{res}} = 0$, die Wellen löschen sich aus *(destruktive Interferenz.*

Stehende Wellen

Laufen die Wellen, die zur Interferenz gelangen, nicht in gleicher Richtung, sondern einander entgegen, so entstehen *stehende Wellen*. Dies kann z. B. durch Reflexion an einem Spiegel geschehen. Die beiden Wellen können bei gleicher Amplitude nach (14.7) durch

$$y_1 = \hat{y} \cdot \sin(\omega t - kx) \tag{14.18}$$

und

$$y_2 = \hat{y} \cdot \sin(\omega t + kx) \tag{14.19}$$

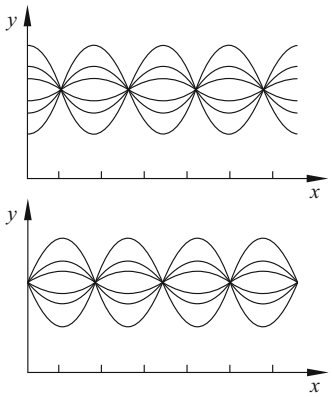


Abb. 14.3 Stehende Wellen

beschrieben werden. Auch hier ergibt sich die resultierende Welle durch Addition der Teilwellen. Unter Anwendung des Additionstheorems folgt:

$$y_{\text{res}} = 2\hat{y}\sin(\omega t)\cos(kx) \tag{14.20}$$

An den Stellen, an denen der Kosinusterm in (14.20) verschwindet, hat auch die resultierende Welle zu allen Zeiten keine Auslenkung. Dies ist der Fall für solche Orte, die um $\lambda/2$ voneinander entfernt sind. Der erste Nullpunkt hängt von der Art der Reflexion ab, an einem dichteren Medium liegt er in der Grenzfläche, sonst um $\lambda/4$ vor der Grenzfläche.

In den Punkten dazwischen ergibt sich je nach Laufzeit eine Auslenkung, die maximal den doppelten Wert der Ursprungsamplitude haben kann. In Abb. 14.3 ist oben die Reflexion an einem dichteren Medium, unten an einem dünneren Medium dargestellt. Dabei muss noch berücksichtigt werden, dass bei Reflexion an einem dichteren Medium ein Phasensprung von $\lambda/2$ eintritt.

14.2 Huygens'sches Prinzip

Die Wellenausbreitung kann nach dem *Huygens'schen Prinzip* (nach *Christian Huygens*) dadurch erklärt werden, dass jeder Punkt, der von einer Wellenfront getroffen wird, wiederum Ausgangspunkt einer sogenannten *Elementarwelle* wird. Diese ist, da es sich um punktförmige Erreger handelt, eine Kugelwelle (siehe Abschn. 14.1). Die neue Wellenfront entsteht als Einhüllende aller dieser Teilwellen. Dieses Prinzip gilt nicht nur für eine Wellenausbreitung in Medien, sondern ganz allgemein, auch für elektromagnetische Wellen, wie z. B. Licht. Mit Hilfe dieses Prinzips lassen sich Vorgänge wie Reflexion, Brechung und Beugung erklären (Abb. 14.4).

14.2.1 Reflexion

Eine ebene Wellenfront W, die sich mit der Geschwindigkeit c bewegt, erreicht zu einem bestimmten Zeitpunkt den Punkt A

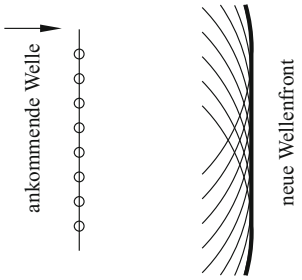


Abb. 14.4 Huygens'sches Prinzip

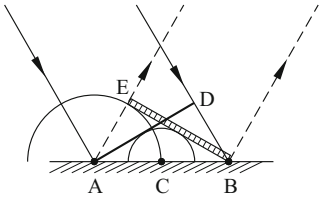


Abb. 14.5 Reflexion

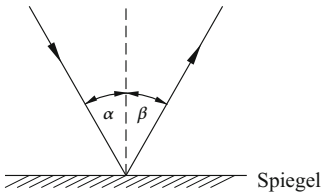


Abb. 14.6 Reflexionsgesetz

eines Spiegels (Abb. 14.5). Für die Strecke $\overline{DB}$ braucht die Wellenfront die Zeit t_1. Vom Punkt A geht eine Elementarwelle aus, die dann, wenn der Punkt B erreicht ist, den Radius $\overline{AE} = c \cdot t_1$ hat. In der Zeit $t_1/2$ ist der Punkt C in der Mitte zwischen A und B erreicht. Die Elementarwelle, die in C angeregt wird, hat den Radius $\overline{AE}/2$, wenn die Wellenfront nach der Zeit t_1 den Punkt B erreicht hat. Die Einhüllende dieser Elementarwellen ist mit W' bezeichnet. Da die Dreiecke AEB und ADB kongruent sind, folgt, dass beide Wellenfronten mit dem Spiegel denselben Winkel bilden. Üblicherweise werden nicht die Wellenfronten, sondern die darauf senkrecht stehenden *Wellennormalen* gezeichnet, um die Richtung einer Welle darzustellen. Auch werden die Winkel nicht zur Spiegelfläche, sondern zur Senkrechten auf der Spiegelfläche gemessen, wie in Abb. 14.6 zu ersehen ist. Es gilt das *Reflexionsgesetz*: *Einfallswinkel* = *Ausfallswinkel*

$$\alpha = \beta \tag{14.21}$$

14.2.2 Brechung

Wenn sich die Welle in Gebieten mit unterschiedlichen Phasengeschwindigkeiten ausbreitet, so ändert sich beim Durchgang durch die Grenzfläche die Richtung der Wellenfronten (Abb. 14.7). Im Medium 1 habe die Welle die Geschwindigkeit c_1 und laufe unter einem Winkel a_1 ein, im Medium 2 sei

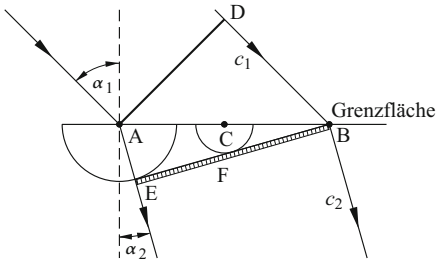


Abb. 14.7 Brechung

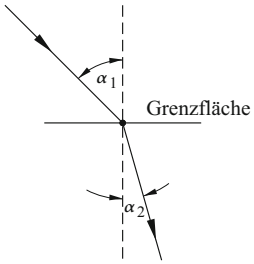


Abb. 14.8 Brechungsgesetz

die Geschwindigkeit c_2 und der Winkel a_2. Beim Erreichen des Punktes A wird hier eine Elementarwelle ausgesendet, die dann, wenn die Welle im Medium 1 nach der Zeit t_1 den Punkt B erreicht hat, den Radius $\overline{AE}$ hat. Die Einhüllende im Medium 2 ist $\overline{BD}$. Im Dreieck ABD gilt:

$$\sin\alpha_1 = \frac{\overline{DB}}{\overline{AB}}$$

und im Dreieck AEB:

$$\sin\alpha_2 = \frac{\overline{AE}}{\overline{AB}}$$

Weiter gilt

$$\overline{DB} = c_1 \cdot t_1, \quad \overline{AE} = c_2 \cdot t_1$$

Hieraus folgt

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{\overline{DB}}{\overline{AE}} = \frac{c_1 \cdot t_1}{c_2 \cdot t_1}$$

das *Brechungsgesetz* (Abb. 14.8)

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{c_1}{c_2} \tag{14.22}$$

Dies Gesetz ist ebenfalls für alle Wellenarten gültig.

14.2.3 Beugung

Trifft eine Welle auf eine Wand, in der sich eine kleine Öffnung befindet, so breitet sich die Welle hinter der Öffnung nicht geradlinig aus, sondern auch in dem Teil des Raumes, der durch die Wand abgedeckt ist, ist die Welle bemerkbar. Dieses Phänomen wird als *Beugung* bezeichnet. Die Beugung lässt sich mit Hilfe des Huygens'schen Prinzips so verstehen, dass von der Öffnung Elementarwellen ausgehen. Sind mehrere Öffnungen vorhanden, so entsteht das Beugungsfeld durch Interferenz der einzelnen Elementarwellen. Dies ist in Abb. 14.10 für den Fall eines Doppelspaltes aufgezeigt. Die Kreise stellen die Maxima der Elementarwellen dar. An den Punkten, an denen sich die Elementarwellen schneiden, erfolgt Verstärkung. Die Verbindungslinien dieser Schnittpunkte liegen auf Hyperbeln, deren Asymptoten zur Ausbreitungsrichtung unter den Winkeln a verlaufen. Für diese Winkel gilt mit ganzen Zahlen n die Bedingung

$$\sin\alpha_n = n\frac{\lambda}{d}, \quad n = 0, 1, 2, \ldots \tag{14.23}$$

erfüllt ist.

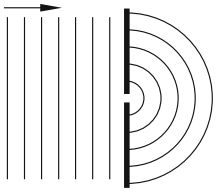


Abb. 14.9 Beugung an einer Öffnung

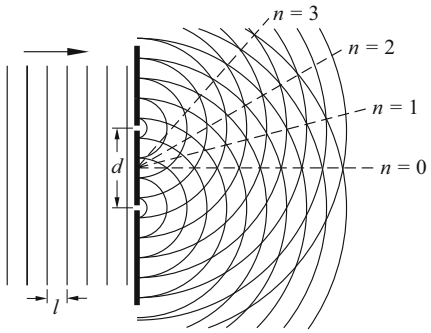


Abb. 14.10 Beugung am Doppelspalt

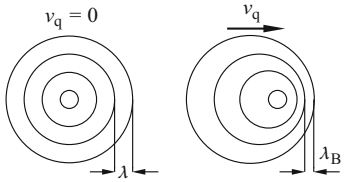


Abb. 14.11 Dopplereffekt

14.3 Dopplereffekt

Wenn sich der Erreger einer Welle, die Quelle, und der Beobachter relativ zum Medium, in dem die Welle übertragen wird, bewegen, so treten Frequenzverschiebungen auf, die *Dopplereffekt* genannt werden. Dabei ist es für die Berechnung wichtig,

Tab. 14.2 Dopplereffekt

Quelle	Beobachter	Formel
→	→	$f_B = f_Q \frac{c-v_B}{c-v_Q}$
→	←	$f_B = f_Q \frac{c+v_B}{c-v_Q}$
←	→	$f_B = f_Q \frac{c-v_B}{c+v_Q}$
←	←	$f_B = f_Q \frac{c+v_B}{c+v_Q}$

ob sich die Quelle oder der Beobachter bewegen. Beobachten kann man den Dopplereffekt beim Herannahen eines hupenden Autos: Bei Annäherung erhöht sich zunächst die Frequenz, der Ton wird höher, beim Entfernen erniedrigt sich die Frequenz und der Ton wird niedriger. Zunächst soll der Fall des ruhenden Beobachters und der sich auf den Beobachter zu bewegenden Quelle untersucht werden. Von einer punktförmigen Quelle werden Kugelwellen ausgesendet. Da sich die Quelle bewegt, liegen die Zentren der zu verschiedenen Zeiten nacheinander ausgesendeten Kugelwellen nicht mehr in einem Punkt. Der Beobachter im Punkt B erkennt eine kürzere Wellenlänge. In Abb. 14.11 ist links eine ruhende Quelle, die eine Welle der Wellenlänge λ aussendet, rechts eine bewegte Quelle dargestellt. Die Quelle bewegt sich nach rechts mit der Geschwindigkeit v_q. Der Beobachter sieht nun eine kleinere Wellenlänge λ_B. In der Zeit $T_q = \frac{1}{f_q}$ hat sich die Quelle um die Strecke $v_q T_q$ nach rechts bewegt, somit ist die Wellenlänge $\lambda_B = \lambda - v_q T_q$ verkleinert, die Frequenz hat sich jedoch vergrößert: $\lambda_B = \lambda - v_q T_q$. Mit $\lambda = \frac{c}{f}$ wird daraus:

$$\frac{c}{f_b} = \frac{c}{f_q} - \frac{v_q}{f_q} \tag{14.24}$$

$$\frac{c}{f_b} = \frac{c}{f_q}\left(1 - \frac{v_q}{c}\right) \tag{14.25}$$

$$f_B = \frac{f_q}{1 - \frac{v_q}{c}} = f_Q \frac{c}{c - v_Q} \tag{14.26}$$

Zu unterscheiden ist, ob und wie sich Quelle und Beobachter relativ zueinander bewegen. In Tab. 14.2 sind die möglichen Fälle angegeben.

Wenn Quelle oder Beobachter ruhen, so ist in der entsprechenden Gleichung diese Geschwindigkeit = 0 zu setzen.

Beispiel

Eine Schallquelle sendet einen Ton von 440 Hz aus. Die Quelle bewegt sich mit 100 km/h an einem ruhenden Beobachter vorbei.

Frage: Welche Frequenz nimmt der Beobachter

a) bei Annäherung
b) bei Entfernung der Quelle wahr?

Die Schallgeschwindigkeit beträgt 330 m/s.

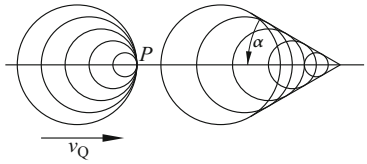


Abb. 14.12 Schallmauer und Machscher Kegel

Lösung: Da die Geschwindigkeit $v_B = 0$ ist, folgt:

a)

$$v_Q = 100\,\frac{\text{km}}{\text{h}} = \frac{100}{3{,}6}\,\frac{\text{m}}{\text{s}} = 27{,}78\,\frac{\text{m}}{\text{s}}$$

$$f_B = f_Q \frac{c}{c - v_Q} = \frac{440 \cdot 330}{330 - 27{,}78} = 480{,}44\,\text{Hz}$$

b)

$$f_B = f_Q \frac{c}{c + v_Q} = \frac{440 \cdot 330}{330 + 27{,}78} = 405{,}8\,\text{Hz} \quad \blacktriangleleft$$

Ein Sonderfall tritt dann ein, wenn sich die Quelle mit derselben Geschwindigkeit bewegt wie die ausgesendete Welle, z. B. eine Schallwelle. Es entsteht das Wellenbild nach Abb. 14.12 links. Alle Kugelwellen addieren sich im Punkt P, es entsteht die sogenannte *Schallmauer*, die mit der Quelle mitläuft. Wird die Geschwindigkeit der Quelle größer als die Schallgeschwindigkeit, so sind die Verhältnisse entsprechend Abb. 14.12 rechts. Für den *Öffnungswinkel* des entstehenden *Mach'schen Kegels* gilt:

$$\text{Öffnungswinkel} \quad \sin\alpha = \frac{c}{v_Q} \tag{14.27}$$

Auf dem Kegelmantel bildet sich eine einheitliche Wellenfront, die von einem Beobachter als Knall wahrgenommen wird. Diese Wellenfront bewegt sich mit der Quelle. In Vorwärtsrichtung ist die Quelle nicht zu hören.

Die Gleichungen aus Tab. 14.2 gelten nur für den Fall, dass sich die Welle in einem Medium als Träger der Welle ausbreitet. Eine Sonderstellung nehmen hier die elektromagnetischen Wellen ein (z. B. Licht), denn diese benötigen kein Medium, um sich auszubreiten. In diesem Fall gibt es bei bewegten Quellen und Beobachtern auch eine Frequenzverschiebung, diese ist aber nur von der Relativgeschwindigkeit zwischen Quelle und Beobachter abhängig.

Bei Annäherung von Quelle und Beobachter gilt:

$$f_B = f_Q \sqrt{\frac{c + v}{c - v}} \tag{14.28}$$

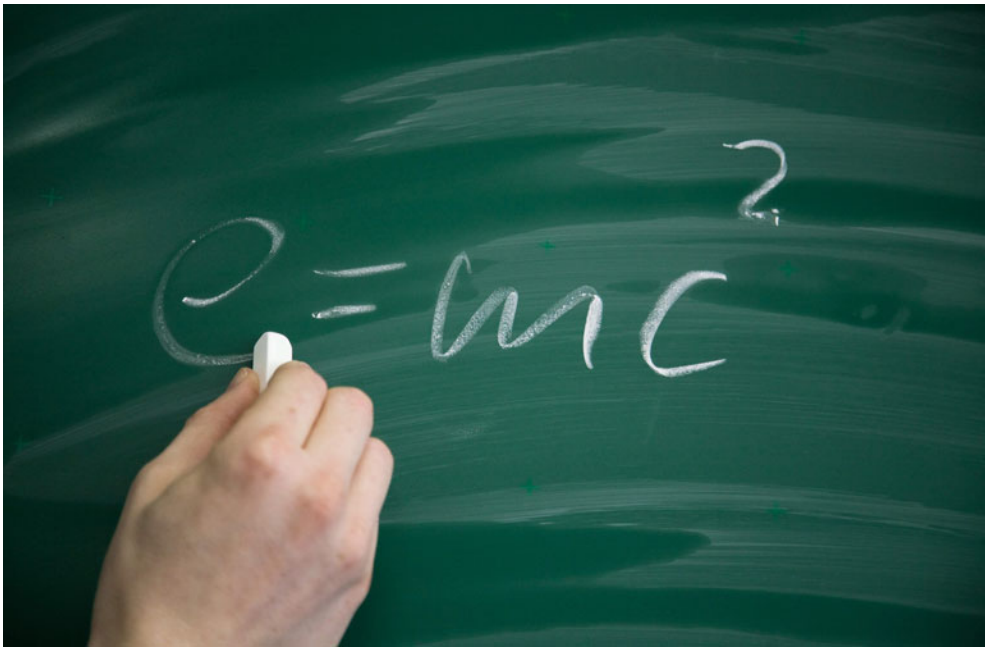
Entfernen sich Quelle und Beobachter, so muss v durch $-v$ ersetzt werden.

$$f_B = f_Q \sqrt{\frac{c - v}{c + v}} \tag{14.29}$$

Akustik

15

Horst Steffen



15.1 Schallausbreitung 234
15.2 Reflexion, Transmission, Absorption 234
15.3 Ultraschall 235

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_15

Tab. 15.1 Schallgeschwindigkeit und Schallkennimpedanz

	Dichte ρ	c	Z
	kg m^{-3}	m s^{-1}	N s m^{-3}
Luft −20 °C	1,396	319	445
Luft 0 °C	1,293	331	427
Luft 20 °C	1,21	344	416
Eis	920	3200	$2{,}94 \cdot 10^6$
Holz	600	4500	$2{,}7 \cdot 10^6$
Glas	2500	5300	$13 \cdot 10^6$
Beton	2100	4000	$8{,}4 \cdot 10^6$
Stahl	7700	5050	$39 \cdot 10^6$

In der Akustik wird die Ausbreitung von Schallwellen in festen, flüssigen und gasförmigen Medien untersucht. In festen Stoffen können diese Wellen Longitudinalwellen und Transversalwellen sein, in Flüssigkeiten und Gasen gibt es nur Longitudinalwellen, weil dort die zur Ausbreitung von Transversalwellen notwendigen rückstellenden Querkräfte fehlen. Im folgenden werden vorzugsweise Schallwellen in Gasen behandelt, weil diese im täglichen Leben eine übergeordnete Rolle spielen.

15.1 Schallausbreitung

Schall breitet sich in Gasen und Flüssigkeiten durch Druckänderungen als Longitudinalwelle aus. Hiermit verbunden ist eine Dichteschwankung, eine Druckschwankung und eine Schwankung der Geschwindigkeit der einzelnen Moleküle um die jeweiligen konstanten Mittelwerte (Index 0), die auch ohne Schall vorhanden sind.

$$\textit{Dichte} \qquad \rho = \rho_0 + \rho_w \tag{15.1}$$

$$\textit{Druck} \qquad p = p_0 + p_w \tag{15.2}$$

$$\textit{Geschwindigkeit} \qquad v = v_0 + v_w \tag{15.3}$$

mit den Bezeichnungen: ρ_w: Wechseldichte, p_w: Schallwechseldruck oder *Schalldruck*, v_w: *Schallschnelle*.

Die Lösungsfunktion als Funktion von Ort und Zeit hängt unter anderem vom Erregertyp ab. Für den einfachsten Fall einer eindimensionalen sinusförmigen Erregung lautet sie für den *Schalldruck*:

$$p(x,t) = p_0 + \hat{p}\sin(\omega t - kx) \tag{15.4}$$

und für die *Schallschnelle*:

$$v_w(x,t) = \hat{v}\sin(\omega t - kx) \tag{15.5}$$

mit

$$\hat{v} = \frac{\hat{p}}{\rho_0 c} \tag{15.6}$$

Hierbei ist c die Schallgeschwindigkeit. Die Größe $\rho_0 \cdot c$ wird als *Schallkennimpedanz* Z, früher auch als Schallwellenwiderstand, bezeichnet.

$$Z = \rho_0 \cdot c \qquad \begin{array}{c|c|c} Z & \rho_0 & c \\ \hline \mathrm{N\,m^{-1}} & \mathrm{kg/m^3} & \mathrm{m/s} \end{array} \tag{15.7}$$

Die Schallkennimpedanz ist eine charakteristische Größe für das jeweilige Ausbreitungsmedium und ist bei ebenen Wellen konstant.

Für die Schallgeschwindigkeit in Gasen gilt (siehe Abschn. 14.1):

$$c = \sqrt{\frac{\chi p}{\rho}} \tag{15.8}$$

oder

$$c = \sqrt{\chi R_i T}. \tag{15.9}$$

In Festkörpern gilt:

$$\textit{Schallgeschwindigkeit} \quad c = \sqrt{\frac{E}{\rho}} \tag{15.10}$$

15.2 Reflexion, Transmission, Absorption

Jedes Medium wird durch seine Schallkennimpedanz beschrieben. Treffen nun Schallwellen auf eine Grenzfläche zwischen zwei Medien mit unterschiedlicher Schallkennimpedanz, so wird ein Teil reflektiert, ein Teil tritt ins Medium ein. Der *Reflexionsfaktor* r ist definiert als das Verhältnis des Schalldruckes p_r der reflektierten Welle zu dem Schalldruck p_e der einfallenden Welle.

$$\textit{Reflexionsfaktor} \quad r = \frac{p_r}{p_e} \tag{15.11}$$

Werden die Intensitäten I_r und I_e ins Verhältnis gesetzt, so erhält man den

$$\textit{Reflexionsgrad} \quad \rho_s = \frac{I_r}{I_e} = \frac{p_r^2}{p_e^2} = r^2 \tag{15.12}$$

Werden die Schallkennimpedanzen für senkrechten Einfall mit Z_1 bzw. Z_2 bezeichnet, so gilt für senkrechten Einfall:

$$\rho_s = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \tag{15.13}$$

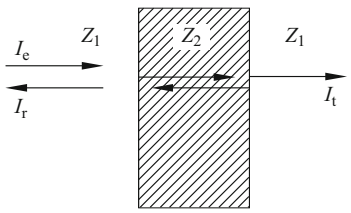


Abb. 15.1 Reflexion und Transmission

Ein großer Reflexionsgrad ρ tritt immer dann auf, wenn die Schallkennimpedanzen der beteiligten Medien sehr unterschiedlich sind. Dabei ist es gleichgültig, welches der beiden Medien die größere Schallkennimpedanz besitzt.

Wird die Schallwelle im Medium 2 gedämpft, so wird die Welle bei genügend dickem Medium 2 vollständig absorbiert. Der *Schallabsorptionsgrad* a_s wird definiert durch:

Schallabsorptionsgrad

$$\alpha_s = \frac{I_e - I_r}{I_e} = 1 - \frac{I_r}{I_e} \tag{15.14}$$

$$\alpha_s = 1 - \rho_s = 1 - r^2 \tag{15.15}$$

Wird nun der Schall im Medium 2 nicht vollständig absorbiert, weil z. B. die Wand zu dünn ist, so tritt ein Teil I_t der einfallenden Schallintensität I_e durch das Medium 2 durch. Dieser Anteil wird als *Schalltransmissionsgrad* τ_s bezeichnet. Auch an der zweiten Grenzfläche wird dann ein Teil der dort ankommenden Intensität reflektiert.

$$\textit{Schalltransmissionsgrad} \quad \tau_s = \frac{I_t}{I_e} \tag{15.16}$$

Der Absorptionsgrad gibt an, welcher Teil der einfallenden Welle nicht reflektiert wurde, der Transmissionsgrad gibt den Anteil an, der durch eine Wand hindurch gelangt ist. Die in der Wand tatsächlich verlorengegangene Schallintensität wird durch den *Schalldissipationsgrad* δ_s angegeben. Somit gilt der Zusammenhang:

$$\rho_s + \tau_s + \delta_s = 1 \tag{15.17}$$

$$\rho_s + \alpha_s = 1 \tag{15.18}$$

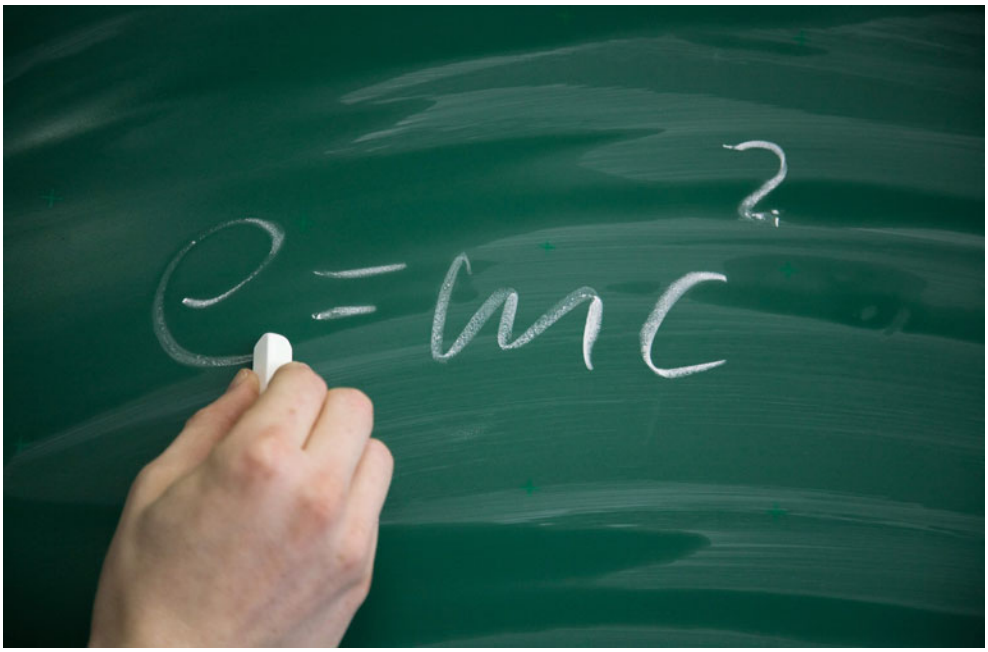
15.3 Ultraschall

Unter Ultraschall wird der Schall verstanden, dessen Frequenz jenseits (= ultra) der menschlichen Hörschwelle von ca. 20 kHz liegt. Die obere Grenze liegt bei etwa 1 GHz. Die zugehörigen Wellenlängen liegen in Luft zwischen etwa 2 cm und 0,34 µm. Für den Ultraschall gibt es in der Praxis eine Vielfalt von Einsatzmöglichkeiten. Dazu nachfolgend einige Beispiele: Ultraschall-Echolot, Ultraschall-Blindenleitgerät, Ultraschall-Alarmanlagen, Werkstoffprüfung, Materialprüfverfahren, Medizin.

Optik

16

Horst Steffen



16.1 Eigenschaften des Lichtes 238
16.2 Geometrische Optik 238
16.3 Wellenoptik 246
16.4 Photometrie 248
16.5 Licht als Korpuskel 251
Literatur 252

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_16

Die Optik ist die Lehre vom Licht, seiner Ausbreitung, dem Aufbau optischer Instrumente und allen den Erscheinungen, die mit dem Auge wahrgenommen werden können. Heute wird zwischen der *klassischen Optik* und der *Quantenoptik* unterschieden. In der klassischen Optik wird das Licht als Welle verstanden. Seit Beginn dieses Jahrhunderts sind allerdings viele Experimente durchgeführt worden, deren Deutung nicht im Rahmen der Wellentheorie möglich ist. Zur Erklärung muss angenommen werden, dass das Licht sich wie ein Teilchen, als *Korpuskel*, dem sogenannten *Lichtquant*, verhält. Dieses Bild ist immer dann anzuwenden, wenn Licht mit Materie in Wechselwirkung tritt, also zum Beispiel bei der Absorption und Emission, aber auch bei der Streuung von Licht.

16.1 Eigenschaften des Lichtes

Die Ansichten über die Natur des Lichtes haben sich im Laufe der Jahrhunderte grundlegend gewandelt. *Newton* stellte im 17. Jahrhundert eine Korpuskulartheorie auf, eine Lichtquelle sendet kleine Korpuskeln aus, die sich geradlinig mit großer Geschwindigkeit fortbewegen, bis sie im menschlichen Auge durch Nervenanregung wahrgenommen werden. Reflexion und Brechung von Licht konnten hiermit erklärt werden, Beugung und Interferenz jedoch nicht. Dies gelang *Huygens* im Jahr 1678 mit der von ihm entwickelten *Wellentheorie*, die von *Young* (1802) durch Experimente untermauert wurde. Nach der Entdeckung der *Polarisation* durch *Malus* (1808) hat *Fresnel* (1815) das Licht als transversale Welle beschrieben, und dass das Licht eine *elektromagnetische Transversalwelle* ist, wurde von *Maxwell* (1865) entdeckt. Die physikalischen Größen, die sich bei elektromagnetischen Wellen zeitlich und räumlich periodisch ändern, sind die elektrische Feldstärke und damit verbunden auch die magnetische Feldstärke. Beide stehen senkrecht aufeinander und senkrecht zur Ausbreitungsrichtung. Das Licht breitet sich mit der Naturkonstanten *Lichtgeschwindigkeit* aus. Die Lichtgeschwindigkeit im Vakuum hängt mit der elektrischen Feldkonstante ε_0 und der magnetischen Feldkonstante μ_0 durch die Gleichung

$$\textit{Lichtgeschwindigkeit} \quad c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{16.1}$$

zusammen.

Die Lichtgeschwindigkeit im Vakuum hat den Wert

$$c = 2{,}99792458 \cdot 10^8 \, \frac{\mathrm{m}}{\mathrm{s}} \tag{16.2}$$

Auch hier gilt, wie bei allen Wellen, der Zusammenhang $c = \lambda \cdot f$ zwischen der Wellenlänge λ und der Frequenz f. Das sichtbare Licht nimmt nur einen kleinen Teil des Spektrums der elektromagnetischen Wellen ein (siehe Abb. 16.1). Alle die dort aufgeführten Wellen sind elektromagnetische Wellen und breiten sich im Vakuum mit der Lichtgeschwindigkeit c aus. Die Frequenzen erstrecken sich von 10^4 bis $10^{20}\,\mathrm{s}^{-1}$.

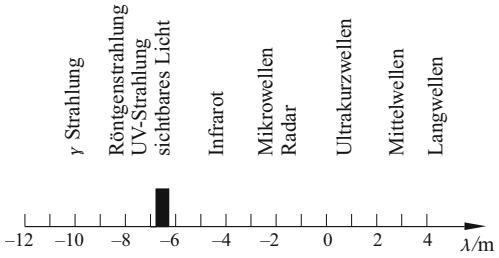


Abb. 16.1 Spektrum der elektromagnetischen Wellen

16.2 Geometrische Optik

Wird nun die Ausbreitung des Lichtes durch Öffnungen, durch Blenden oder an Gegenständen vorbei betrachtet, und sind diese Gegenstände groß gegenüber der Wellenlänge des Lichtes, so kann die Ausbreitung des Lichtes nach den Gesetzen der geometrischen Optik bestimmt werden, im anderen Fall muss die Wellenoptik herangezogen werden. Die Lichtwellen werden in ihren Ausbreitungsrichtungen durch die Strahlen gekennzeichnet, die senkrecht auf den Wellenfronten stehen (siehe Kap. 14). Bei ebenen Wellen sind die Strahlen parallel. Kreuzen sich zwei Lichtstrahlen, so beeinflussen sie sich gegenseitig nicht.

16.2.1 Reflexion des Lichtes

Wird ein einfallender Lichtstrahl an einer ebenen Fläche reflektiert, so gilt das *Reflexionsgesetz* (Abb. 16.2):

Einfallender Strahl, reflektierter Strahl und Einfallslot liegen in einer Ebene.

Der Einfallswinkel α und der Reflexionswinkel β sind gleich.

Das Einfallslot ist hier das Lot auf der Spiegelfläche. Dies Gesetz wurde bereits in Kap. 14 mit der Wellentheorie abgeleitet.

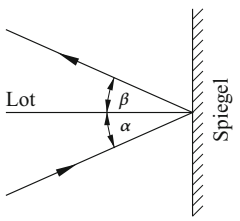


Abb. 16.2 Reflexionsgesetz

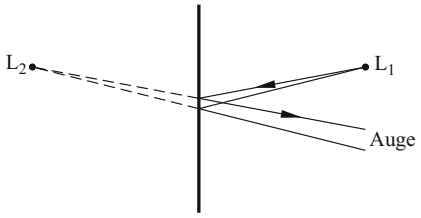


Abb. 16.3 Bildentstehung

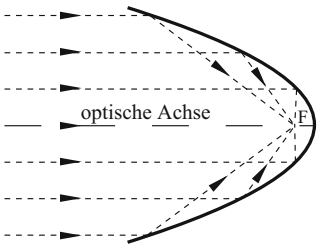


Abb. 16.4 Parabolspiegel

Ein Beobachter, der vor einem ebenen Spiegel steht, sieht sein Bild hinter dem Spiegel. Das menschliche Auge in Zusammenarbeit mit dem Gehirn kann keine Richtungsumkehrung oder Abweichungen von der geradlinigen Ausbreitung des Lichtstrahles wahrnehmen. Eine Lichtquelle L_1, die vor einem Spiegel steht und deren Licht in das Auge fällt, erscheint dem Beobachter hinter dem Spiegel, dabei wird der tatsächlich vorhandene Strahlenverlauf durch das menschliche Gehirn in den nicht vorhandenen, gestrichelt gezeichneten Verlauf umgedacht. Das Bild L_2 der Lichtquelle erscheint dort, wo sich die gedachten Strahlen kreuzen (Abb. 16.3). Bilder, die durch einen nicht wirklich vorhandenen Schnittpunkt, sondern einen gedachten Schnittpunkt entstehen, heißen *virtuelle* Bilder. Schneiden sich zwei Lichtstrahlen tatsächlich, so entsteht ein *reelles* Bild.

Im folgenden soll die Reflexion an nicht ebenen Spiegeln untersucht werden. Dabei hat der Spiegel im Querschnitt eine Parabelform; das räumliche Gebilde wird *Paraboloid* genannt. Diese *Parabolspiegel* haben die Eigenschaft, alle Strahlen, die parallel zur Symmetrielinie, der *optischen Achse*, einfallen, so zu reflektieren, dass sie sich in einem Punkt, dem Brennpunkt F, schneiden (Abb. 16.4). Da Parabolspiegel schwer herzustellen sind, werden für einfache Anwendungen Spiegel genommen, deren Oberfläche aus einer Kugel geschnitten wurde. Diese haben zwar Abbildungsfehler, beschränkt man sich jedoch auf Strahlen, die nahe der optischen Achse einfallen, so können diese Fehler vernachlässigt werden. Der Brennpunkt hat im Rahmen dieser Näherung den Abstand

$$f = \frac{r}{2} \tag{16.3}$$

Diese Größe f ist die *Brennweite*. Zur Bildkonstruktion werden mindestens zwei Strahlen benötigt. Es werden die Strahlen verwendet, deren Verlauf sich einfach konstruieren lässt, nämlich ein achsenparalleler Strahl 1 und ein Brennpunktstrahl 2. Strahl 1 verläuft nach der Reflexion durch den Brennpunkt, Strahl 2 wird achsenparallel reflektiert. Hier können je nach Lage des Gegenstandes G virtuelle oder reelle Bilder entstehen (siehe Abb. 16.5).

Wird der Abstand des Gegenstandes vom Spiegel mit der *Gegenstandsweite* g, der Abstand des Bildes mit der *Bildweite* b bezeichnet, so gilt die

Abbildungsgleichung

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b} \tag{16.4}$$

Die Gegenstandsgröße G und die Bildgröße B lassen sich aus einfachen geometrischen Überlegungen bestimmen, und es gilt:

Abbildungsmaßstab

$$\beta = \frac{B}{G} = -\frac{b}{g} \tag{16.5}$$

In diesem System sind die einzelnen Größen mit Vorzeichen behaftet, die positiven Richtungen weisen nach oben und vom Spiegel zum Brennpunkt, in diesem Fall nach links.

Ob das Bild reell oder virtuell, verkleinert oder vergrößert wird, hängt von der Gegenstandsweite ab. Es gilt:

$g < f$	virtuell	aufrecht	vergrößert
$f < g < 2f$	reell	höhenverkehrt	vergrößert
$g > 2f$	reell	höhenverkehrt	verkleinert

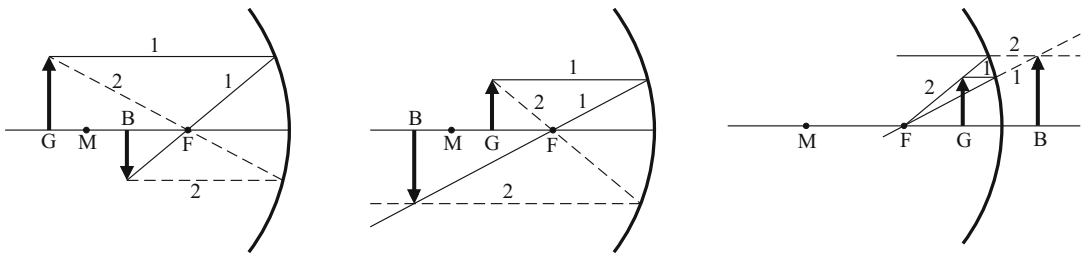


Abb. 16.5 Bildkonstruktion am Hohlspiegel. M: Mittelpunkt des Hohlspiegels; F: Brennpunkt; G: Gegenstand; B: Bild

Beispiel

Vor einem Hohlspiegel mit dem Krümmungsradius $r = 10\,\text{cm}$ steht in 3 cm Abstand ein 2 cm großer Gegenstand. Wie groß sind Bildweite und Bildgröße?

Lösung:

$$f = \frac{r}{2} = 5\,\text{cm}$$

$$\frac{1}{b} = \frac{1}{f} - \frac{1}{g} = \frac{1}{5} - \frac{1}{3}$$

$$b = -7{,}5\,\text{cm}$$

$$B = -G\frac{b}{g} = -2\,\text{cm} \cdot \frac{-7{,}5\,\text{cm}}{3\,\text{cm}} = 5\,\text{cm}$$

Da b negativ ist, ist das Bild ein virtuelles Bild, B ist positiv, somit ist das Bild aufrecht, was dem Abb. 16.5c entspricht. ◀

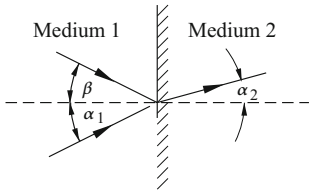


Abb. 16.6 Brechungsgesetz

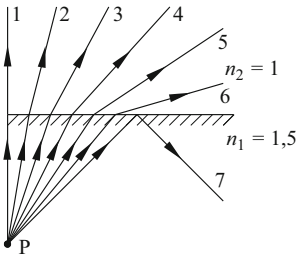


Abb. 16.7 Totalreflexion

16.2.2 Brechungsgesetz

Trifft ein Lichtstrahl schräg auf die Grenzfläche zwischen zwei nicht absorbierenden Stoffen, in denen das Licht unterschiedliche Lichtgeschwindigkeit hat, so wird ein Teil des einfallenden Lichtes reflektiert, ein anderer Teil tritt in das zweite Medium ein. Dabei ändert sich die Ausbreitungsrichtung an der Grenzfläche. Dies wird als *Brechung* bezeichnet.

Es gelten folgende Beziehungen zwischen den Winkeln in der Abbildung VII-6:

$$\textit{Reflexionsgesetz} \qquad \alpha_1 = \beta \tag{16.6}$$

$$\textit{Brechungsgesetz} \qquad \frac{\sin\alpha_1}{\sin\alpha_2} = \frac{c_1}{c_2} \tag{16.7}$$

Jedem Medium wird eine Zahl zugeordnet, die durch das Verhältnis der Lichtgeschwindigkeit im Medium zu der im Vakuum definiert ist (Tab. 16.1).

$$\textit{Brechungsindex} \quad n = \frac{c_{\text{Vac}}}{c} \tag{16.8}$$

Tab. 16.1 Brechungsindizes für $l = 589\,\text{nm}$, 20 °C und 1013 hPa

Feste Stoffe	n	Flüssigkeiten und Gase	n
Eis	1,31	Wasser	1,333
Flussspat	1,4338	Äthylalkohol	1,3618
Steinsalz	1,5443	Leinöl	1,486
Optische Gläser		Benzol	1,5012
Borkron BK1	1,51		
Flint F3	1,6128	Luft	1,000292
Schwerflint SF4	1,7550	Sauerstoff	1,000271
Quarzglas	1,4588	Stickstoff	1,000298
Plexiglas	1,50 … 1,52	Kohlendioxid	1,000449
Diamant	2,4173		

Der *Brechungsindex* des betreffenden Materials hängt allerdings von der Wellenlänge des Lichtes ab. Die Tatsache, dass der Brechungsindex von der Wellenlänge des Lichtes abhängt, heißt *Dispersion* (Tab. 16.2). Im optischen Bereich ist die Abhängigkeit normalerweise so, dass mit steigender Wellenlänge der Brechungsindex kleiner wird. Mit Hilfe der Brechungsindizes lässt sich das Brechungsgesetz auch schreiben:

$$\textit{Brechungsgesetz} \qquad \frac{\sin\alpha_1}{\sin\alpha_2} = \frac{n_2}{n_1} \tag{16.9}$$

Stoffe mit kleinerem Brechungsindex sind *optisch dünner*, solche mit größerem Index *optisch dichter*.

Totalreflexion

Das Brechungsgesetz (16.9) gilt unabhängig davon, ob das Licht aus einem optisch dichteren oder optisch dünnerem Medium austritt. Tritt das Licht aus einem optisch dichteren Medium in ein optisch dünneres Medium, so verläuft der Strahl wie in Abb. 16.7 dargestellt, er wird vom Einfallslot weggebrochen. Wenn der Lichtstrahl aus einem optisch dichteren Medium in z. B. Luft tritt, so ist $n_2 < n_1$, und es folgt für den Ausfallswinkel nach (16.9)

$$\sin\alpha_2 = \frac{n_1}{n_2} \cdot \sin\alpha_1 \tag{16.10}$$

Da $n_1 > n_2$, kann der Fall eintreten, dass bei einem bestimmten Einfallswinkel a_1 der Wert für $\sin\alpha_2 > 1$ wird. Ab diesem Grenzwinkel kann dann der Lichtstrahl nicht mehr ins Medium 2 eintreten, er wird vollständig reflektiert. Dies Phänomen heißt *Totalreflexion*. In Abb. 16.7 sind einige Strahlenverläufe, ausgehend vom Punkt P, gezeichnet. Der Grenzwinkel ergibt sich in

Tab. 16.2 Dispersion einiger Stoffe

Brechungsindex							
	Farbe und Wellenlänge λ						
Stoff	Violett 396,8 nm	Blau 430,7 nm	Blaugrün 486,1 nm	Grün 527 nm	Gelb 589 nm	Rot 656,2 nm	Dunkelrot 760,8 nm
Wasser	1,3435	1,3406	1,3371	1,3352	1,333	1,3312	1,3289
Flint F3	1,6542	1,6355	1,6246	1,6190	1,6128	1,6081	1,6029

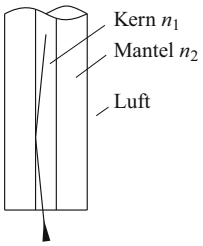


Abb. 16.8 Lichtleiter

diesem Beispiel zu

$$\sin\alpha_1 = \frac{n_2}{n_1} = \frac{1}{1{,}5}$$
$$\alpha_1 = 41{,}81^\circ$$

und liegt zwischen den Strahlen 6 und 7.

Technische Anwendung findet dieses Phänomen bei der Nachrichtenübertragung in *Lichtleitern*. Ein Lichtleiter ist im Prinzip so aufgebaut, dass ein optisch dichteres Material, der Kern, von einem optisch dünneren Material, dem Mantel, umgeben ist. Auf die Stirnfläche eintretendes Licht wird an der Grenzfläche Kern–Mantel totalreflektiert (Abb. 16.8).

Prisma

Ein *Prisma* ist ein von geraden, gegeneinander geneigten Flächen begrenzter Körper aus einem durchsichtigen Material, normalerweise aus Glas. Die Kante $\overline{CF}$ ist die *brechende Kante* K und der Winkel ε der *brechende Winkel*. Die Fläche $ABDE$ heißt die Basis des Prismas. Im Abb. 16.9 ist ein symmetrischer Strahlengang gezeichnet, die Winkelhalbierende von ε steht senkrecht auf dem Lichtstrahl. In diesem Fall des *symmetrischen Strahlengangs* gilt für den *Ablenkwinkel* δ, dass dieser minimal wird. Ist das Prisma von Luft umgeben und hat den Brechungsindex n, so gilt die Gleichung:

$$\sin\frac{\delta+\varepsilon}{2} = n\cdot\sin\frac{\varepsilon}{2} \qquad (16.11)$$

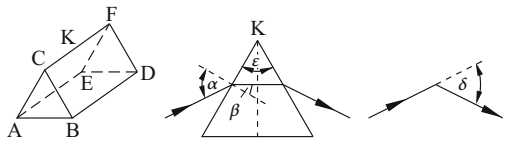


Abb. 16.9 Prisma

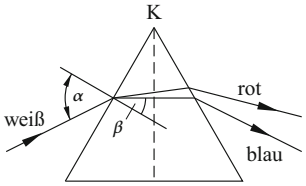


Abb. 16.10 Spektrale Zerlegung

Aus einer Messung des Ablenkwinkels δ kann, bei bekanntem brechenden Winkel ε, der Brechungsindex des Prismas bestimmt werden.

$$n = \frac{\sin\frac{\delta+\varepsilon}{2}}{\sin\frac{\varepsilon}{2}} \qquad (16.12)$$

Dies Verfahren kann auch bei Flüssigkeiten angewendet werden. Die zu untersuchende Flüssigkeit wird dann in ein Hohlprisma gefüllt und der Ablenkwinkel bestimmt.

Da der Brechungsindex von der Wellenlänge abhängt, kann mit einem Prisma Licht spektral zerlegt werden (Abb. 16.10). Einfallendes weißes Licht, in dem alle Wellenlängen bzw. Farben vertreten sind, wird durch ein Prisma nach Farben zerlegt. Für rotes Licht ist der Brechungsindex kleiner als für blaues Licht, daher ist der Ablenkwinkel für rotes Licht kleiner als für blaues Licht (siehe Tab. 16.2).

Linsen

In allen optischen Systemen und Geräten werden Linsen verwendet. Dieses sind Glaskörper, die durch zwei Kugelflächen begrenzt sind. In der technischen Optik (nach DIN 1335) werden für die Brechung an einer Kugelfläche Vorzeichen festgelegt. Es gilt: Die Achse durch den Kugelmittelpunkt ist die optische Achse, zugleich z-Achse des Koordinatensystems (Abb. 16.11 und 16.12). Die positive z-Richtung wird durch die

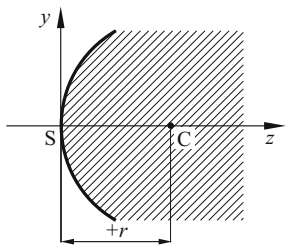


Abb. 16.11 Vorzeichenkonvention

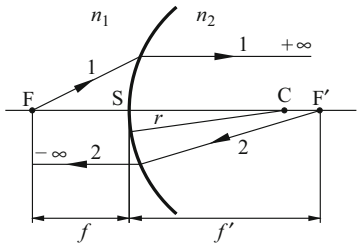


Abb. 16.12 Brennpunkte einer konvexen Kugelfläche

Laufrichtung des Lichtes festgelegt. Die y-Achse steht senkrecht auf der z-Achse und weist von unten nach oben. Der Radius der Kugel ist positiv, wenn das Zentrum C der Kugel in positiver z-Richtung vom Scheitelpunkt S liegt, hier also nach rechts. Sämtliche Strecken, die von S in negativer z-Richtung gemessen werden, erhalten ein negatives Vorzeichen. Da durch eine Kugelfläche Medien mit unterschiedlichen Brechungsindizes n_1 und n_2 getrennt werden, ist der Strahlenverlauf auf beiden Seiten der Fläche unterschiedlich. Die Brennpunkte F und F' sind so definiert, dass Strahlen, die von ihnen ausgehen, so gebrochen werden, dass sie achsenparallel weiter verlaufen. Für die Brennweiten f und f' gilt:

$$\textit{Brennweite} \quad f = \frac{n_1 r}{n_2 - n_1} \tag{16.13}$$

$$f' = \frac{n_2 r}{n_2 - n_1} \tag{16.14}$$

Kommen die Strahlen nicht aus dem Brennpunkt, sondern aus einem weiter entfernten Punkt im Abstand g, so schneiden sich alle Strahlen im Abstand b. Für diese Abstände gilt die

$$\textit{Abbildungsgleichung} \quad \frac{f'}{b} - \frac{f}{g} = 1 \tag{16.15}$$

Dabei ist g gemäß der Definition negativ einzusetzen.

Der Einfachheit halber soll angenommen werden, dass die Linsendicke vernachlässigbar klein ist, es sich also um *dünne Linsen* handelt. An beiden Grenzflächen zwischen der Linse und dem umgebenden Medium, im allgemeinen Luft, werden die Lichtstrahlen gebrochen. Bei dünnen Linsen kann man den Strahlengang konstruieren, indem beide Brechungen durch eine Brechung an der Mittellinie ersetzt werden. Es gibt *Sammellinsen*, oder *konvexe* Linsen, und *Zerstreuungslinsen*, oder *konkave* Linsen. Der Strahlenverlauf ist in Abb. 16.13 dargestellt. Die tatsächlichen Strahlen sind durchgezogen gezeichnet, bei der Zerstreuungslinse wird der Brennpunkt durch die gestrichelt gezeichnete gedachte rückwärtige Verlängerung konstruiert. Da der Brennpunkt auf der linken Seite liegt, hat die Zerstreuungslinse eine negative Brennweite. Die Linse sei nun in Luft und besitze den Brechungsindex n_L. Es gilt für die Brennweite f':

$$\frac{1}{f'} = (n_L - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{16.16}$$

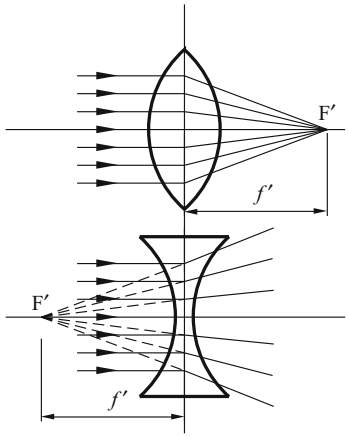


Abb. 16.13 Sammel- und Zerstreuungslinse

r_1 Radius der linken Kugelfläche, r_2 Radius der rechten Fläche. Sind beide Brennweiten gleich, so wird (16.15) zur

$$\textit{Linsenformel} \quad \frac{1}{f} = \frac{1}{b} - \frac{1}{g} \tag{16.17}$$

Den Kehrwert der Brennweite f nennt man die Brechkraft oder *Dioptrie* D.

$$D = \frac{1}{f} \qquad \begin{array}{c|c} D & f \\ \hline \mathrm{m}^{-1} & \mathrm{m} \end{array} \tag{16.18}$$

Der *Abbildungsmaßstab* ist:

$$\beta' = \frac{B}{G} = \frac{b}{g} \tag{16.19}$$

Zur Konstruktion der Abbildung durch Linsen verwendet man zweckmäßigerweise folgende Strahlen:

1. Strahl parallel zur optischen Achse: durch den Brennpunkt,
2. Strahl durch die Linsenmitte: keine Ablenkung,
3. Strahl durch den Brennpunkt: parallel zur optischen Achse.

In Abb. 16.14 sind für zwei Fälle die Bildkonstruktionen durchgeführt. Links: Gegenstand außerhalb der doppelten Brennweite: Es ergibt sich ein reelles, höhenverkehrtes und verkleinertes Bild. Rechts: Gegenstand innerhalb der Brennweite: Es ergibt sich ein virtuelles, aufrechtes und vergrößertes Bild.

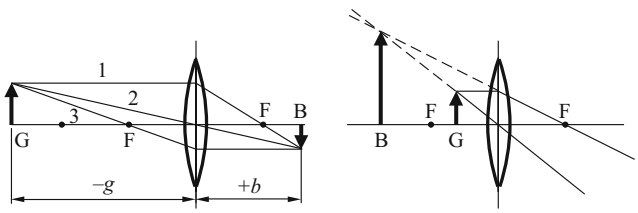


Abb. 16.14 Abbildung durch eine Sammellinse

Beispiel

Ein Gegenstand der Größe 4 cm steht 25 cm entfernt von einer Sammellinse der Brennweite $f = 5\,\text{cm}$. Wie groß sind a) Bildweite, b) Abbildungsmaßstab und c) Bildgröße?

Lösung:

a) Bildweite

$$\frac{1}{b} = \frac{1}{f} + \frac{1}{g} = \frac{1}{5\,\text{cm}} + \frac{1}{-25\,\text{cm}}$$
$$b = 6{,}25\,\text{cm}$$

b) Abbildungsmaßstab

$$\beta = \frac{b}{g} = \frac{6{,}25\,\text{cm}}{-25\,\text{cm}} = -0{,}25$$

c) Bildgröße

$$B = \beta \cdot G = -0{,}25 \cdot 4\,\text{cm} = -1\,\text{cm}$$

Da b positiv ist, liegt das Bild rechts von der Linse, B ist negativ, somit ist das Bild höhenverkehrt, dies entspricht dem Abb. 16.14a. ◀

Beispiel

Ein Gegenstand der Größe 4 cm steht 3 cm entfernt von einer Sammellinse der Brennweite $f = 5\,\text{cm}$. Wie groß sind a) Bildweite, b) Abbildungsmaßstab und c) Bildgröße?

Lösung:

a) Bildweite

$$\frac{1}{b} = \frac{1}{f} + \frac{1}{g} = \frac{1}{5\,\text{cm}} + \frac{1}{-3\,\text{cm}}$$
$$b = -7{,}5\,\text{cm}$$

b) Abbildungsmaßstab

$$\beta = \frac{b}{g} = \frac{-7{,}5\,\text{cm}}{-3\,\text{cm}} = 2{,}5$$

c) Bildgröße

$$B = \beta \cdot G = 2{,}5 \cdot 4\,\text{cm} = 10\,\text{cm}$$

b ist jetzt < 0, $B > 0$, somit liegt ein virtuelles, vergrößertes Bild vor. ◀

Bei der Abbildung durch Zerstreuungslinsen werden im Prinzip die gleichen Strahlen verwendet wie bei Sammellinsen, nur entstehen hier grundsätzlich virtuelle Bilder, da sich die Strahlen ja nie schneiden können (Abb. 16.15).

Tab. 16.3 Bauformen von Linsen

				
	Bi-konvex	Plan-konvex	Bi-konkav	Plan-konkav
Radien	$r_1 > 0$ $r_2 < 0$	$r_1 = \infty$ $r_2 < 0$	$r_1 < 0$ $r_2 > 0$	$r_1 = \infty$ $r_2 > 0$
Brennweite	$f > 0$	$f > 0$	$f < 0$	$f < 0$

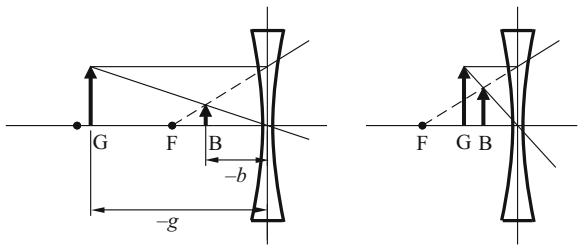


Abb. 16.15 Abbildung durch Zerstreuungslinsen

Werden zwei Linsen kombiniert, und zwar so, dass sie in einem Abstand e voneinander stehen, so ergibt sich für das System eine Gesamtbrennweite nach folgender Formel:

Brennweite einer Linsenkombination

$$\frac{1}{f_{\text{ges}}} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{e}{f_1 \cdot f_2} \tag{16.20}$$

Beispiel

Ein Linsensystem besteht aus einer Sammellinse der Brennweite $f_1 = 10\,\text{cm}$ und einer Zerstreuungslinse der Brennweite $f_2 = -6\,\text{cm}$. Beide haben einen Abstand von $e = 5\,\text{cm}$. Wie groß ist die Brennweite des Systems?

Lösung:

$$\frac{1}{f_{\text{ges}}} = \frac{1}{10\,\text{cm}} + \frac{1}{-6\,\text{cm}} - \frac{5\,\text{cm}}{-6 \cdot 10\,\text{cm}^2} = 0{,}0167\,\text{cm}^{-1}$$
$$f_{\text{ges}} = 60\,\text{cm}$$ ◀

In Linsensystemen werden die verschiedensten Formen verwendet, die gängigsten Bauformen von Linsen sind in Tab. 16.3 aufgeführt.

16.2.3 Optische Geräte

In optischen Geräten werden je nach Anwendung Linsen unterschiedlichster Bauformen zu Linsensystemen zusammengesetzt. Sie dienen dazu, den *Sehwinkel* zu vergrößern.

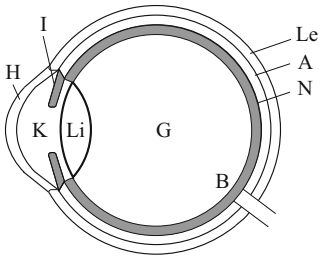


Abb. 16.16 Das menschliche Auge

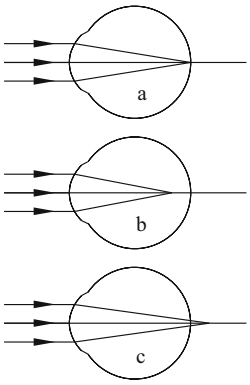


Abb. 16.17 Augenfehler

Das menschliche Auge

Da jede optische Erscheinung mit dem Auge wahrgenommen wird, soll hier kurz die Wirkungsweise des menschlichen Auges dargestellt werden. Das Prinzip ist in Abb. 16.16 dargestellt. Das Auge wird von der *Lederhaut Le* eingehüllt, darunter liegt die *Aderhaut A* und die lichtempfindliche *Netzhaut N*. Das Licht fällt durch die Hornhaut *H* und die *Augenkammer K*, tritt durch die *Linse Li* und dann auf die Netzhaut *N*. Der Augennerv tritt am *Blinden Fleck B* durch das Auge. Durch die variable Krümmung und damit variable Brennweite der Linse kann das Auge Gegenstände in verschiedenen Entfernungen scharf auf die Netzhaut abbilden. Die Iris *I* regelt als Blende die eintretende Lichtmenge. Der Abstand, in dem man noch eben scharf sehen kann, ist der *Nahpunkt*, er liegt beim jungen Menschen bei ca. 10 cm, beim älteren Menschen weiter entfernt. Als *deutliche Sehweite* ist ein Abstand von $s_0 = 25\,\text{cm}$ definiert.

Der *Fernpunkt* liegt beim normalsichtigen Menschen im Unendlichen. Fehlsichtigkeit tritt dann auf, wenn parallel einfallende Strahlen nicht auf der Netzhaut (Abb. 16.17a), sondern beim *kurzsichtigen* Auge vor der Netzhaut (Abb. 16.17b), beim *weitsichtigen* Auge hinter der Netzhaut (Abb. 16.17c) fokussiert werden. Die lichtempfindlichen Sinneszellen sind die in der Netzhaut eingebetteten *Stäbchen* und *Zäpfchen*. Die Zäpfchen sind farbempfindlich, während die Stäbchen nur Grautöne unterscheiden können, dafür sind sie aber wesentlich lichtempfindlicher. Von einem reellen Gegenstand wird durch das optische System des Auges ein umgedrehtes, reelles Bild auf der Netzhaut erzeugt. Das menschliche Gehirn dreht auf Grund seiner Erfahrung dieses Bild um, sodass wir von einem aufrechten

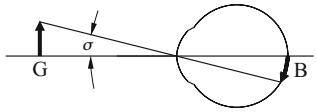


Abb. 16.18 Sehwinkel

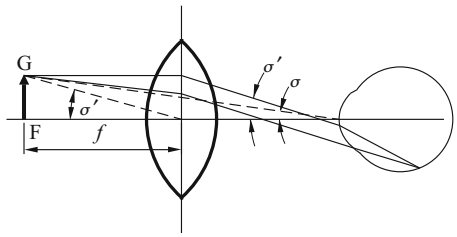


Abb. 16.19 Lupe

Gegenstand auch ein aufrechtes Bild „sehen“. Der Gegenstand erscheint dem Auge unter dem *Sehwinkel* σ (Abb. 16.18). Soll der Gegenstand vergrößert betrachtet werden, so muss der Sehwinkel vergrößert werden, man muss näher an den Gegenstand herangehen. Dies geht aber nur bis zum Nahpunkt. Eine weitere Vergrößerung ist nur mit Hilfe optischer Instrumente erreichbar. Wird der Sehwinkel mit Instrument als σ' bezeichnet, so ist die *Vergrößerung* Γ definiert zu

$$\textit{Vergrößerung} \quad \Gamma = \frac{\tan\sigma'}{\tan\sigma} \tag{16.21}$$

Lupe

Eine Lupe als einfachstes optisches Instrument ist eine Sammellinse hoher Brechkraft (Abb. 16.19). Die Vergrößerung hängt von der Brennweite der Linse, aber auch vom Abstand des Gegenstandes von der Linse ab. Um eine vergleichbare Angabe zu erzielen, wird festgelegt, dass sich der Gegenstand in der Brennebene der Lupe befindet und sich das Auge auf unendlich eingestellt, *akkomodiert*, hat. Diese Vergrößerung bezeichnet man als *Normalvergrößerung*. Verglichen wird der sich dann ergebende Sehwinkel mit dem, bei dem sich der Gegenstand ohne Lupe in der deutlichen Sehweite s_0 befindet.

Mit $\tan\sigma' = \frac{G}{f}$ und $\tan\sigma = \frac{G}{-s_0}$ folgt für die

$$\textit{Vergrößerung einer Lupe} \quad \Gamma = -\frac{s_0}{f} \tag{16.22}$$

Wird der Gegenstand näher an die Lupe herangebracht, so entsteht ein virtuelles, aufrechtes vergrößertes Bild.

Mikroskop

Noch bessere Vergrößerungen erzielt man mit einem *Mikroskop* (Abb. 16.20). Dies ist im Prinzip eine Kombination aus zwei Sammellinsen, dem Objektiv *Obj* mit den Brennpunkten F_1 und dem Okular *Ok* mit den Brennpunkten F_2. Das Objektiv entwirft vom Gegenstand *G* ein reelles *Zwischenbild ZB* fast im Brennpunkt F_2. Dieses Zwischenbild wird mit dem Okular, das wie

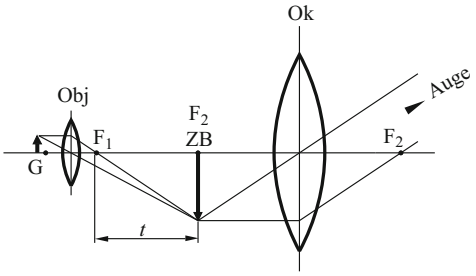


Abb. 16.20 Mikroskop

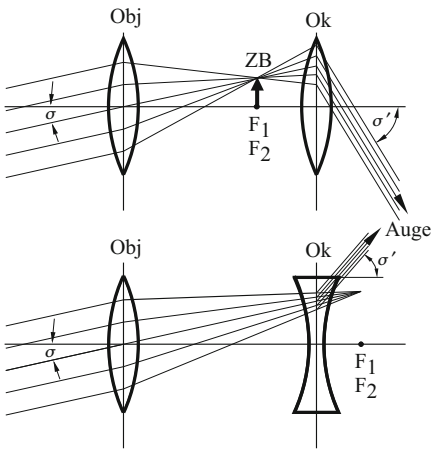


Abb. 16.21 Fernglas

eine Lupe wirkt, betrachtet. Die nahezu parallel austretenden Strahlen werden durch die Augenlinse auf der Netzhaut gebündelt. Eingezeichnet ist die sogenannte *Tubuslänge t*, der Abstand zwischen beiden Brennpunkten der Linsen. Die Gesamtvergrößerung ergibt sich zu:

Vergrößerung eines Mikroskops

$$\Gamma_M = \frac{t \cdot s_0}{f_1 \cdot f_2} \tag{16.23}$$

Fernglas

Beim Fernrohr liegt der zu betrachtende Gegenstand sehr weit entfernt, es treten parallele Strahlen in das Linsensystem. Es gibt verschiedene Anordnungen, solche, bei denen das Bild höhenverkehrt erscheint, diese werden in der Astronomie verwendet (Kepler'sches Fernrohr), und solche mit aufrechten Bildern (Galilei'sches Fernrohr). Der grundsätzliche Strahlenverlauf ist in Abb. 16.21 gezeichnet. Die Brennpunkte von Objektiv F_1 und Okular F_2 fallen zusammen. Die parallel austretenden Strahlen werden auch hier im Auge auf der Netzhaut fokussiert. Die Gesamtvergrößerung berechnet sich zu

$$\textit{Vergrößerung eines Fernrohrs} \quad \Gamma_F = -\frac{f_1}{f_2} \tag{16.24}$$

Werden die Brennweiten mit richtigem Vorzeichen eingesetzt, so ergibt sich beim Kepler'schen Fernrohr eine negative Vergrößerung, also eine Bildumkehr, beim Galilei'schen Fernrohr eine positive Vergrößerung, also ein aufrechtes Bild. Durch den Einbau einer weiteren Linse kann allerdings auch im Kepler'schen Fernrohr ein aufrechtes Bild erzielt werden; das Fernrohr wird dadurch nur länger. Eine Verlängerung des Fernrohres kann vermieden werden, indem der Strahl durch geeignete Prismen umgelenkt wird. Technische Angaben eines Fernglases sind die Vergrößerung und der Durchmesser D_E der Eintrittsöffnung, normalerweise des Objektivs. Die Angabe 7×50 besagt, dass die Vergrößerung $G_F = 7$ ist und die Eintrittsöffnung einen Durchmesser von 50 mm hat. Für den Durchmesser D_A der Austrittsöffnung gilt:

$$|\Gamma_F| = \frac{D_E}{D_A} \tag{16.25}$$

Die Leistungsfähigkeit bei Dämmerung wird nach DIN 58386 durch die *Dämmerungszahl Z* gekennzeichnet.

$$Z = \sqrt{\Gamma_F \cdot D_E} \tag{16.26}$$

Bei dem Fernglas 7×50 ist

$$D_A = \frac{50\,\text{mm}}{7} = 7{,}14\,\text{mm}$$

und

$$Z = \sqrt{7 \cdot 7{,}14} = 7$$

hingegen bei 8×30 gilt:

$$D_A = \frac{30\,\text{mm}}{8} = 3{,}75\,\text{mm}$$

$$Z = \sqrt{8 \cdot 3{,}75} = 5{,}47$$

somit ist das Glas 7×50 in der Dämmerung besser geeignet.

Die Augenpupille ist maximal 8 mm im Durchmesser, es hat also keinen Sinn, ein Fernrohr zu benutzen, bei dem $D_A > 8\,\text{mm}$ ist.

Fotoapparat

Ein Fotoapparat bildet – wie das Auge – den Gegenstand auf einer lichtempfindlichen Schicht, dem Film, ab. Das Objektiv besteht, um Linsenfehler zu vermeiden, aus Vielfach-Linsensystemen. Im Gegensatz zum Auge muss für unterschiedliche Entfernungen des Gegenstandes vom Objektiv der Abstand Film–Objektiv variiert werden, um eine scharfe Abbildung zu erreichen. Wie beim Auge kann die einfallende Lichtmenge durch eine variable Blende geregelt werden. Als technische Angabe für Objektive dient nach DIN 4521 das Verhältnis vom Durchmesser der Eintrittsöffnung zur Brennweite

$$\textit{relative Öffnung} = \frac{D_E}{f} \tag{16.27}$$

Steht z. B. auf einem Objektiv $1 : 2{,}8, f = 50\,\text{mm}$, so bedeutet dies bei einer Brennweite von 50 mm eine maximale Eintrittsöffnung von 17,85 mm. Der Kehrwert der relativen Öffnung ist die *Blendenzahl k* oder kurz *Blende*.

$$\textit{Blende} \quad k = \frac{f}{D_E} \tag{16.28}$$

Die nach DIN 4522 genormten Blendenzahlen sind: 1; 1,4; 2; 2,8; 4; 5,6; 8; 11; 16; 22. Sie sind so gewählt, dass sich bei Vergrößerung der Blendenzahl um einen Wert die Lichtintensität halbiert, daher ist der Faktor zwischen benachbarten Blendenwerten $\sqrt{2}$.

16.3 Wellenoptik

Die Eigenschaften der Wellen sind bereits im Kap. 14 besprochen worden, hier sollen spezielle Eigenschaften der Lichtwellen dargestellt werden.

16.3.1 Interferenz

Bei Überlagerung von Wellen gleicher Frequenz können sich die Wellen je nach Gangunterschied auslöschen oder verstärken. Da Licht eine sehr kleine Wellenlänge hat, ist dieser Effekt nicht so leicht zu sehen. Außerdem braucht man zur Interferenz Licht gleicher Wellenlänge und fester Phasenbeziehung φ. Normale Temperaturstrahler senden Licht nicht kontinuierlich aus, sondern in räumlich begrenzten Wellenpaketen. Zwei Wellenpakete, die von unterschiedlichen Orten einer Lichtquelle emittiert wurden, können nur dann interferieren, wenn sie zu gleicher Zeit am Interferenzort eintreffen. Entscheidend für die Interferenzfähigkeit ist, dass die Wegdifferenz nicht länger als die Wellenzüge ist. Der größte Gangunterschied, bei dem noch Interferenz möglich ist, ist die *Kohärenzlänge* l. In Tab. 16.4 sind Kohärenzlängen einiger typischer Lichtquellen aufgelistet.

Da aber bei normaler Lichtemission einzelne Bereiche von Lichtquellen das Licht ohne feste Phasenbeziehung abstrahlen, müssen bei Interferenzversuchen einige experimentelle Tricks angewendet werden. Licht gleicher Frequenz und fester Phasenbeziehung wird *kohärentes Licht* genannt. Sehr gut geeignete Lichtquellen für Interferenzversuche stehen heutzutage durch die *Laser* zur Verfügung. Diese liefern besonders scharf gebündelte Lichtstrahlen mit sehr guter Kohärenz. Interferenzerscheinungen können in speziellen Fällen auch im normalen Tageslicht beobachtet werden, z. B. rühren die schillernden Oberflächen von mit Öl bedecktem Wasser von Interferenz her. Dies soll an einer dünnen Schicht erläutert werden. In einer solch dünnen Schicht wird einfallendes Licht an den Grenzflächen vielfach reflektiert, ein Teil auch immer transmittiert. Die vom Beobachter aufgenommenen Lichtstrahlen haben dann unterschiedliche Wegstrecken zurückgelegt. Da weißes Licht

Tab. 16.4 Kohärenzlängen

Lichtquelle	Kohärenzlänge
Weißes Licht	1–2 µm
Spektrallampe	20 cm
Halbleiter-Laser	150 m
He-Ne-Laser	2 km

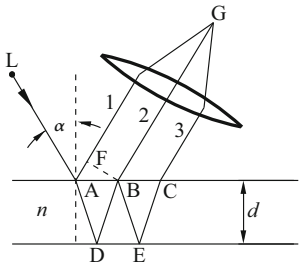


Abb. 16.22 Interferenz

einfällt, sind alle Wellenlängen vorhanden. Die Interferenzbedingung für Auslöschung kann unter einem festen Winkel nur für eine Wellenlänge erfüllt sein, daher wird, wenn diese Farbe ausgelöscht wird, aus dem weißen Licht farbiges Licht. Der Strahlengang ist in Abb. 16.22 dargestellt. Der von der Lichtquelle L ausgehende Lichtstrahl wird im Punkt A teilreflektiert und trifft als Strahl 1 im Punkt G auf. Ein Teil wird zum Punkt D gebrochen, dort teilreflektiert und vom Punkt B als Strahl 2 ebenfalls nach G gelenkt. Im Punkt G interferieren die Strahlen $1, 2, 3, \ldots$ Der geometrische Gangunterschied zwischen Strahl 1 und 2 beträgt $\overline{AD} + \overline{DB} - \overline{AF}$, der optische $n(\overline{AD} + \overline{DB}) - \overline{AF}$. Mit Hilfe des Brechungsgesetzes und der Tatsache, dass bei Reflexion an einem dichteren Medium ein Phasensprung von $\frac{\lambda}{2}$ auftritt, kann für den Gangunterschied Δ zwischen den Strahlen 1 und 2 abgeleitet werden:

$$\Delta = 2d\sqrt{n^2 - \sin^2\alpha} - \frac{\lambda}{2} \tag{16.29}$$

Auslöschung tritt dann ein, wenn die Bedingung

$$2d\sqrt{n^2 - \sin^2\alpha} = (m+1)\lambda, \quad m = 0, 1, 2, \ldots \tag{16.30}$$

erfüllt ist.

16.3.2 Beugung

Fällt auf eine Öffnung ein Bündel aus parallelem monochromatischem Licht, so ist hinter der Öffnung auch in den Bereichen, die im geometrischen Schattenbereich liegen, Lichtintensität zu beobachten. Diese Lichtbeugung lässt sich ebenfalls nur mit Hilfe der Wellennatur des Lichtes erklären. Bei Lichtwellen ist in der Regel die Öffnung größer als die Wellenlänge des eingestrahlten Lichtes.

Spalt

Als Öffnung wird ein Spalt der Breite b betrachtet (Abb. 16.23). Jeder Punkt des Spaltes, der von dem einfallenden Licht getroffen wird, ist nach dem *Huygens'schen Prinzip* (siehe Abschn. 14.2) Ausgangspunkt einer *Elementarwelle*. Die von den einzelnen Punkten ausgehenden Strahlen haben gegeneinander unterschiedliche Gangunterschiede. Wenn der Gangunterschied

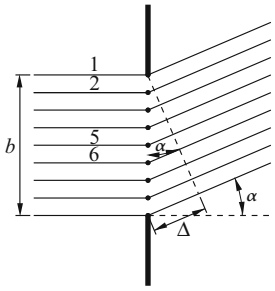


Abb. 16.23 Beugung am Spalt

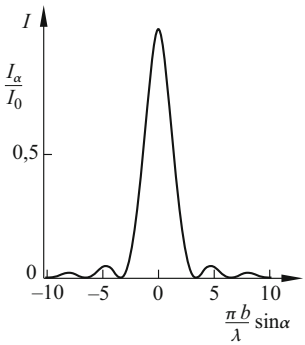


Abb. 16.24 Intensitätsverteilung hinter einem Spalt

zwischen den beiden Randstrahlen genau λ beträgt, ist der Gangunterschied zwischen einem Randstrahl und dem Mittenstrahl, also z. B. Strahl 1 und Strahl 5, $\lambda/2$, und diese beiden Strahlen löschen sich aus. Aber auch zwischen 2 und 6 besteht dann der Gangunterschied $\lambda/2$, und diese beiden löschen sich ebenfalls aus. So lässt sich in diesem Fall zu jedem Strahl immer ein anderer mit dem Gangunterschied $\lambda/2$ finden, somit löschen sich alle Strahlen aus und in dieser bestimmten Richtung herrscht Dunkelheit. Auslöschung tritt immer dann auf, wenn die Bedingung

$$\sin\alpha_{\min} = \frac{m\lambda}{b}, \quad m = 1, 2, 3, \ldots \tag{16.31}$$

erfüllt ist. Es kann gezeigt werden, dass die Intensität in einer bestimmten Richtung hinter einem Spalt nach der Formel

$$I_{\alpha\,\text{Spalt}} = I_0 \frac{\sin^2(\frac{\pi b}{\lambda}\sin\alpha)}{(\frac{\pi b}{\lambda}\sin\alpha)^2} \tag{16.32}$$

berechnet werden kann. Im Abb. 16.24 ist diese Intensitätsverteilung aufgetragen.

Die Maxima liegen näherungsweise in der Mitte zwischen zwei Nullstellen, für sie gilt

Maxima Beugungsbild

$$\sin\alpha_{\max} \approx \left(m + \frac{1}{2}\right)\frac{\lambda}{b}, \quad m = 1, 2, 3, \ldots \tag{16.33}$$

Die Beugung am Spalt liefert keine sehr scharfen Intensitätsmaxima.

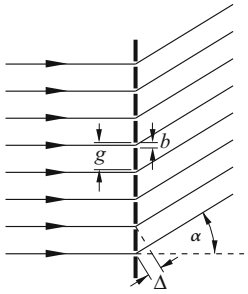


Abb. 16.25 Beugung am Gitter

Gitter

Ein optisches Gitter ist eine Anordnung von vielen gleichen Spalten. Zur Interferenz gelangen jetzt einmal die Strahlen aus einem Spalt, aber auch Strahlen aus benachbarten Spalten (s. Abb. 16.25). Das Beugungsbild hat jetzt ebenfalls Maxima und Minima, die Schärfe hängt von den Größen b und g ab, aber auch davon, wie viele Strahlen p zur Interferenz gelangen, d. h., wie viele Spalte beleuchtet werden. Die *Hauptmaxima*, das sind die Maxima, die durch Interferenz aus benachbarten Spalten entstehen, liegen so, dass die Bedingung

$$\sin\alpha_{\max} = \frac{m\lambda}{g}, \quad m = 0, 1, 2, \ldots \tag{16.34}$$

erfüllt ist. Die gesamte Intensitätsverteilung ist eine Funktion, die aus Überlagerung von Interferenz am einzelnen Spalt und aus benachbarten Spalten entsteht. Sie lautet

$$I_{\alpha\,\text{Gitter}} = I_{\alpha\,\text{Spalt}} \cdot \frac{\sin^2(\frac{p g\pi}{\lambda}\sin\alpha)}{\sin^2(\frac{g\pi}{\lambda}\sin\alpha)} \tag{16.35}$$

Die Lage der Maxima ist von der Wellenlänge abhängig, somit können Gitter zur Trennung von verschiedenen Wellenlängen bzw. Farben verwendet werden, indem hinter dem Gitter durch geeignet angebrachte Blenden bis auf einen Wellenlängenbereich $\Delta\lambda$ das Licht ausgeblendet wird. Dies wird in *Gittermonochromatoren* angewendet.

Polarisation

Da Licht eine *elektromagnetische Transversalwelle* ist, ist es möglich, die Schwingungsrichtung so zu beeinflussen, dass der elektromagnetische Feldvektor $\vec{E}$ nur in einer Ebene schwingt. Da der magnetische Feldvektor $\vec{H}$ immer senkrecht zum $\vec{E}$-Vektor steht, schwingt auch dieser in einer Ebene. Solch ein Licht wird *linear polarisiertes Licht* genannt. Erzeugen kann man linear polarisiertes Licht u. a. durch *Polarisationsfilter*. Dies sind Filter, in denen lange organische Molekülketten liegen. In Richtung der Ketten kann dem einfallenden natürlichen Licht durch Energieübertragung an die Elektronen in den Molekülketten Energie entzogen werden, diese Schwingungsrichtung fehlt dann im Licht hinter dem Filter. Da jede Schwingungsrichtung als vektorielle Summe von zwei ausgezeichneten senkrecht zueinander stehenden Richtungen zusammengesetzt werden kann, liegt bei Fehlen einer dieser beiden Richtungen nur

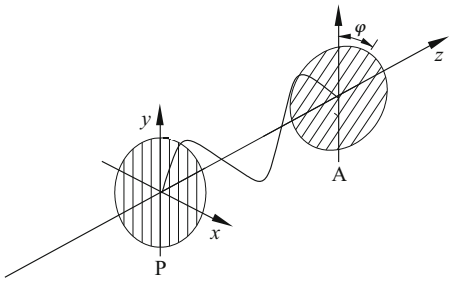


Abb. 16.26 Linear polarisiertes Licht

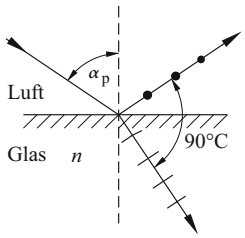


Abb. 16.27 Polarisation durch Reflexion

noch eine Schwingungsrichtung in einer Ebene vor. Trifft Licht der Intensität I_0 auf einen solchen Polarisator P, so wird die Intensität nach dem Polarisator I_P halbiert, da ja eine Schwingungsrichtung fehlt.

$$I_P = \frac{I_0}{2} \tag{16.36}$$

Wird dieses linear polarisierte Licht durch ein zweites Polarisationsfilter A (*Analysator*) geschickt, so ist die Intensität hinter dem Analysator abhängig von der Verdrehung des Analysators gegen den Polarisator.

Intensität hinter dem Analysator

$$I_A = I_P \cdot \cos^2 \varphi \tag{16.37}$$

Ist die Ausbreitungsrichtung des Lichtes die z-Richtung, die Schwingungsrichtung des $\vec{E}$-Vektors die x-Richtung, dann schwingt der $\vec{H}$-Vektor in y-Richtung. Die in Abb. 16.26 eingezeichneten Linien im Polarisator geben die Durchlassrichtung an; diese steht senkrecht auf den Molekülketten.

Auch durch Reflexion an einer Glasplatte kann Licht linear polarisiert werden (Abb. 16.27). Durch den einfallenden und reflektierten Strahl wird eine Ebene festgelegt, die *Einfallsebene*. Fällt natürliches Licht unter einem bestimmten Winkel, dem *Brewster'scher Winkel* oder auch *Polarisationswinkel* α_p, auf die Glasplatte, so fehlt dann im reflektierten Strahl eine Schwingungsebene, und zwar die Ebene parallel zur Einfallsebene. Dies ist dann der Fall, wenn reflektierter und gebrochener Strahl senkrecht zueinander stehen. Für den Polarisationswinkel gilt:

$$\sin \alpha_p = n \cdot \sin(90° - \alpha_p) \tag{16.38}$$

$$\sin \alpha_p = n \cdot \cos \alpha_p \tag{16.39}$$

Polarisationswinkel $$n = \tan \alpha_p \tag{16.40}$$

Das reflektierte Licht ist dann senkrecht zur Einfallsebene linear polarisiert. Dieser Effekt wird z. B. in der Fotografie benutzt, um störende Reflexionen zu unterdrücken: Durch ein Polarisationsfilter vor dem Kameraobjektiv (Analysator) kann die linear polarisierte reflektierte Intensität durch richtige Drehung des Analysators beeinflusst werden.

16.4 Photometrie

In der geometrischen Optik wird allein die Frage nach Abbildungen und Strahlengängen untersucht, nicht aber Größen wie Lichtintensität, Helligkeit oder Strahlleistung. Diese Begriffe werden in der Photometrie definiert und angewendet. Dabei interessiert die gesamte Strahlungsleistung sowie deren räumliche und spektrale Verteilung. In der *objektiven Photometrie* wird die Strahlungsleistung mit einem Messinstrument gemessen, die Größen erhalten den Index „*e*“ (energetisch). Wird die Strahlungsleistung mit dem Auge bewertet, spricht man von *subjektiver Photometrie*, und die Größen erhalten den Index „*v*“ (visuell).

16.4.1 Strahlungsphysikalische Größen

Unter Licht wird die elektromagnetische Strahlung im Wellenlängenbereich von 380 bis 780 nm verstanden. Die Begriffe sind in der DIN 5031 festgelegt. Es gibt verschiedene Effekte, Strahlungsenergie in andere Energieformen umzuwandeln und so der Messung zugänglich zu machen, z. B. Erwärmung, Änderung des Widerstandes von Halbleitern oder Entstehung eines *Fotostromes* in *Fotodioden*. Die auf einen Detektor treffende *Strahlungsleistung* Q_e, auch *Strahlungsfluss* oder *Lichtstrom* genannt, ist die pro Zeiteinheit Δt auftreffende *Strahlungsenergie* ΔQ_e.

Strahlungsleistung

$$\Phi_e = \frac{\Delta Q_e}{\Delta t} \qquad \begin{array}{c|c|c} \Phi_e & \Delta Q_e & \Delta t \\ \hline \mathrm{W} & \mathrm{J} & \mathrm{s} \end{array} \tag{16.41}$$

Die Strahlungsleistung, die auf einen Empfänger trifft, hängt bei konstantem Sender von der Fläche des Empfängers, von der Richtung der Empfängerfläche und vom Abstand vom Sender ab. Da die Strahlung sich räumlich ausbreitet, ist der *Raumwinkel* ein Maß für den erfassten Bereich. Zur Definition des Raumwinkels kann man sich eine punktförmige Lichtquelle im Zentrum einer Kugel vorstellen. Die gesamte Strahlungsleistung verteilt sich auf die gesamte Kugeloberfläche. Wird eine Teilfläche A bestrahlt, so ist der zugehörige *Raumwinkel* Ω nach der Gleichung (Abb. 16.28)

Raumwinkel $$\Omega = \frac{A}{r^2} \qquad \begin{array}{c|c|c} \Omega & A & r^2 \\ \hline \mathrm{sr} & \mathrm{m^2} & \mathrm{m^2} \end{array} \tag{16.42}$$

definiert.

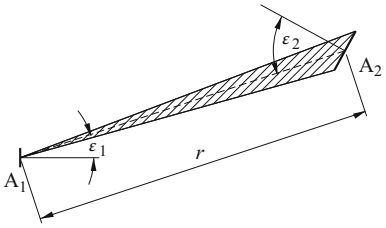


Abb. 16.28 Raumwinkel

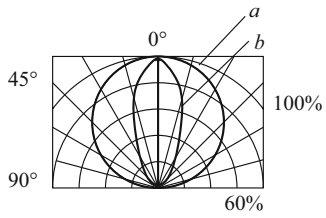


Abb. 16.29 Strahlcharakteristik

Der Raumwinkel ist dimensionslos, aber der Deutlichkeit halber wird er in der Einheit 1 sr, 1 *Steradiant*, angegeben.

$$\Omega = \frac{A}{r^2} \cdot \Omega_0 \tag{16.43}$$

mit $\Omega_0 = 1\,\text{sr}$. Da die Kugeloberfläche $O = 4\pi r^2$ ist, gilt für den gesamten Raumwinkel:

gesamter Raumwinkel

$$\Omega_{\max} = \frac{4\pi r^2}{r^2} = 4\pi\ sr \tag{16.44}$$

Streng genommen gilt diese Definition für eine kugelförmige Empfängerfläche. Ist jedoch der Abstand Sender–Empfänger groß und ist die Empfängerfläche klein, so kann der Empfänger als ebene Fläche aufgefasst werden. Die in den folgenden Formeln benötigten Winkel sind die zwischen der Flächennormalen und der Strahlrichtung. Der Sender hat die Fläche A_1, der Empfänger A_2. Der schraffiert eingetragene Raumwinkel ist

$$\Omega = \frac{A_2 \cdot \cos\varepsilon_2}{r^2} \cdot \Omega_0 \tag{16.45}$$

Der Faktor $A_2 \cdot \cos\varepsilon_2$ gibt die wirksame senkrecht zur Strahlrichtung stehende Fläche an. Ein Sender strahlt die *Strahlstärke* I_e ab. Die Einheit der Strahlstärke I_e ist 1 $\frac{\text{W}}{\text{sr}}$. Für die gesamte vom Empfänger aufgenommene Strahlungsleistung gilt dann:

Strahlungsleistung

$$\Phi_e = I_e \cdot \Omega \qquad \begin{array}{c|c|c} \Phi_e & I_e & \Omega \\ \hline \text{W} & \text{W/sr} & \text{sr} \end{array} \tag{16.46}$$

Die gesamte vom Sender in einen bestimmten Raumwinkel ausgesendete Strahlstärke und somit der Strahlungsfluss hängt von der wirksamen Fläche von A_1 und von der pro Flächeneinheit ausgesendeten Strahlleistung, der *Strahldichte* L_e, ab.

Strahldichte

$$L_e = \frac{\Phi_e}{A\Omega} \qquad \begin{array}{c|c|c|c} L_e & \Phi_e & A & \Omega \\ \hline \text{W/(m}^2\,\text{sr)} & \text{W} & \text{m}^2 & \text{sr} \end{array} \tag{16.47}$$

Strahlstärke

$$I_e(\varepsilon_1) = L_e A_1 \cos\varepsilon_1 \qquad \begin{array}{c|c|c} I_e & L_e & A \\ \hline \text{W/sr} & \text{W/(m}^2\,\text{sr)} & \text{m}^2 \end{array} \tag{16.48}$$

Die Strahlstärke als Funktion des Abstrahlungswinkels wird oft in Datenblättern angegeben. In Abb. 16.29 sind im Polardiagramm für zwei unterschiedliche Strahlertypen die Strahlcharakteristik dargestellt, Kurve *b* ist die einer Leuchtdiode mit starker Vorwärtsstrahlung, Kurve *a* gehört zu einem Strahler mit konstanter Strahldichte L_e, einem sogenannten *Lambert-Strahler*. Für diesen gilt

Lambert'sches Gesetz

$$I_e(\varepsilon_1) = I_e(\varepsilon_0) \cdot \cos\varepsilon_1 \tag{16.49}$$

Zu diesem Strahlertyp gehören alle Körper mit rauen, diffus reflektierenden Flächen. Sie erscheinen aus allen Richtungen gleich hell.

Für die auf den Empfänger treffende Strahlungsleistung gilt dann unter Zusammenfassung aller bisherigen Gleichungen:

$$\Phi_e = L_e \frac{A_1 \cos\varepsilon_1 A_2 \cos\varepsilon_2}{r^2} \Omega_0 \tag{16.50}$$

Unter der spezifischen Ausstrahlung wird die pro Flächeneinheit emittierte Strahlungsleistung definiert:

$$M_e = \frac{\Phi_e}{A_1} = L_e \cos\varepsilon_1 \Omega \tag{16.51}$$

Auf der Empfängerseite ist die Bestrahlungsstärke E_e, der auf die Empfängerfläche bezogene Strahlungsfluss, wichtig.

$$E_e = \frac{\Phi_e}{A_2} \tag{16.52}$$

Ist die Strahlung über einem größeren Wellenlängenbereich vorhanden, so werden spektrale Größen zur Beschreibung nötig sein. Dies wird durch den Index λ gekennzeichnet, so ist z. B. die spektrale Strahlungsdichte durch $L_{e,\lambda}$ gekennzeichnet und definiert durch

spektrale Strahldichte

$$L_{e,\lambda} = \frac{\Delta L_e}{\Delta\lambda} \qquad \begin{array}{c|c|c} L_{e,\lambda} & \Delta L_e & \Delta\lambda \\ \hline \text{W/(m}^2\,\text{sr nm)} & \text{W/(m}^2\,\text{sr)} & \text{nm} \end{array} \tag{16.53}$$

$L_{e,\lambda}$ ist also die im Wellenlängenintervall $\Delta\lambda$ emittierte Strahldichte ΔL_e. Besonderes Interesse hat die spektrale Verteilung der Wärmestrahlung von Körpern. Jeder Körper sendet Wärmestrahlung aus, deren spektrale Verteilung (also deren Farbe) von

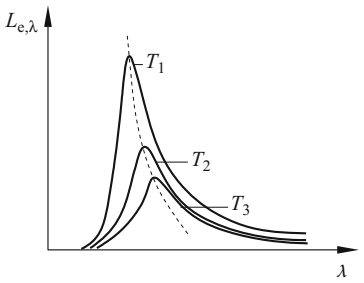


Abb. 16.30 Schwarzer Strahler

der Temperatur abhängt. Für einen idealisierten Körper, nämlich einen solchen, der alle auftreffende Strahlung vollständig absorbiert, kann die spektrale Strahldichte berechnet werden. Diese Körper werden *schwarze Körper* genannt. Für diese Körper gilt die *Planck'sche Strahlungsformel* für die Temperaturabhängigkeit der Strahldichte:

Planck'sche Strahlungsformel

$$L_{e,\lambda}(\lambda, T) = \frac{c_1}{\lambda^5} \frac{1}{e^{c_2/\lambda T} - 1} \frac{1}{\Omega_0} \tag{16.54}$$

Die Konstanten c_1 und c_2 hängen mit der Lichtgeschwindigkeit c und einer weiteren Naturkonstanten, dem *Planck'schen Wirkungsquant* h und der *Boltzmann-Konstante* k zusammen:

$$c_1 = 2hc^2 = 1{,}191 \cdot 10^{-16}\,\mathrm{W\,m^2} \tag{16.55}$$

$$c_2 = \frac{hc}{k}^2 = 1{,}439 \cdot 10^{-2}\,\mathrm{m\,K} \tag{16.56}$$

Planck'sches Wirkungsquant

$$h = 6{,}626 \cdot 10^{-34}\,\mathrm{J\,s} \tag{16.57}$$

Boltzmann-Konstante

$$k = 1{,}381 \cdot 10^{-23}\,\mathrm{J/K} \tag{16.58}$$

Ein typischer Verlauf ist in Abb. 16.30 dargestellt. Es sind hier die prinzipiellen Kurvenverläufe für drei Temperaturen dargestellt, wobei $T_1 > T_2 > T_3$ ist. Das Maximum der spektralen Strahldichte $l_{\max}$ verschiebt sich mit höheren Temperaturen zu kleineren Wellenlängen entlang der gestrichelt gezeichneten Hyperbel. Es gilt das

Wien'sche Verschiebungsgesetz:

$$\lambda_{\max} \cdot T = const = 2898\,\mu\mathrm{m\,K} \tag{16.59}$$

Betrachtet man nicht die spektrale Strahldichte, sondern die insgesamt von einem schwarzen Strahler ausgesendete Leistung, so gilt das *Stefan-Boltzmann'sche Gesetz*:

$$M_e(T) = \sigma \cdot T^4 \tag{16.60}$$

mit der Konstanten

$$\sigma = 5{,}670 \cdot 10^{-8} \frac{\mathrm{W}}{\mathrm{m^2\,K^4}} \tag{16.61}$$

16.4.2 Lichttechnische Größen

Werden Strahlungen nicht mit objektiven Instrumenten, sondern mit dem Auge beurteilt, so muss die spektrale Lichtempfindlichkeit des Auges berücksichtigt werden. Strahlen z. B. eine rote und eine grüne Lichtquelle mit gleicher Strahlungsleistung, so empfindet das Auge die grüne Quelle ca. 16mal heller. Das Auge hat sich im Lauf der Entwicklung dem Sonnenlicht angepasst, und somit für Grün die größte Empfindlichkeit entwickelt. Die Augenempfindlichkeit ist bei Tageslicht eine andere als bei Nacht, hier ist sie zu kleineren Wellenlängen, also zu Blau hin, verschoben. Die Augenempfindlichkeit für Tageslicht wird mit dem *Hellempfindlichkeitsgrad* $V(\lambda)$ und die für Nachtlicht mit $V'(\lambda)$ bezeichnet. Bei Tageslicht sind die Zäpfchen (*fotopische Anpassung*), bei Nachtsehen die Stäbchen (*skoptopische Anpassung*) angesprochen. Wird mit dem Lichtstrom Φ_v der Helligkeitseindruck bezeichnet, so gilt bei einfarbigen Lichtquellen:

$$\Phi_v = K_m \Phi_e V(\lambda) \tag{16.62}$$

Die Konstante K_m wird als Maximalwert des *photometrischen Strahlungsäquivalents* bezeichnet. Sie ist in einer noch zu definierenden Einheit:

$$K_m = 683 \frac{\mathrm{lm}}{\mathrm{W}} \quad (1\,\mathrm{lm} = 1\ \mathrm{Lumen}) \tag{16.63}$$

Diese Größe ist mit der SI-Einheit Candela (s. Kap. 10) verknüpft. Licht mit der Frequenz $f = 540 \cdot 10^{12}$ Hz hat eine Wellenlänge von 555 nm. Der Hellempfindlichkeitsgrad $V(555) = 1$, somit gilt für die Lichtstärke 1 Candela:

$$I_v = 1\,\mathrm{cd} = K_m I_e = K_m \frac{1}{683} \frac{\mathrm{W}}{\mathrm{sr}} \tag{16.64}$$

und damit der o. a. Wert für K_m. Als abgeleitete Einheiten sind für den *Lichtstrom* das *Lumen* $1\,\mathrm{lm} = 1\,\mathrm{cd\,sr}$ und für die *Beleuchtungsstärke* E_v das *Lux* $1\,\mathrm{lx} = 1\,\mathrm{lm/m^2}$ erlaubt.

Zusammenfassung in Tab. 16.5.

Beispiel

Welchen Lichtstrom empfängt eine 20 cm × 30 cm große Fläche, auf die das Licht einer 2,40 m entfernten punktförmigen Lichtquelle von 80 cd senkrecht auftrifft?

Lösung:

Raumwinkel

$$\Omega = \frac{A}{r^2} = \frac{0{,}2\,\mathrm{m} \cdot 0{,}3\,\mathrm{m}}{(2{,}4\,\mathrm{m})^2} = 0{,}0104\,\mathrm{sr}$$

$$\Phi_v = I_e \cdot \Omega = 80\,\mathrm{cd} \cdot 0{,}0104\,\mathrm{sr} = 0{,}833\,\mathrm{lm} \quad \blacktriangleleft$$

Tab. 16.5 Photometrische Größen

Strahlungsphysikalische Größen			Lichttechnische Größen		
Größe	Zeichen	Einheit	Größe	Zeichen	Einheit
Strahlungsenergie	Q_e	W s	Lichtmenge	Q_v	lm s
Strahlungsleistung	Φ_e	W	Lichtstrom	Φ_v	lm
Spez. Ausstrahlung	M_e	$\mathrm{W/m^2}$	Spez. Lichtausstrahlung	M_v	$\mathrm{lm/m^2}$
Strahlstärke	I_e	W/sr	Lichtstärke	I_v	cd = lm/sr
Strahldichte	L_e	$\mathrm{W/(m^2\,sr)}$	Leuchtdichte	L_v	$\mathrm{cd/m^2}$
Bestrahlungsstärke	E_e	$\mathrm{W/m^2}$	Beleuchtungsstärke	E_v	$\mathrm{lx = lm/m^2}$
Bestrahlung	H_e	$\mathrm{W\,s/m^2}$	Belichtung	H_v	lx s

Beispiel

Welche Lichtstärke I_v muss eine Lichtquelle haben, damit an einem $r = 1{,}5\,\mathrm{m}$ entfernten Arbeitsplatz die Beleuchtungsstärke $E_v = 500\,\mathrm{lx}$ beträgt?

Lösung:

$$I_v = \frac{E_v r^2}{\Omega_0} = \frac{500\,\mathrm{lx}\cdot(1{,}5\,\mathrm{m})^2}{1\,\mathrm{sr}} = 1125\,\mathrm{cd} \quad \blacktriangleleft$$

16.5 Licht als Korpuskel

Trifft Licht auf eine negativ geladene Metallplatte, so können unter geeigneten Bedingungen Elektronen ausgelöst werden. Dieser Effekt ist der *Photoeffekt* oder auch *lichtelektrische Effekt*. Entdeckt wurde dieser Effekt bereits 1887, richtig erklärt konnte er aber erst werden, nachdem die Korpuskel-Eigenschaften des Lichtes verstanden waren. Nach der Wellentheorie ist es zwar ebenfalls möglich, dass eine Lichtwelle aus einem Metall Elektronen herausschlagen kann, nur muss dann die Bewegungsenergie der Elektronen im Vakuum von der Intensität des eingestrahlten Lichtes abhängen. Gemessen wurde aber, dass die Energie der Elektronen von der Frequenz des Lichtes abhängt. Die Aussage der *Quantenphysik* ist, dass jedem Materie-Teilchen eine Welle und jeder Welle ein Materie-Teilchen zugeordnet werden kann. Diese *Lichtquanten* heißen *Photonen*. Die Photonen haben die Eigenschaft, dass sie sich im Vakuum mit Lichtgeschwindigkeit bewegen, eine Bewegungsenergie und eine Bewegungsgröße oder auch Impuls haben. Der Zusammenhang zwischen beiden Betrachtungsweisen erfolgt über die Grundgleichungen

$$\textit{Energie} \qquad W = h\cdot f \tag{16.65}$$

$$W = \frac{hc}{\lambda} \tag{16.66}$$

$$\textit{Impuls} \qquad p = \frac{h}{\lambda} \tag{16.67}$$

h ist das *Planck'sche Wirkungsquant*, W und p sind Energie und Impuls, die Teilchengrößen f und λ sind Frequenz und Wellenlänge des Lichtes, also die Wellengrößen. Trifft nun eine Lichtwelle der Frequenz f auf eine geladene Metallplatte, so kann das entsprechende Photon die Energie W an die Elektronen des Metalls abgeben. Ist die Energie groß genug, um ein Metallelektron aus der Metalloberfläche herauszuschlagen, so wird das Metall entladen. Dazu ist es notwendig, dass das Photon eine Mindestenergie hat, nämlich die Energie, die benötigt wird, um ein Elektron aus dem betreffenden Material austreten zu lassen. Diese Energie heißt *Austrittsenergie* oder *Austrittsarbeit*. Die Bewegungsenergie des Elektrons im Vakuum ist dann gleich der Energie des eingestrahlten Lichtes, vermindert um die Austrittsarbeit.

$$W_{\mathrm{kin}} = h\cdot f - W_A \tag{16.68}$$

Der Photoeffekt ist erst dann bemerkbar, wenn das Licht eine Grenzenergie oder Grenzfrequenz f_{Gr} bzw. Grenzwellenlänge λ_{Gr} hat. Die hier gebräuchliche Einheit der Energie ist das *Elektronenvolt*.

Aus der Definition der Energie in der Elektrotechnik

$$\textit{elektrische Energie} \quad W_{\mathrm{el}} = Q\cdot U \tag{16.69}$$

ist die Energie einer Ladung abhängig von der Spannung, mit der sie beschleunigt wurde. Die Ladung eines Elektrons ist die *Elementarladung*:

$$e_0 = 1{,}602\cdot 10^{-19}\,\mathrm{A\,s} \tag{16.70}$$

Wird dieses Elektron mit der Spannung 1 V beschleunigt, so hat es die Energie:

$$W_{\mathrm{kin}} = e_0\cdot U \tag{16.71}$$

$$W_{\mathrm{kin}} = 1{,}602\cdot 10^{-19}\,\mathrm{W\,s} \tag{16.72}$$

$$1\,\mathrm{eV} = 1{,}602\cdot 10^{-19}\,\mathrm{W\,s} \tag{16.73}$$

Die Austrittsarbeit hängt vom Material und von der kristallographischen Orientierung der Oberfläche ab. Typische angenäherte Werte sind in Tab. 16.6 angegeben.

Die Grenzwellenlänge errechnet sich nach (16.66) zu

Grenzwellenlänge

$$\lambda_{\mathrm{Gr}} = \frac{hc}{W_A} \tag{16.74}$$

$$\lambda_{\mathrm{Gr}} = \frac{6{,}627\cdot 10^{-34}\,\mathrm{J\,s}\cdot 3\cdot 10^{8}\,\mathrm{m/s}}{1{,}602\cdot 10^{-19}\,\mathrm{J/eV}}\cdot\frac{1}{W_A} \tag{16.75}$$

$$\lambda_{\mathrm{Gr}} = 1241\cdot\frac{1}{W_A/\mathrm{eV}} \quad (\lambda_{\mathrm{Gr}}\text{ in nm}, W_A\text{ in eV}) \tag{16.76}$$

Tab. 16.6 Austrittsarbeit

Metall	W_A in eV	λ_{Gr} in nm
W	5,0	248
Ni	4,6	270
Ag	4,5	275
Au	5,4	229
Cu	4,6	270
Cs	2,0	620
Cs auf W	1,35	920

Dabei ist die Austrittsarbeit in eV anzugeben, die Wellenlänge wird in nm berechnet. Da die Ladungsträger ins Vakuum austreten, nennt man diesen Effekt auch den *äußeren Photoeffekt*. Daneben gibt es noch den *inneren Photoeffekt*. Hierbei wird ebenfalls die Energie der einfallenden Photonen an Elektronen abgegeben, nur treten diese z. B. bei Fototransistoren, Fotodioden oder Photowiderständen nicht ins Vakuum, sondern werden im Festkörper auf andere Energiezustände gehoben und verändern somit die Leitfähigkeit des Bauteils.

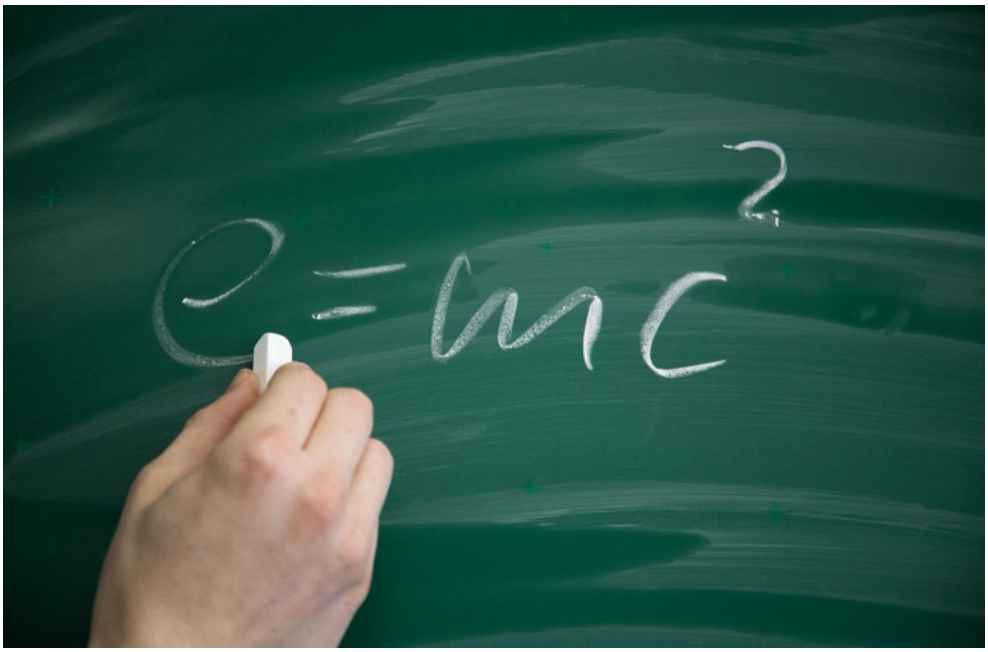
Literatur

1. *Hering, E.; Martin, R.; Stohrer, M.:* Physik für Ingenieure: Springer, 2007
2. *Meschede, D. (Hrsg.):* Gerthsen Physik: Springer, 2006
3. *Böge, A.; Eichler, J.:* Physik: Vieweg, 2005
4. *Lindner, H.:* Physik für Ingenieure: Fachbuchverlag Leipzig im Carl-Hanser-Verlag, 2006
5. *Dobrinski, P.; Krakau, G.; Vogel, A.:* Physik für Ingenieure: Teubner, 2007

Anhang

17

Horst Steffen



17.1 Physikalische Größen und Einheiten 254
17.2 Zahlenwerte physikalischer Größen 254

Teil II

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_17

17.1 Physikalische Größen und Einheiten

Tab. 17.1 Abgeleitete Größen

Bezeichnung	Symbol	Einheit
Arbeit, Energie	W	$1\,\mathrm{J} = 1\,\mathrm{N\,m} = 1\,\mathrm{W\,s} = 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^2}}$
Beleuchtungsstärke	E	$1\ \mathrm{Lux} = 1\ \mathrm{lx}$
Beschleunigung	a	$1\,\frac{\mathrm{m}}{\mathrm{s^2}}$
Bewegungsgröße Impuls	p	$1\,\frac{\mathrm{kg\,m}}{\mathrm{s}}$
Massendichte	ρ	$1\,\frac{\mathrm{kg}}{\mathrm{m^3}}$
Drehimpuls	L	$1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s}}$
Drehmoment	M	$1\,\mathrm{N\,m} = 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^2}}$
Druck	p	$1\,\mathrm{Pa} = 1\,\frac{\mathrm{N}}{\mathrm{m^2}} = 1\,\frac{\mathrm{kg}}{\mathrm{m\,s^2}}$
Elastizitätsmodul	E	$1\,\frac{\mathrm{N}}{\mathrm{m^2}} = 1\,\frac{\mathrm{kg}}{\mathrm{m\,s^2}}$
Frequenz	f	$\frac{1}{\mathrm{s}} = 1\ \mathrm{Hz}$
Geschwindigkeit	v	$1\,\frac{\mathrm{m}}{\mathrm{s}}$
Kraft	F	$1\,\mathrm{N} = 1\,\frac{\mathrm{kg\,m}}{\mathrm{s^2}}$
Leistung	P	$1\,\mathrm{W} = 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^3}}$
Lichtstrom	Φ	$1\ \mathrm{Lumen} = 1\ \mathrm{lm} = 1\ \mathrm{cd\,sr}$
Raumwinkel	Ω	$1\ \mathrm{Steradiant} = 1\ \mathrm{sr}$
Spez. Wärmekapazität	c	$1\,\frac{\mathrm{J}}{\mathrm{kg\,K}} = 1\,\frac{\mathrm{m^2}}{\mathrm{s^2\,K}}$
Trägheitsmoment	J	$1\ \mathrm{kg\,m^2}$
Winkelgeschwindigkeit	ω	$\frac{1}{\mathrm{s}}$

17.2 Zahlenwerte physikalischer Größen

Tab. 17.2 Naturkonstanten

Avogadro-Konstante	N_A	$6{,}0221367 \cdot 10^{23}\ \mathrm{mol^{-1}}$
Boltzmann-Konstante	k	$1{,}380658 \cdot 10^{-23}\ \mathrm{J/K}$
Elektrische Feldkonstante	ε_0	$8{,}8541878 \cdot 10^{-12}\ \mathrm{A\,s/(V\,m)}$
Elementarladung	e	$1{,}60218 \cdot 10^{-19}\ \mathrm{A\,s}$
Gravitationskonstante	γ	$6{,}67259 \cdot 10^{-11}\ \mathrm{N\,m^2/kg^2}$
Lichtgeschwindigkeit im Vakuum	c	$2{,}99792458 \cdot 10^{8}\ \mathrm{m/s}$
Magnetische Feldkonstante	μ_0	$4\pi \cdot 10^{-7}\ \mathrm{(V\,s)/(A\,m)}$
Planck'sches Wirkungsquantum	h	$6{,}62607 \cdot 10^{-34}\ \mathrm{J\,s}$
Ruhemasse des Elektrons	m_{0e}	$9{,}1093897 \cdot 10^{-31}\ \mathrm{kg}$
Ruhemasse des Protons	m_{0P}	$1{,}6726231 \cdot 10^{-27}\ \mathrm{kg}$
Stefan-Boltzmann-Konstante	σ	$5{,}67051 \cdot 10^{-8}\ \mathrm{W/(m^2\,K^4)}$
Universelle Gaskonstante	R	$8{,}31441\ \mathrm{J/(mol\,K)}$

Nach dem von *K. von Klitzing* 1980 entdeckten quantisierten Halleffekt (Nobelpreis 1985) lässt sich auch für den elektrischen Widerstand R aus Naturkonstanten eine exakte Basisgröße bestimmen.

$$R = \frac{h}{ie^2}, \quad i = 1,2,3,\ldots \tag{17.1}$$

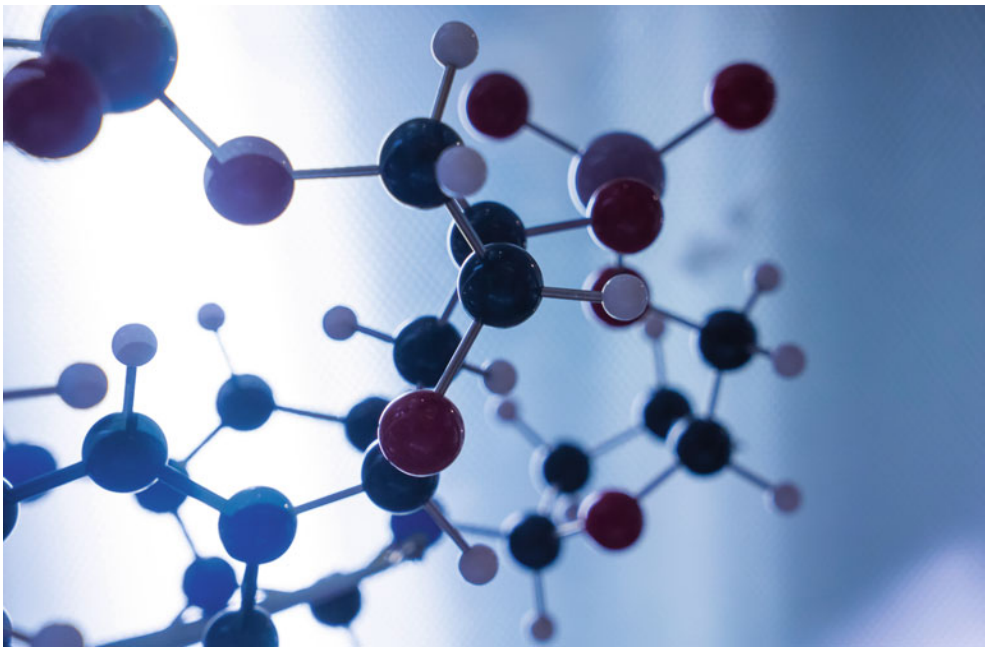
$$R = 25812{,}8\,\Omega \tag{17.2}$$

Tab. 17.3 Zahlenwerte

Molvolumen idealer Gase	$22{,}41383 \cdot 10^{-3}\ \frac{\mathrm{m^3}}{\mathrm{mol}}$
Radius der Erde, Äquators	$r_E = 6378\ \mathrm{km}$
Radius der Erde, Pole	$r_P = 6356\ \mathrm{km}$
Masse der Erde	$m_E = 5{,}977 \cdot 10^{24}\ \mathrm{kg}$
Masse des Mondes	$m_M = 7{,}35 \cdot 10^{22}\ \mathrm{kg}$
Entfernung Erde–Mond	$384{,}4 \cdot 10^{3}\ \mathrm{km}$
Entfernung Erde–Sonne	$149{,}6 \cdot 10^{6}\ \mathrm{km}$
Erdbeschleunigung	$g = 9{,}80665\ \frac{\mathrm{m}}{\mathrm{s^2}}$
Tripelpunkt des Wassers	$T_{Tr} = 273{,}16\ \mathrm{K}$
Dichte der Luft	$\rho_{\mathrm{Luft}} = 1{,}2928\ \frac{\mathrm{kg}}{\mathrm{m^3}}$
Absolute Temperatur	$-273{,}15\ ^\circ\mathrm{C}$
Normaldruck	$p_0 = 1{,}01325 \cdot 10^{5}\ \mathrm{Pa}$

Werkstoffkunde

III

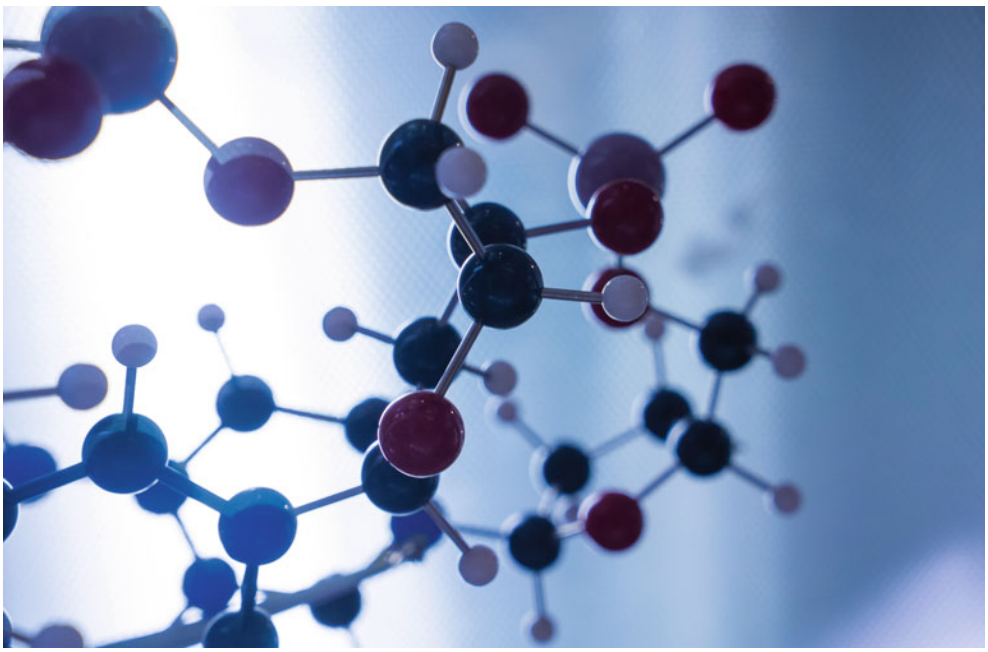


18 Stoffe 257
19 Elektrische Leitfähigkeit 267
20 Elektrische Leiter 273
21 Magnetische Leitfähigkeit 279
22 Magnetika 285
23 Dielektrische Eigenschaften 293
24 Dielektrika 297

Stoffe

18

Egon Döring



18.1 Eigenschaften der Stoffe 258
18.2 Atombau und Periodensystem 258
18.3 Aufbau der festen Körper 260
18.4 Chemische Grundzusammenhänge 262
18.5 Elektrochemie 264

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_18

18.1 Eigenschaften der Stoffe

Der Begriff des Stoffes entstammt der Chemie, ist aber für den technischen Gebrauch als *Werkstoff* auf die praktisch nutzbaren Stoffe begrenzt. Die Anforderungen an die Eigenschaften eines Werkstoffes und ihre praktische Erfüllung bestimmen dann seine Nutzbarkeit. In der Elektrotechnik erscheint die Antwort hierzu recht einfach: Da „nur“ elektrische und magnetische Felder auftreten können, müssten elektrische und magnetische Eigenschaften ausreichen. Elektrische und besonders auch magnetische Vorgänge führen aber dazu, dass ein zusätzliches Anforderungsprofil entsteht wie es in der Mechanik üblich ist (Tab. 18.1).

Dem Anforderungskatalog stehen die Eigenschaften der, vorzugsweise festen, Werkstoffe gegenüber (Tab. 18.2). Für die Praxis müssen diese Eigenschaften durch messbare Größen charakterisiert werden, um die richtige Auswahl treffen zu können. Wegen ihrer Grundeigenschaften *elektrische Leitfähigkeit* und hohe *Festigkeit* bei guter *Verformbarkeit* nehmen die Metalle eine Sonderstellung unter den Werkstoffen ein. Darüber hinaus sind von den zurzeit bekannten 111 Elementen die meisten Metalle. Nur 15 Elemente zählen zu den *Nichtmetallen*, und ungefähr 8 Elemente fallen unter die heute besonders interessanten *Halbmetalle* oder *Übergangselemente*. Andererseits soll nicht verkannt werden, dass die Chemie eine Vielzahl von „maßgeschneiderten“ Sonderwerkstoffen geschaffen hat.

Tab. 18.1 Beanspruchung von Werkstoffen (prinzipielle Übersicht)

Bereich	Beanspruchungen	
Mechanik	Kräfte	z. B. Zug, Druck, Torsion
	Korrosion	z. B. Feuchtigkeit, (elektro-)chemische Prozesse
	Temperatur	z. B. Hoch- und Tieftemperatur-Prozesse
Elektrik	Felder	z. B. Stromfluss, Magnetisierung, dielektrische Verschiebung
	Kräfte	z. B. Zug, Druck, Torsion
	Korrosion	z. B. Feuchtigkeit, (elektro-)chemische Prozesse
	Temperatur	z. B. Hoch- und Tieftemperatur-Prozesse

Tab. 18.2 Eigenschaftsanforderungen an Materialien bei verschiedenen Einsatzgebieten (Auswahl)

Gebiet	Eigenschaften
Physik	Dichte, Schmelzpunkt, Siedepunkt, Wärmedehnzahl, Wärmeleitfähigkeit usw.
Chemie	Atomgewicht, Wertigkeit, Affinität, Korrosionsbeständigkeit usw.
Mechanik	Härte-, Zug-, Druck-, Biege-, Torsions-, Dauerfestigkeit usw.
Elektrik	Elektrische und magnetische Leitfähigkeit, Permittivität, Permeabilität, Remanenz usw.
Technologie	Biege-, Härte-, Schweiß-, Löt-, Gieß-, Platifizier-, Tiefziehbarkeit usw.

18.2 Atombau und Periodensystem

Stoffe, die sich auf chemischem Wege nicht weiter in andere Stoffe zerlegen lassen, nennt man *Elemente*. Die kleinsten, noch mit den charakteristischen Eigenschaften eines Elementes ausgestatteten Teilchen heißen *Atome*. Sie bestehen aus dem Atomkern mit Pro*tonen* und (Ausnahme Wasserstoff) *Neutronen* und der Atomhülle mit *Elektronen*, daher auch Elektronenhülle. Die Massen der Atome liegen im Bereich von $10^{-22} \ldots 10^{-24}$ g die Durchmesser bei $10^{-9} \ldots 10^{-10}$ m (Lichtwellenlänge ungefähr $500 \cdot 10^{-9}$ m). Protonen und Neutronen haben fast die gleiche Masse von $1{,}67 \cdot 10^{-24}$ g. Das Neutron hat keine elektrische Ladung. Das Proton hat mit $1{,}6 \cdot 10^{-19}$ A s (Amperesekunden) die gleiche Ladung wie das negative Elektron, jedoch mit positivem Vorzeichen. Die Ruhemasse des Elektrons ist mit $9{,}1 \cdot 10^{-28}$ g verschwindend gegenüber der Protonen- oder Neutronenmasse.

Die Anordnung dieser Materieteilchen wird im *Bohr'schen Atommodell* als kugelförmiger Kern mit Protonen, Neutronen und Kugelschalen für die Elektronen beschrieben. Nach dem genaueren *Orbitalmodell* bewegen sich die Elektronen auf komplizierten Bahnen, wie sie durch räumliche Schwingungen vorgegeben werden (dreidimensionaler Oszillator). Daraus resultieren Aufenthaltsräume mit hoher Besetzungswahrscheinlichkeit und bestimmter Orientierung (Abb. 18.1a–d). Die *Orbitale* entstehen bei der räumlichen Schwingung durch Knotenflächen in Form konzentrischer Kugeln (Abb. 18.1a, b) und räumlicher Knotenflächen, die durch (gedachte) Schnittflächen (Abb. 18.1c) die Schwingungsbereiche in räumliche Bezirke, Orbitale (engl.: orbit = Planetenbahn) aufteilen. In jedem Orbital können dann

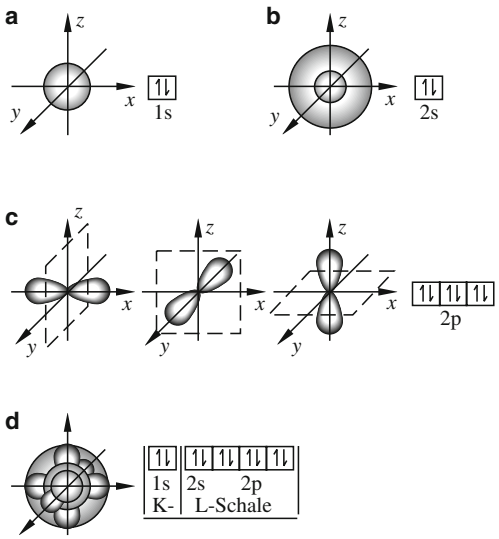


Abb. 18.1 Orbitalmodell. **a** 1s-Zelle (K-Schale). Besetzung mit 2 Elektronen möglich. **b** 2s-Zelle (L-Schale, teilw.). Die Kugel ist innen von einer Knotenfläche begrenzt (Hohlkugel), 2 Elektronen möglich. **c** 2p-Zelle (L-Schale, forts.) mit den möglichen 3 × 2 Elektronen und jeweils einer (gedachten) Trennwand in den 3 Ebenen. **d** Neon als Beispiel 1s, 2s und 2p, d. h. K- und L-Schale sind voll besetzt

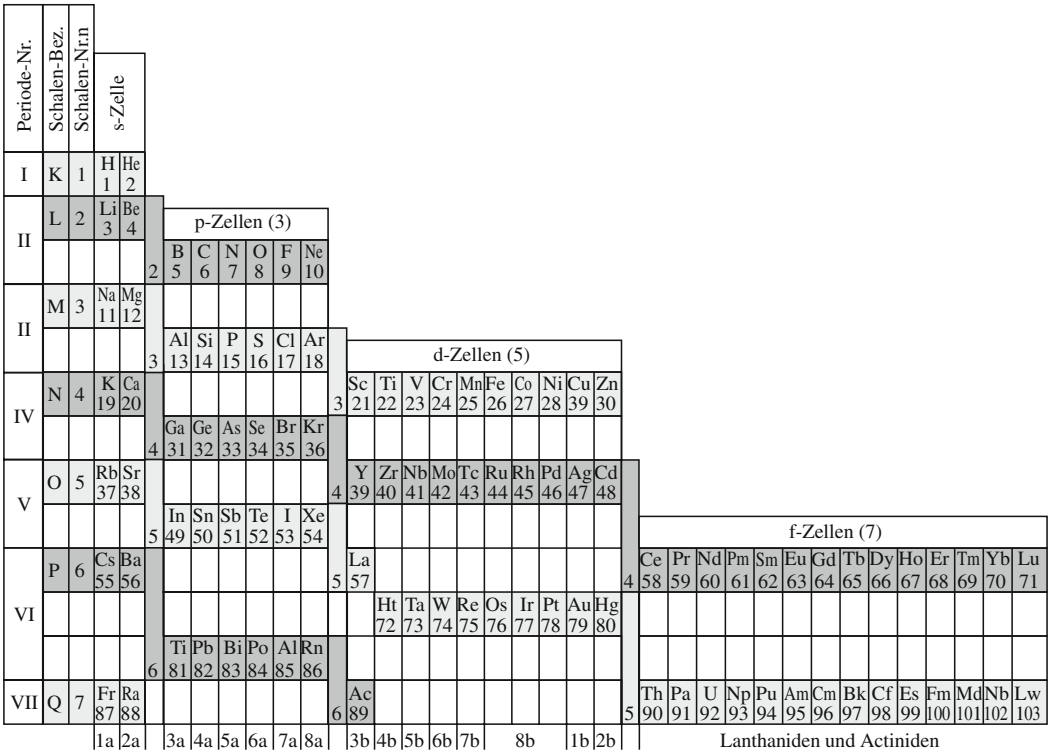


Abb. 18.2 Periodensystem der Elemente (PSE) mit Perioden Nr. I ... VII, Schalenbezeichnungen K ... Q, Schalen Nr. $n = 1 \ldots 7$, Hauptgruppen Nr. 1a ... 8a und Nebengruppen Nr. 1b ... 8b sowie den Lanthaniden und Actiniden

wieder 2 Elektronen schwingen, die sich nach *Pauli* durch ihre Eigenrotation (Spin) voneinander unterscheiden.

Das dynamische Verhalten des gesamten Systems rechtfertigt dann wieder die einfachere Betrachtung als kugelförmige Atomhüllen (Abb. 18.1d) mit der klassischen Bezeichnung der 7 möglichen „Schalen" von K bis Q oder mit den sogenannten *Hauptquantenzahlen* 1 bis 7. Ihnen werden auch die *Hauptenergieniveaus* zugeordnet. Die weitere Unterteilung in *Unterenergieniveaus* oder „*Zellengruppen*" erfolgt durch die *Nebenquantenzahlen*, die mit den Buchstaben s, p, d und f bezeichnet werden.

Den *Spin*, d. h. den Eigendrehimpuls der Elektronen, kann man sich mechanisch als entgegengesetzte Rotation, magnetisch als entgegengesetzte magnetische Momente vorstellen. Für die bildhafte Darstellung, wie z. B. in Abb. 18.1, benutzt man dazu gegeneinander gerichtete *Halbpfeile* (↿⇂). Ordnet man die Elemente, repräsentiert durch die in der Chemie üblichen Kurzzeichen, nach der Zahl der Protonen im Kern und damit auch nach der Zahl der Elektronen in der *Außenhülle an*, so gelangt man zum *Periodensystem der Elemente* (PSE Abb. 18.2). Die Anordnung ist so getroffen, *dass* nicht nur die Ordnungszahl (unterhalb der Symbole) von oben nach unten zunimmt, sondern auch die Energie. Ungefähr gleiche Energiestufen sind auf gleiche Höhe gesetzt. Durch Energiezufuhr können Elektronen auf höhere *Energieniveaus* gebracht werden, die sie dann unter Energieabgabe, z. B. Lichtemission, wieder verlassen. Werden ein oder mehrere Elektronen durch Energiezufuhr vollständig vom Atom entfernt, entstehen *ionisierte* Atome oder *Atomionen*.

Die *Ionisierungsenergie* (gemessen in Elektronenvolt = eV mit $1\,\text{eV} = 1{,}6 \cdot 10^{-19}\,\text{W s}$), nimmt durch die Abschirmwirkung der Schalen mit der Annäherung an den Atomkern zu.

Durch den Austausch von Hüllelektronen sind die Atome – von den Edelgasen abgesehen – zu chemischen Reaktionen fähig. Die hieran beteiligten Elektronen heißen *Valenzelektronen*. Aus einer Hülle können bis zu 8 Elektronen abgegeben werden, das verbleibende positive Atomion heißt *Kation*. Es können aber auch bis zu 4 Valenzelektronen aufgenommen werden; das so entstandene Atomion heißt *Anion*.

Die Ladung der Atomionen wird rechts oben am Elementsymbol gekennzeichnet, z. B. H^+, AI^{3+}, Cl^-, S^{2-} oder S^-. Die Metalle bilden nur positive Atomionen, die Nichtmetalle vorzugsweise negative Atomionen.

Unten rechts am Elementsymbol erscheint die Anzahl der Atome (oder Ionen), z. B. H_2O, H_2SO_4. In der Isotopentechnik erscheinen auf der linken Seite der Elementsymbole noch unten die Ordnungszahl (= Protonenzahl) und oben die Massenzahl (= Nukleonenzahl). So hat z. B. das Kohlenstoff(bezugs)isotop für die Massenbestimmung mit der Ordnungszahl 6 und der Massenzahl 12 die Schreibweise $^{12}_{6}C$.

Das Periodensystem (Abb. 18.2) wird in *Perioden* 1 ... 7 entsprechend den 7 Schalen K ... Q senkrecht und 8 *Hauptgruppen*, welche der Auffüllung der jeweils äußersten Elektronenschale entsprechen, waagerecht aufgeteilt (*x*a in der untersten Zeile von Abb. 18.2) In den *Nebengruppen* werden jeweils die zweit- und drittäußerste Schale aufgefüllt (*x*b in der untersten

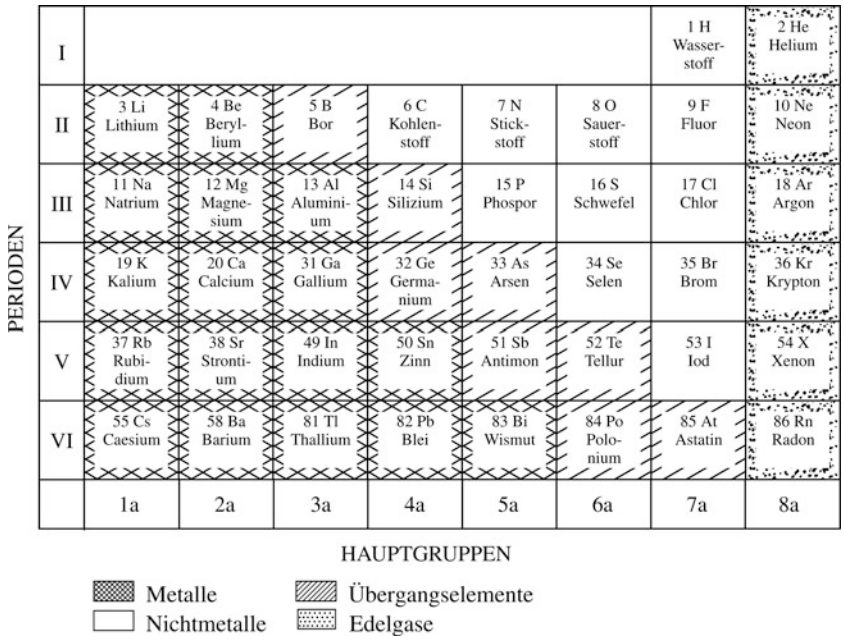


Abb. 18.3 Kurzform des Periodensystems der Elemente (nur Hauptgruppen)

Zeile von Abb. 18.2). Die Metalle, als Mehrzahl der Elemente, sind links und unten im PSE und die Nichtmetalle rechts oben zu finden. Die Halbmetalle wie etwa B, Si, As, Te usw., auch Übergangselemente genannt, bilden die „Grenze“ (Abb. 18.3).

In der Richtung von unten links nach oben rechts, d. h. Metalle – Nichtmetalle steigt auch die *Elektronegativität*. Es ist dies die (relative) Fähigkeit der Atome, bei der Bindung Elektronen aufzunehmen.

18.3 Aufbau der festen Körper

Die in der Praxis besonders wichtigen festen Stoffe basieren auf den Bindungskräften der Atomhüllen. Die *Ionenbindung*, auch *heteropolare* Bindung genannt (Abb. 18.4a), kommt durch Ladungsaustausch (Elektronen) zustande. Die Atomionen aus chemischen Reaktionen führen zu starken Anziehungskräften und stabilen Konfigurationen. Typische Vertreter sind demnach die anorganischen Verbindungen. Bei der *Atombindung*, auch *homöopolare* oder *kovalente Bindung* (Abb. 18.4b) genannt, haben benachbarte Atome gemeinsame Elektronen(paare). Diese Bindungsart ist typisch für nichtmetallische Elemente und organische Verbindungen. Liegt eine *metallische Bindung* vor (Abb. 18.4c), haben die Atome ein Elektron abgegeben und stellen positive (Atom-)Ionen dar. Die abgegebenen Elektronen sind als bewegliche *Leitungselektronen* (*Elektronengas*) allen Atomrümpfen gemeinsam zugeordnet. Diese Bindungsart ist die Grundlage der elektrischen Leitfähigkeit und typisch für alle Metalle. Die *Van der Waals-* oder *zwischenmolekulare Bindung* wird durch die Dipolbildung benachbarter Teilchen (Atome oder Moleküle) bewirkt. Es ergeben sich daraus schwache anziehende Kräfte (Abb. 18.4d). Im allgemeinen überwiegt zwar die eine oder andere Bindungsart, in der Praxis liegt jedoch immer eine Kombination vor. Im Regelfall führen die Bindungskräfte zu regelmäßigen, räumlich periodischen Strukturen der kleinsten Teile (Atome, Ionen oder Moleküle), den *Kristallen*. Stoffe, die nicht regelmäßig, also nicht kristallin, aufgebaut sind, bezeichnet man als *amorph*, das heißt gestaltlos.

Ein Kristall kann sehr klein, z. B. nur mikroskopisch sichtbar, sein, dann spricht man von *Kristallit* oder *Korn*. Die Zusammenlagerung vieler Kristallite liefert dann erst ein größeres Gebilde, den *Polykristall*. Kann ein Kristallit in den makroskopischen Bereich wachsen, entsteht ein *Einkristall*. Die kleinste Einheit, die bereits die Kristallstruktur aufweist und im Kristall periodisch wiederkehrt, ist die *Elementarzelle*. Die räumlich regelmäßige Anordnung einer großen Anzahl von Elementarzellen führt zum *Raum-* oder *Kristallgitter*.

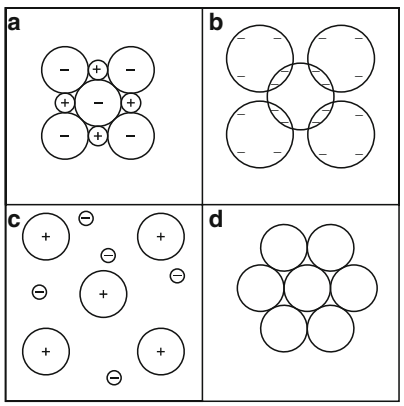


Abb. 18.4 Bindungsarten. **a** Ionenbindung oder heteropolare Bindung, z. B. Na^+Cl^-, **b** Atombindung, homöopolare oder kovalente Bindung, z. B. Ge, **c** metallische Bindung, z. B. Cu, **d** Van der Waals- oder Molekülbindung, z. B. Ar (kristallisiert)

Tab. 18.3 Gitterfehler

Ausdehnung (atomar)	Nulldimensional	Eindimensional	Zweidimensional
Bezeichnung	1. Leerstelle 2. Zwischengitteratom (Schottky-Fehlordnung) 3. Frenkel-Fehlordnung (Kombination von 1 und 2)	Versetzungen	Korngrenzen
Konzentration (in Metallen)	Leerstellenkonzentration $= f(\text{Temp.}) \approx 10^{-7} \ldots 10^{-11}$	Versetzungsdichte $= f(\text{Verformung, Wärmebehandlung})$ $\approx 10^{6} \ldots 10^{12}\ \text{cm/cm}^3$	Korngrenzenflächen $= f(\text{Korngröße}) \approx 10^{2}\ \text{cm}^2/\text{cm}^3$

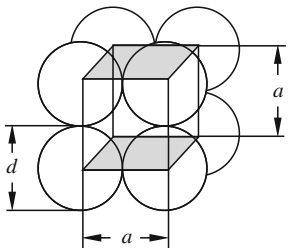


Abb. 18.5 Kubisch primitive Elementarzelle (kp) als Anordnung von Kugeln im Raum, a = Gitterkonstante, d = Atom(kugel)durchmesser

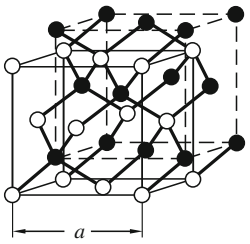


Abb. 18.7 Diamantgitter. Entstehung des Diamantgitters aus zwei ineinander gestellten kfz-Gittern. Der Versatz der kfz-Gitter beträgt 1/4 Raumdiagonale

Mit der Kugelform der Atome wird der Aufbau der Kristallgitter aus Elementarzellen eine Frage der Anordnung von Kugeln im Raum. Die einfachste Form ist die *kubisch-primitive* (*kp*) Elementarzelle, ein Würfel, dessen Ecken mit sich berührenden „Kugeln" besetzt ist (Abb. 18.5). Der Abstand längs einer Würfelkante ist die *Gitterkonstante a*. Die vereinfachte Darstellung zeigt Abb. 18.6a. Ein zusätzliches Atom in der Raummitte (Abb. 18.6b) führt zum *kubisch-raumzentrierten Gitter* (*krz*). Wird auf jeder Flächenmitte ein Atom eingefügt, entsteht das *kubisch-flächenzentrierte Gitter* (*kfz*) (Abb. 18.6c). Die beiden letztgenannten sind die häufigsten metallografischen Gittertypen. Als hexagonale Struktur ist noch die Elementarzelle mit der *hexagonal dichtesten Packung* (*hdp*) (Abb. 18.6d) bedeutsam. Die Halbleiterwerkstoffe Germanium und Silizium kristallisieren in Form eines *Diamantgitters* (*dia*) (Abb. 18.7). Die Elementarzelle des Diamantgitters ist kubischflächenzentriert mit einer tetraedrischen Unterstruktur.

Die idealisiert dargestellten Gitter sind in Wirklichkeit durch *Gitterfehler* und *Fremdatome* verzerrt. Die in reinen Stoffen möglichen Gitterfehler sind in der Tab. 18.3 zusammengestellt. Die von einzelnen Atomen herrührenden *nulldimensionalen Fehler* werden auch *Punktdefekte* genannt. Die möglichen Fehler zeigt Abb. 18.8. Eine Leerstelle oder ein Zwischengitteratom haben jeweils einen Fehler zur Folge. Aus der Kombination von beiden, dem *Frenkel-Defekt*, folgen zwei Fehler.

Bei *eindimensionalen Fehlern*, auch *Versetzungen* genannt, haben die Fehler eine linienförmige Ausdehnung. Abb. 18.9a zeigt eine *Stufenversetzung*, die das Ende einer zusätzlichen (Halb-) Ebene darstellt. In Abb. 18.9b sind Teilbereiche gegeneinander geglitten und haben eine *Schraubenversetzung* bewirkt. Als Maßzahl wird die Anzahl der „Durchstoßpunkte pro Flächeneinheit" in $1/\text{cm}^2$ angegeben. Bei normalen Fertigungswerkstoffen liegen die Zahlenwerte für Versetzungen im Bereich von $10^{6} \ldots 10^{12}\ \text{cm}^{-2}$, wohingegen Halbleiter-Einkristalle Werte von weniger als $10\,\text{cm}^{-2}$ aufweisen.

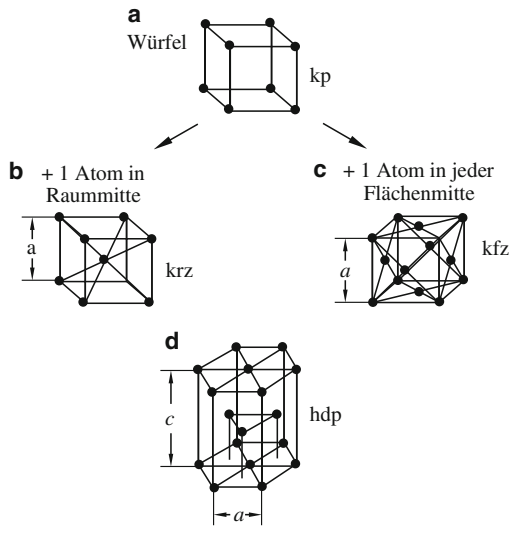


Abb. 18.6 Elementarzellen. **a** kp = kubisch-primitiv, **b** krz = kubisch-raumzentriert, **c** kfz = kubisch-flächenzentriert, **d** hdp = hexagonal dichteste Packung. a, b, c = Gitterkonstanten

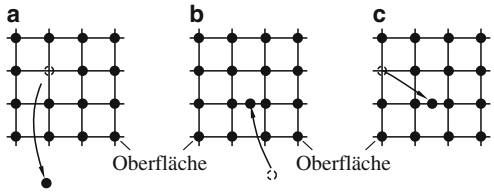


Abb. 18.8 Nulldimensionale Fehlstellen. **a** Leerstelle, **b** Zwischengitteratom (Atom falsch eingebaut), **c** Frenkel-Defekt, Atom wandert im Gitter und erzeugt zwei Fehler

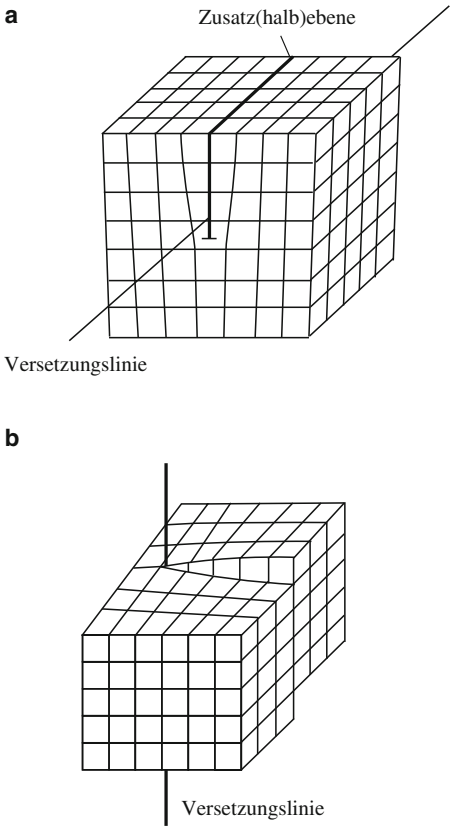


Abb. 18.9 Eindimensionale Fehler. **a** Stufenversetzung, **b** Schraubenversetzung

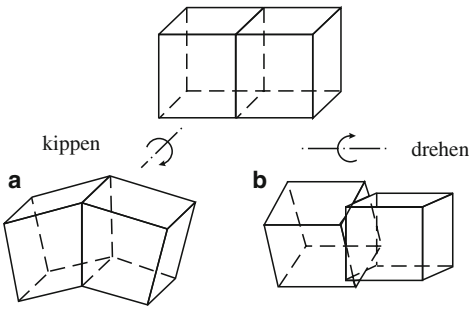


Abb. 18.10 Zweidimensionale Fehler. **a** Kippung, **b** Drehung

Treffen Kristallgitter verschiedener Orientierung aufeinander, so können sie (Abb. 18.10) gegeneinander gekippt, gedreht oder beides zusammen sein. Als Folge davon müssen sich Übergänge bilden, die zu sog. *Korngrenzen* führen. Die Kristalle bilden dabei regellose *Korngrenzflächen* gegeneinander aus. Mechanische Beanspruchungen können dazu führen, dass eine gewisse Ordnung in die Orientierung hineinkommt. Es ergeben sich *Kornorientierungen*, *Vorzugsrichtungen* oder *Texturen*. So liegen z. B. die Kanten der Elementarzellen nach dem Walzen von Blechen weitgehend in der Walzrichtung.

Werden in einem Kristall Fremdatome eingelagert, ist er nicht mehr rein. Man spricht von *Mischkristallen*, *festen Lösungen* oder *Legierungen*. Die Einlagerung kann auf Gitterplätzen der Wirtsatome erfolgen, dann entsteht ein *Substitutions-Mischkristall* (Abb. 18.11). Werden Fremdatome auf Zwischengitterplätzen eingelagert, ergeben sich *Einlagerungs-* oder *interstitielle Mischkristalle*.

18.4 Chemische Grundzusammenhänge

Wie bei den Bindungsarten kurz angedeutet, vermögen die Atome – über ihre Valenzelektronen – miteinander zu reagieren. Es entstehen dann neue Stoffe, die dem Bereich *anorganische Chemie* zugeordnet werden, wenn es sich beim Stoffaufbau weitgehend um Ionenbindungen handelt. Bei der *organischen Chemie* steht das Kohlenstoffatom mit überwiegend kovalenten Bindungen im Mittelpunkt. Das, wie das Kohlenstoffatom, gleichfalls vierwertige Silicium ist Basis einer eigenständigen Reihe von Kunststoffen, den *Siliconen*.

Für den Aufbau von Stoffen benutzt die Chemie die *Synthese*, während die Untersuchung des Aufbaues mit Hilfe der *Analyse* erfolgt. Zur Beschreibung dieser Vorgänge dienen *Formelsprache* und *Reaktionsgleichung* unter Benutzung der Elementsymbole.

So ist mit H_2SO_4 Schwefelsäure gemeint, und C_2H_5OH kennzeichnet eindeutig Alkohol als *Ethylalkohol* (Ethanol). Reagieren Stoffe miteinander, so beschreibt die Reaktionsgleichung die Ausgangsstoffe und Endprodukte nach Art und relativer Menge. Die Darstellung entspricht einer mathematischen (Additions-) Gleichung, bei der an die Stelle des Gleichheitszeichens ein

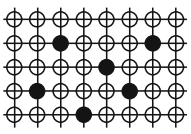
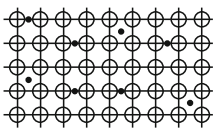
Ort des Fremdatoms	Gitterplatz	Zwischengitterplatz
Art des Mischkristalls	Substitions-Mischkristall (substitutioneller Mkr.)	Einlagerungs-Mischkristall (interstitieller Mkr.)
Durchmesser	Fremdatom ≈ Wirtsatom	Fremdatom < Wirtsatom
Mischkristallbildung bei ebenem Gitter (schematisch) ○ Wirtsatome ● Fremdatome		

Abb. 18.11 Mischkristalle in ebener Darstellung (schematisch)

Tab. 18.4 pH-Skala

⟵ Zunahme der Säurewirkung ⟵								⟶ Zunahme der Basenwirkung ⟶						
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Pfeil die Reaktionsrichtung kennzeichnet. Ist der Reaktionsablauf nicht umkehrbar *(irreversibel)*, zeigt dies ein einfacher Pfeil an.

Beispiel

Die Verbrennung von Ethanol

$$C_2H_5OH + 3\,O_2 \rightarrow 2\,CO_2 + 3\,H_2O$$

1 mol Ethanol + 3 mol Sauerstoff → 2 mol Kohlendioxid + 3 mol Wasser

◀

Mol steht hierbei für die Stoffmenge, die aus genauso vielen Teilchen besteht wie 12 g des Kohlenstoffisotops ^{12}C, nämlich $6{,}023 \cdot 10^{23}$ Teilchen (Avogadro'sche Konstante N_A oder Loschmidtzahl N_L).

Bei umkehrbaren *(reversiblen)* Reaktionen kennzeichnet ein Doppelpfeil die beiden möglichen Reaktionsrichtungen, z. B.

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$

1 mol Stickstoff + 3 mol Wasserstoff ⇌ 2 mol Ammoniak

Die Verbindungen des Sauerstoffs mit anderen Elementen, die *Oxide*, sind von besonderer Bedeutung, die Schwermetalloxide als Erze zur Gewinnung von Metallen, die Leichtmetalloxide wegen ihrer hohen Schmelzpunkte als (oxid)keramische Werkstoffe.

Die zur Oxidbildung gehörige chemische Reaktion ist die *Oxidation*, allgemeiner als Elektronenabgabe eines Moleküls oder Ions definiert. Die damit verbundene Aufnahme eines Elektrons durch ein anderes Molekül oder Ion ist die *Reduktion*. Die zwangsweise miteinander verkoppelten Oxidations- und Reduktions-Prozesse führen die Bezeichnung *Redox-Reaktionen*.

Metalloxide können ihrerseits wieder mit Wasser reagieren und bilden *Hydroxide*, sie enthalten die einwertige OH-Gruppe (Hydroxyl-Gruppe) und bilden bei den Metallen der 1. und 2. Hauptgruppe (Alkali- und Erdalkalimetalle) *Basen* oder *Laugen*. Sie sind in Metall- und OH-Ionen gespalten und stellen sich als seifige, die Haut ätzende Flüssigkeiten dar.

Beispiel

$$CaO + H_2O \rightarrow Ca(OH)_2$$

Metalloxid + Wasser → Hydroxid

◀

Die Oxide dreiwertiger Metalle sind amphoter, d. h. sie können sowohl schwache Basen als auch schwache Säuren bilden. Oxide von Nebengruppenelementen in höheren Wertigkeiten bilden – mit Wasser – Säuren.

Nichtmetalloxide bilden mit Wasser Säuren.

Beispiel

$$SO_2 + H_2O \rightarrow 2\,H_2SO_3$$

Nichtmetalloxid + Wasser → Säure

◀

Ihre Moleküle bauen sich aus dem Säurewasserstoff und dem Säurerest auf. Der Säurerest ist eine nicht selbständig existenzfähige Atomgruppe aus Nichtmetall und Sauerstoff. Da allein nicht beständig, sind es auch keine Moleküle. Derartige Atomgruppen werden Radikale genannt, im Beispiel SO_2. Die Säuren sind gegenüber den meisten Metallen aggressiv, da in vielen Fällen die Wasserstoffatome gegen Metallatome, die dann als Ionen in Lösung gehen, ausgetauscht werden können. Säuren schmecken – verdünnt (!) – sauer, ätzen die Haut und verfärben, ähnlich wie Basen, Farbstoffe. Dieser Effekt wird für Indikatoren benutzt.

Nach „Brönsted"* lautet die umfassendere Definition: *Säuren* sind *Protonenspender* (Moleküle oder Ionen), *Basen* sind *Protonenfänger* (Moleküle oder Ionen). Die Protonenübergangsreaktion nennt man *Protolyse*. Dieser Vorgang findet auch in geringem Umfang im Wasser statt und wird dann *Autoprotolyse* genannt. Nach „Brönsted" liegt dem folgende Reaktion zugrunde

$$H_2O + H_2O \rightarrow H_3O^+ + OH^-$$

Wasser → Hydroniumion + Hydroxidion

In dieser Gleichgewichtsreaktion ist die Konzentration der beiden Ionenarten vom Prinzip her gleich groß, aber mit 10^{-7} mol/l sehr gering. Andererseits muss das Produkt dieser beiden Ionenkonzentrationen als *Ionenprodukt des Wassers* mit $K_w = 10^{-14}\,\text{mol}^2/\text{l}^2$ als Gleichgewichtsbedingung bei Zugabe von Säuren oder Basen konstant bleiben. Es verschieben sich lediglich die Konzentrationsverhältnisse von H_3O^+ zu OH^-. „Drehpunkt" dieses Geschehens ist der Zahlenwert 10^{-7} der *Hydroniumionenkonzentration*. Der negative Exponent dieser Konzentrationsangabe (lt. Definition der negative dekadische Logarithmus der Hydroniumionenkonzentration) ist der *pH-Wert*. Er hat große Bedeutung, da mit diesem Zahlenwert die Angaben „sauer", $pH < 7$, „neutral", $pH = 7$ und „basisch", $pH > 7$ präzise beurteilt werden können. Zwangsweise liegen die Grenzen für sehr starke Säuren dann bei $pH = 0$ und für sehr starke Basen bei $pH = 14$ (s. Tab. 18.4). Die Bestimmung ist elektrisch oder mit Indikatoren sehr genau möglich.

* Brönsted, Johannes Nicolaus, Dänischer Chemiker, lebte von 1879 bis 1947.

Tab. 18.5 Elektronegativitätsskala nach „Pauling" (Zahlenwert relativ mit Maximum 4 bei Fluor)

K	Na, Ba	Li, Ca	Mg	Al	Zn	Si	H	P	C	S	N	Cl	O	F
0,8	0,9	1,0	1,2	1,5	1,7	1,8	2,1	2,1	2,5	2,5	3,0	3,0	3,5	4,0
← elektropositiver								elektronegativer →						

Treffen Säuren und Basen in solchen Konzentrationen zusammen, dass pH = 7 erreicht wird, so spricht man von *Neutralisation*. Das Ergebnis ist Wasser und ein Salz.

Beispiel

$$HCl + NaOH \rightarrow NaCl + H_2O$$

Säure + Base → Salz + Wasser ◀

Salze im klassischen Sinne bestehen aus positiven Metallionen und negativen Säurerestionen. Nach „Brönsted" sind alle Stoffe Salze, die im *festen Zustand Ionengitter* bilden.

Bei der Namensbildung steht an erster Stelle das elektropositivere Element (Metall siehe auch Tab. 18.5) meist unverändert. Dann folgt der Name des elektronegativeren Bestandteiles mit einer entsprechenden Endung (z. B. NaCl = Natriumchlorid). Die Endung **-id** weist darauf hin, dass es sich um ein Salz mit einem Säurerest (Cl^-) aus einer *sauerstofffreien* Säure handelt. Bei Säureresten aus stabilen Säuren mit *normalem* Sauerstoffgehalt lautet die Endung **-at** (z. B. Na_2SO_4 = Natriumsulfat). Entstammt der Säurerest einer sauerstoffärmeren Säure, ist die Endung **-it** (z. B. Na_2SO_3 = Natriumsulfit). Rührt der Säurerest von einer besonders sauerstoffarmen Säure her, kommen der Zusatz **-hypo-** und die Endung **-it** dazu (z. B. NaClO = Natriumhypochlorit). Bei Säureresten von besonders sauerstoffreichen Säuren lautet der Zusatz **-per-** und die Endung **-at** (z. B. $NaClO_4$ = Natriumperchlorat). Wurden die Säure-Wasserstoffatome nicht vollständig durch Metallatome ersetzt, ergeben sich *saure* oder *hydrogene* Salze (z. B. $NaHCO_3$ = Natriumhydrogencarbonat, früher Natriumbicarbonat oder – trivial – Natron).

18.5 Elektrochemie

Innerhalb eines Moleküles ziehen sich die Kationen und Anionen an. In wässeriger Lösung gehen diese Kräfte durch die *Permittivität* des Wassers ($\varepsilon_r = 80$) auf 1/80 zurück. Die Moleküle sind in Ionen dissoziiert, die eine elektrische Stromleitung als *Leiter 2. Ordnung* ermöglichen. Die Lösung wird zum *Elektrolyten*, in dem unter der Einwirkung einer elektrischen Spannung ein Materietransport erfolgt. Die positiven Kationen wandern zur Katode, die negativen Anionen zur Anode. Nach Ladungsausgleich, Elektronenabgabe bzw. -aufnahme, ist der zum Ion gehörige Stoff elementar verfügbar. Der Vorgang heißt *Elektrolyse* und ermöglicht die Zerlegung eines Stoffes durch

Tab. 18.6 Elektrochemische Spannungsreihe nach „Pauling". a) Metalle (Normalpotentiale bezogen auf die Normal-Wasserstoffelektrode), b) einige Nichtmetalle

a)

Met.	Ox.	+ e⁻	Volt
Cs	$= Cs^+$	$+ e^-$	−3,02
Li	$= Li^+$	$+ e^-$	−3,02
Rb	$= Rb^+$	$+ e^-$	−2,99
K	$= K^+$	$+ e^-$	−2,92
Ba	$= Ba^{++}$	$+ 2e^-$	−2,90
Sr	$= Sr^{++}$	$+ 2e^-$	−2,89
Ca	$= Ca^{++}$	$+ 2e^-$	−2,87
Na	$= Na^+$	$+ e^-$	−2,71
La	$= La^{+++}$	$+ 3e^-$	−2,37
Mg	$= Mg^{++}$	$+ 2e^-$	−2,34
Be	$= Be^{++}$	$+ 2e^-$	−1,70
Al	$= Al^{+++}$	$+ 3e^-$	−1,67
Mn	$= Mn^{++}$	$+ 2e^-$	−1,05
Zn	$= Zn^{++}$	$+ 2e^-$	−0,76
Cr	$= Cr^{+++}$	$+ 3e^-$	−0,71
Fe	$= Fe^{++}$	$+ 2e^-$	−0,44
Cd	$= Cd^{++}$	$+ 2e^-$	−0,40
Co	$= Co^{++}$	$+ 2e^-$	−0,28
Ni	$= Ni^{++}$	$+ 2e^-$	−0,25
Sn	$= Sn^{++}$	$+ 2e^-$	−0,14
Pb	$= Pb^{++}$	$+ 2e^-$	−0,13
H_2	$= 2\,H^+$	$+ 2e^-$	0,00
Cu	$= Cu^{++}$	$+ 2e^-$	+0,34
$2\,J^-$	$= J_2$	$+ 2e^-$	+0,53
Ag	$= Ag^+$	$+ e^-$	+0,80
Pd	$= Pd^{++}$	$+ 2e^-$	+0,83
Hg	$= Hg^{++}$	$+ 2e^-$	+0,85
$2\,Br^-$	$= Br_2$ (fl)	$+ 2e^-$	+1,06
Pt	$= Pt^{++}$	$+ 2e^-$	+1,2
$2\,H_2O$	$= O_2 + 4\,H^+$	$+ 4e^-$	+1,23
$2\,Cl^-$	$= Cl_2$	$+ 2e^-$	+1,36
Au	$= Au^+$	$+ e^-$	+1,68
$2\,F^-$	$= F_2$	$+ 2e^-$	+2,85

b)

Red.	⇌ Ox.	+ e⁻	Volt
Te^{--}	⇌ Te	$+ 2e^-$	−0,92
Se^{--}	⇌ Se	$+ 2e^-$	−0,78
S^{--}	⇌ S	$+ 2e^-$	−0,51
$2\,J^-$	⇌ J_2	$+ 2e^-$	+0,53
$2\,Br^-$	⇌ Br_2	$+ 2e^-$	+1,07
$2\,Cl^-$	⇌ Cl_2	$+ 2e^-$	+1,36
$2\,F^-$	⇌ F_2	$+ 2e^-$	+2,85

den elektrischen Strom. Wasserfrei ist der gleiche Vorgang im Schmelzen von Ionenkristallen als *Schmelzfluss-Elektrolyse* möglich.

Die Neigung der Metalle zur Abgabe von Elektronen (Oxidation) ist verschieden stark und durch die unterschiedliche Bindung der Valenzelektronen an den Atomkern bedingt. Das Bestreben eines Metalles, Elektronen abzugeben und als Ion in Lösung zu gehen, nennt man *Lösungsdruck*. Er kann als elektrische Spannung in Volt gemessen werden und liefert die *elektrochemische Spannungsreihe* (Tab. 18.6). Der Nullpunkt ist dabei (willkürlich) für Wasserstoff angesetzt. Die *unedlen*, leicht reagierenden Elemente haben ein zunehmend negatives Potential gegenüber Wasserstoff; umgekehrt werden die Elemente mit wachsendem positivem Potential *edler* und reaktionsträger. Die negativeren, d. h. weiter oben stehenden (Tab. 18.6) Metalle haben den größeren Lösungsdruck gegenüber den positiveren, also weiter unten stehenden Metallen. Als Folge davon reduzieren die „oberen“ die „unteren“ Metalle und gehen selbst in Lösung.

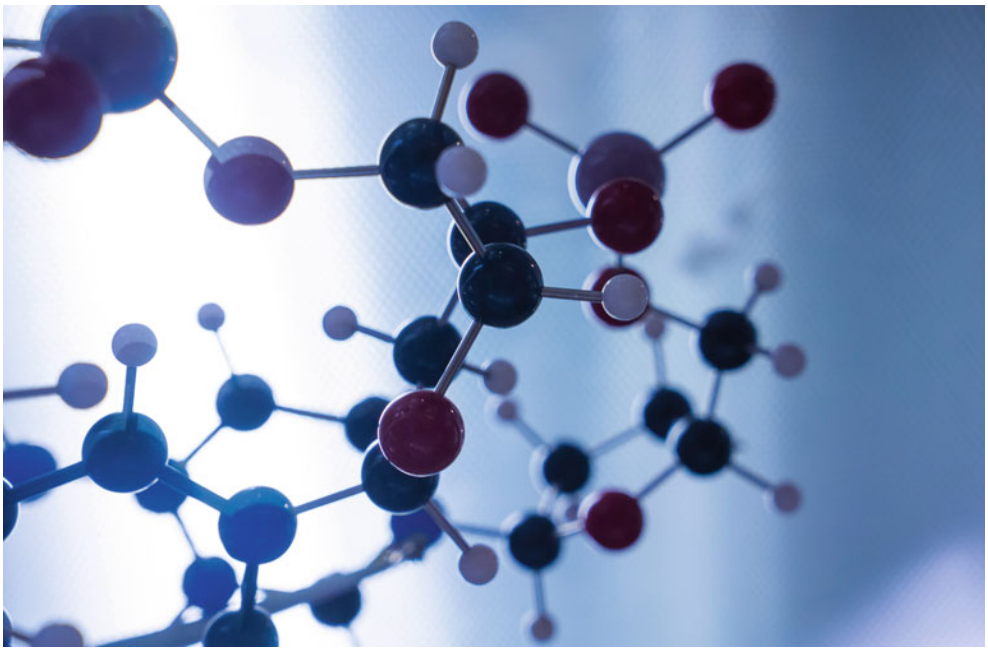
Nutzt man den Lösungsdruck zwischen unterschiedlichen Elementen durch Einbringen in einen Elektrolyten aus, so entstehen *Galvanische Elemente* oder *Primärelemente*.

Der Stoffumsatz bei der Elektrolyse transportiert je Ladung $1{,}6 \cdot 10^{-19}$ A s, d. h. bei einem Mol mit $6 \cdot 10^{23}$ Teilchen $=$ 96.500 A s $= F =$ *Faraday-Konstante*. Bezieht man die transportierte und abgeschiedene Masse auf 1 A s, so ergibt sich das für jeden Stoff charakteristische *elektrochemische Äquivalent* (c). Diese Größe mit 1,118 mg/A s bei Silber diente früher zur Festlegung von 1 A.

Elektrische Leitfähigkeit

19

Egon Döring



19.1 Leitungsmechanismus 268
19.2 Isolator 268
19.3 Halbleiter 269
19.4 Normalleiter 269
19.5 Supraleiter 270
19.6 Halleffekt 270

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_19

19.1 Leitungsmechanismus

Metallische Leiter sind Leiter 1. *Ordnung*, denn mit dem Stromfluss ist kein Materietransport verbunden. Die elektrische Leitfähigkeit ist dabei an die Existenz freier Elektronen, dem sogenannten Elektronengas, gebunden. Für die Leitfähigkeit eines Stoffes ist die Anzahl der verfügbaren Teilchen (n), ihre Beweglichkeit (μ) und die Teilchenladung (e) maßgebend. Da der Leitungsmechanismus sowohl auf negative (Index n) als auch positive (Index p) Ladungsträger zurückgehen kann, gilt für die Leitfähigkeit γ:

$$\gamma = e(n_n \cdot \mu_n + n_p \cdot \mu_p) \qquad \begin{array}{c|c|c|c} \gamma & e & n & \mu \\ \hline \mathrm{S/cm} & \mathrm{A\,s} & \mathrm{cm^{-3}} & \mathrm{cm/V\,s} \end{array} \qquad (19.1)$$

Für den normalen elektrischen Leiter liegen nur die negativen Elektronen als Ladungsträger vor; damit entfällt der Term mit den p-Indizes. Die *Beweglichkeit* μ der Ladungsträger ist die pro Einheit der elektrischen Feldstärke E (meist in V/cm) bewirkte *Driftgeschwindigkeit*, d. h. die der allgemeinen Teilchenbewegung in einer bestimmten Richtung überlagerte Geschwindigkeit. Die in der Praxis auftretenden Werte von Leitfähigkeit, Teilchenzahl und Beweglichkeit sind für einige charakteristische Materialien in der Tab. 19.1 zusammengestellt.

Im Festkörper mit seinem Kristallgitter haben die Atome eine periodische Anordnung mit definierten Abständen. Zwischen den Atomen machen sich *Austauschkräfte* bemerkbar, die die Besetzung der einzelnen Energieniveaus beeinflussen. Gemäß dem Pauli-Prinzip, nach dem in einem Atom keine zwei Elektronen in Bezug auf ihren Zustand völlig übereinstimmen können, dürfen in einem gegebenen System die Energieniveaus und Spinorientierungen nur einmal belegt sein. Da die Beeinflussung über den ganzen Kristall reicht, gilt das Pauli-Prinzip entsprechend weiträumig. Die bei Einzelatomen diskreten Energieniveaus der Elektronen spreizen sich zu *Energiebändern* auf (Abb. 19.1). Neben den mit Ladungsträgern besetzten Bändern treten energetisch unbesetzte oder *verbotene Zonen* auf. Für den Leitungsmechanismus sind das letzte noch mit Elektronen besetzte Band, das *Valenzband*, und das nächste für eine Besetzung mögliche Band, das *Leitungsband*, bedeutsam. Nur im letzteren verfügen Elektronen über die für einen Stromtransport nötige Beweglichkeit. Aus der Ähnlichkeit der statistischen Energieverteilung im Valenzband mit der bei Gasen rührt auch die Bezeichnung Elektronengas her.

Tab. 19.1 Elektrische Leitfähigkeit g, Elektronenkonzentration n_n und Elektronenbeweglichkeit m_n einiger Werkstoffe bei Raumtemperatur

Material	γ	n_n	μ_n
	S/cm	$\mathrm{cm^{-3}}$	$\mathrm{cm^2/V\,s}$
Cu	$58 \cdot 10^4$	$8{,}45 \cdot 10^{22}$	43
Al	$34{,}5 \cdot 10^4$	$6{,}0 \cdot 10^{22}$	36
Ag	$61{,}4 \cdot 10^4$	$5{,}87 \cdot 10^{22}$	65
Ge	$2{,}3 \cdot 10^{-2}$	$2{,}4 \cdot 10^{13}$	3600
Si	$4{,}35 \cdot 10^{-6}$	$1{,}5 \cdot 10^{10}$	1400
InSb	$3{,}5 \cdot 10^2$	$2{,}8 \cdot 10^{16}$	78.000

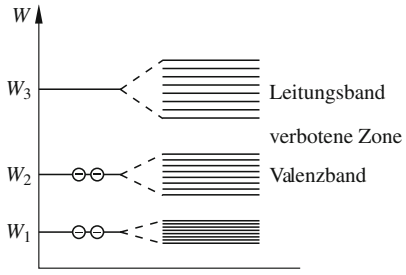


Abb. 19.1 Energiebänder durch Aufspaltung der Energieniveaus. ⊖⊖ durch Elektronen besetzte Niveaus

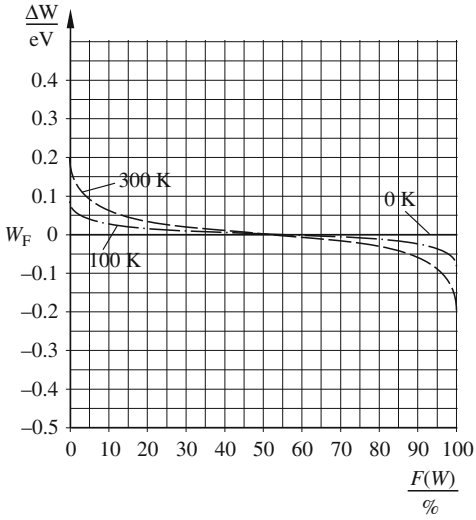


Abb. 19.2 Fermi-Funktion $F(W)$ für die Temperaturen 0 K (—), 300 K (– – –) und 100 K (–·–·–)

Wegen der statistischen Energieverteilung können Teilchen die verbotene Zone überwinden und aus dem Valenzband in das Leitungsband gelangen und dort Leitfähigkeit bewirken. Die Wahrscheinlichkeit, mit der die Energie eines Elektrons einen gewissen vorgegebenen Wert über- oder unterschreitet, wird durch den Verlauf der *Fermi-Funktion* $F(W)$ beschrieben. In Abb. 19.2 ist sie für die 3 Temperaturen 0, 100 und 300 K dargestellt. Das *Fermi-Niveau* W_F ist dabei willkürlich auf den (Bezugs-)Wert 0 gelegt, und die Energie ist als Differenz ΔW gegenüber diesem Wert in eV aufgetragen. Der tatsächliche Wert des Fermi-Niveaus gibt die Energie an, über welche die Teilchen auch noch beim absoluten Nullpunkt verfügen. Es ist offensichtlich, dass die Lage des Fermi-Niveaus in der verbotenen Zone entscheidenden Einfluss auf die Leitfähigkeit eines Stoffes hat.

19.2 Isolator

Abb. 19.3a zeigt ein Bandschema mit einer breiten verbotenen Zone von $\Delta W_B = 2\,\mathrm{eV}$ und daneben die Energieverteilung bei 300 bzw. 600 K für ein mittig liegendes Fermi-Niveau. Offensichtlich verfügt kein Teilchen über eine ausreichende Energie, um in das Leitungsband zu gelangen. Im vollbesetzten Valenz-

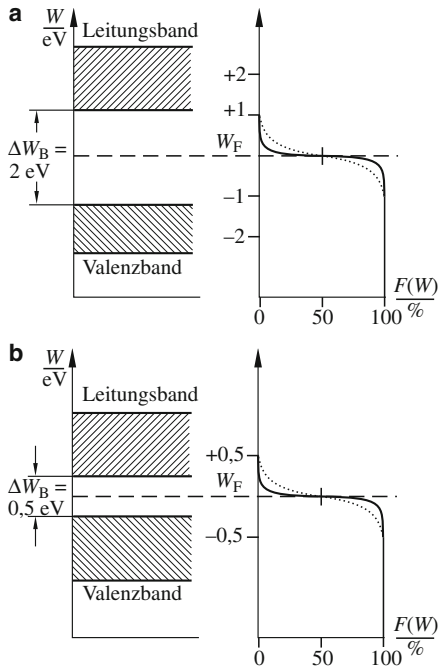


Abb. 19.3 Bandschemata (Prinzip-Beispiele). **a** Isolator, Bandlücke $\Delta W_B = 2$ eV, **b** Halbleiter, Bandlücke $\Delta W_B = 0{,}5$ eV, W_F = Fermi-Niveau, — ≙ 300 K und ··· ≙ 600 K

band ist auch keine Beweglichkeit der Teilchen und damit keine Leitfähigkeit vorhanden. Es liegt das typische Bandschema eines Isolators vor.

Die punktierte Linie für 600 K zeigt aber auch eine andere Tendenz auf. Mit steigender Temperatur wächst die Wahrscheinlichkeit, dass Ladungsträger im Leitungsband verfügbar werden. Der Isolator wird leitend, d. h. er schlägt durch!

19.3 Halbleiter

Wird der Bandabstand oder die Energielücke vom englischen „energy gap„ (Formelzeichen E_g) merklich kleiner als 1 eV, so tritt eine ausreichende Anzahl von Teilchen auf, die in das Leitungsband gelangen können (Abb. 19.3b). Es ist dies der typische Fall des Halbleiters, bei dem im reinen Zustand das Fermi-Niveau genau mittig liegt. Leitfähigkeit ist jetzt sowohl über das Leitungsband mit seinen thermisch gewonnenen Elektronen (n) als auch im teilweise leer gewordenen Valenzband über *Defektelektronen* oder *Löcher* (p) möglich. Es ist sofort ersichtlich, dass die Leitfähigkeit einerseits durch Temperaturerhöhung (z. B. beim Stromdurchgang) oder durch Verschiebung des Fermi-Niveaus zum Leitungs- bzw. Valenzband (durch Dotierung) erhöht werden kann. Trotz zunehmender Beweglichkeit der Teilchen mit sinkender Temperatur verschwindet beim absoluten Nullpunkt die Leitfähigkeit von Halbleitern.

19.4 Normalleiter

Sind andererseits Bänder unvollständig besetzt oder überlappen Valenz- und Leitungsband (die verbotene Zone ist dann negativ), so liegt der Fall des metallischen Leiters vor. Das Fermi-Niveau liegt jetzt innerhalb des Leitungsbandes. In Abb. 19.4 sind die beiden Fälle dargestellt, einmal für Natrium mit einem unvollständig besetzten 3s-Band (Abb. 19.4a) und für Kupfer mit überlappendem 3d- und 4s-Band (Abb. 19.4c) (siehe auch Abschn. 18.2 „Atombau und Periodensystem“).

Mit sinkender Temperatur und damit steigender Beweglichkeit der Teilchen nimmt die Leitfähigkeit der metallischen Leiter bis zum absoluten Nullpunkt um etwa 4 Zehnerpotenzen zu.

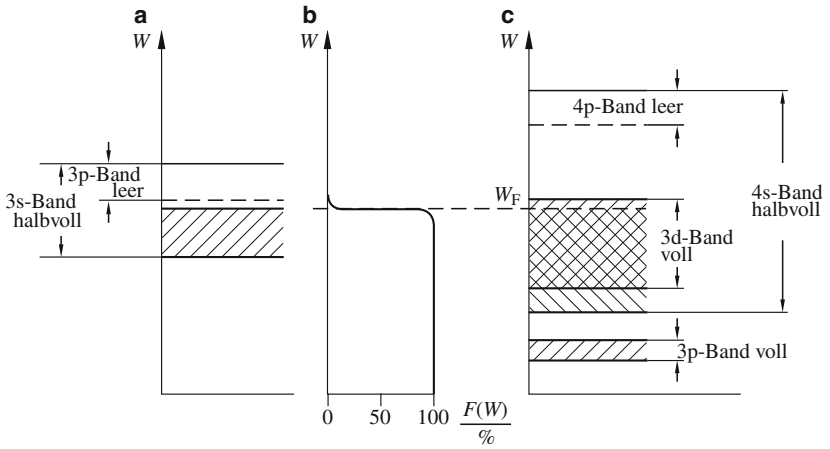


Abb. 19.4 Bandschemata von Leitern. **a** Unvollständig besetztes Band, als Beispiel Natrium. Das 3s-Band ist nur zur Hälfte besetzt, **b** Fermi-Funktion für 300 K mit Lage zu **a** und **c**, **c** Überlappende Bänder, als Beispiel Kupfer. Das voll besetzte 3d-Band liegt innerhalb des nur zur Hälfte besetzten 4s-Bandes

Tab. 19.2 Sprungtemperaturen einiger Basismaterialien für Supraleiter

Material	Ga	Al	Sn	Pb	Nb	NbTi	Nb_3Sn	Nb_3Ga	Nb–Al–Ge	Y–Ba–Cu–O	Ti–Ba–Ca–Cu–O	Hg–Ba–Ca–Cu–O
T_c/K	1,1	1,1	3,7	7,3	9,2	10	18	20,3	21	93	125	133

19.5 Supraleiter

Eine größere Zahl von Metallen und Legierungen zeigt bei sehr tiefen Temperaturen ein schlagartiges Ansteigen der Leitfähigkeit um mehr als 20 Zehnerpotenzen. Der Effekt heißt *Supraleitung* und die Temperatur, bei der er eintritt, *Sprungtemperatur* T_c. Die Höhe der Sprungtemperatur ist materialabhängigund liegt für reine Metalle unter 10 K (Heliumkühlung), während Legierungen Werte um 20 K erreichen und damit bereits in den Bereich der Wasserstoffkühlung (Siedepunkt von Wasserstoff: 20 K bzw. −253 °C) gelangen. Ein entscheidender Schritt gelang in den letzten Jahren mit Mischoxiden auf der Basis Seltene-Erden-Barium-Kupfer (SE-Ba-Cu-O), mit denen Sprungtemperaturen bei Stickstoffkühlung (Siedepunkt Stickstoff. 77 K bzw. −196 °C möglich sind (Tab. 19.2).

Da die Supraleitung durch magnetische Felder – auch durch das eigene Feld eines im Supraleiter fließenden Stromes – aufgehoben werden kann, unterscheidet man *weiche Supraleiter* oder *Supraleiter 1. Art*, die gegen magnetische Felder bis zu einer kritischen Feldstärke $H_c = 0{,}1\,\mathrm{MA/m}$ beständig sind, und *harte Supraleiter* oder *Supraleiter 2. Art*, die noch sehr starke magnetische Felder ($H_c > 10\,\mathrm{MA/m}$) vertragen. Für den praktischen Gebrauch ist die Angabe der *kritischen Stromdichte* S_c bei der die Supraleitfähigkeit durch das eigene Magnetfeld des fließenden Stromes aufgehoben wird, geeigneter.

Nach dem derzeitigen Kenntnisstand wird die Supraleitung von paarweise *spinkompensierten* Elektronen, die nicht mehr dem Pauli-Prinzip gehorchen, den sog. *Cooper-Paaren*, mit einer sehr hohen Beweglichkeit bewirkt. Die Cooper-Paare verursachen durch Ringströme an der Leiteroberfläche ein magnetisches Kompensationsfeld für das Leiterinnere, was zu einem magnetfeldfreien Raum, d. h. $\mu_r = 0$ (!) führt.

In den harten Supraleitern bestehen bei Stromfluss im Inneren normal- und supraleitende Bereiche in Form sogenannter *Flussschläuche* nebeneinander, die allerdings durch den fließenden Strom in Bewegung geraten und damit Instabilitäten bewirken. Durch Inhomogenitäten (Gitterfehler) im Aufbau des Supraleiters lassen sich sogenannte *Pinning-Zentren* schaffen, die eine Stabilisierung herbeiführen. Derart stabilisierte Systeme führen die Bezeichnung *Hochfeld-Supraleiter* oder *Supraleiter 3. Art*.

Die Leitfähigkeit eines Supraleiters ist gegenüber Kupfer ungefähr um soviel größer wie die Leitfähigkeit von Kupfer gegenüber einem Isolator (etwa 20 Zehnerpotenzen)!

19.6 Halleffekt

Die Driftgeschwindigkeit bewegter Ladungsträger führt zu deren Ablenkung in einem Magnetfeld (Abb. 19.5). Die quer zum Stromfluss auftretende *Hallspannung* U_H ist

$$U_H = R_H \frac{B}{d} I \qquad \begin{array}{c|c|c|c|c} U & B & d & I & R_H \\ \hline \mathrm{V} & \mathrm{T} & \mathrm{mm} & \mathrm{A} & \mathrm{cm^3/A\,s} \end{array} \qquad (19.2)$$

mit B = Flussdichte, I = Strom, d = Leiterdicke (Abb. 19.5) und R_H = Hallkonstante $= Ale \cdot n = A \cdot \mu/\gamma$. Die Hallkonstante ist damit von der Beweglichkeit μ und der Leitfähigkeit γ bzw. Teilchenkonzentration n und Elementarladung e abhängig. Bei normalen Leitern ist die Konstante $A = 1$, bei Halbleitern $A = 3\pi/8$.

Die Ablenkung der Strombahnen durch das Magnetfeld führt zu einem Winkel gegenüber der ursprünglichen (Längs-)Richtung, dem Hallwinkel Q_H

$$\tan \Theta_H = \mu \cdot B \qquad \begin{array}{c|c|c} \Theta & \mu & B \\ \hline ^\circ & \mathrm{cm^2/V\,s} & \mathrm{T} \end{array} \qquad (19.3)$$

d. h. der Hallwinkel ist nur von der Beweglichkeit der Ladungsträger (μ_n oder μ_p) und der magnetischen Flussdichte abhängig.

Weil bei vorgegebenen Abmessungen die Hallspannung außer von einem eingeprägten Strom I nur noch vom einwirkenden Magnetfeld B abhängt, steht damit als *Hallgenerator* ein einfacher Messsensor für magnetische Felder zur Verfügung.

In Abb. 19.6a ist die auf den Ruhestrom normierte Ausgangsspannung eines Hallgenerators über der magnetischen Flussdichte B leerlaufend (R_∞) und mit einem für beste Linearisierung der Kennlinie (R_{lin}) ermittelten Abschlusswiderstand

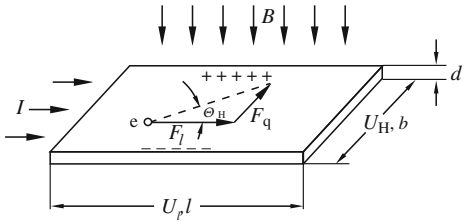


Abb. 19.5 Halleffekt (Prinzip). Stromdurchflossener Leiter (Länge l, Breite b und Dicke d) im Magnetfeld. I = Strom (Betriebsstrom), B = Magnetfeld, F_l, F_q = Längs- bzw. Querkraft, U_l = Längsspannung (Betriebsspannung), U_H = Hallspannung und Q_H = Hallwinkel

Teil III

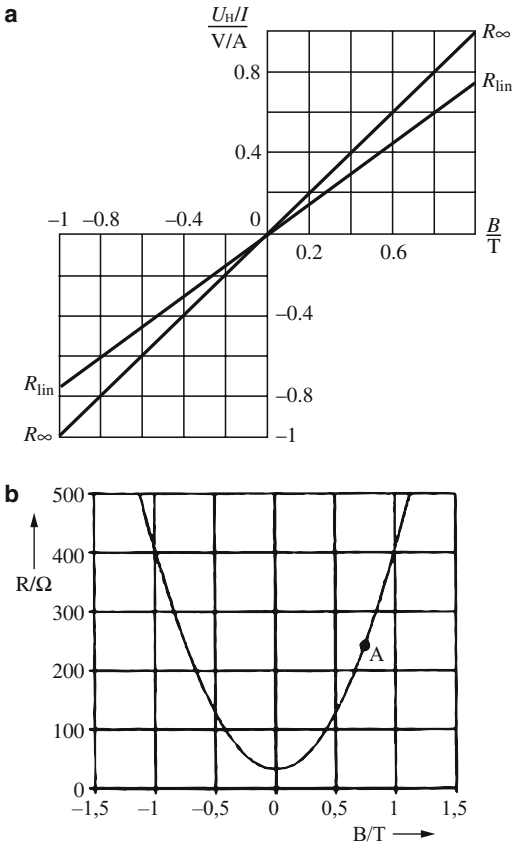


Abb. 19.6 **a** Normierte Hallspannung in Abhängigkeit vom anliegenden Magnetfeld. Im Leerlauf (R_∞) und mit Abschlusswiderstand R_{lin} für optimal lineare Kennlinie. **b** Widerstand in Abhängigkeit vom anliegenden Magnetfeld für eine InSb-Feldplatte. Der Widerstand ist von der Polarität des Magnetfeldes unabhängig und verläuft nichtlinear. Eine magnetische „Vorspannung" (Punkt *A*) liefert eine bessere lineare Kennlinie und eine Feldpolaritätserkennung

dargestellt. Wie sofort ersichtlich und wie es auch dem Prinzip der Entstehung der Hallspannung entspricht, ist ihr Vorzeichen von der Polarität des Magnetfeldes abhängig. Magnetische Wechselfelder erzeugen dementsprechend eine – vorzeichenrichtige – Wechselspannung.

Die Ruheströme für Hallgeneratoren liegen zwischen einigen und einigen hundert mA. Bei Flussdichten von 1 T erreichen die Hallspannungen im Leerlauf Werte von 0,1 bis 1 V.

Der Hallwinkel hat eine Verlängerung der Strombahnen im Leiter zur Folge, woraus ein erhöhter Widerstand resultiert. Dieser

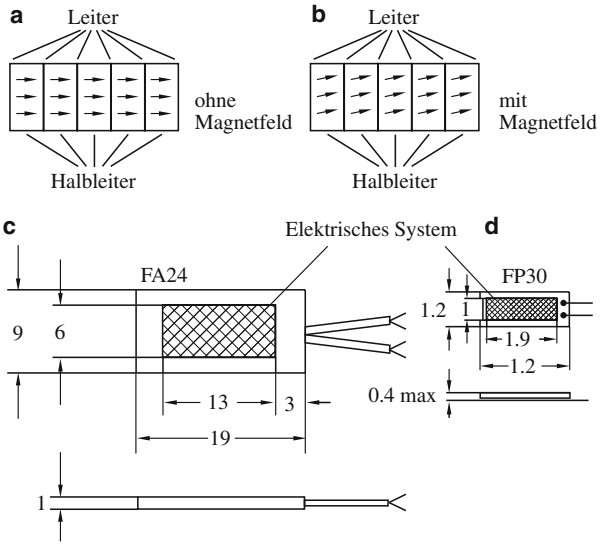


Abb. 19.7 Feldplatte. **a** ,**b** prinzipieller Aufbau (Erläuterungen im Text), **c** Abmessungen eines Hallgenerators (Siemens FA 24), **d** Abmessungen einer Feldplatte (Siemens FP 30)

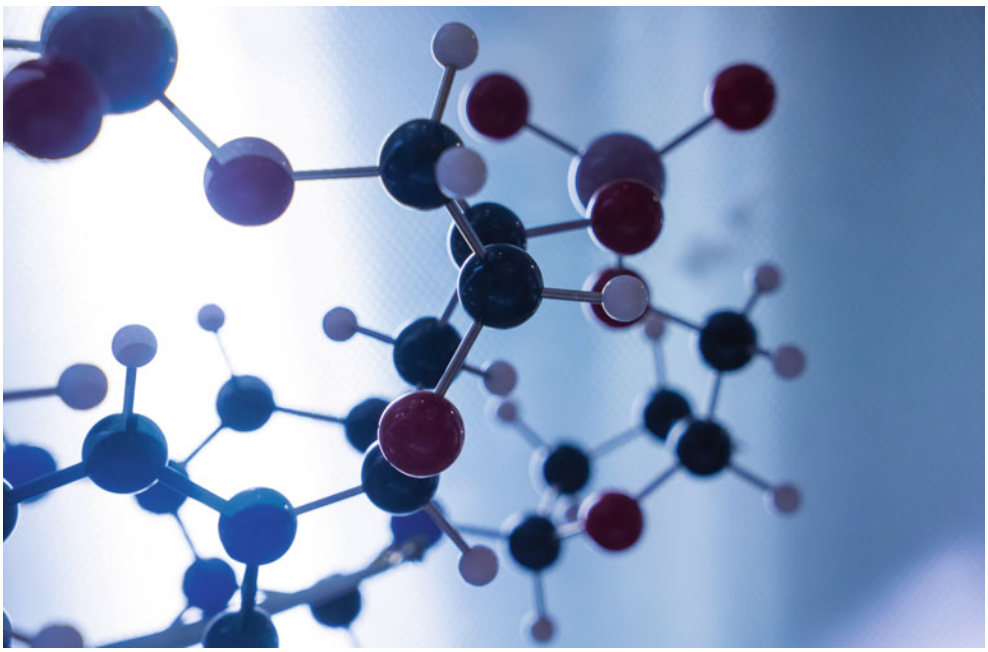
Effekt wird in der *Feldplatte* ausgenutzt. Im Prinzip wechseln dazu ein Material mit großer Hallkonstante (z. B. InSb oder InAs) mit einem guten Leiter, d. h. Material mit verschwindender Hallkonstante, ab (Abb. 19.7a, b). Die Pfeile symbolisieren in den Abbildungen die Strombahnen. Unter dem Einfluss eines Magnetfeldes weisen sie einen längeren Weg auf. Die entstehenden Hallspannungen werden durch die querliegenden – Äquipotenzialflächen erzwingenden – Leiter kurzgeschlossen. Damit liegt ein *magnetfeldabhängiger Widerstand* vor. Wie aus dem Diagramm der relativen Widerstandsänderungen in Abb. 19.6b hervorgeht, sind die Zusammenhänge nur stark nichtlinear. Auch ist die Polarität des Magnetfeldes nicht erkennbar, da nur die Stärke des Feldes in die Widerstandsänderung eingeht. Wenn nicht mit einer Vormagnetisierung gearbeitet wird, bewirkt ein magnetisches Wechselfeld eine Ausgangspannung doppelter Frequenz am „Arbeitswiderstand" Feldplatte.

In der Praxis werden diese Elemente im mm-Bereich als Hall- bzw. Magnetoresistive-Sensore (MRS) in Rund- bzw. Flachstäben oder IC-Gehäusen mit Stromversorgung und (Temperatur-) Stabilisierung beschaltet ausgeliefert. Einfacher Anschluss an eine Betriebsspannung von etwa 3–12 V ergibt dann an den Ausgangsklemmen die gewünschte Information über das zu messende Magnetfeld.

Elektrische Leiter

20

Egon Döring



20.1 Normalleiter 274
20.2 Halbleiter 274
20.3 Supraleiter 276

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_20

20.1 Normalleiter

Die Daten der wichtigsten Leitermaterialien sind in der Tab. 20.1 zusammengestellt. Für Leitkupfer ist der Mindestwert der elektrischen Leitfähigkeit (DIN EN 13604:2002) $58 \cdot 10^6$ S/m bzw. der spezifische Widerstand $0{,}01724 \cdot 10^{-6}\ \Omega$ m. Je nach Vorbehandlung und Reinheit kann die elektrische Leitfähigkeit des Kupfers zwischen $60 \cdot 10^6$ S/m bei extremer Reinheit und $55 \cdot 10^6$ S/m bei geringfügigen Verunreinigungen schwanken. Aus diesem Grunde kommt in der Elektrotechnik ausschließlich *elektrolytisch raffiniertes E-Kupfer* (Nummer CW003A, Kurzzeichen Cu-ETP1) oder das *sauerstofffreie SE-Kupfer* (Nummer CW007A, Kurzzeichen Cu-OF1) in Betracht (EN 13604:2002). Mit zunehmender Temperatur sinkt die Leitfähigkeit bzw. steigt der Widerstand der Leiterwerkstoffe. Generell sind metallische Leiter als *Kaltleiter* oder PTC (engl. positive temperature coefficient) zu klassifizieren. Eine geringe Temperaturabhängigkeit weisen die wegen ihrer guten Langzeitkonstanz zunehmend eingesetzten Metallfilmwiderstände auf. Preiswert galvanisch hergestellt, erreichen sie Temperaturkoeffizienten $\alpha \cong 0{,}02 \ldots 0{,}005\ \%/°C$ (s. a. Tab. 20.1).

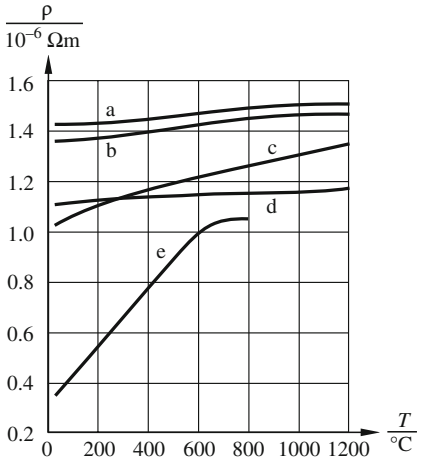


Abb. 20.1 Spezifischer Widerstand einiger Heizleitermaterialien in Abhängigkeit von der Temperatur (nach Thyssen-Krupp). **a** Nicrofer 6023 H (NiCr23Fe), **b** Nicrofer 7216 (NiCr15Fe), **c** Nicrofer 3228 NbCe (X6NiCrNbCe 32–27), **d** Crofer 22 APU (X1CrTiLa22), **e** Nickeleisen (NiFe29Cr1) (Standardlegierung)

Für heiztechnische Zwecke sind die Anforderungen an den Temperaturgang der Heizleiter (DIN 46460, 46463, Ni-Widerstandslegierungen) wesentlich geringer, dafür ist der Temperaturbereich wesentlich größer. Abb. 20.1 zeigt den spezifischen Widerstand für einige Heizleitermaterialien in Abhängigkeit von der Temperatur. Auffällig ist hierbei das Verhalten der NiFe-Legierung mit zunächst Kaltleitercharakteristik, wie bei Metallen üblich, die beim Erreichen der Curie Temperatur (siehe Abschn. 21.2 „Verhalten von Materie im Magnetfeld“) in einen praktisch konstanten Widerstand übergeht.

20.2 Halbleiter

Beim reinen Halbleiter ist, wegen des mittig in der verbotenen Zone liegenden Fermi-Niveaus, die Anzahl der ins Leitungsband übergebenden Elektronen n_n. gleich der Anzahl der im Valenzband frei werdenden Defektelektronen n_p. Die so rein thermisch erzeugte Teilchendichte heißt *Intrinsicdichte* oder auch *Intrinsiczahl* n_i. Allgemein gilt der Zusammenhang

$$n_n \cdot n_p = n_i^2 \quad \frac{n}{\mathrm{cm}^{-3}} \tag{20.1}$$

Aus der Intrinsiczahl folgt die *Intrinsic-* oder *Eigenleitfähigkeit*

$$\gamma_i = e n_i (\mu_n + \mu_p) \quad \begin{array}{c|c|c|c} \gamma & e & n & \mu \\ \hline \mathrm{S/cm} & \mathrm{A\,s} & \mathrm{cm}^{-3} & \mathrm{cm}^2/\mathrm{V\,s} \end{array} \tag{20.2}$$

Die Grunddaten für Germanium (Ge) und Silizium (Si) sind in der Tab. 20.2 zusammengestellt. Daraus ist sofort ersichtlich, dass zum Erreichen der Eigenleitfähigkeit die Konzentration von Fremdteilchen kleiner als ca. $10^{13}\ \mathrm{cm}^{-3}$ bei Ge und kleiner als ca. $10^{10}\ \mathrm{cm}^{-3}$ bei Si sein muss. Bei ca. 10^{20} Teilchen pro cm^3 für Metalle bedeutet dieses eine Reinheit von besser 10^{-9} (= 1 ppb = 1 part per billion = 1 Fremdteilchen auf 1 Milliarde (amerik. billion!) Teilchen) bei Ge und besser 10^{-12} bei Si. Wegen der Abhängigkeit der Leitfähigkeit vom Reinheitsgrad kann umgekehrt über die Messung der Leitfähigkeit eines Halbleitermaterials dessen Reinheit bestimmt werden.

Dotieren ist die Änderung der Leitfähigkeit eines hochreinen Halbleiters durch Verschieben des Fermi-Niveaus. Es erfolgt über das Einbringen von Störungen durch fremde Atome mit

Tab. 20.1 Daten einiger Leiterwerkstoffe. Der spezifische Widerstand ρ in $10^{-6}\ \Omega$ m ist zahlenmäßig gleich der gebräuchlichen Angabe in $\Omega\ \mathrm{mm}^2/\mathrm{m}$

Metall	Leitfähigkeit gegenüber Silber = 100 %	Elektrische Leitfähigkeit γ	Spezifischer Widerstand ρ	Temperaturkoeffizient α
	%	10^6 S/m	$10^{-6}\ \Omega$ m	%/°C
Silber	100	61	0,0163	0,41
Kupfer	96	58	0,0173	0,43
Gold	73	41	0,025	0,39
Aluminium	62	38	0,026	0,43
Zink	31	17	0,059	0,42
Nickel	25	15	0,067	0,69
Eisen	18	10	0,098	0,56
Zinn	15	9	0,115	0,46
Blei	8	4,5	0,222	0,42

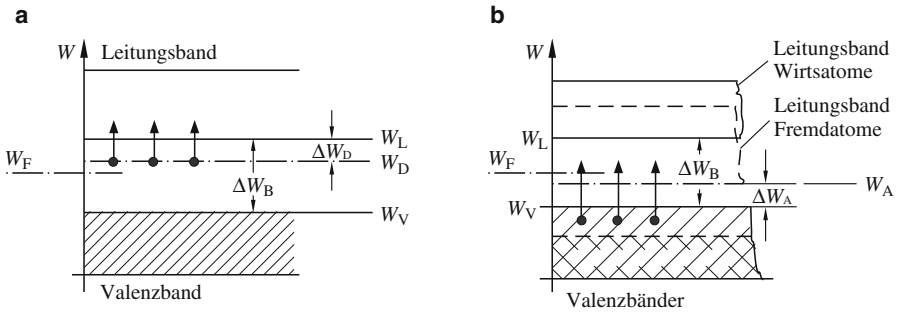


Abb. 20.2 Dotierung (schematisch). **a** Bandschema eines Überschussleiters, n-Leitung, **b** Bandschema eines Defektelektronenleiters, p-Leitung, ▭ Leitungsband, ▨ Valenzband Wirtsatome, ▨ Valenzband Fremdatome, • Elektronen, W_L, W_V Leitungs- bzw. Valenzbandkanten der Wirtsatome, W_F Fermi-Niveau, ΔW_B Bandabstand der Wirtsatome, W_D Donatorniveau, ΔW_D Donatorabstand, W_A Akzeptorniveau, ΔW_A Akzeptorabstand

Tab. 20.2 Intrinsiczahlen und Ladungsträgerbeweglichkeiten von Germanium und Silicium

Material	n_i	μ_n	μ_p
	cm^{-3}	$cm^2/V\,s$	$cm^2/V\,s$
Ge	$2{,}4 \cdot 10^{13}$	3900	1900
Si	$1{,}5 \cdot 10^{10}$	1350	850

einem anderen Fermi-Niveau als dem des Halbleiters. Liegt das eingebrachte Fermi-Niveau nahe dem Leitungsband, stehen zusätzliche Elektronen für den Leitungsmechanismus zur Verfügung, sog. *n-Dotierung* (Abb. 20.2a). Wegen der vom *Donator* im Übermaß vorhandenen Elektronen auch *Überschuss-* oder *n-Leitung* genannt. Liegt andererseits das Fermi-Niveau des störenden Gitters nahe dem Valenzband, sog. *p-Dotierung*, werden durch den *Akzeptor* vermehrt Elektronen aus dem Valenzband abgezogen, und es entsteht die *p-Leitung* oder *Defekt(elektronen)leitung* (Abb. 20.2b). Die Dotierung erfolgt durch Diffusion oder Ionenimplantation.

Da (20.1) unverändert gültig ist, wird mit der Zunahme einer Ladungsträgerart die andere automatisch vermindert und führt damit zu einem *bipolaren* Leitungsmechanismus. Halbleiter finden sich vorzugs-weise in der 4. Hauptgruppe (IVa) des Periodensystems. Die Tab. 20.3 zeigt einen Ausschnitt daraus. Darin sind die als Akzeptoren (A) geeigneten Elemente der 3. Hauptgruppe und die Donatoren (D) der 5. Hauptgruppe entsprechend gekennzeichnet. Der Bandabstand (Bandlücke) gibt die Breite der verbotenen Zone in eV an.

Verbindungshalbleiter haben neben den Elementhalbleitern eine besondere Bedeutung. Bringt man Elemente der III. und V. Hauptgruppe des Periodensystems im stöchiometrischen Verhältnis zusammen, so ist bei derartigen *III-V-Verbindungen* wieder eine ähnliche Konfiguration der Valenzelektronen wie bei Elementhalbleitern möglich. Bei geeigneten Bandstrukturen kann der Leitungsmechanismus ähnlich wie bei Elementhalbleitern durch Dotieren gesteuert werden. In der Tab. 20.4 sind an einigen ausgewählten Beispielen die wesentlichen Daten verschiedener Verbindungs-Halbleiter aufgezeigt, wobei nicht nur

Tab. 20.3 Ausschnitt aus dem Periodensystem der Elemente. ▨ Halbleiter, (A) Akzeptoren und (D) Donatoren (gegenüber IV), a Hauptgruppenelemente, b Nebengruppenelemente

Gruppe	3 Valenzelektronen	4 Valenzelektronen	5 Valenzelektronen	6 Valenzelektronen	7 Valenzelektronen
	III a b	IV a b	V a b	VI a b	VII a b
Symbol	B	C	N	O	F
Ordn.-Zahl	5	6	7	8	9
B-abst. eV	(A) 1,1	6			
Symbol	Al	Si	P	S	Cl
Ordn.-Zahl	13	14	15	16	17
B-abst. eV	(A)	1,09	(D) 0,33		
Symbol	Ga	Ge	As	Se	Br
Ordn.-Zahl	31	32	33	34	35
B-abst. eV	(A)	0,67	(D) 1,1	2,05	
Symbol	In	Sn	Sb	Te	J
Ordn.-Zahl	49	50	51	52	53
B-abst. eV	(A)	0,09	(D)	0,34	1,3

Tab. 20.4 Daten einiger Verbindungshalbleiter. Zum Vergleich sind die Elementhalbleiter Si und Ge mit aufgeführt. E_G Bandlücke (Bandabstand), r_i Intrinsicwiderstand, *Smp* Schmelzpunkt, μ_n Elektronenbeweglichkeit, μ_p Löcherbeweglichkeit

Name	Allgem. Formel	Element-gruppen	E_g	ρ_i	*Smp*	μ_n	μ_p	Bemerkungen
			eV	Ω cm	°C	$cm^2/V\,s$	$cm^2/V\,s$	
Silizium	Si		1,106	$2{,}3 \cdot 10^5$	1420	1350	480	
Germanium	Ge		0,67	46	936	3900	1900	
Binäre Systeme								
Cd-Tellurid	CdTe	II–VI	1,56		1041	1050	100	
Zn-Oxid	ZnO	II–VI	3,35			200	180	
Zn-Sulfid	ZnS	II–VI	3,68		1185	165	5	
Ga-Arsenid	GaAs	III–V	1,42	$5 \cdot 10^7$	1238	8500	400	Elektrolumineszens, LED
In-Antimonid	InSb	III–V	0,17	0,0045	523	75000	780	Hallgenerator, Feldplatte
In-Arsenid	InSb	III–V	0,354	0,16	942	$< 4 \cdot 10^4$	$< 5 \cdot 10^2$	
Ga-Phophid	GaP	III–V	2,26	–	1475	< 250	< 150	Grüne LED
Ga-Antimonid	GaSb	III–V	0,726	10^3	712	< 3000	< 1000	
Ga-Nitrid	GaN	III–V	3,2	0,1	2500	< 1000	< 350	
In-Nitrid	InN	III–V	1,97	0,5	1373	250	35	
Al-Nitrid	AlN	III–V	6,2	10^{12}	3273	300	14	
Bor-Nitrid	BN	III–V	6,4	–	2973	< 200	< 500	
Si-Karbid	SiC	IV–IV	2,4	–	–	60	8	Blaue LED
Ternäre Systeme								
Aluminium-GaAs	Al-Ga-As	III–III–V	$> 0{,}45$	10^{12}	1240	10^4	900	
Indium-GaAs	Ga-In-As	III–III–V	0,74	–	$\simeq 1100$	$< 12 \cdot 10^3$	< 300	
Indium-GaSb	Ga-In-Sb	III–III–V	–	–	–	$< 10^4$	$< 10^3$	
IndiumAsSb	In-As-Sb	III–III–V	–	–	–	$\simeq 10^5$	< 500	
Quaternäre Systeme								
Indium-Antimon-GaAs	Ga-In-As-Sb	III–III–V–V	0,36	–	$\simeq 1100$	$\simeq 3000$	< 100	Weiße LED
Indium-Phosphor-GaAs	Ga-In-As-P	III–III–V–V	0,35 … 2,27	–	–	< 5400	< 200	Weiße LED

einfache III-V-Verbindungen – *binäre Systeme* – berücksichtigt wurden sondern auch drei – *ternäre Systeme* – und vier Komponenten „Mischungen" –*quaternäre Systeme*. Ternäre und Quaternäre Systeme spielen eine bedeutende Rolle für LEDs – light emitting diodes – Leuchtdioden bei farbigen aber besonders auch weißen Beleuchtungen. Weiße LEDs – zunehmend auf GaN Basis – arbeiten entweder mit einer blau leuchtenden LED und einem fluoreszierenden Leuchtstoff oder aber mit mehreren unterschiedlich farbigen LEDs deren Summenfarbe Weiß ergibt.

Vergleichsweise sind auch Germanium und Silizium noch mit aufgeführt.

Im Bereich der Beleuchtung gewinnen organische LEDs – OLEDs – zunehmend an Bedeutung, da sie mit organischen Halbleitern, vorzugsweise Poly-Phenylen-Vinylen in Verbindung mit Metallen, als metall-organische Halbleiter flexible, leuchtende und steuerbare Anwendungen ermöglichen.

Zur Herstellung der für Halbleiterbauelemente notwendigen Sperrschichten, d. h. Übergänge zwischen p- und n-leitenden Materialien, ist noch die *Umdotierung* wichtig. Durch eine stärkere als die vorher eingebrachte Dotierung kann die p- oder n-Leitung umgekehrt werden. Mit Hilfe dieser Umdotierung können auch Mehrfachfolgen von p- und n-Schichten erzeugt werden. Nur so ist die heute übliche Planartechnologie zur Herstellung integrierter Schaltungen möglich.

20.3 Supraleiter

Das grundsätzlich bereits über eine recht hohe Sprungtemperatur verfügende Niob (9,2 K, siehe Tab. 18.2 Nb) liefert als NbTi bzw. Nb_3Sn zurzeit die gebräuchlichsten Supraleitermaterialien. NbTi liegt als leicht zu fertigender Legierungswerkstoff vor, wohingegen Nb_3Sn als Verbindungssupraleiter einen bestimmten Formierungsprozess erfordert. NbTi ermöglicht bei einer Temperatur von 2 K (Heliumkühlung) Flussdichten von 10 T und Nb_3Sn solche bis 16 T.

Hochtemperatursupraleiter (HTS) auf Metalloxid-Basis haben inzwischen Sprungtemperaturen bis 133 K (Tab. 19.2) erreicht.

Eine völlig andere Art von Supraleitern könnten die Fullerene, eine kugelförmige C_{60} Modifikation des Kohlenstoffes, als organische SL liefern. Die aus fünf- und sechseckigen Flächen bestehenden-fußballähnlichen-Oberflächenstrukturen haben ihnen den Namen Fußballmoleküle eingebracht. C_{60} Fullerene wird mit eingebauten Alkali Atomen bei etwa 10 … 40 K supraleitend. J.H. Schön, C. Kloc und B. Batlogg erreichten durch Einbau von Chloroform- bzw. Bromoform-Moleküle eine Sprungtemperatur von 117 K.

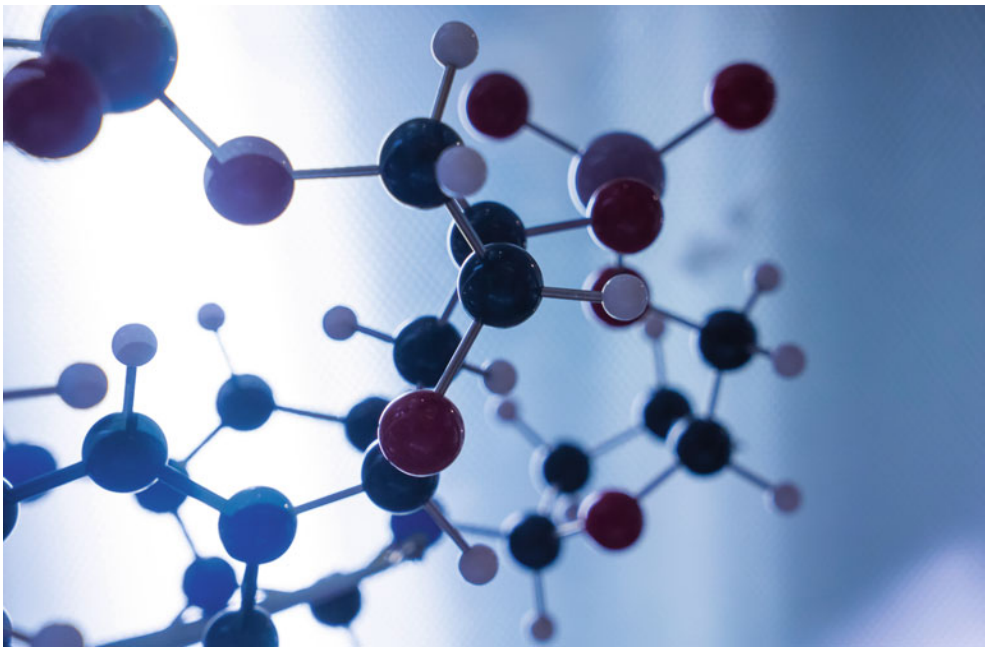
Supraleitendes Material wird zur Stabilisierung in einen normalleitenden Träger (Substrat) auf Cu-Basis (Cu-Sn- oder Cu-Ni-Matrix je nach Herstellprozess) eingebracht. Die dadurch bedingte Aufteilung des Leiters in Normal- und Supraleiter wird

durch das *Querschnittsverhältnis* $\alpha = q_{Cu}lq_{SL}$ beschrieben, mit q_{Cu} = Kupferquerschnitt und q_{SL} = Supraleiterquerschnitt. Selbst wenn die zulässigen Stromdichten für Normalleiter mit 10^3 A/cm^2 angesetzt werden, ist die Stromtragfähigkeit von Supraleitern immer noch um mehr als 10^2 größer als die von Normalleitern. Praktisch brauchbare Formen von Hochtemperatur-Supraleitern (Stickstoffkühlung) haben Fertigungs- und Anwendungsprobleme da sie nicht einfach als Drähte sondern nur als (schmale) Bänder hergestellt werden können, die einen geeigneten Träger benötigen. Der Einsatz von Supraleitern für die Anwendung in Energieübertragungssystemen, Motoren und Generatoren befindet sich daher noch im Versuchsstadium.

Magnetische Leitfähigkeit

21

Egon Döring



21.1 Modellvorstellung 280
21.2 Verhalten von Materie im Magnetfeld 280
21.3 Magnetisierung 281
21.4 Magnetisierungskurve 282
21.5 Permeabilität 283

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_21

21.1 Modellvorstellung

Die zur elektrischen Elementarladung äquivalente magnetische Größe ist das *Bohr'sche Magneton*

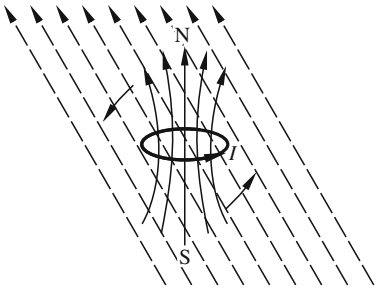
$$\mu_B = 9{,}27 \cdot 10^{-24}\,\mathrm{A\,m^2} \quad (1\,\mathrm{A\,m^2} = 1\,\mathrm{J/T})$$

Im Gegensatz zur „einpoligen" Elementarladung ist das Bohr'sche Magneton zweipolig und damit ein *magnetischer Dipol* mit Nord- und Südpol. Die gleichfalls benutzte Bezeichnung *magnetisches Moment* geht auf die Tatsache zurück, dass die magnetischen Dipole nur verdreht, aber nicht verschoben werden können.

Zur Modellvorstellung des magnetischen Dipols ist in Abb. 21.1 die Magnetpolbildung eines *Ringstromes* dargestellt (die unvollständig gezeichneten Feldlinien sollen die Vorstellung erleichtern). Gelangt die Anordnung in ein magnetisches Feld (gestrichelt in Abb. 21.1), so wird der Dipol eine Drehbewegung ausführen, ohne den Platz zu verlassen. Die Größe des für die Drehbewegung notwendigen Drehmomentes M folgt aus

$$M = m_B \cdot B \cdot \sin\alpha \qquad \begin{array}{c|c|c|c} M & m_B & B & \sin\alpha \\ \hline \mathrm{J} & \mathrm{J/T} & \mathrm{T} & 1 \end{array} \qquad (21.1)$$

Hierbei ist B die Flussdichte, m_B, das *magnetische (Dipol-) Moment* und α der (kleinste) Winkel zwischen „Magnet" und Feld. Wichtig ist das magnetische Moment m_B des Modells, dessen Betrag sich als Strom multipliziert mit der umrandeten Fläche ergibt. Betrachtet man den Stromring als Elektronenbahn und die Fläche als zum Bahnradius gehörig, so ist das damit verkoppelte magnetische Moment als *Bahnmoment* zu deuten. Der mechanisch als Eigenrotation gedeutete Spin der Elektronen führt dementsprechend zum *Spinmoment*. Alle magnetischen Erscheinungen der Materie, die für den praktischen Gebrauch bedeutsam sind, gehen damit auf die Überlagerung von Spin- und Bahnmomenten zurück.



Magnetfeld
von einem Nordpol kommend

Abb. 21.1 Kreisstrom (stromdurchflossener Ring) als magnetischer Dipol

21.2 Verhalten von Materie im Magnetfeld

Messzahl für die Reaktion eines Werkstoffes auf ein Magnetfeld, d. h. seine magnetische Leitfähigkeit ist entweder die *Permeabilität* μ oder die (*magnetische*) *Suszeptibilität* $\varkappa_m$. Für den freien Raum gilt – unter Verzicht auf die Kennzeichnung des Vektorcharakters von Flussdichte B und magnetischer Feldstärke H

$$B = \mu_0 \cdot H \qquad \begin{array}{c|c|c} B & H & \mu_0 \\ \hline \mathrm{T} & \mathrm{A/m} & \mathrm{V\,s/A\,m} \end{array} \qquad (21.2)$$

mit $\mu_0 = 1{,}26 \cdot 10^{-6}\,\mathrm{V\,s/A\,m}$ (bzw. H/m). Für den materieerfüllten Raum ist eine multiplikative oder additive Beschreibung möglich:

$$B = \mu_0 \cdot \mu_r \cdot H \quad \text{bzw.} \qquad (21.3a)$$

$$B = \mu_0 \cdot H + J \qquad (21.3b)$$

B ist die erreichte Flussdichte in $\mathrm{V\,s/m^2}$ oder T und H das erregende magnetische Feld in A/m. J ist die mit B dimensionsgleiche *magnetische Polarisation*. Die relative Permeabilitätszahl μ_r, ist der Multiplikationsfaktor für die Flussdichte im materieerfüllten Raum gegenüber dem Vakuum. Die Polarisation J gibt den auf die Materie entfallenden Flussdichteanteil an, der zusätzlich zum Anteil $\mu_0 H$ des Vakuums auftritt. Aus (21.3a) und (21.3b) folgt:

$$\mu_r = 1 + \frac{J}{\mu_0 H} = 1 + \frac{M}{H} = 1 + \varkappa_m \qquad \begin{array}{c|c|c|c} J, M & H & \mu_0 & \mu_r, \varkappa_m \\ \hline \mathrm{T} & \mathrm{A/m} & \mathrm{V\,s/A\,m} & 1 \end{array} \qquad (21.4)$$

mit $\varkappa_m$ = *Suszeptibilität*(szahl) und $M = J/\mu_0$ = *Magnetisierung*. M ist der Feldstärkeanteil, der für die zusätzliche Flussdichte in der Materie aufzubringen ist. Der für die Praxis wichtige Zusammenhang zwischen J und B folgt aus (21.3a) und (21.3b) zu

$$J = B - \mu_0 H \qquad \begin{array}{c|c|c|c} J & B & H & \mu_0 \\ \hline \mathrm{T} & \mathrm{T} & \mathrm{A/m} & \mathrm{V\,s/A\,m} \end{array} \qquad (21.5)$$

Diamagnetismus: Trifft ein äußeres Magnetfeld auf die Kreisströme der Elektronenbahnen, so wird – gemäß Lenz'scher Regel (siehe Grundlagen der Elektrotechnik) – ein Feld induziert, das dem erzeugenden entgegen gerichtet ist. Es entsteht ein sog. *induziertes magnetisches Moment*. Dies ist die einfachste Form des Magnetismus, die grundsätzlich in allen Stoffen vorkommt. Das resultierende Feld ist schwächer als das erregende. Materialien, die ausschließlich diesen Effekt zeigen, heißen *diamagnetisch*. Die Suszeptibilität $\varkappa_m$ ist negativ, etwa -10^{-6}. Die Permeabilität μ_r ist dann praktisch gleich 1 (Abb. 21.2).

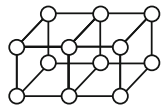
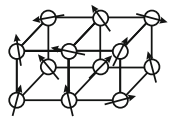
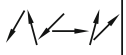
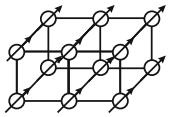
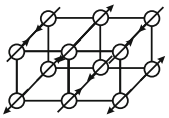
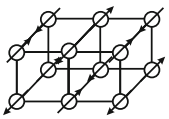
Magnetische Momente im kubischen Gitter	vereinfachte Darstellung	Bezeichnung	μ_r	$\varkappa_m$
		dia-magnetisch	< 1 (≈ 1)	< 0 (≈ 0)
		para-magnetisch	> 1 (≈ 1)	> 0 (≈ 0)
		ferro-magnetisch	$\gg 1$	$\gg 0$
		antiferro-magnetisch	> 1 (≈ 1)	> 0 (≈ 0)
		ferri-magnetisch	$\gg 1$	$\gg 0$

Abb. 21.2 Ordnungszustände magnetischer Momente (schematische Übersicht). ↑ magnetisches Moment mit relativer Größe und Richtung

Zu den diamagnetischen Stoffen gehören die Edelgase, die nichtmetallischen Elemente – außer O_2 und S_2 – sowie die Metalle Ag, Au, Cd, Cu, Hg, Pb und Zn, ferner die Halbleitermaterialien Ge, Si und Se. Normalerweise ist die Suszeptibilität temperaturunabhängig. Ausnahmen hiervon machen Sb, Bi, Ga, Tl und C (als Graphit).

Paramagnetismus liegt vor, wenn sich die magnetischen Momente im atomaren Bereich unvollständig kompensieren. Die resultierenden magnetischen Momente mit ihren willkürlichen Orientierungen werden durch ein äußeres Magnetfeld ausgerichtet. Die resultierende Flussdichte ist etwas größer als im Vakuum. Die Suszeptibilität ist positiv mit Werten von etwa $\varkappa_m = 10^{-5} \ldots 10^{-3}$. Die Permeabilitätszahl μ_r ist damit praktisch nur unwesentlich größer als 1.

Da die Wärmebewegung der Teilchen der ausrichtenden Wirkung des Magnetfeldes entgegenwirkt, ist der Paramagnetismus temperaturabhängig.

Ferromagnetismus: Kompensieren sich im atomaren Bereich erst magnetische Momente größerer, gleich ausgerichteter Gebiete, den sog. *Weiss'schen Bezirken, Elementarbereichen* oder *Domänen*, so liegt *Ferromagnetismus* vor. Die Suszeptibilität $\varkappa_m$ und Permeabilität μ_r erreichen sehr hohe Werte. Dieser Ausnahmezustand in der Ordnung der Spinmomente wird durch *Austauschkräfte* hervorgerufen, denen mit steigender Temperatur die Wärmebewegung der Atome(-rümpfe) entgegenwirkt, so dass ab einer bestimmten Temperatur – der *Curie-Temperatur* – wieder der ungeordnete, d. h. paramagnetische Zustand, vorherrscht.

Antiferromagnetismus: Die Austauschkräfte können im Gitter auch zu einer paarweise antiparallelen Ordnung der magnetischen Momente führen. Dieser Ordnungszustand heißt *antiferromagnetisch*. Nach außen hin ist ein solches System durchweg unmagnetisch und verhält sich bei den üblichen Untersuchungen wie ein paramagnetischer Stoff. Auch dieser Ordnungszustand wird thermisch aufgehoben. Die zugehörige Temperatur heißt *Néel-Temperatur*.

Ferrimagnetismus: Sind bei antiparalleler Ausrichtung die magnetischen Momente ungleich groß, so verbleibt in einer Richtung ein resultierendes magnetisches Moment, mit ähnlicher Wirkung wie bei ferromagnetischen Materialien. Dieser unvollständig kompensierte Antiferromagnetismus wird *Ferrimagnetismus* genannt.

In Abb. 21.2 sind die möglichen Ordnungszustände der magnetischen Momente schematisiert zusammengestellt.

21.3 Magnetisierung

Die Weiss'schen Bezirke sind durch Austauschkopplungen bis zur Sättigung magnetisiert. Diese spontane Magnetisierung erfolgt im Kristallgitter immer nur in den Richtungen der leichten Magnetisierbarkeit, den *(magnetisch) leichten Richtungen*. Beim kubischen Gitter des Eisens sind dies z. B. die Würfelkanten. Die Diagonalen dazu sind schwerer zu magnetisieren, d. h. es liegen *(magnetisch) schwere Richtungen* vor.

Die Ausrichtung der magnetischen Momente erfolgt wegen der Weiss'schen Bezirke durch *Wandverschiebungen* und *Drehprozesse*. (Bloch-)Wandverschiebungen (Abb. 21.3b) sind für Magnetisierungen mit kleinen Feldstärken wichtig. Mit zunehmender Feldstärke müssen auch Bezirke magnetisiert werden, die mit dem aufmagnetisierenden Feld Winkel $> 90°$ einschließen. Hierbei treten *irreversible Drehungen* auf, wobei die spontan magnetisierten Bezirke ihre Magnetisierungsrichtung um 180° ändern (Abb. 21.3c). Die weitere Ausrichtung bis zur Sättigung erfolgt über einfache *reversible Drehungen* (Abb. 21.3d). Die plötzlichen Richtungswechsel der Magnetisierung durch die irreversiblen Drehungen sind auch unter dem Namen *Barkhausensprünge* bekannt.

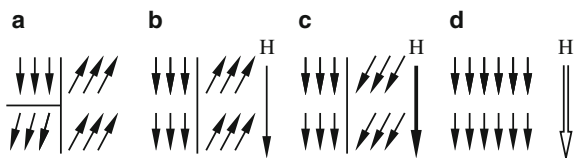


Abb. 21.3 Stufen der Magnetisierung. **a** ungestörter Zustand, **b** (Bloch-) Wandverschiebung, **c** irreversible Drehungen, **d** reversible Drehung, *H* magnetisierendes Feld zunehmender Stärke

21.4 Magnetisierungskurve

Der Zusammenhang zwischen der erregenden Feldstärke H und der daraus im Inneren des Werkstoffes resultierenden Flussdichte B, d. h. der Addition von J und $\mu_0 H$ ist bei Werkstoffen mit hoher Permeabilität im Allgemeinen nicht linear. Der nur experimentell zu ermittelnde Zusammenhang zwischen H und B wird durch die *Magnetisierungs-* oder *Hysteresekurve* (*-schleife*) beschrieben (Abb. 21.4). Für einen noch nie magnetisierten (oder einwandfrei entmagnetisierten) Werkstoff beginnt die *Neu-* oder *Kommutierungskurve* im Koordinatenursprung und wird im Anfangsbereich durch Wandverschiebungen (Abb. 21.4a), im Hauptteil durch irreversible Drehungen (Abb. 21.4b) und, im Bereich bis zur Sättigungsflussdichte B_s (Abb. 21.4c), durch reversible Drehungen bestimmt. Bei gegen Null absinkender Feldstärke sinkt die Flussdichte auf einem höher liegenden Kurvenast (Abb. 21.4d) gegen die *Remanenzflussdichte (Remanenzinduktion)* B_r ab. Dies ist die Auswirkung der irreversiblen Drehungen, die erst durch eine entgegen gerichtete *Koerzitivfeldstärke* $-H_c$ wieder bis auf $B = 0$ gebracht werden können. Eine weitere Steigerung von H in negativer Richtung führt zu einer negativen Sättigungsflussdichte, die ihrerseits erst wieder durch eine positive Koerzitivfeldstärke auf Null gebracht werden kann. Eine Rückkehr zum Koordinatenursprung ist auf diese Weise nicht möglich. Der Werkstoff verfügt über eine Hysterese.

Zur Entmagnetisierung muss von einer Sättigungsfeldstärke, Punkt 1 oder 2 in Abb. 21.5 mit allmählich absinkenden Feldstärkewerten ständig ummagnetisiert werden. Die gleichfalls mögliche thermische Entmagnetisierung durch Überschreiten der Curietemperatur dürfte nur in Sonderfällen anwendbar sein.

Verluste: Die Fläche der Hysterese Schleife, d. h. das Integral der Funktion B von H, stellt eine „Verlustfläche“ A dar. Diese Verluste treten bei jedem Durchlauf der Kurve auf und sind damit der Frequenz f proportional. Ist die Fläche der Hysterese Schleife gleich den Verlusten A in $\mathrm{W\,s/m^3}$ und das Materialvolumen gleich V, so ergibt sich die, auf die Masse bezogene, *Hysterese-Verlustleistung* p_h zu

$$p_h = \frac{A \cdot f}{\rho} \qquad \begin{array}{c|c|c|c} p_h & A & f & \rho \\ \hline \mathrm{W/kg} & \mathrm{W\,s/m^3} & \mathrm{s^{-1}} & \mathrm{kg/m^3} \end{array} \qquad (21.6)$$

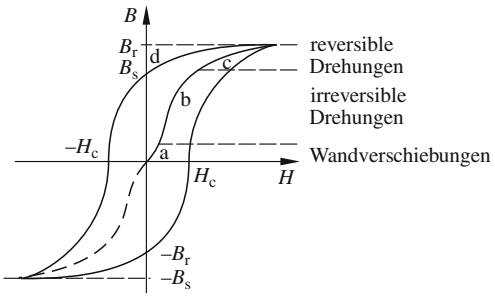


Abb. 21.4 Grundsätzlicher Verlauf und Entstehung einer Magnetisierungskurve oder Hystereseschleife (Erläuterungen im Text)

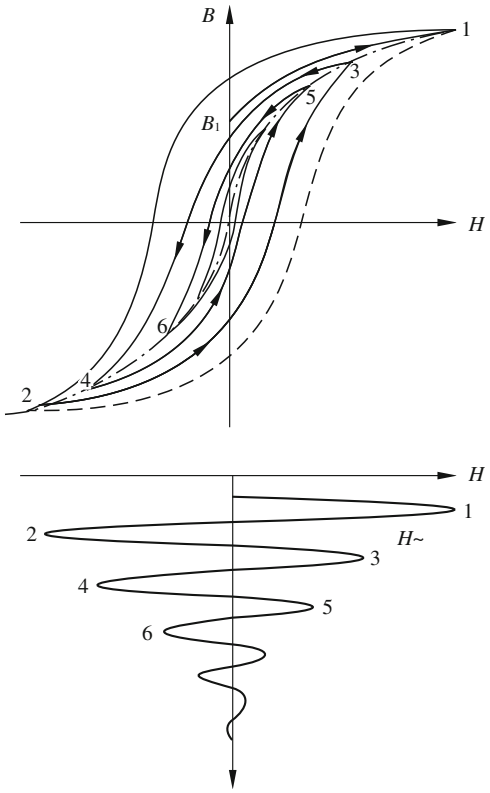


Abb. 21.5 Entmagnetisierung (schematisch). B_1 (Rest-)Vormagnetisierung, $H_\sim$ abklingendes, entmagnetisierendes Wechselfeld, 1, 2, 3, ... Umkehrpunkte auf der Kommutierungskurve, d. h. abklingende Scheitelwerte von $H_\sim$, –·–·– Kommutierungskurve und – – – Ergänzung der (äußersten) Hystereseschleife

mit ϱ = Dichte $(\mathrm{kg/m^3})$. Die elektrische Leitfähigkeit γ der meisten magnetischen Werkstoffe führt zu *Wirbelstromverlusten* p_w, die sich, bei der Verwendung von dünnen Blechen der Dicke d und sinusförmiger Flussdichte B, nach der klassischen Wirbelstromtheorie ergeben zu:

$$p_w = 1{,}64 \frac{\gamma}{\rho} (\hat{B} f d)^2 \qquad \begin{array}{c|c|c|c|c|c} p_w & \gamma & \rho & B & f & d \\ \hline \mathrm{W/kg} & \mathrm{S/m} & \mathrm{kg/m^3} & \mathrm{T} & \mathrm{s^{-1}} & \mathrm{m} \end{array} \qquad (21.7)$$

$\hat{B}$ = Scheitelwert der sinusförmigen Flussdichte und ρ = Dichte des Werkstoffes. Damit werden die im Allgemeinen zusammengefassten Verluste p

$$p = p_h + p_w \qquad \frac{p}{\mathrm{W/kg}} \qquad (21.8)$$

Bei konstanter Netzfrequenz $f = 50\,\mathrm{Hz}$ sind die Verluste nur noch von der Blechdicke d und der Aussteuerung B abhängig. Da die Blechdicke für den allgemeinen Fall variabel bleiben muss, die Aussteuerung aber bis in den nichtlinearen Bereich durchgeführt wird, ist die Flussdichte mit $\hat{B} = 1{,}0\,\mathrm{T}$, $\hat{B} = 1{,}5\,\mathrm{T}$ und $\hat{B} = 1{,}7\,\mathrm{T}$ genormt. Daraus folgen die *Verlustkennziffern* $p_{1,0}$, $p_{1,5}$ und $p_{1,7}$, meist in W/kg angegeben.

21.5 Permeabilität

Eine der wichtigsten Kenngrößen magnetischer Werkstoffe ist die Permeabilität μ. Sie gibt den zahlenmäßigen Zusammenhang zwischen Flussdichte B und Feldstärke H an:

$$\mu = \frac{\Delta B}{\Delta H} \tag{21.9}$$

Δ ist die zur jeweiligen Definition gehörige Differenz. Die Permeabilität μ ist in die absolute Permeabilität μ_0 mit der Einheit H/m und eine weiter zu definierende unbenannte Zahl μ_r aufzuspalten.

$$\mu = \mu_0 \cdot \mu_r \tag{21.10}$$

mit

$$\mu_0 = 4\pi \cdot 10^{-7}\,\mathrm{V\,s/A\,m} = 1{,}256 \cdot 10^{-6}\,\mathrm{H/m}$$

Anfangspermeabilität = Steigung im Nullpunkt. Die *Anfangspermeabilität* μ_A, μ_{1a} oder μ_i (engl.: i = initial) entspricht der Steigung der Neukurve im Ursprung (Abb. 21.6a). Die formale Definition lautet:

$$\mu_A = \mu_i = \frac{1}{\mu_0} \cdot \frac{\hat{B}}{\hat{H}}\bigg|_{\hat{H}\to 0} \tag{21.11}$$

Da mit $\hat{H} = 0$ keine Messungen mehr durchführbar sind, ist hierfür ein sehr kleiner Wert (einige mA/cm) festgelegt. Ältere Angaben benennen p_5 d. h. $\hat{H} = 5\,\mathrm{mOe} = 4\,\mathrm{mA/cm}$. In Anlehnung daran findet man oft auch als Angabe μ_4.

Wechselfeldpermeabilität = Steigung bei großer Aussteuerung. Die Wechselfeldpermeabilität $\mu_\sim$ entspricht der üblichen Permeabilitätszahl μ_r mit der Definition (Abb. 21.6a):

$$\mu_\sim = \frac{1}{\mu_0} \cdot \frac{\hat{B}}{\hat{H}}\bigg|_{\hat{H}\ \text{groß}} \tag{21.12}$$

$\hat{H}$ wird zweckmäßig von $\hat{H} = 0$ bis zu einem maximalen Wert, bei dem B_s erreicht wird, geführt und das Ergebnis als Diagramm über H aufgetragen (Abb. 21.6b). Für $H = 0$ geht $\mu_\sim$ dann in μ_A über. Ist zusätzlich zur Wechselfeldaussteuerung noch ein magnetisierendes Gleichfeld $H_=$ vorhanden, so ergibt sich die *Überlagerungspermeabilität* (s. a. μ_{rev}).

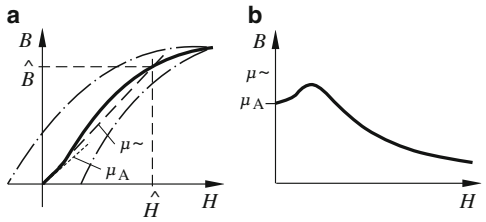


Abb. 21.6 Wechselfeldpermeabilität. **a** Definition der Anfangs- (m_A) und Wechselfeldpermeabilität ($m_\sim$) an der Kommutierungskurve, **b** Wechselfeldpermeabilität $m_\sim$ in Abhängigkeit von der magnetischen Feldstärke

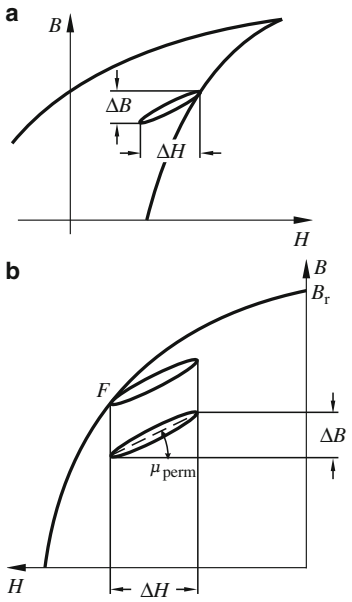


Abb. 21.7 Permeabilitäten innerhalb der Hysteresekurve. **a** Reversible Permeabilität, **b** permanente Permeabilität

Reversible Permeabilität = Steigung innerhalb der Hystereseschleife. Wird ein magnetischer Werkstoff mit einem Gleichfeld $H_=$ vormagnetisiert und mit kleinen Wechselfeldstärken ausgesteuert, so ergeben sich die in Abb. 21.7a dargestellten Verhältnisse. Einer inneren Magnetisierungskennlinie folgend, verläuft die Aussteuerung flacher gemäß:

$$\mu_{\mathrm{rev}} = \frac{1}{\mu_0} \cdot \frac{\Delta B}{\Delta H}\bigg|_{\Delta H\to 0, H_= \neq 0} \tag{21.13}$$

ΔB und ΔH liegen dabei immer innerhalb der Hysterese Schleife. μ_{rev} sollte meist groß sein, damit trotz Gleichstromvormagnetisierung der magnetische Kreis noch einen genügend großen induktiven Widerstand aufweist.

Ein Sonderfall der reversiblen Permeabilität ist die *permanente Permeabilität* μ_{perm}, die in der Dauermagnetik wichtig ist. Der Arbeitspunkt liegt im 2. (oder 4.) Quadranten (Abb. 21.7b). Da bei Dauermagneten die Flussdichte – und deren Konstanz – die entscheidende Größe ist, sollte $\mu_{\mathrm{perm}} \sim 1$ sein, um bei einem störenden Wechselfeld keine Änderungen des magnetischen Arbeitspunktes zu verursachen.

Die Frequenzabhängigkeit der Permeabilität wird durch die *komplexe Permeabilität* $\bar{\mu}$ beschrieben. Sie ist definiert als

$$\bar{\mu} = \mu' - j\mu'' \tag{21.14}$$

Mit μ' = Permeabilitäts-Permeabilität und μ'' = Verlust-Permeabilität. Der Betrag von $\bar{\mu}$ liefert die rechnerisch gebräuchliche Permeabilität μ_r oder, um auf die Verwendung für die Bestimmung der induktiven Impedanz Z hinzudeuten, μ_{rZ}.

Der Quotient aus μ'' und μ' liefert den Verlustfaktor

$$\tan\delta = \frac{\mu''}{\mu'} \tag{21.15}$$

Die Division durch die Anfangspermeabilität μ_i oder μ_4 liefert den *relativen* oder *bezogenen Verlustwinkel*

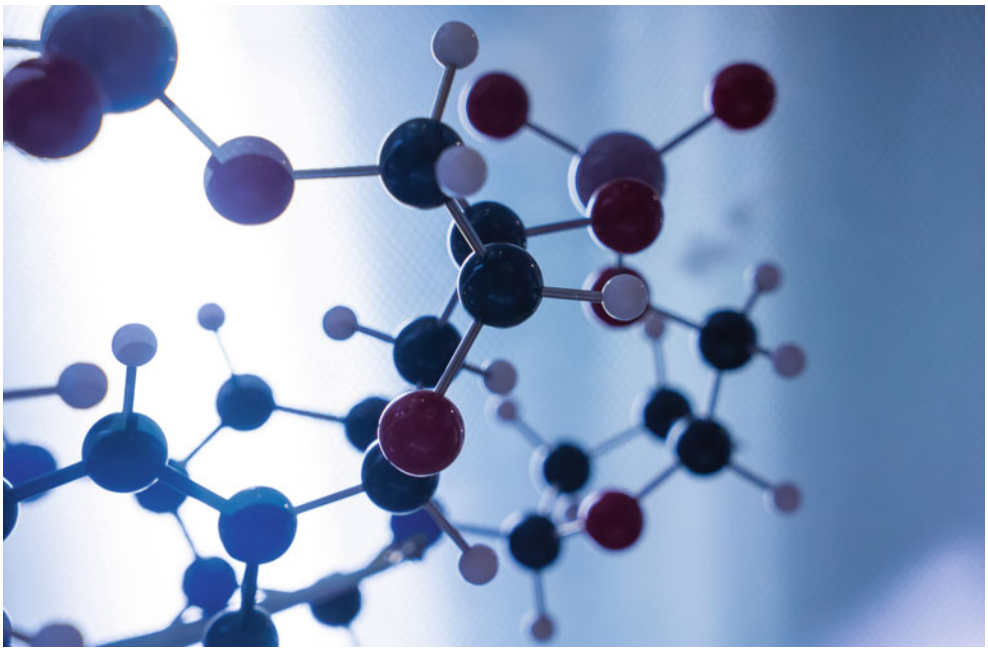
$$\frac{\tan \delta}{\mu_i} \tag{21.16}$$

Weniger bedeutende Eigenschaften von Werkstoffen im Magnetfeld sind einmal die *Magnetostriktion*, das ist die Änderung der Abmessungen durch ein Magnetfeld, wobei zwischen *Längen-* und *Volumenmagnetostriktion* unterschieden wird, und *multiferroisch* was einer gleichzeitigen ferromagnetischen und ferroelektrischen Eigenschaft des Werkstoffes entspricht. Das Interessante ist dabei die mögliche wechselweise Steuerung der Effekte „Spannung“ und „Magnetisierung“. Materialen mit diesem Effekt werden unter dem Namen *Multiferroika* geführt.

Magnetika

22

Egon Döring



22.1 Metalloxide (Ferrite) 286
22.2 Weichmagnetika 286
22.3 Hartmagnetika (Dauermagnete) 290

Teil III

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_22

Tab. 22.1 Übersicht über Magnetika (Ohne Anspruch auf Vollständigkeit, Metalloxide ohne Fe_2O_3)

Metalle				Metalloxide (Ferrite)	
reine Metalle		Legierungen			
weich	hart	weich	hart	weich	hart
Fe		Fe–Si	Fe–Al–Ni	Mn–Zn–O	Ba–O
Ni		Fe–Ni–Fe	Al–Ni–Co	Ni–Zn–O	Sr–O
Co		Fe–Co	Fe–Ni–Cu	Mg–Mn–O	
(Gd)		Fe–Al	Fe–Co–V	Y–O (Granat)	
		Cu–Mn–Al	Co–Sm		
			Nd–Fe–B		

Bei den Magnetika werden vom Werkstoff her *Metalle* und deren Legierungen sowie *Metalloxide (Ferrite)* und, nach den Anwendungseigenschaften, *Weich-* und *Hartmagnetika* (Dauermagnete) unterschieden. In Tab. 22.1 ist eine grundsätzliche, aber nicht vollständige, Übersicht der magnetischen Werkstoffe zusammengestellt.

22.1 Metalloxide (Ferrite)

Ferrite der Elektrotechnik sind chemische Verbindungen des Eisen(III)-oxids Fe_2O_3 mit zweiwertigen Metalloxiden (M″O) vom allgemeinen Typus

$$M''O \cdot Fe_2O_3$$

Die Gitterstruktur der am meisten verbreiteten kubischen Ferrite entspricht der des Halbedelsteins *Spinell*. In der normalen kubischen Spinellstruktur mit ihrem flächenzentrierten Sauerstoffionen-Gitter (Abb. 22.1) befinden sich die zweiwertigen Metallionen (Punkte in Abb. 22.1) auf *Tetraederplätzen* und die dreiwertigen (Dreiecke in Abb. 22.1) auf *Oktaederplätzen*. Vertauschen die zwei- und dreiwertigen Ionen ihre Plätze, entstehen die *inversen Spinelltypen*. Nur diese und die aus beiden bestehenden Mischtypen liefern magnetische Ferrite. Sie sind vom Prinzip her ferrimagnetisch. Als Metalloxide sind die Ferrite keramische Werkstoffe und daher *elektrische Nichtleiter*.

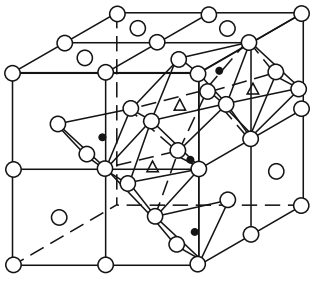


Abb. 22.1 Kristallgitter eines Spinells. ○ Sauerstoffion, ● zweiwertiges Metallion und △ dreiwertiges Metallion (zur besseren Übersicht sind die Gitterplätze nur teilweise besetzt)

22.2 Weichmagnetika

Weichmagnetische Werkstoffe sind durch eine Koerzitivfeldstärke von $H_c < 1\,\mathrm{kA/m}$ gekennzeichnet.

Unter den Elementen kommen nur weichmagnetische Stoffe vor, und zwar die Ferromagnetika Eisen, Nickel, Kobalt sowie das für die Praxis bedeutungslose Gadolinium (Curietemperatur 16 °C).

Metalle: Eisen wird als *Weicheisen* für Relais und als *Karbonyl-Eisen(pulver)* in geringem Umfang für Hochfrequenzkerne und zur Abbildung von Magnetfeldern (Bitter'sche Methode) benutzt. Nickel findet rein gelegentlich (im Sättigungsbereich) Anwendung als magnetostriktiver Ultraschallgeber, ansonsten dient es, ähnlich wie Kobalt, als Legierungsbestandteil für verschiedene magnetische Werkstoffe. Von den übrigen Metallen gilt in erster Näherung, dass die Schwermetalle (z. B. Cu, Ag, Bi, Au usw.) diamagnetisch und die Leichtmetalle (z. B. Al, Mg, Na, Ca usw.) paramagnetisch sind.

Legierungen: Eisen ist mit verschiedenen Legierungszusätzen wie Si, Ni, Co, Mo u. a. der wichtigste magnetische Werkstoff. Neben den Zusätzen haben auch die Art der Behandlung z. B. warm- oder kaltgewalzt, nachgeglüht usw. erheblichen Einfluss auf die Verluste ($p_{1,0}$, $p_{1,5}$ bzw. $p_{1,7}$) eines Elektrobleches. Höherwertige Bleche sind durchweg kornorientiert.

Für die Unterscheidung von Elektroblechen ist gelegentlich noch die veraltete Bezeichnung *Dynamoblech* mit den Abstufungen *Dyn.-Bl. I*… *IV* zu finden. Die heutigen enger tolerierten Anforderungen bedienen sich gemäß DIN 46400 bzw. EN 10027-1:2005 einer Kombination aus Buchstaben und Zahlen nach folgendem Schema: Kennbuchstabe, Verluste, Blechdicke und Behandlung wie in Tab. 22.2 dargestellt.

In der Tab. 22.3 sind für einige nach dem genannten Schema (EN 10027) bezeichnete Werkstoffe, neben der Werkstoffnummer, die Grunddaten zusammengestellt. Bevorzugtes Anwendungsgebiet sind Transformatoren, Motoren und Generatoren.

Werkstoffe für Übertrager behandelt DIN IEC 740-2. Der *Kurzname* besteht nach DIN aus einem Buchstaben und Ziffern,

Tab. 22.2 Bezeichnung von Elektroband

M	Europaweit Kennbuchstabe für Elektrobleche
mmm	100facher höchstzulässiger Magnetisierungsverlust
dd	100fache Blechdicke
Behandlung	A nicht kornorientiert N kornorientiert P kornorientiert, niedrige Ummagnetisierungsverluste D unlegiert, nicht schlussgeglüht E legiert, nicht schlussgeglüht

wobei, mit im Alphabet fortschreitenden Buchstaben, die Legierungsanteile wachsen. A bedeutet gar nicht oder schwach Si legiert, C definiert mit Si legiert, D sind Ni-Legierungen, E und F sind hoch nickelhaltig und noch weiter legiert. Die Tab. 22.4 gibt eine Übersicht über einige Werkstoffe dieser Klasse. IEC benutzt 3 Buchstaben, beginnend mit Y und Ziffern. Zusätzlich zu den DIN-Bezeichnungen haben alle Magnetwerkstoffe noch Eigennamen der Hersteller, z. B. Vacofer für *Werkstoffe der* Klasse A, Mumetall aus Klasse E oder Vacuflux bei Klasse F von der Firma *Vacuumschmelze GmbH*.

Nickel ist, wie Tab. 22.4 zeigt, von größtem Einfluss auf die magnetischen Eigenschaften von Eisen. Während ca. 75 % Ni einen hochpermeablen Werkstoff mit $\mu_A > 10^4$ liefert, sinkt bei ca. 25 … 30 % Ni die Curie-Temperatur unter die Raumtemperatur, so dass *unmagnetische Stähle* entstehen. Für die Technologie ist die gute Auswalzbarkeit von Fe-Ni-Legierungen bis in den Bereich von $d \simeq 50\,\mu\text{m}$ wichtig. Für extrem dünne Schichten im µm-Bereich der Halbleitertechnologie sind hochpermeable Legierungen mit 78 % Ni aufdampfbar.

Amorphe Metalle: Unter dieser Bezeichnung oder auch unter dem Namen *Metallgläser* werden weichmagnetische Werkstoffe mit extremen Eigenschaften geliefert. Es handelt sich dabei um Legierungen auf Fe-Ni- bzw. Co-Ni-Fe-Basis, die nicht kristallin sind! Diese Eigenschaft wird durch eine außerordentlich hohe Abkühlgeschwindigkeit von ca. $10^6\,\text{K/s}$ d. h. $1000\,\text{K/ms}$, beim Erstarren aus der Schmelze mit bis zu 25 % kristallisationsverzögernde Zusätze erreicht. So ist z. B. die Co-Ni-Fe-Legierung der Fa. Vacuumschmelze GmbH als $O_{50}Ni_{20}Fe_6Si_{12}B_{12}$ aufgebaut. Die bestechendste Eigenschaft derartiger, in Folienstärke um 20 µm anfallender, Werkstoffe ist die hohe Permeabilität. Im Gegensatz zu anderen Weichmagnetika weisen diese Werkstoffe aber erhebliche Härte- ($\sim 1000\,\text{HV02}$) und Zugfestigkeitswerte ($R_m \sim 2\,\text{kN/mm}^2$) auf. Die thermische Belastbarkeit ist nicht durch die Curie-Temperatur, sondern durch die Kristallisationstemperatur von ca. 400 … 500 °C begrenzt. Die bei anderen hochpermeablen Materialien gefürchtete Empfindlichkeit gegen mechanische Beanspruchungen tritt bei amorphen Metallen nicht auf.

Wird die Kristallisationsgeschwindigkeit in den Bereich von etwa 10 K/s „verlangsamt“ oder ein amorphes Material geeignet wärmebehandelt entsteht *nanokristallines* Material mit Kristallen in der Größenordnung von etwa 10 nm. Die Materialstärke beträgt unverändert etwa 20 µm. Der hochpermeable Zustand ergibt sich erst nach einer Wärmebehandlung die ein starkes Verspröden des Materials bewirkt. Eine mechanische Behandlung ist danach effektiv nicht mehr möglich. Je nach Wärmebehandlung und mit oder ohne Magnetfeldanwendung entstehen unterschiedliche Hysteresekurven mit rechteckigem (Z), rundem (R) oder flachem (F) Verlauf (Abb. 22.2a). Die daraus bei der Magnetisierung resultierenden unterschiedlichen Wechselfeldpermeabilitäten zeigt Abb. 22.2b.

Kernbleche: Während für Großtransformatoren die Kernbleche aus Plattenmaterial dem jeweiligen Bedarf entsprechend zu-

Tab. 22.3 Elektroband, ausgewählte Beispiele nach Thyssen Krupp AG, Vacuumschmelze, Voestalpine

Kurzname	Werkstoff-nummer	Firmen-bezeichnung	d	Verluste p_{max}			Polarisation J/T bei			
				$p_{1,0}$	$p_{1,5}$	$p_{1,7}$	H25	H50	H80	H100
			mm	W/kg	W/kg	W/kg	A/cm	A/cm	A/cm	A/cm
Nicht kornorientiert, nicht schlussgeglüht										
M1300-100A	1.0897		1,00	5,80	13,0		1,62	1,70		1,79
M450-65K	1.0843		0,65	1,92	4,50		1,56	1,64		1,74
M340-50K	1.0841		0,50	1,42	3,4		1,54	1,62		1,72
Nicht kornorientiert, schlussgeglüht										
M530-65A	1.0824		0,65	2,30	5,30		1,54	1,64		1,74
M350-50A	1.0810		0,50	1,50	3,50		1,50	1,60		1,70
M250-35A	1.0800		0,35	0,95	2,50		1,49	1,60		1,70
Kornorientiert, geglüht										
M150-35S	1.0861	PowerCore C150-35	0,35		1,05	1,50			1,78	
M140-27S	1,0865	PowerCore C140-27	0,27		0,89	1,40			1,75	
M110-23S	1.0863	PowerCore C110-23	0,23		0,73	1,10			1,78	
M105-30P	1.0886	PowerCore H105-30	0,30			1,05			1,88	
M95-27-P	1.0839	PowerCore H095-27	0,27			0,95			1,88	
M85-23-P	1.0882	PowerCore H085-23	0,23			0,85			1,88	

Tab. 22.4 Hochpermeable Legierungen. Typische Beispiele aus Unterlagen von Thyssen Krupp AG und VAC Vacuumschmelze

Kurzname	Firmenname	Chemischer Aufbau	d	Verluste $p_{1,0}$	H_c	B_s	ρ	T_c	μ_4	μ_{max}
			mm	W/kg	A/m	T	µΩ m	°C		
A12	Vacofer S1	Fe (Reineisen)	0,35	1,0	< 12	2,03		770	2000	40.000
C1-24	Trafoperm N3	$Si_{3\,\%}Fe$	0,35	0,85	< 24	2,03	0,4	750	1000	30.000
D1a	Magnifer 36	$Ni_{36\,\%}Fe$	0,35	1,1	< 15	1,3	0,75	260	2300	20.000
E4	Magnifer 75	$FeNi_{36\,\%}Cu_{5\,\%}$	0,2	1,1	< 2	0,8	0,55	400	40.000	115.000
E11	Magnifer 7904	$FeNi_{80\,\%}Mo_{5\,\%}$	0,2		< 1	0,8	0,45	410	200.000	350.000
E11-60	Mumetall	$FeNi_{\simeq 80\,\%}$	0,2	0,1	0,04	0,8	0,55	400	60.000	150.000
F11	Vacoflux 50	$FeCo_{\simeq 50\,\%}$	0,35	1,0	< 0,8	2,1	0,44	950	1000	13.000
Amorph	Vitrovac 6030	$FeCo_{80\,\%}$	0,02	$40_{100\,kHz}$	0,03	0,8	1,30	365	20.000	100.000
Nanokristallin	Vitroperm	$Fe_{73\,\%}Si_{15\,\%}B_{7\,\%}Nb_{3\,\%}$	0,02	$35_{100\,kHz}$	< 3	1,2	1,1	600	10.000	150.000

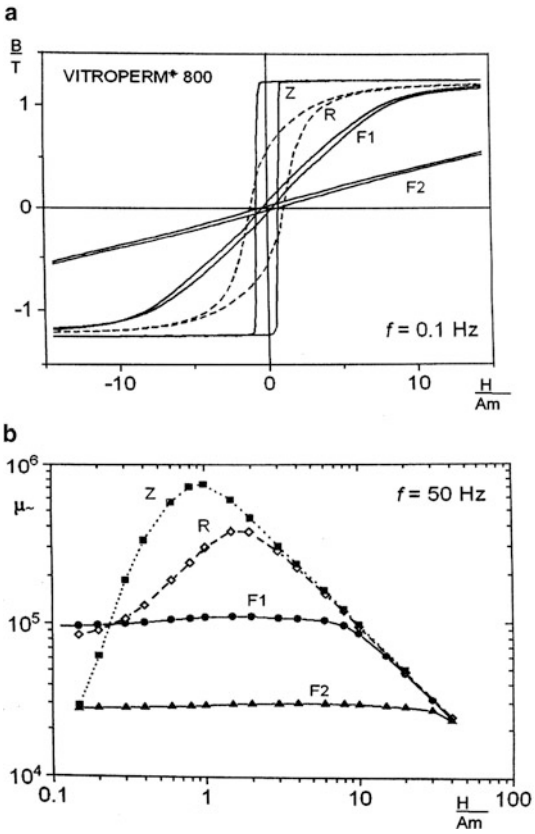


Abb. 22.2 Nanokristallines Material. **a** Hysteresekurven unterschiedlich wärmebehandelten nanokristallinen Materials, **b** Wechselfeldpermeabilität der Materialien nach **a** (Aus Unterlagen der Vacuumschmelze nach Elsevier „Handbook of Magnetic Materials" bzw. „Soft Magnetic Alloys")

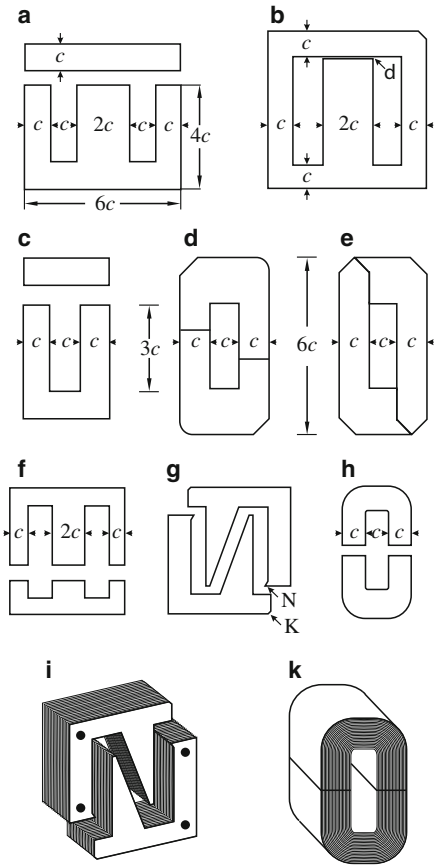


Abb. 22.3 Blechschnitte für Kleintransformatoren. **a** YEI EI-Kernblech (abfallarm), **b**–**e** E-Kernblechvariationen, **b** YEE2, EE-Bleche unterschiedlicher Länge, besonders auch kürzere Formen, **c** YEE2L EE-Kernbleche doppelter Länge, **d** YEF F-Form der Schenkel des E-Bleches, **e** YEL L-Form der Schenkel eines E-Bleches, **f** YM M-Kernblech oder Mantelkernblech mit oder (fast) ohne Luftspalt δ, **g** YUI UI-abfallarme UI-Kernbleche, **h** Philbertkernblech PU, **i** Philbertkernblech PL, **j**, **l** Steckkern(blechpaket), **k**, **m** Schnittbandkern

geschnitten werden, ist für Kleintransformatoren die Anwendung vorgefertigter Kerne bzw. Kernbleche zweckmäßiger (EN 60740-1:2005). Material- und fertigungsmäßig sind die abfalllosen *YEI*- (Abb. 22.3a) bzw. *YUI-Schnitte* (Abb. 22.3g) besonders günstig. In Varianten zu diesen Schnitten werden die *Is″* durch ein mehr oder weniger vollständiges *E* (Abb. 22.3b) bzw. *U* ersetzt. Es entstehen dann EE- bzw. UU-Formen, bei denen die Luftspalte der Kerne innerhalb der Wicklungen liegen. In Abb. 22.3d und e (nach EN 60740-1:2005) weitere Variationen zu EE-Kernblechen dargestellt. Die Kennzeichnung mit YEF bzw. YEL lehnt an die modifizierte Form des „zweiten" Es an. Eine spezielle Variante der E-Kernform ist der Drehstrom-E-Kern, mit drei gleich breiten Stegen. Eine verarbeitungsmäßig aufwendige, für geringe Luftspalte aber beste Kernblechform ist der *YM-Schnitt* (Abb. 22.3f). Der Luftspalt δ kann von „klemmender Zunge", d. h. verschwindendem Luftspalt, bis zu einer Breite von mehre-

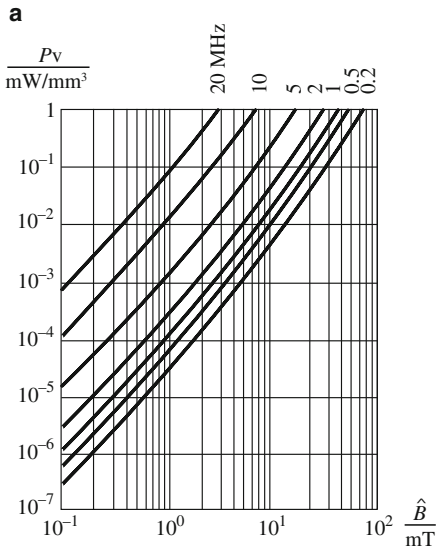


Abb. 22.4 Darstellung der Kernverlustleistung p_v bei Ferriten. **a** Verlustleistung in mW/cm³ in Abhängigkeit von der Spitzenflussdichte mit der Frequenz als Parameter (Ferroxcube 2P ... von Philips), **b** Verlustleistung in mW/cm³ in Abhängigkeit von der Frequenz mit der Spitzenflussdichte als Parameter (Ferritmaterial B1 von Neosid)

ren Millimetern variiert werden. Um die Lage der Bleche auch im fertigen Transformator kontrollieren zu können, sind die Bleche an einer Ecke (oben rechts in Abb. 22.3f) abgeschrägt. *Philbertschnitte* in PU- (Abb. 22.3h) oder in PL-Form (Abb. 22.3i), mit den verbreiterten Querstegen in den magnetisch schweren Richtungen, nutzen bei *Gosstexturen* (leichte Richtungen längs, schwere Richtungen quer und in Dickenrichtung der Bleche) das Material besser aus. Die beste Ausnutzung der Gosstextur liefert der herstellungs- und verarbeitungsmäßig aufwändige *Schnittbandkern* (Abb. 22.3k und m). Der für das Einbringen in die Wicklungen geteilte Kern ist an den Trennflächen sorgfältig geschliffen und muss in der Fertigung immer wieder mit den einander zugehörigen Flächen zusammengebracht werden.

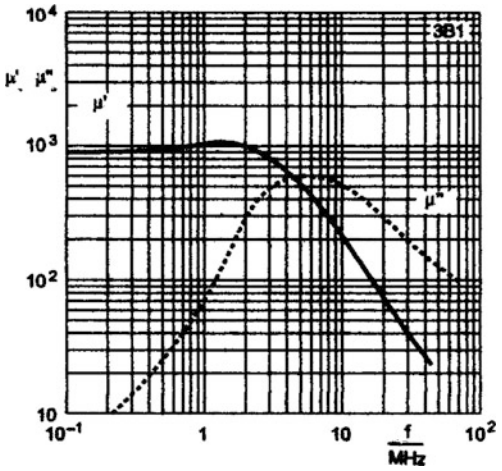


Abb. 22.5 Komplexe Permeabilität eines MnZn-Ferrites (Ferroxcube 3B1 von Philips)

Für rationelle Fertigungen sind fertig geklebte Blechpakete in EE-Form oder als *Steckkerne*, die über Kerben (K) und Vorsprünge (N) einrastend zusammenhalten (Abb. 22.3j und l), trotz magnetisch ungünstigerer Eigenschaften (Luftspalt) vorteilhaft.

Um die durch dünne Bleche niedrig gehaltenen Wirbelströme auch in einem Blechpaket klein zu halten, sind die einzelnen Bleche elektrisch voneinander zu isolieren. Diese Maßnahme hat einen Verlust an Eisenquerschnitt zur Folge, der durch den *Eisenfüllfaktor* gekennzeichnet wird. Er gibt den relativen Anteil des Eisens an der gesamten magnetischen Querschnittsfläche an. Mit sinkender Blechdicke sinkt auch der Füllfaktor, da die Isolation nicht beliebig dünn gemacht werden kann. Für die üblichen Blechdicken von 0,25 ... 0,5 mm liegt der Eisenfüllfaktor bei 95 %. Er kann bei sehr dünnen Blechen bis etwa 35 % absinken.

Weichmagnetische Ferrite: Wegen der Nichtleitereigenschaften der Ferrite, und den damit fehlenden Wirbelstromverlusten, sind sie besonders für hohe Frequenzen und die Nachrichtentechnik interessant. Die praktisch genutzten Werkstoffe sind alle sog. *Mischferrite*. Dementsprechend auch die Bezeichnungen wie Mn–Zn-Ferrit o.ä. Für Ferrite typisch sind die geringen Sättigungsflussdichten von ca. 0,3 ... 0,5 T. Als keramische Werkstoffe sind sie mit $\rho = 3 \ldots 5\,\mathrm{g/cm^3}$ von geringer Dichte, absolut *korrosionsbeständig* und *schlechte Wärmeleiter*. Die bei hohen Aussteuerungen auftretende Verlustwärme ist demgemäß schwierig abzuführen. Wegen der Frequenzabhängigkeit werden die Verluste in Diagrammen über der Aussteuerung (Abb. 22.4a) mit der Frequenz als Parameter oder über der Frequenz (Abb. 22.4b) mit der Aussteuerung als Parameter dargestellt. Zu beachten ist noch die bei tiefen Frequenzen oft sehr hohe *Permittivitätszahl* der Ferrite, die bei hohen Frequenzen stark absinkt.

Wegen des möglichen Einsatzes von Ferriten auch bei höheren Frequenzen ist die komplexe Permeabilität bedeutsam. Abb. 22.5 zeigt hierzu das Beispiel eines Mn-Zn-Ferrites, wobei deutlich wird, dass die Verlustpermeabilität zwischen 4 und 5 MHz die Permeabilitätspermeabilität übersteigt und der

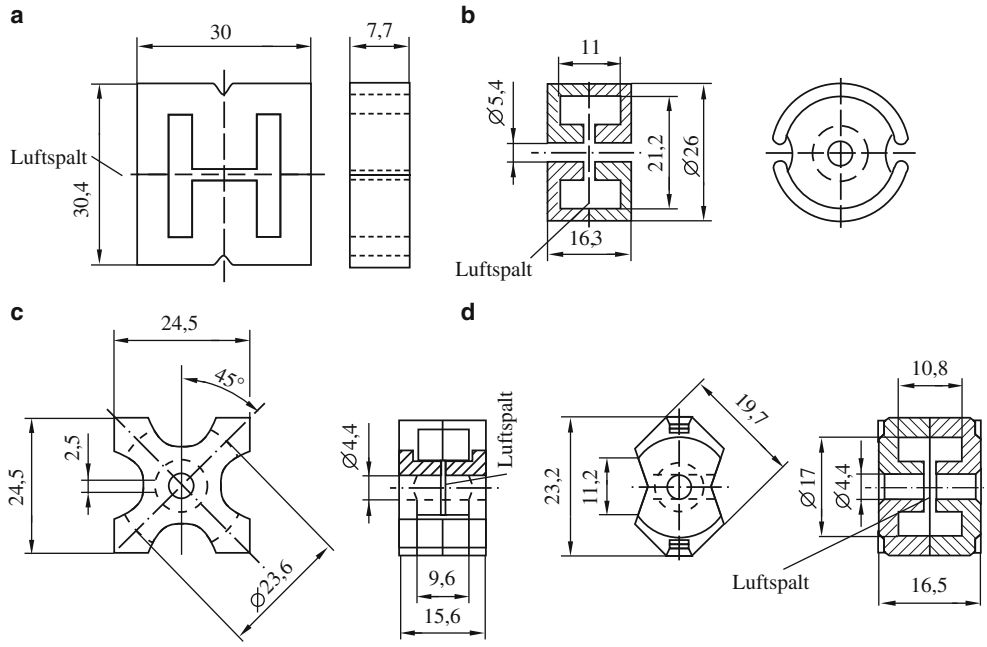


Abb. 22.6 Einige Ferritkerne mit ihren Hauptabmessungen. **a** E-Kern: E30, **b** EER Kern EER28 (E-Kern für Rundspulen), **c** Schalenkern S26 × 16, **d** X-Kern: X26, **e** RM-Kern: RM8

Werkstoff mehr Verluste, d. h. Dämpfung verursacht als Induktivitätsanstieg.

Mn-Zn-Ferrite sind hochpermeabel mit μ_i-Werten von 1000 bis gegen 10.000. Ihre elektrische Leitfähigkeit ist mit $\gamma = 1\,\mathrm{S/m}$ für Ferrite hoch, ebenso die Sättigungsflussdichte mit 0,5 T. Sie sind aus diesen Gründen in geringem Maße auch für Leistungsübertrager geeignet.

Ni-Zn-Ferrite sind die verbreitetsten Ferrite überhaupt. Die Anfangspermeabilität liegt bei $\mu_i \simeq 100 \ldots 1000$. Der nutzbare Frequenzbereich reicht bis oberhalb 100 MHz. Wegen der geringen elektrischen Leitfähigkeit von $\gamma \simeq 10^{-5}\,\mathrm{S/m}$ kann in vielen Fällen die Wicklung direkt ohne Zwischenisolation auf den Ferritkörper aufgebracht werden.

Ba-Co-Ferrite können noch, allerdings mit erheblichen Verlusten, bis zu sehr hohen Frequenzen genutzt werden. Sie finden daher bevorzugt für Entstördrosseln Verwendung.

Kernformen: Die Bauformen der ferritischen E-Kerne (Abb. 22.6a) schließen an die entsprechenden Blechschnitte an und sind so ausgelegt, dass sie in die Wickelkörper der Blechkerne passen. Für höhere Frequenzen kommen dann die Spulenkörper in Form von *Schalen-*, X- und *RM-Kernen* in Frage (Abb. 22.6c, d und e). X- und RM-Kerne sind Bauformen, die dem geringen Platzangebot in gedruckten Schaltungen besser entsprechen als die Rundformen. EER-Kerne ermöglichen eine Rundspule für die Wicklung auf einem quasi EE Kern. Darüber hinaus gibt es noch eine große Anzahl verschiedener Kernformen wie Stäbe, Schraubkerne, Ein- und Mehrlochkerne usw.

22.3 Hartmagnetika (Dauermagnete)

Hartmagnetisch ist ein Werkstoff mit $H_c \gtrsim 1\,\mathrm{kA/m}$ (10 A/cm). Da die in einem Dauermagneten gespeicherte Energie über die Fläche der Hystereseschleife zum Ausdruck kommt, sollte diese und damit B_r und H_c groß sein. Die Energiedichte, das Produkt aus korrespondierenden B- und H-Werten, weist ein Maximum, das *Energieprodukt* $(B \cdot H)_{\max}$ auf (Abb. 22.7b). Es ist ein für einen Magnetwerkstoff charakteristischer Wert.

Entmagnetisierungskurve (Abb. 22.7a): Sie ist mit dem reversiblen Bereich I, dem Knickbereich II und dem irreversiblen Bereich III in unterschiedliche Nutzungsbereiche aufgeteilt. Im reversiblen Bereich, mit seinem linearen Teil, erfolgen Ar-

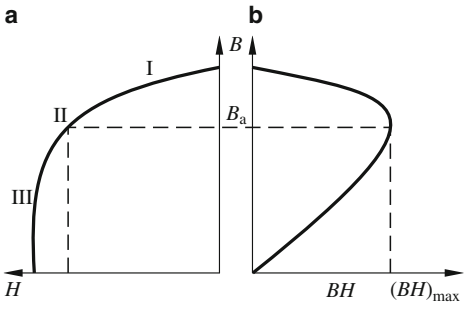


Abb. 22.7 Entmagnetisierungskurve und Energiedichte. **a** Entmagnetisierungskurve im 2. Quadranten, I reversibler Bereich, II Knickbereich, III irreversibler Bereich, B_a und H_a Flussdichte bzw. Feldstärke im Arbeitspunkt, **b** Energiedichte $B \cdot H$ in Abhängigkeit von der Flussdichte, mit höchster Energiedichte $(B \cdot H)_{\max}$

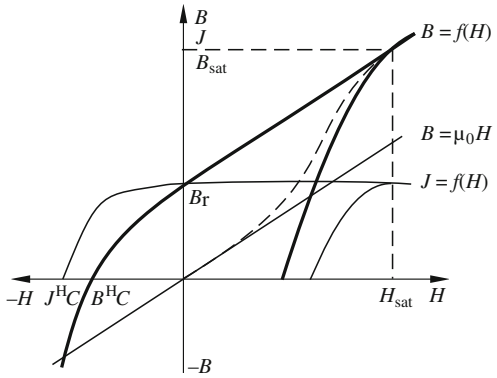


Abb. 22.8 Zusammenhang zwischen Flussdichte *B* und Polarisation *J* (Erläuterungen im Text)

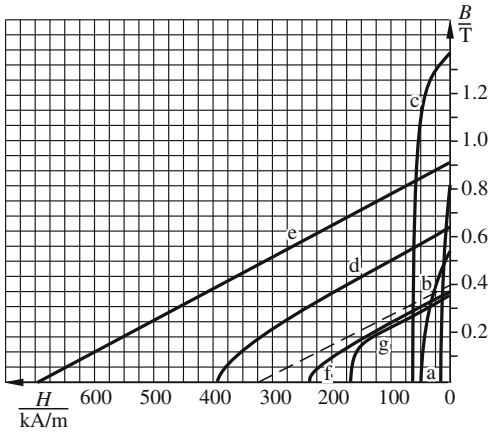


Abb. 22.9 Entmagnetisierungskurven einiger Hartmagnetika. **a** AlNi120, **b** AlNiCo700

beitspunktänderungen – z. B. durch äußere Felder – auf dieser Kennlinie. Im Knickbereich und auch im irreversiblen Bereich verlagern äußere Felder den Arbeitspunkt auf innere Kennlinien der Entmagnetisierungskurve. Diese ist bei Dauermagnetwerkstoffen häufig als Polarisationskurve, d. h. $J = f(H)$ dargestellt. Abb. 22.8 zeigt die Zusammenhänge hierzu, indem von der bekannten $B = f(H)$-Kurve (dick ausgezogen) die Vakuumflussdichte, $\mu_0 H$ abgezogen ist. Während der Remanenzpunkt für beide Kurven gleich bleibt, ist die Koerzitivfeldstärke unterschiedlich. Bei der Koerzitivfeldstärke ${}_BH_c$ (oder H_{cB}) wird die Flussdichte *B*, d. h. die Flussdichte des Systems auf Null gebracht. Bei der Koerzitivfeldstärke ${}_JH_c$ (oder H_{cJ}) wird die dem Werkstoff zugehörige Polarisation auf Null gebracht. Aus diesem Grunde ist für die Werkstoffcharakterisierung bei Dauermagneten die Kurve $J = f(H)$ wichtig. Im Idealfall würde ${}_JH_c \to \infty$ und die Sättigungspolarisation J_s gleich der Remanenzflussdichte B_r werden, die einem äußeren Feld $\mu_0 H$ entgegenwirkt. Die Flussdichte *B*des (idealen) Gesamtsystems wird dann

$$B = B_r - \mu_0 H \tag{22.1}$$

Damit fällt die „ideale Entmagnetisierungskennlinie“ linear von B_r bis ${}_BH_c = B_r/\mu_0 H$ mit der Steigung der idealen *permanenten Permeabilität* $\mu_{perm} = 1$ ab.

Die Entmagnetisierungskurve als $B = f(H)$ stellt die im Allgemeinen nichtlineare *Innenleitwertkennlinie* des Werkstoffes dar. Der jeweilige Innenleitwert Λ_i folgt dann zu

$$\Lambda_i = \mu_0 \cdot \mu_p \tag{22.2}$$

μ_p = Permeabilität des Magnetwerkstoffes im Arbeitspunkt. Für ein Dauermagnetfeld in Luft gilt aber $\mu_r = 1$, d. h. bei Anpassung eines Dauermagnetwerkstoffes an Luft muss ein Arbeitspunkt gefunden werden, bei dem $\mu_p = \mu_r = 1$ ist! Ein guter Dauermagnetwerkstoff muss also eine Entmagnetisierungskennlinie aufweisen, die möglichst weit oberhalb und gut parallel zur Geraden $B = \mu_0 H$ (in Abb. 22.9 gestrichelt und zum besseren Vergleich aus dem Nullpunkt verschoben) verläuft.

FeAlNi- und *FeAlNiCo-Legierungen* (Abb. 22.9a, b) mit Energieprodukten von $10 \ldots 30\,\mathrm{kJ/m^3}$ sind auch heute noch die wichtigsten Magnetwerkstoffe. Es handelt sich hierbei um Ausscheidungslegierungen, die erst durch eine Wärmebehandlung (Tempern), gegebenenfalls unter Einwirkung eines Magnetfeldes (Formanisotropie), die gewünschten Eigenschaften erhalten. Diese Legierungen sind sehr hart und nur durch Schleifen zu bearbeiten.

Selten-Erd-Magnete oder SECo-Magnete (Abb. 22.9d), mit Energieprodukten von $200\,\mathrm{kJ/m^3}$ und mehr, ist eine Kurzbezeichnung für Magnetwerkstoffe auf der Basis Seltene Erdmetalle-Kobalt. Wichtigster Vertreter der Seltenen Erden ist dabei Samarium (Sm). Die permanente Permeabilität erreicht bei diesen Materialien sehr genau den Wert 1.

Aus Seltenerd-Magneten entwickelten sich die Neodym-Eisen Systeme, die als NdFeB noch höhere Energiedichten erreichen und die zurzeit „besten Magnete“ liefern. Die theoretische Grenze dieser Systeme liegt bei $485\,\mathrm{kJ/m^3}$. Erreicht sind zurzeit etwa $400\,\mathrm{kJ/m^3}$ (Abb. 22.9f).

Hartmagnetische Metalloxide (Ferrite): Die hartmagnetischen Ferrite haben, genau wie die weichmagnetischen, eine geringe Dichte, hohe Korrosionsbeständigkeit, geringe Remanenzflussdichte und niedrige Curie-Temperatur. Abb. 22.9c zeigt eine Entmagnetisierungskurve von Ba-Ferrit. Neben der geringen Remanenzflussdichte fallen die hohe Koerzitivfeldstärke und die Steigung mit $\mu_{perm} \simeq 1$ auf. Als keramische Werkstoffe sind Ferritmagnete nicht nur mechanisch hart, sondern auch bruchempfindlich. Der spezifische Widerstand aller Ferritmagnete ist mit $r > 10\,\Omega\,\mathrm{m}$ sehr hoch, wodurch sie selbst in hochfrequenten Feldern eingesetzt werden können.

Ba-Ferrit: Hier kommt eigentlich nur der hexagonale Ba-Ferrit vom Typus $BaO \cdot 6Fe_2O_3$ mit einem Energieprodukt von $25\,\mathrm{kJ/m^3}$ für die technische Anwendung in Betracht. Er liegt sowohl isotrop als auch anisotrop, meist noch mit geringen Zusätzen von Sr, Pb und/oder Co vor. In der Tab. 22.5 sind die Daten einiger typischer Dauermagnetwerkstoffe zusammengestellt.

Isotrop bedeutet bei magnetischen Werkstoffen, dass sie in allen Richtungen etwa gleiche magnetische Eigenschaften haben. An-

Tab. 22.5 Daten einiger typischer Dauermagnetwerkstoffe, nach IBS Magnet, Meder electronic, ThyssenKrupp und VAC Vacuumschmelze, (gem. DIN IEC 60404-8-1, bzw. DIN 17410)

Name	Dichte ρ	$(BH)_{max}$	B_r	H_{cB}	H_{cJ}	m_{perm}	T_{max}	T_c	Bemerkungen
	g/cm^3	kJ/m^3	T	kA/m	kA/m		°C	°C	
Legierungen									
Isotrop									
AlNico 8/4	7,2	8	0,52	42	43		440		Gesintert
AlNiCo 18/8	7,2	18	0,65	81	82	5	525	800	Gesintert
AlNiCo 500	7,4	35	1,12	47	48	24	450	860	Feinguss
CROVAC 16/160	7,6	15	0,85	39	40		480	640	Verformbar
Anisotrop									
AlNico 35/5	7,2	35	1,12	47	48	$\simeq 5$	550	800	Gesintert
AlNiCo 52/6	7,2	52	1,3	55	56	$\simeq 5$	525	800	Gegossen
LNGT80	7,3	84	1,07	122	125		550	860	Gegossen
RE_2Co_{17}	8,3	200	10,3	720	1600	1,1	300	800	Gesintert
REFeB400/88	$\simeq 7,5$	400	1,44	800	> 875	1,1	> 80	310	Gesintert
REFeB85/35p	5,9	85	0,66	430	830				Kunststoffgepresst
Metalloxide									
Isotrop									
Hartferrit 8/22	4,8	8,0	0,21	135	220	1,2	250	450	
CER5 (Din 24/18)	5,0	28	0,4	160	185		250	450	
Anisotrop									
Hartferrit 28/26	4,85	28	0,38	250	265	1,5	250	450	
Hartferrit 32/26	4,8	32	0,41	250	260	1,5	250	450	

Tab. 22.6 Werkstoffauswahlübersicht nach Meder Electronic

	niedrig → hoch			
Kosten	Ferrite	AlNiCo	NdFeB	SmCo
Energieprodukt ($B{\cdot}H_{max}$)	Ferrite	AlNiCo	SmCo	NdFeB
Einsatztemperatur	NdFeB	Ferrite	SmCo	AlNiCo
Korrosionsbeständigkeit	NdFeB	SmCo	AlNiCo	Ferrite
Gegenfeld-Resistenz	AlNiCo	Ferrite	NdFeB	SmCo
Mechanische Festigkeit	Ferrite	SmCo	NdFeB	AlNiCo
Temperaturkoeffizient	AlNiCo	SmCo	NdFeB	Ferrite

isotrope Magnete haben durch ihre Herstellung im Magnetfeld eine Vorzugsrichtung des Magnetfeldes.

Die stark unterschiedlichen Werte von Flussdichte und Feldstärke der Magnetwerkstoffe führt zu „kleinen“ Magnetflächen bei hoher Flussdichte B_r und „geringer“ Magnetlänge bei hoher Feldstärke H_c.

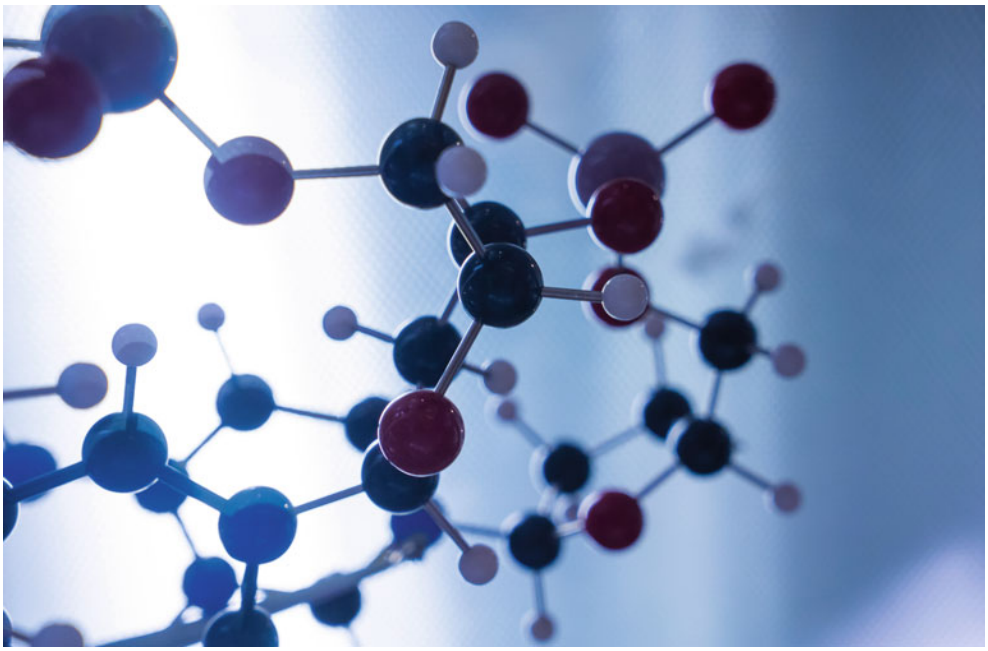
Bei den grundsätzlichen Werkstoffeigenschaften ist zu beachten, dass AlNiCo-Systeme praktisch „Stahl“ sind, während Ferrite als Metalloxide bruchempfindliche „Keramik“ darstellen. Bei den Seltenerd-Systemen sind SmCo sehr hart und spröde, NdFeB Systeme nicht so sehr. Im Gegensatz zu den gegen äußere Einflüsse praktisch unempfindlichen Ferriten oxidieren die Seltenerd-Magnete in feuchter Umgebung. NdFeB etwas stärker als SmCo. Eine Oberflächenbehandlung, z. B. Galvanisierung, ist daher in jedem Falle notwendig.

Wegen der doch recht unterschiedlichen Eigenschaften auf den verschiedenen Gebieten zum Abschluss eine kleine Vergleichstabelle als Entscheidungshilfe (Tab. 22.6).

Dielektrische Eigenschaften

23

Egon Döring



23.1 Modellvorstellungen zur dielektrischen Polarisation 294
23.2 Dielektrische Materialeinteilung 295
23.3 Elektrische Materialeinteilung 295

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_23

23.1 Modellvorstellungen zur dielektrischen Polarisation

In einem elektrischen Feld zwischen zwei Plattenelektroden (Abb. 23.1) sammeln sich auf jeder Platte (freie) Ladungen an (nicht schraffierte Kreise in Abb. 23.1). Wird der Raum zwischen den Platten mit einem *Dielektrikum* gefüllt, so wirken auch darin die Kräfte des elektrischen Feldes. Sie verursachen im atomaren bzw. molekularen Bereich eine *Polarisation*, d. h. eine Verschiebung positiver bzw. negativer Teilchen, die als Beeinflussung vorhandener oder durch Ladungsverschiebungen infolge der Wirkung des elektrischen Feldes gebildeter Dipole verstanden werden kann (Abb. 23.1). Die Folge davon sind zusätzliche Ladungen auf den Kondensatorbelägen, sogenannte *gebundene Ladungen* oder auch *influenzierte Ladungen*. Die Ladung Q auf dem, durch die Elektrodenplatten gebildeten, *Kondensator* mit Dielektrikum ergibt sich zu

$$Q = \varepsilon_0 \varepsilon_r A E \qquad \begin{array}{|c|c|c|c|} \hline Q & \varepsilon_0 & A & E \\ \hline \mathrm{A\,s} & \mathrm{F/m} & \mathrm{m^2} & \mathrm{V/m} \\ \hline \end{array} \qquad (23.1)$$

A = (Kondensator-)Plattenfläche, E = Feldstärke, ε_0 = *elektrische Feldkonstante* oder *absolute Dielektrizitätskonstante* = $8{,}85 \cdot 10^{-12}\,\mathrm{F/m}$ und ε_r = *Permittivitätszahl* oder *relative Dielektrizitätszahl*. Die Größe

$$\varkappa_e = \varepsilon_r - 1 \qquad (23.2)$$

heißt *elektrische Suszeptibilität*. Sie gibt, ähnlich wie beim Magnetfeld, die durch die Materie zusätzlich mögliche Feld(linien)dichte an.

Elektronenpolarisation ist die Verschiebung der Ladungsschwerpunkte von Atomkern und Elektronenhülle. Wegen der geringen Trägheit der Elektronen folgt die Elektronenpolarisation Wechselfeldern bis zu den Frequenzen der UV-Strahlung. In Materialien mit instabilem Charakter der Elektronenverteilung ist die Elektronenpolarisierbarkeit besonders groß.

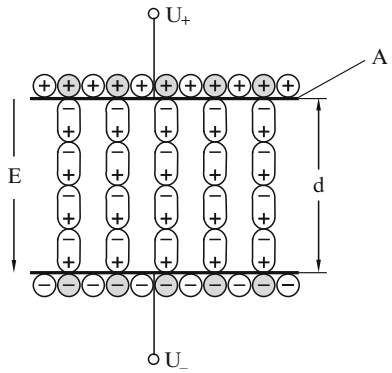


Abb. 23.1 Plattenkondensator. Ladung bei Spannung U bzw. Feldstärke E ohne und mit Dielektrikum (schematisch)

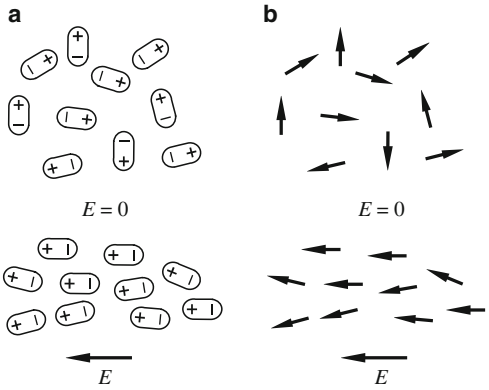


Abb. 23.2 Ordnungspolarisation (schematisch). Die zunächst regellos orientierten Dipole ordnen sich unter dem Einfluss eines elektrischen Feldes E. **a** Dipol-Darstellung, **b** Schema-Darstellung

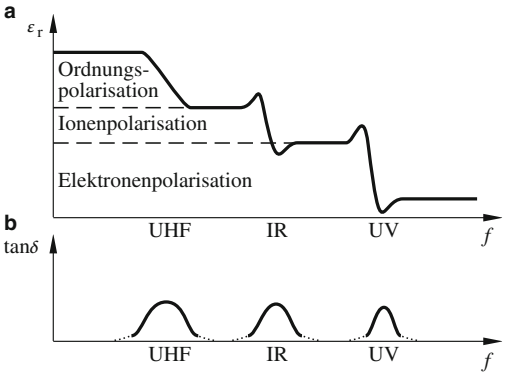


Abb. 23.3 Frequenzabhängigkeit der Polarisation und der Verluste (prinzipieller Verlauf). **a** Beiträge der verschiedenen Polarisationsmechanismen zur Permittivität, **b** Bereiche ausgeprägter Verluste durch Resonanzerscheinungen. $\tan\delta$ = Verlustfaktor, f = Frequenz, UHF = Ultrahochfrequenz, IR = Infrarot(strahlung) und UV = Ultraviolett(strahlung)

Ionenpolarisation verschiebt die Ionen unter dem Einfluss des elektrischen Feldes und bewirkt eine Deformation des Gitters. Der Stoff ist polar. Ionenpolarisation ist bis zu Frequenzen der IR-Strahlung möglich. Ionen- und Elektronenpolarisation sind von ähnlicher Größenordnung, kaum temperaturabhängig und können unter dem Oberbegriff *Deformiationspolarisation* zusammengefasst werden.

Ordnungspolarisation setzt Dipolmomente im molekularen Aufbau voraus. Die einzelnen Dipole ordnen sich weitgehend in Feldrichtung (Abb. 23.2), daher *Ordnungs-* oder *Orientierungspolarisation*. Der Stoff ist (di)polar. Der ordnenden Kraft des elektrischen Feldes wirken thermische Kräfte entgegen. Die Ordnungspolarisation und Grenzfrequenz (bis etwa 10^9 Hz) sind stark temperaturabhängig.

Grenzflächenpolarisation tritt bei Dielektrika mit Einschlüssen guter Leitfähigkeit auf. Frei bewegliche Ladungsträger der Einschlüsse können zu den Grenzen wandern und wie gebundene Ladungen wirken. Diese Grenzflächenpolarisation führt zu einer höheren Polarisation als dem homogenen Material entspräche und bewirkt ein ε_r über dem des eigentlichen Dielektrikums. Der Grenzfall tritt auf, wenn das Dielektrikum eine gewisse

Leitfähigkeit zeigt. Eine Unterscheidung zwischen Ladungsbewegung zur Grenzflächenpolarisation und Leitfähigkeit ist dann nicht mehr möglich.

Abb. 23.3 zeigt schematisch den aus den verschiedenen Mechanismen resultierenden Frequenzgang der Polarisation. Die Endbereiche der verschiedenen Polarisationsarten sind durch Steilabfälle der Permittivitäten und wegen der dabei auftretenden Resonanzerscheinungen durch starke Verluste gekennzeichnet (Abb. 23.3b).

23.2 Dielektrische Materialeinteilung

Die Elektronenpolarisation, mit der Verschiebung der Elektronen gegenüber dem Atomkern, ist in allen Stoffen von Null verschieden. Daraus folgt grundsätzlich für alle Dielektrika eine Permittivität $\varepsilon_r > 1$ bzw. eine positive elektrische Suszeptibilität, allgemein *Paraelektrika*.

Piezoelektrika reagieren mit einer Polarisation(sänderung) auf eine mechanische Deformation und umgekehrt. Deformationsrichtung und Polarisationsrichtung – meist senkrecht zueinander – sind dabei einander fest zugeordnet.

Bei einer Reihe von Materialien mit in kleinen Bereichen geordneten Dipolen (auch hier Weiss'sche Bezirke genannt) kann eine bei erhöhter Temperatur erfolgte Ausrichtung der Dipolmomente, über einen größeren Bereich durch Abkühlung „eingefrorene", und starke remanente Polarisation erzeugt werden. Derartige künstliche Piezoelektrika zeigen stärkere Effekte als die natürlichen.

Die bekanntesten natürlichen Piezoelektrika sind der Quarz und der Turmalin. Künstliche Piezoelektrika basieren auf Mischoxiden von Ba, Ti, Pb, Nb u. a.

Pyroelektrika sind polare piezoelektrische Materialien mit einer spontanen Polarisation, die ihre Ursache in getrennten Schwerpunkten der positiven und negativen Ladungen in der Kristallstruktur hat. Weil die Polarisation zwar mit der Temperatur veränderlich, aber selbst bis zum Schmelzpunkt des Materials nicht aufhebbar ist, heißt dieser Effekt *pyroelektrisch*.

Ferroelektrika. Gewisse Stoffe weisen ein elektrisches Dipolmoment ohne Anlegen eines äußeren elektrischen Feldes auf. Der ferroelektrische Effekt kommt nur in Kristallen vor, in denen die kristalline Symmetrie eine polare Achse zulässt. Ferroelektrika sind deshalb auch immer piezo- und pyroelektrisch. Die Vorsilbe „ferro" bezieht sich nur auf die Analogie zu „ferromagnetisch", wegen des Auftretens einer Hystereseschleife $D = f(E)$ und einer damit verbundenen Koerzitivfeldstärke E_c. Ursache sind auch hier Domänen (gleichfalls Weiss'sche Bezirke genannt), wobei die Übergänge, Domänenwände, „dünner" sind als die Übergänge bei Ferromagnetika. In Anlehnung an den Begriff Magnete nennt man derartige Stoffe *Elektrete*. Die spontan vorhandene Polarisation kann bei Ferroelektrika durch eine äußere Spannung um 180° umgepolt werden, was bei reinen Piezoelektrika nicht möglich ist. Bei hohen Temperaturen, auch hier (ferroelektrische) Curie-Temperatur genannt, verschwindet die Polarisation, das Material wird paraelektrisch. Ferroelektrika sind einerseits Materialien mit großer Permittivität ($\varepsilon_r \simeq 10^4$) andererseits sind die Deformationen durch elektrische Felder als *Aktuatoren* anwendbar oder umgekehrt durch Deformation – Schlag – als (Funken)Spannungsquelle. Eine andere Anwendungsmöglichkeit sind, wegen der hohen Dielektrizitätskonstanten, ferroelektrische Arbeitsspeicher.

23.3 Elektrische Materialeinteilung

Isolatoren: Die wesentlichste Kenngröße eines Isolators ist der Isolationswiderstand mit spezifischen Widerständen von $\rho = 10^{10} \ldots 10^{20}\,\Omega$m. In der technischen Anwendung ist dabei noch zwischen *Durchgangs-* und *Oberflächenwiderstand* zu unterscheiden.

Durchgangswiderstand ist eine von den inneren Eigenschaften des Stoffes abhängige und für ihn charakteristische Größe. Er wird aus dem Querstrom I bestimmt, den eine angelegte Spannung U (Abb. 23.4a) durch den Werkstoff treibt. Wegen einer stets vorhandenen Anzahl freier Ladungsträger wird dieser Querstrom niemals absolut zu Null. Unmittelbar nach dem Anlegen einer Gleichspannung tritt noch ein *Polarisationsstrom* auf, der u. U. durch die Trägheit der Dipolmomente eine merkliche *Abklingzeit (Relaxationszeit)* aufweist. Wird der Querstrom durch Ionen verursacht (Materialtransport), so ist über längere Zeit mit einer Zerstörung des Isolators zu rechnen.

Der Durchgangswiderstand nimmt mit steigender Temperatur ab. Der daraus folgende erhöhte Querstrom trägt seinerseits zu einem Temperaturanstieg bei. Falls sich kein Gleichgewicht durch Ausgleich mit der Umgebungstemperatur einstellt, kommt es zum Wärmedurchschlag.

Oberflächenwiderstand (Abb. 23.4b) ist im strengen Sinne keine Materialeigenschaft, da Oberflächenbeschaffenheit und -verunreinigungen die entscheidenden Faktoren sind. Verunreinigungen auf der Oberfläche eines Isolators ermöglichen *Kriechströme*. Die dadurch verursachten Kriechspuren können die Oberfläche des Materials angreifen und die Leitfähigkeit weiter erhöhen. Der Widerstand gegen die Kriechspurbildung, die *Kriechstromfestigkeit*, ist in der Praxis meist wichtiger als der Oberflächenwiderstand selbst.

Durchschlagfestigkeit ist die beim *Spannungsdurchschlag* eines Isolierstoffes wirksame elektrische Feldstärke E_0 in kV/cm oder kV/mm angegeben. Die Durchschlagfestigkeit ist nur unter gleichen Bedingungen vergleichbar, da sie sowohl von der Elektrodenform als auch Materialdicke abhängt (DIN 53481 bzw. VDE 0303). Im Allgemeinen sinkt die Durchschlagfestigkeit mit zunehmender Materialdicke. Die Durchschlagfeldstärke

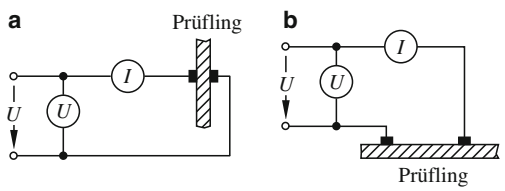


Abb. 23.4 Isolationswiderstand. **a** Durchgangswiderstand, **b** Oberflächenwiderstand, U = Spannung und I = Strom

Tab. 23.1 Wärmebeständigkeitsklassen nach DIN/IEC

Klasse	Grenztemperatur (°C)	Isolierstoffcharakteristika
Y	90	Organische Stoffe, ungetränkt
A	105	Organische Stoffe mit organischen Bindemitteln
E	120	Lackdrähte
B	130	Anorganische Stoffe mit organischen Bindemitteln
F	155	Anorganische Stoffe mit besonderen organischen Bindemitteln
H	180	Anorganische Stoffe z. B. mit Silikon
C	> 180	Anorganische Stoffe ohne Bindemittel

Tab. 23.2 Eigenschaften dielektrischer Werkstoffe

Stoff	Unpolar	Polar	Ferroelektrisch
ε_r	< 10	10 ... 100	> 100
Temperaturgang	Gering	Gering	Stark
Verluste	Gering	Gering	Stark frequenzabhängig

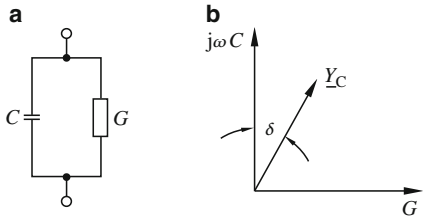


Abb. 23.5 Verlustbehafteter Kondensator. **a** (Parallel-)Ersatzschaltung, C = idealer Kondensator, G = Verlustleitwert, **b** kapazitiver Leitwert $\underline{Y}_c$ in der komplexen Ebene, δ = Verlustwinkel

ist auch noch von der Anstiegsgeschwindigkeit der angelegten Spannung abhängig.

Wegen der Zeitabhängigkeit des thermischen oder Wärmedurchschlages muss die Prüfung über Zeit (> 40 s) erfolgen.

Für die Zuverlässigkeit elektrischer Geräte ist die Isolationseigenschaft der Werkstoffe auch bei langzeitiger Temperaturbelastung wichtig. Man unterscheidet daher nach VDE bzw. IEC *Wärmebeständigkeitsklassen*, wie sie in der Tab. 23.1 aufgeführt sind.

Erfahrungsgemäß genügt für die *Halbierung/Verdoppelung der Lebensdauererwartung* bei einer Isolation nach:

Klasse A eine um 8 °C erhöhte/verminderte Betriebstemperatur
Klasse B eine um 10 °C erhöhte/verminderte Betriebstemperatur
Klasse H eine um 12 °C erhöhte/verminderte Betriebstemperatur

Kondensatoren: Mit $C = Q/U$ folgt für Kondensatoren die Kapazität C aus (23.1) zu

$$C = \varepsilon_0 \varepsilon_r \frac{A}{d} \qquad \begin{array}{c|c|c|c} C & \varepsilon_0 & A & d \\ \hline \mathrm{F} & \mathrm{F/m} & \mathrm{m^2} & \mathrm{m} \end{array} \qquad (23.3)$$

Als Werkstoffgröße ist die Permittivitätszahl ε_r des Dielektrikums und der damit verbundene Verlustfaktor $\tan\delta$ für die

Tab. 23.3 Permittivitätszahlen einiger Werkstoffe

Material	ε_r
Al_2O_3-Keramik	9,8
TiO_2-Keramik	85
Quarzglas	3,8
Bor-Silikat-Glas	4–6
Kalk-Alkali-Glas	6–8
Porzellan	4–8
Glimmer	5–6
Diamant (C)	5,6
Silizium (Si)	11,9
Germanium (Ge)	16,2
Steatit ($SiO_2/Al_2O_3/MgO$)	6–7
Papier	4–5,6
Polyvinylchlorid (PVC)	≈3
Polyäthylen (PE)	2,3
Polypropylen (PP)	2,2
Polystyrol (PS)	3
Polycarbonat (PC)	2,8
Polytetrafluoräthylen (PTFE)	2,1

Beschreibung der Verluste maßgebend. In der Ersatzschaltung (Abb. 23.5) liegt der Kapazität C der Leitwert G parallel. Definitionsgemäß ist der dielektrische *Verlustfaktor* das Verhältnis von Wirk- zu Blindleistung, daraus folgt

$$\tan\delta = \frac{G}{\omega C} = \frac{1}{\omega RC} \qquad \begin{array}{c|c|c|c} G & R & C & \omega \\ \hline \mathrm{S} & \Omega & \mathrm{F} & \mathrm{s^{-1}} \end{array} \qquad (23.4)$$

mit ω = Kreisfrequenz und $R = 1/G$. Das Produkt des Verlustfaktors $\tan\delta$ mit der Permittivität ε_r liefert die, besonders in der Kabeltechnik wichtige, (dielektrische) *Verlustzahl* oder *Verlustziffer*

$$\varepsilon'' = \varepsilon_r \cdot \tan\delta \qquad (23.5)$$

Eine weitere für Gleichstrom und niedrige Frequenzen angewandte Verlustdefinition ist über das Produkt $R \cdot C$ (23.4) als *Isolationszeitkonstante* oder *Isolationsgüte* möglich. Sie wird in MΩ µF oder zahlen- und dimensionsgleich in s angegeben.

Aus den verschiedenen Arten der Polarisation folgen für die Dielektrika unterschiedliche Größenordnungen der Permittivität, die in der Tab. 23.2 einmal prinzipiell zusammengestellt sind.

Bei den keramischen Materialien wird in der Praxis noch zwischen *niederdielektrischer Keramik*, NDK-Materialien, und *hochdielektrischer Keramik*, HDK-Materialien, unterschieden.

NDK-Materialien basieren auf nicht-ferroelektrischen Titanaten. Temperaturgang, Verlustfaktor (10^{-4} bis 10^{-5}) und Spannungsabhängigkeit von ε_r sind gering. Die elektrische Langzeitstabilität ist gut.

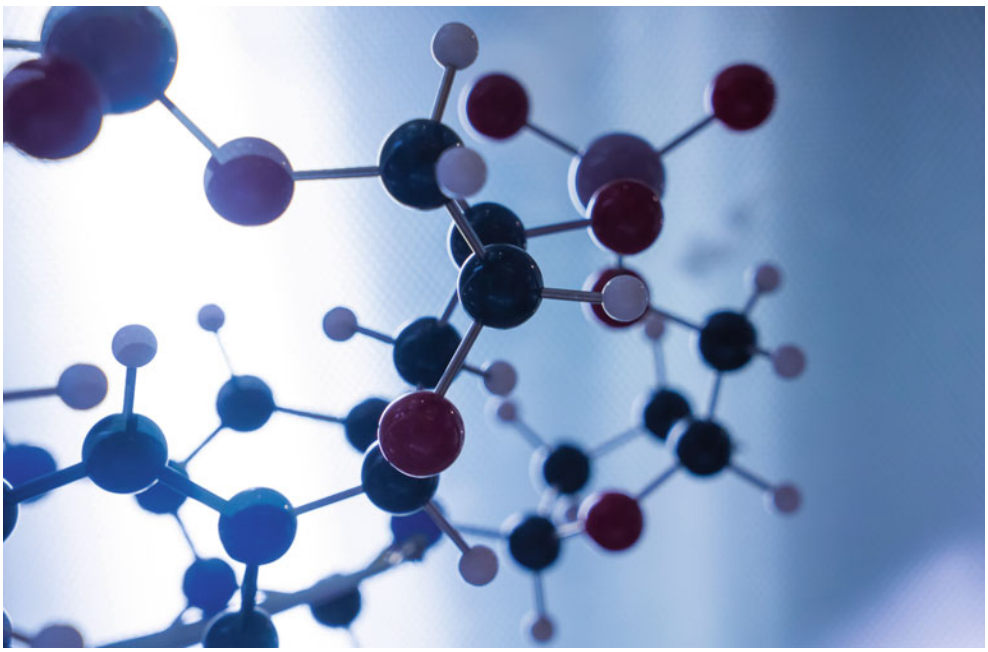
HDK-Materialien basieren auf ferroelektrischem Bariumtitanant mit stark temperatur- und spannungsabhängiger Permittivität. Der Verlustfaktor ist mit $10^{-2} \ldots 10^{-3}$ recht hoch.

In der Tab. 23.3 sind abschließend die Permittivitätszahlen einiger ausgewählter Materialien zusammengestellt.

Dielektrika

24

Egon Döring



24.1 Natürliche anorganische Dielektrika 298
24.2 Natürliche organische Dielektrika 299
24.3 Künstliche anorganische Dielektrika 299
24.4 Künstliche organische Dielektrika 300
24.5 Silikone 301
Literatur 301

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_24

Die Dielektrika lassen sich nach ihrem Vorkommen in natürliche und künstliche und nach ihrem chemischen Aufbau in anorganische und organische unterteilen.

24.1 Natürliche anorganische Dielektrika

Glimmer ist ein leicht spaltbares, schneid- und stanzbares Naturprodukt mit hervorragenden elektrischen Eigenschaften. Glimmerreste werden mit Kunstharzen verpresst (Mikanit) oder nach einem Spezialverfahren zu einem glimmerähnlich nutzbaren Vlies verarbeitet. E_0 bis zu 1000 kV/cm, $\varepsilon_r \simeq 6 \ldots 8$, $\tan\delta \simeq 2 \cdot 10^{-4}$.

Quarz ist reines Siliziumdioxid, hochtemperaturfest, mit geringer Wärmedehnung und guter UV-Durchlässigkeit bis 185 nm. Abgeleitete Produkte sind Quarzglas, Quarzgut, gezüchtete Quarzkristalle und Quarzmehl (Füllstoff). $E_0 = 300 \ldots 400\,\mathrm{kV/cm}$, $\varepsilon_r = 3 \ldots 4$, $\tan\delta \simeq 10^{-2} \ldots 10^{-4}$ (stark von der Reinheit abhängig).

Wegen seines Piezoeffektes hat Quarz besondere Bedeutung als *elektromechanischer Wandler* und *Resonator* zur Frequenzstabilisierung in der Nachrichtentechnik. Die durch Umkehrung des Piezoeffektes sehr stabile mechanische Resonanz des Quarzes führt zur Anwendung als „Bandfilter“ mit geringer Bandbreite.

Glas ist Siliziumdioxid das durch schmelzpunkterniedrigende Zusätze – sog. Flussmittel – aus dem Bereich der Carbonate, Nitrate und Sulfonate der Alkalimetalle von seinem Schmelzpunkt von 1860 °C auf Erweichungstemperaturen oberhalb von 600 °C gebracht wird. Wobei zu beachten ist, dass Glas zunächst bei steigender Temperatur erweicht und erst allmählich in einen flüssigen Zustand etwa oberhalb von 1200 °C übergeht. Die allgemeinen dielektrischen Daten sind: $\varepsilon_r \simeq 5 \ldots 16$, $E_0 \simeq 10\,\mathrm{kV/mm}$ und $\tan\delta \to 0$. Letztere Eigenschaft hat zu *Glaskondensatoren* geführt, die, zunächst als Glimmerersatz, inzwischen für die Raumfahrt wichtig sind. Sie sind auch bei langen Raumflügen gut ladungskonstant und gegen Strahlungen praktisch aller Art unempfindlich.

Die andere bedeutende Anwendung ist die Lichtleitung in dünnen Fasern, sog. *Lichtwellenleiter*. Wie Abb. 24.1 zeigt hat

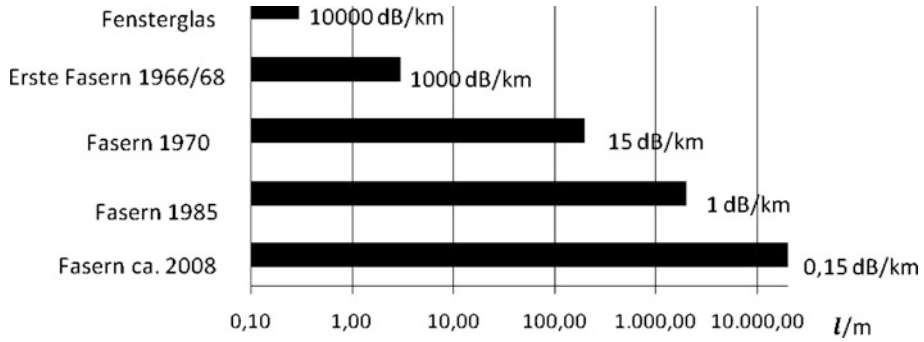


Abb. 24.1 Änderung der Übertragungsdämpfung in dB/km und der daraus folgenden Dämpfungs-„Halbwertslänge“ *l* im Laufe der Zeit. Nach ANT und Schott

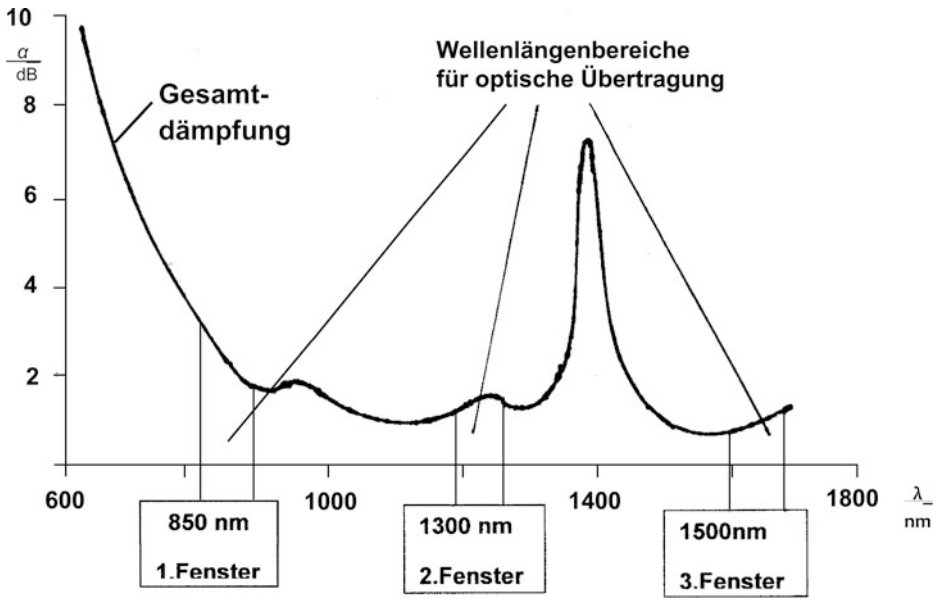


Abb. 24.2 Verlauf der Gesamtdämpfung α in Lichtwellenleitern und der Definition *optischer Fenster*. Sie liegen bei 850, 1300 und 1500 nm

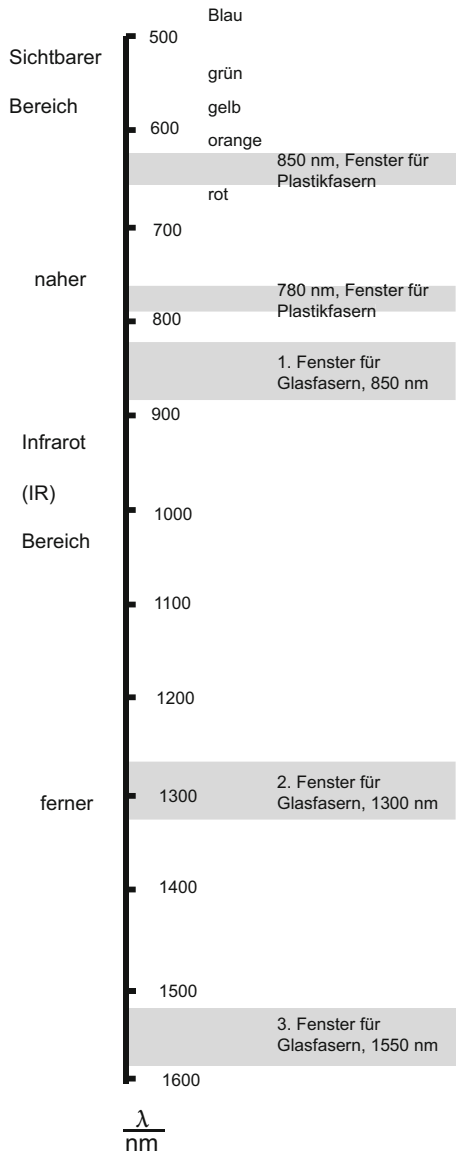


Abb. 24.3 Übersicht über die Wellenlängenbereiche und die Zuordnung der zur Übertragung genutzten optischen Fenster, λ = Wellenlänge (1000 nm = 1 µm ≙ 300 THz)

Glas seine „Lichtdurchlässigkeit“ im Laufe der letzten Jahrzehnte durch die Entwicklung geeigneter Glassorten erheblich verbessert. Die Lichtübertragungsstruktur in den verschiedenen Faserarten, monomode, multimode usw., ist in Abschn. 103.2 „Grundmodelle von Lichtwellenleitern“ in Kap. 103 „Nachrichtenübertragung über Lichtwellenleiter (LWL)“ in Teil XIII erläutert.

Bei den Fasern der Lichtwellenleiter bilden sich durch Streuung und Absorption unterschiedliche Dämpfungsbereiche für die verschiedenen Wellenlängen aus. Wie Abb. 24.2 zeigt, bilden diese optimalen Bereiche sog. *optische Fenster* mit den Nummern 1,2 und 3 für 850, 1300 und 1550 nm. In Abb. 24.3 ist noch einmal die wellenlängenmäßige Zuordnung der Fenster und auch die Lage der Farblichtstrahlungen, zusammengestellt.

Gase: Hierbei ist Luft mit einer Durchschlagfestigkeit von 20 … 30 kV/cm der am häufigsten angewandte Isolator. Technisch sind außerdem noch Stickstoff (N, $E_0 \sim 20\,\mathrm{kV/cm}$), Kohlendioxid ($CO_2$, $E_0 \simeq 24\,\mathrm{kV/cm}$) und Schwefelhexafluorid (SF_6, $E_0 \simeq 100\,\mathrm{kV/cm}$) bedeutsam. Wegen der durch erhöhten Druck verminderten Teilchenbeweglichkeit steigt die *Ionisierungsfeldstärke* für den Durchbruch und damit auch die Durchschlagfestigkeit mit einer Druckerhöhung. Erschwert ionisierbar sind auch *elektronegative Gase*, die bevorzugt als Halogenverbindungen – z. B. CCl_4, CCl_2F_2, CF_4, SF_6 u. a. – mit großer Ionenmasse und dementsprechend niedriger Wanderungsgeschwindigkeit genutzt werden. So hat das technisch bedeutsame Schwefelhexafluorid (SF_6) eine um etwa den Faktor 2,5 höhere Durchschlagfestigkeit als Luft.

24.2 Natürliche organische Dielektrika

Papier weist als *Isolierpapier* je nach Behandlung und Tränkung sehr unterschiedliche Eigenschaften auf. Durch Tränkung entstehen die Öl-, Öllack, Schellack- und Kunstharzpapiere mit E_0 etwa 25 … 50 kV/mm gegenüber ca. 10 kV/mm bei einfachen Papieren. Vorteilhaft ist die geringe Dicke von Papierisolationen, nachteilig sind die hohen Verluste mit $\tan\delta = 0{,}5 \ldots 10^{-2}$. ε_r liegt bei 2 … 8.

Textilstoffe werden meist getränkt verarbeitet, wobei besonders Lackseide (Ölseide) wegen der geringen Dicke und der guten elektrischen und mechanischen Eigenschaften wichtig ist. Baumwolle und Seide werden als Fäden und Bänder getränkt und ungetränkt als Umhüllungen für blanke und lackierte Drähte benutzt.

Öle sind als *Isolieröle* (DIN EN 60296, VDE 0370) hochsiedende Destillationsprodukte des Erdöls. Isolationswiderstand und Durchschlagfestigkeit (5 … 30 kV/mm) sind besonders in dünnen Schichten günstig. $\tan\delta = 10^{-3}$ und $\varepsilon_r = 2 \ldots 3$ sind niedrig. Nachteilig sind die starke Wärmedehnung, Temperaturabhängigkeit der Viskosität, Entflammbarkeit, Feuchtigkeitsaufnahme aus der Luft und Alterung durch den Luftsauerstoff. Die Alterungsprodukte (Säuren, Metallseifen, Schlamm u. ä.) erhöhen die Verluste und mindern die Wärmeleitfähigkeit des Öls. Eine Regeneration ist möglich.

24.3 Künstliche anorganische Dielektrika

Vorzugsweise Metalloxide, die als *keramische Massen* teigig geformt, getrocknet und gebrannt ihre endgültigen Eigenschaften erhalten. Nachbearbeitung ist wegen der Härte durchweg nur (nass-)schleifend möglich. Für die praktische Handhabung sind noch verschiedene Oberflächenbehandlungen notwendig:

Metallisieren ist für die Kontaktierung bei Dielektrika oder als Basismaterial für Hochfrequenzschaltungen notwendig.

Glasieren und *Lackieren* dient zum Schutz der meist rauen oder porösen Oberfläche gegen Verschmutzung und Feuchtigkeitsaufnahme.

Hydrophobieren ist eine besondere Art des Feuchtigkeitsschutzes durch Aufbringen dünnster (bis monomolekularer) Silikonschichten, die durch ihre niedrigen Oberflächenspannungen wasserabweisend sind.

Metalloxide von Aluminium (Al_2O_3) und Titan (TiO_2) werden hochrein wegen ihrer geringen Verluste besonders bei hohen Frequenzen als Trägermaterial eingesetzt. $E_0 \simeq 10\,\mathrm{kV/mm}$, Aluminiumoxid: $\varepsilon_r = 9{,}8$ und $\tan\delta = 10^{-4}$ ($\ldots 10^{-3}$), Titandioxid: $\varepsilon_r = 85$ und $\tan\delta = 4 \cdot 10^{-3}$.

Porzellan hat als Ausgangsmaterialien Kaolin (Aluminiumsilikat $Al_2O_3 \cdot SiO_2$, H_2O, Porzellanerde), Feldspat ($Al_2O_3 \cdot CaO_2 \cdot SiO_2$) und Quarz ($SiO_2$).

Das Endprodukt ist durch die Zusammensetzung und geringe weitere Zusätze in weiten Grenzen beeinflussbar. Porzellane gehören zu den dichten keramischen Massen. Sie sind bei Raumtemperatur beständig gegen Basen und Säuren (Ausnahme: Flusssäure) und hochwiderstandsfähig gegen elektrische Funken. Für das bei Isolatoren wichtige Hartporzellan betragen $E_0 = 30 \ldots 35\,\mathrm{kV/mm}$, $\varepsilon_r = 4 \ldots 8$ und $\tan\delta \simeq 0{,}02 \ldots 0{,}1$.

Titanate mit Barium ($BaTiO_3$), Kalzium ($CaTiO_3$) und Strontium ($SrTiO_3$) erreichen als HDK-Materialien sehr hohe Permittivitäten (ε_r bis 10^4 bei Bariumtitanat, das auch piezoelektrisch ist). Für NDK-Materialien auf Titandioxid-Basis (Rutil) kann der Temperaturkoeffizient TK in weiten Grenzen zwischen positiven und negativen Werten eingestellt werden. $\varepsilon_r = 10^2 \ldots 10^4$ und $\tan\delta = 10^{-3} \ldots 10^{-5}$.

Eine andere wichtige Anwendung der künstlichen piezoelektrischen Materialien sind Resonanzen zwischen den durch elektrische Spannungen hervorgerufenen mechanischen Verformungs(resonanzen) und den damit möglichen *keramischen* oder auch *dielektrischen Filtern*. Im einfachsten Fall quasi ein Kondensator mit einem „Resonator-Dielektrikum“ bis hin zu Oberflächenwellenfiltern (OFW) (engl. SAW surface acoustic waves). Weitere Einzelheiten hierzu in Abschn. 94.4.9 „Dielektrische Filter“ in Teil XIII.

24.4 Künstliche organische Dielektrika

Zellulosekunststoffe sind durch chemische Behandlung von Zellulose erhaltene Fasern und Folien. Für Isolationen sind die blau eingefärbten Triazetat-($T_{\max} = 115\,^\circ\mathrm{C}$) und Azetobutyratfolien ($T_{\max} = 120\,^\circ\mathrm{C}$) wichtig. $E_0 = 20 \ldots 50\,\mathrm{kV/mm}$, $\varepsilon_r = 4$ und $\tan\delta \simeq 0{,}01 \ldots 0{,}02$.

Chlophene sind thermisch und chemisch gut beständige chlorierte Diphenyle, die jedoch bei Bränden durch die Bildung von polychlorierten Biphenylen (PCB) umweltgefährlich sind. Kunststoffe, z. B. phenolhaltige, können von Chlophenen angegriffen werden! Bei der Entsorgung sind die Sondermüllvorschriften zu beachten. $E_0 \sim 200\,\mathrm{kV/cm}$, $\varepsilon_r = 4{,}5 \ldots 6$ und $\tan\delta = 10^{-3}$. Als biologisch ungefährliche synthetische Isolierflüssigkeit kann Pentaerythrit-Tetraester eingesetzt werden. $E_0 \simeq 200\,\mathrm{kV/cm}$, $\varepsilon_r = 3{,}3$ und $\tan\delta = 10^{-3}$.

Polykondensationsprodukte treten vorwiegend als *Duroplaste* auf, d. h. als Kunststoffe, die beim Erhitzen vorübergehend schmelzen oder erweichen und dann in einen harten, unlöslichen und unschmelzenden Endzustand übergehen. Es gibt sie rein, modifiziert und plastifiziert. Die wichtigsten sind:

Phenoplaste (PF) in selbst- oder nichtselbsthärten- der Ausführung. Sie finden Anwendung bei Lacken, Pressmassen und Schichtstoffen (Hartpapiere und -gewebe). $E_0 = 5 \ldots 20\,\mathrm{kV/mm}$, $\varepsilon_r = 4 \ldots 9$, $\tan\delta = 0{,}05 \ldots 0{,}3$ und $T_{\max} \simeq 160\,^\circ\mathrm{C}$.

Aminoplaste (HF) ähneln den Phenoplasten, sind im Gegensatz dazu jedoch farbecht und lichtbeständig. Sie haben glatte Oberflächen mit geringer Haftfähigkeit und werden für Lacke, Schichtstoffe und Pressmassen verwandt. $E_0 = 10 \ldots 20\,\mathrm{kV/mm}$, $\varepsilon_r = 5 \ldots 7$, $\tan\delta \simeq 0{,}1$ und $T_{\max} \simeq 150\,^\circ\mathrm{C}$.

Polyester(-harze) (PETF) sind flüssige bis elastische oder auch splitterharte Kunstharze. $E_0 \simeq 30\,\mathrm{kV/mm}$, $\varepsilon_r = 3$, $\tan\delta \simeq 0{,}03$ und $T_{\max} = 110\,^\circ\mathrm{C}$.

Ungesättigte Polyesterharze (UP) nehmen eine Mittelstellung zwischen Kondensations- und Polymerisationsharzen ein, da die über Katalysatoren erfolgende Aushärtung bereits eine Polymerisation darstellt. Thermisch nicht härtbare Typen von Polyestern dienen zur Herstellung mechanisch und elektrisch hochwertiger Folien und Fasern. Mit Glasfasern als Füllstoff ergeben die Polyesterharze die *glasfaserverstärkten Kunststoffe (GFK)*.

Epoxidharze (EP) werden wie Polyesterharze eingesetzt, haben ihnen gegenüber jedoch eine höhere mechanische Festigkeit und ein sehr hohes Haftvermögen auf beinahe allen Untergründen (Klebstoff). $E_0 = 40\,\mathrm{kV/mm}$, $\varepsilon_r = 3 \ldots 5$, $\tan\delta \simeq 0{,}02$ und $T_{\max} = 130\,^\circ\mathrm{C}$.

Polyurethane (PU) sind lackartig als Lötlacke interessant. $T_{\max} = 120\,^\circ\mathrm{C}$. Thermisch höher belastbar ist *Polyesterimidlack*. Die Durchschlagspannung derartig isolierter Drähte liegt bei etwa $4 \ldots 6\,\mathrm{kV}$, $\tan\delta \simeq 0{,}01$ (stark temperaturabhängig).

Polymerisationsprodukte: Vorwiegend Thermoplaste, d. h. in der Wärme weich und in der Kälte zähhart bis spröde.

Polystyrol (PS) ist glasklar und spröde, hat niedrige Verlustfaktoren noch bis zu sehr hohen Frequenzen und eine geringe Wasseraufnahmefähigkeit. Nachteilig sind die geringe Wärmefestigkeit ($T_{\max} = 75\,^\circ\mathrm{C}$), schlechte Licht- und Wetterbeständigkeit sowie die Gefahr von Spannungskorrosionen bei Kontakt mit Lösungsmitteln. Anwendung findet Polystyrol für Spritzgussmassen (meist schlagfestes PS), elektrisch hochwertige Formteile und Folien für Kondensatoren und HF-Kabel. $E_0 \simeq 100\,\mathrm{kV/mm}$, $\varepsilon_r \simeq 3$ und $\tan\delta = 3 \cdot 10^{-4}$.

Polyethylen (PE) hat Ähnlichkeit mit Paraffinen, eine wachsartige Oberfläche, sehr geringe Wasseraufnahmefähigkeit, ist trüb und quillt wenig in polaren, aber stark in unpolaren Lösungsmitteln und in der Hitze. Durch Vernetzung (VPE = vernetztes

PE), d. h. Bildung von Riesenmolekülen aus den durch Polymerisation entstandenen langen PE-Fadenmolekülen, lassen sich die mechanischen und thermischen Eigenschaften weiter verbessern. Der Isolationswiderstand ist mit $10^{17}\,\Omega\,\mathrm{cm}$ sehr hoch. PE findet Anwendung für Spritzgussmassen, Folien, elastische, sehr korrosionsbeständige Formteile und Hochspannungskabel. $E_0 \simeq 100\,\mathrm{kV/mm}$, $\varepsilon_r = 2{,}3$, $\tan\delta \simeq 10^{-4}$ und $T_{\max} = 70\,^\circ\mathrm{C}$ (unvernetzt) bzw. 90 °C (vernetzt, erhöhte Kurzschlussbelastbarkeit bei Kabeln).

Polyvinylchlorid (PVC) ist ein weitverbreiteter, abriebfester und zäher Kunststoff mit guten Isolationseigenschaften. Achtung, beim Verbrennen oder Verschwelen wird HCI abgespalten! PVC findet für Formteile, Folien und Isolationen vielfältige Anwendungen. $E_0 \simeq 40\,\mathrm{kV/mm}$, $\varepsilon_r = 3$, $\tan\delta = 0{,}02\ldots0{,}1$ und $T_{\max} = 70\,^\circ\mathrm{C}$.

Polytetrafluorethylen (PTFE) (Teflon) ist das gegen Chemikalien beständigste synthetische Material. Es ist außerordentlich korrosions-, feuchtigkeits- und temperaturbeständig (Betriebstemperaturen von $-200\ldots+260\,^\circ\mathrm{C}$). $E_0 \simeq 30\,\mathrm{kV/mm}$, $\varepsilon_r \simeq 2$, $\tan\delta = 2\cdot10^{-4}$ und $T_{\max} = 280\,^\circ\mathrm{C}$.

24.5 Silikone

Silikone sind Kunststoffe, bei denen die Kettenbildung über das vierwertige Silizium erfolgt. Hervorstechend sind die hohe Temperaturbeständigkeit und die geringen Temperaturbeiwerte der mechanischen und elektrischen Daten. Silikone sind absolut tropenfest und neigen nicht zu Schimmelansatz. Sie zeigen gutes Kriechstrom- und Lichtbogenverhalten. Durch elektrische Funken entsteht isolierendes Siliziumdioxid und kein leitender Kohlerückstand wie bei organischen Stoffen. Nachteilig ist der grundsätzlich höhere Preis gegenüber anderen Werkstoffen.

Silikonöle weisen nur eine geringe Änderung der Viskosität zwischen $-60\ldots+300\,^\circ\mathrm{C}$ auf. Anwendung (teuer) als Imprägniermittel und flüssiges Dielektrikum. $E_0 \simeq 10\,\mathrm{kV/mm}$, $\varepsilon_r \simeq 3$, $\tan\delta \simeq 10^{-3}$ und $T_{\max} \simeq 300\,^\circ\mathrm{C}$.

Silikonharze dienen zur Herstellung von warm- und lichtbogenfesten, unbrennbaren Pressstücken. $E_0 = 10\ldots30\,\mathrm{kV/mm}$, $\varepsilon_r \simeq 3$, $\tan\delta = 2\ldots5\cdot10^{-3}$ und $T_{\max} = 100\ldots200\,(\mathrm{u.\,U.}\ldots300)\,^\circ\mathrm{C}$.

Silikonelastomere (Silikonkautschuke) sind gummielastische Massen mit Temperaturbeständigkeit und Biegsamkeit zwischen ca. $-80\ldots+250\,^\circ\mathrm{C}$. Sie finden Anwendung als Isolierungen und dauerelastische, hochwarmfeste Verbindungen zwischen praktisch beliebigen Werkstoffen. $E_0 = 10\ldots30\,\mathrm{kV/mm}$, $\varepsilon_r = 3\ldots9$, $\tan\delta = 0{,}01\ldots0{,}1$ und $T_{\max} \simeq 180\,^\circ\mathrm{C}$.

Diagramme und Daten wurden aus Veröffentlichungen folgender Unternehmen übernommen:

EBG Gesellschaft für elektromagnetische Werkstoffe mbH, Bochum

EPCOS AG München

Meder Electronic AG Singen

Murata Manufacturing Co. Ltd. Japan

Neosid Pemetzrieder GmbH&Co KG Halver

Philips Deutschland GmbH Hamburg

RFS Radio Frequency Systems GmbH, Hannover

Schott AG Mainz

ThyssenKrupp AG, Düsseldorf

VAC Vacuumschmelze, Hanau

Voestalpine AG Linz, Österreich

Literatur

1. *Boll, R.:* Weichmagnetische Werkstoffe, Vacuumschmelze, Hanau 1977
2. *Boll, R. u. a.:* Magnettechnik, expert Verlag Grafenau 1980
3. *Döring, E.:* Werkstoffkunde der Elektrotechnik, Verlag Vieweg, Wiesbaden 1988
4. *Domke, W.:* Werkstoffkunde und Werkstoffprüfung, Verlag Girardet, Essen 1989
5. *Fasching, G.*: Werkstoffe für die Elektrotechnik, Springer Verlag, Wien 2005
6. *Fischer, H.:* Werkstoffe der Elektrotechnik, Hanser Verlag, München 2007
7. *Guillery, P. ; Hezel, R.; Reppich, B.:* Werkstoffkunde für Elektroingenieure, Verlag Vieweg, Wiesbaden 1985
8. *Heck, C .:* Magnetische Werkstoffe und ihre technische Anwendung, Hüthig-Verlag, Heidelberg 1975
9. *Hofman, Spindler:* Werkstoffe in der Elektrotechnik, Hanser Fachbuchverlag
10. *Meinke, H.; Gundlach, F. W.:* Taschenbuch der Hochfrequenztechnik, Springer Verlag, Berlin, Heidelberg, New York 1992
11. *Ivers-Tiffee, E.; Münch, W. v.:* Werkstoffe der Elektrotechnik, B.G. Teubner Stuttgart 2007
12. *Müller, K.; Scherf, U. (Hgg):* Organic Light Emitting Devices, Verlag, Wiley-VCH 2006
13. *Michalowsky, L. u. a.:* Weichmagnetische Ferrite Expert Verlag 2006
14. *Michalowsky, L.; Schneider, J. (Hgg):* Magnettechnik, Verlag Vulkan 2006
15. *Münch, W. v.:* Elektrische und magnetische Eigenschaften der Materie B. G. Teubner, Studienskripten, Stuttgart 1987
16. *Philippow, E. (Hrsg.):* Bauelemente und Bausteine der Informationstechnik, Verlag Technik, Berlin 1989
17. *Seitz, D.; Ross, G.; Cassing, W.; Kuntze, K.:* Dauermagnete, Expert Verlag 2006
18. *Spickermann, D.:* Werkstoffe der Elektrotechnik und Elektronik, Verlag Schlembach 2002
19. *Zinke, O. ; Seither, H.:* Widerstände, Kondensatoren, Spulen und ihre Werkstoffe, Springer Verlag, Berlin, Heidelberg, New York 1982

Grundlagen der Elektrotechnik

IV



25 Grundbegriffe 305
26 Der Gleichstromkreis 311
27 Das Elektrische Feld 323
28 Das Magnetische Feld 335
29 Induktion 353
30 Wechselstrom 361
31 Drehstrom 383

Grundbegriffe

25

Reinhard von Liebenstein



25.1 Aufbau der Atome 306
25.2 Ladungsträger 306
25.3 Spannung 306
25.4 Strom 306
25.5 Das Ohm'sche Gesetz 307
25.6 Spezifischer Widerstand, Leitfähigkeit 307
25.7 Temperaturabhängigkeit des elektrischen Widerstandes von Metallen 308

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_25

25.1 Aufbau der Atome

Die Eigenschaften aller Elemente lassen sich durch den Aufbau der Atome erklären. Die Atome bestehen aus einem Kern und einer Hülle. Der Atomkern besitzt einen Durchmesser von ca. 10^{-14} m, die Ausdehnung der Atomhülle hängt von der Ordnungszahl im Periodensystem ab und beträgt einige 10^{-10} m. Im positiv geladenen Atomkern befindet sich nahezu die gesamte Masse des Atoms. Er besteht aus Neutronen und Protonen. Die Hülle ist negativ geladen und wird durch Elektronen gebildet. Im neutralen Atom sind die Anzahl der Elektronen in der Hülle und die Protonenanzahl im Kern gleich. Die Elektronen in der Hülle können sich nur in bestimmten Schalen, die durch ihre Energie unterschieden sind, bewegen. In jeder Schale ist nur eine bestimmte Anzahl von Elektronen maximal möglich. Die Elektronen der in der Regel nicht vollständig besetzten äußeren Schalen sind die *Valenzelektronen*. Jedes Elektron besitzt die negative *Elementarladung*

$$e_0 = 1{,}60218 \cdot 10^{-19}\,\mathrm{A\,s} \tag{25.1}$$

Die Masse eines Elektrons ist wesentlich kleiner als die eines Protons oder Neutrons. Das Verhältnis von Elektronenmasse zu Protonenmasse ergibt sich durch

$$\frac{m_\mathrm{el}}{m_p} \approx \frac{1}{1836} \tag{25.2}$$

Wird von einem Atom ein Elektron entfernt, ist der Rest positiv geladen. Dieses geladene Atom wird *Ion* genannt. Bei Entfernung eines Elektrons entsteht ein einfach positiv geladenes Ion. Wird ein Elektron hinzugefügt, liegt ein einfach negativ geladenes Ion vor.

25.2 Ladungsträger

Ein Stoff ist nur dann elektrisch leitfähig, wenn er bewegliche Ladungsträger besitzt. Die Ladungsträger können negativ oder positiv sein. In Metallen liegen immer negative Ladungsträger, also Elektronen vor. Hier können sich die Valenzelektronen sehr leicht vom Atom trennen, sie werden dann *freie Elektronen* genannt. Auf den festen Plätzen im Kristallgitter bleiben positive Ionen. Da sie nicht beweglich sind, liefern sie keinen Beitrag zum Ladungstransport. In Halbleitern können aber auch *Defektelektronen* oder *Löcher* als Ladungsträger vorkommen. In Gasen und Flüssigkeiten sind auch die Ionen am Ladungstransport beteiligt. Hier ist dann mit dem Ladungstransport auch ein Massentransport verbunden.

25.3 Spannung

Werden elektrische Ladungen getrennt, muss hierzu Arbeit verrichtet werden. Die Größe der Arbeit ist proportional zur Ladungsmenge und kann durch die Gleichung

$$W = U \cdot Q \tag{25.3}$$

beschrieben werden. Die Größe U wird als Spannung bezeichnet.

$$\textit{Spannung} \quad U = \frac{W}{Q} \qquad \begin{array}{c|c|c} U & W & Q \\ \hline \mathrm{V} = \frac{\mathrm{J}}{\mathrm{A\,s}} & \mathrm{J} & \mathrm{A\,s} \end{array} \tag{25.4}$$

Die Einheit $1\,\mathrm{J/A\,s}$ wird als 1 V (Volt) bezeichnet. Durch diese Spannung wird auf eine Ladung eine Kraft ausgeübt. Die Spannung ist somit Ursache für eine Bewegung von Ladungsträgern.

25.4 Strom

25.4.1 Bewegung von Ladungsträgern

Da man in der Elektrotechnik hauptsächlich metallische Leiter einsetzt, wird die Bewegung der Ladungsträger in Metallen untersucht. Die Ladungsträger führen bei Temperaturen $> 0\,\mathrm{K}$ ständig statistisch verteilt unregelmäßige Bewegungen durch. Aus dieser Bewegung ergibt sich eine mittlere Geschwindigkeit.

$$\textit{mittlere Geschwindigkeit} \quad v_m = 0\,\frac{\mathrm{m}}{\mathrm{s}} \tag{25.5}$$

Wird dieser statistischen Bewegung eine einseitig gerichtete Bewegung überlagert, bewegen sich alle Elektronen zusätzlich mit einer *Strömungsgeschwindigkeit* oder auch *Driftgeschwindigkeit* v_d (Abb. 25.1).

Mit dieser Bewegung ist ein Ladungstransport verbunden. An einer bestimmten Stelle des Leiters können in der Zeiteinheit Δt alle die Elektronen durch den Querschnitt fließen und somit zum Ladungstransport beitragen, deren Abstand kleiner als l ist. Im Volumen lA befinden sich $n_\mathrm{el} lA$ Elektronen, wobei n_el die Dichte der Elektronen (Anzahl pro Volumeneinheit) ist. Da jedes Elektron die Elementarladung besitzt, ist die in diesem Volumen vorhandene Ladung ΔQ.

$$\Delta Q = n_\mathrm{el} l A e_0 \tag{25.6}$$

und mit $l = v_0 \cdot \Delta t$.

$$\textit{Ladung pro Volumeneinheit} \quad \Delta Q = n_\mathrm{el} A e_0 v_d \Delta t \tag{25.7}$$

Die Elektronendichte ist eine Materialkonstante und beträgt z. B. für Kupfer $n_\mathrm{el} = 8{,}47 \cdot 10^{19}\,\mathrm{mm}^{-3}$.

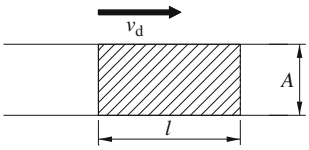


Abb. 25.1 Driftgeschwindigkeit

25.4.2 Stromstärke

Die Ladungsmenge ΔQ, die pro Zeiteinheit Δt durch einen Leiterquerschnitt tritt, ist die *Stromstärke* I.

$$\textit{Stromstärke} \quad I = \frac{\Delta Q}{\Delta t} \qquad \begin{array}{c|c|c} I & \Delta Q & \Delta t \\ \hline \mathrm{A} & \mathrm{A\,s} & \mathrm{s} \end{array} \qquad (25.8)$$

$$I = n_{\mathrm{el}} A e_0 v_d \qquad (25.9)$$

Im Internationalen Einheitensystem (SI-System) ist die Stromstärke als Basisgröße durch die Kraft zwischen zwei stromdurchflossenen Leitern definiert:

> 1 Ampere ist die Stärke eines zeitlich unveränderlichen Stromes, der, durch zwei im Vakuum parallel im Abstand von 1 Meter voneinander angeordnete, geradlinige, unendlich lange Leiter von vernachlässigbar kleinem kreisförmigen Querschnitt fließend, zwischen diesen Leitern je 1 Meter Leiterlänge die Kraft $2 \cdot 10^{-7}$ Newton hervorruft.

Beispiel

Ein Kupferleiter von $A = 1{,}5\,\mathrm{mm}^2$ wird von einem Strom $I = 15\,\mathrm{A}$ durchflossen. Wie groß ist die Driftgeschwindigkeit der Elektronen?

$$\begin{aligned} v_d &= \frac{I}{n_{\mathrm{el}} A e_0} \\ &= \frac{15\,\mathrm{A}}{8{,}47 \cdot 10^{19}\,\mathrm{mm^{-3}} \cdot 1{,}5\,\mathrm{mm^2} \cdot 1{,}602 \cdot 10^{-19}\,\mathrm{A\,s}} \\ &= 0{,}74\,\frac{\mathrm{mm}}{\mathrm{s}} \end{aligned}$$

Diese Driftgeschwindigkeit ist relativ gering. Dagegen breitet sich ein elektrisches Signal mit Lichtgeschwindigkeit aus. ◀

25.4.3 Stromdichte

In der Technik wichtig für die Belastbarkeit von Leitungen ist die *Stromdichte* S, also die auf den Querschnitt bezogene Stromstärke.

$$\textit{Stromdichte} \quad S = \frac{I}{A} \qquad \begin{array}{c|c|c} S & I & A \\ \hline \frac{\mathrm{A}}{\mathrm{mm^2}} & \mathrm{A} & \mathrm{mm^2} \end{array} \qquad (25.10)$$

Da die Querschnittsfläche in der Regel in mm^2 angegeben wird, ist die in der Technik gebräuchliche Einheit der Stromdichte $[S] = 1\,\frac{\mathrm{A}}{\mathrm{mm^2}}$.

25.5 Das Ohm'sche Gesetz

Wird an einen Leiter in einem geschlossenen Stromkreis eine Spannung U angelegt, fließt in dem Leiter ein Strom I. Strom und Spannung sind einander proportional. Die Abhängigkeit lässt sich beschreiben durch:

$$\textit{Ohm'sches Gesetz} \quad U = R \cdot I \qquad (25.11)$$

Die materialabhängige Größe R bezeichnet man als *elektrischen*

$$\textit{Widerstand} \quad R = \frac{U}{I} \qquad \begin{array}{c|c|c} R & U & I \\ \hline \Omega = \frac{\mathrm{V}}{\mathrm{A}} & \mathrm{V} & \mathrm{A} \end{array} \qquad (25.12)$$

Die Einheit ist 1 *Ohm* $= 1\,\Omega$.

Der Kehrwert des elektrischen Widerstandes ist der Leitwert.

$$\textit{Leitwert} \quad G = \frac{1}{R} \qquad \begin{array}{c|c} G & R \\ \hline \mathrm{S} = \frac{1}{\Omega} & \Omega \end{array} \qquad (25.13)$$

Seine Einheit ist 1 *Siemens* $= 1\,\mathrm{S}$.

25.6 Spezifischer Widerstand, Leitfähigkeit

Der elektrische Widerstand eines metallischen Leiters hängt vom Material, der Länge l und dem Querschnitt A ab. Er wird dadurch verursacht, dass die Ladungsträger im Leiter durch Stöße mit den auf festen Plätzen sitzenden Ionen in ihrer Bewegung gehemmt werden.

$$\textit{Widerstand eines Drahtes} \quad R = \frac{\rho l}{A} = \frac{l}{\varkappa A} \qquad (25.14)$$

Die Größe ρ ist der spezifische Widerstand.

spezifischer Widerstand

$$\rho = \frac{RA}{l} \qquad \begin{array}{c|c|c|c} \rho & R & A & l \\ \hline \frac{\Omega\,\mathrm{mm^2}}{m} & \Omega & \mathrm{mm^2} & \mathrm{m} \end{array} \qquad (25.15)$$

Die Einheit von ρ kann umgerechnet werden in

$$[\rho] = 1\,\frac{\Omega\,\mathrm{mm^2}}{\mathrm{m}} = 10^{-6}\,\Omega\mathrm{m} \qquad (25.16)$$

Analog zum Leitwert ist der Kehrwert des spezifischen Widerstandes die *Leitfähigkeit* $\varkappa$

$$\textit{Leitfähigkeit} \quad \varkappa = \frac{1}{\rho} \qquad \begin{array}{c|c} \varkappa & \rho \\ \hline \frac{1}{\Omega\,\mathrm{m}} & \Omega\,\mathrm{m} \end{array} \qquad (25.17)$$

In Tab. 25.1 sind Zahlenwerte des spezifischen Widerstandes und der Leitfähigkeit aufgeführt. Der Wert für Kupfer kann je nach Reinheit schwanken. Der hier angegebene Wert ist der internationale Standardwert.

Tab. 25.1 Spezifischer Widerstand und Leitfähigkeit

Leiterwerkstoff	ρ [W mm^2/m]	$\varkappa$ [S/m]
Silber	0,0163	$61 \cdot 10^6$
Kupfer	0,0173	$58 \cdot 10^6$
Aluminium	0,0263	$38 \cdot 10^6$
Messing	0,0752	$13{,}3 \cdot 10^6$
Manganin	0,435	$2{,}3 \cdot 10^6$

Tab. 25.2 Temperaturkoeffizient a_{20}

Leiterwerkstoff	a_{20} [K^{-1}]
Silber	$3{,}7 \cdot 10^{-3}$
Kupfer	$3{,}9 \cdot 10^{-3}$
Aluminium	$3{,}7 \cdot 10^{-3}$
Messing	$1{,}6 \cdot 10^{-3}$
Manganin	$0{,}01 \cdot 10^{-3}$

Beispiel

Der Widerstand eines Kupferdrahtes der Länge 160 m bei einem Querschnitt von 1,5 mm^2 soll berechnet werden.

$$R = \frac{\rho l}{A} = \frac{0{,}0173\,\Omega\,\text{mm}^2 \cdot 160\,\text{m}}{\text{m} \cdot 1{,}5\,\text{mm}^2} = 1{,}845\,\Omega \quad \blacktriangleleft$$

Beispiel

Aus Manganindraht von $A = 1\,\text{mm}^2$ soll ein Widerstand von 60 Ω gewickelt werden. Wie lang muss der Draht sein?

$$l = RA\varkappa = 60\,\Omega \cdot 10^{-6}\,\text{m}^2 \cdot 2{,}3 \cdot 10^6\,\frac{\text{S}}{\text{m}} = 138\,\text{m} \quad \blacktriangleleft$$

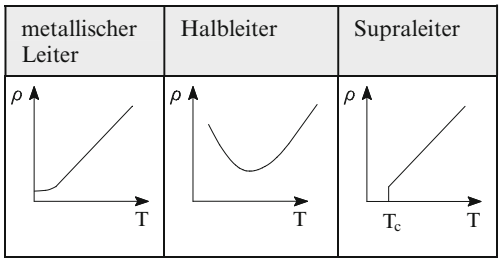


Abb. 25.2 Prinzipieller Verlauf des spezifischen Widerstandes

25.7 Temperaturabhängigkeit des elektrischen Widerstandes von Metallen

Die elektrische Leitung in Festkörpern ist mit der Bewegung von Ladungsträgern verbunden. Diese stoßen mit den auf festen Plätzen sitzenden Atomen oder Ionen zusammen. Die Schwingung der Atome und Ionen um ihren Platz ist temperaturabhängig. Deshalb ist auch der elektrische Widerstand temperaturabhängig. Bei höheren Temperaturen ist die freie Weglänge und die Beweglichkeit der Ladungsträger kleiner und somit der Widerstand größer. In Metallen ist die Anzahl der freien Ladungsträger temperaturunabhängig, in Halbleitern nicht. Daher haben Metalle und Halbleiter unterschiedliche Temperaturabhängigkeiten.

Der spezifische elektrische Widerstand von Metallen und damit auch der Widerstand steigt mit zunehmender Temperatur. Im Temperaturbereich von ca. 200 bis 400 K lässt sich diese Abhängigkeit bei Metallen in linearer Näherung durch eine Gerade darstellen.

lineare Näherung
$$R_\vartheta = R_{\vartheta_0}[1 + \alpha(\vartheta - \vartheta_0)] \qquad (25.18)$$

Dabei ist R_ϑ der Widerstand bei der Temperatur ϑ und R_{ϑ_0} der Widerstand bei der Bezugstemperatur ϑ_0. Die Größe a ist der *Temperaturkoeffizient*. Da (25.18) eine Näherung ist, hängt der Wert für a von der Bezugstemperatur ab. Als Bezugstemperatur wird meistens 20 °C genommen.

Beispiel

Bei 10 °C ist der Widerstand eines Kupferdrahtes 200 W. Wie groß ist er bei 100 °C?

Zunächst muss der Widerstand $R_{\vartheta 0}$ bestimmt werden.

$$R_{\vartheta_0} = \frac{R_\vartheta}{1 + \alpha(\vartheta - 20°)} = \frac{200\,\Omega}{1 + 3{,}9 \cdot 10^{-3}\,\text{K}^{-1}(10 - 20)\,\text{K}} = 208{,}1\,\Omega$$

Dann kann hiermit R_ϑ berechnet werden.

$$R_\vartheta = R_{\vartheta_0}[1 + \alpha(\vartheta - \vartheta_0)] = 208{,}1\,\Omega[1 + 3{,}9 \cdot 10^{-3}(100 - 20)] = 273\,\Omega$$

Die lineare Näherung ist recht grob. Genauer kann die Temperaturabhängigkeit durch eine quadratische Näherung angepasst werden. ◀

Quadratische Näherung

$$R_\vartheta = R_0(1 + A(\vartheta - \vartheta_0) + B(\vartheta - \vartheta_0)^2) \qquad (25.19)$$

A und B sind Materialkonstanten. Als Bezugstemperatur wird $\vartheta_0 = 0\,°\text{C}$ angenommen. R_0 ist dann der Widerstand bei 0 °C. Ein oft benutzter Metallwiderstand zur Temperaturmessung ist der Platinwiderstand Pt100, der 100 Ω bei 0 °C hat. Für Pt sind die Konstanten $A = 3{,}90802 \cdot 10^{-3}\,\text{K}^{-1}$, $B = -0{,}5802 \cdot 10^{-6}\,\text{K}^{-2}$.

Bei Halbleitern fällt der spezifische Widerstand mit steigender Temperatur zunächst, da hier mehr Ladungsträger frei gesetzt werden. Anschließend steigt der Widerstand wie bei Metallen, da die Beweglichkeit verringert wird. Supraleiter zeigen unterhalb einer charakteristischen Temperatur, der Sprungtemperatur T_c, überhaupt keinen messbaren Widerstand mehr. Werkstoffe, bei denen der Widerstand mit steigender Temperatur zunimmt, haben einen *positiven Temperaturkoeffizienten* (PTC). Nimmt der Widerstand ab, besitzen sie einen *negativen Temperaturkoeffizienten* (NTC), (Abb. 25.2).

Der Gleichstromkreis

26

Reinhard von Liebenstein



26.1 Zählpfeilsysteme 312
26.2 Kirchhoff'sche Gesetze 312
26.3 Schaltung von Widerständen 313
26.4 Reale Spannungsquelle 315
26.5 Ersatzstromquelle 316
26.6 Netzwerkberechnung 317
26.7 Energie, Leistung, Wirkungsgrad 320

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_26

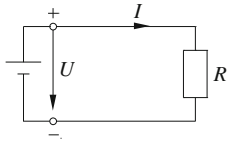


Abb. 26.1 Zählpfeilsystem

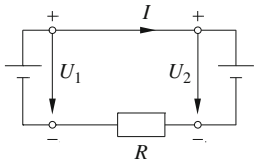


Abb. 26.2 Schaltung

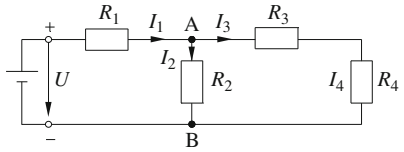


Abb. 26.3 Knotenregel

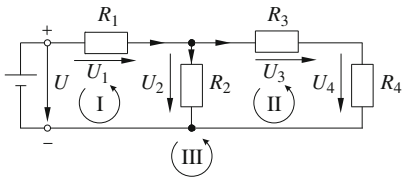


Abb. 26.4 Maschenregel

26.1 Zählpfeilsysteme

Damit in einem Leiter ein elektrischer Strom fließen kann, muss ein geschlossener Stromkreis vorliegen. In elektrischen Schaltungen werden die Spannungs- und Stromrichtungen allgemein durch Pfeile gekennzeichnet. Die Richtung der Spannung wird in Schaltbildern durch einen Pfeil dargestellt, der von + nach − verläuft. Die Stromrichtung wird ebenfalls durch einen Pfeil angegeben, der beim Verbraucher von + nach − gezeichnet wird. Diese Stromrichtung wird als *technische Stromrichtung* bezeichnet; die tatsächliche Bewegung der Elektronen ist dieser Richtung entgegengesetzt. In der linken Seite von Abb. 26.1 sind Spannungs- und Strompfeil entgegengesetzt, man spricht hier vom *Erzeuger-Zählpfeilsystem*. Auf der Verbraucherseite sind Spannungs- und Strompfeil gleichgerichtet; hier liegt das *Verbraucher-Zählpfeilsystem* vor. In den meisten Fällen wird die Pfeilzuordnung im Verbraucher- Zählpfeilsystem vorgenommen.

Liegt nun eine Schaltung nach Abb. 26.2 vor, so hängt die tatsächliche Stromrichtung von der Höhe der Spannungen U_1 und U_2 ab. In diesem Fall gibt man für den Strom willkürlich eine Richtung vor und bezeichnet diese als Bezugsrichtung oder Zählrichtung. Fließt der technische Strom tatsächlich in der willkürlich gewählten Richtung, wird er positiv gezählt, im anderen Fall negativ.

26.2 Kirchhoff'sche Gesetze

Eine elektrische Schaltung besteht in der Regel aus mehreren Spannungsquellen und Widerständen, die sehr unterschiedlich miteinander verbunden sein können. Das gesamte System wird auch als *Netzwerk* bezeichnet. Für die Berechnung von Strömen und Spannungen in einem solchen Netzwerk sind die beiden *Kirchhoff'schen Gesetze*, die auch als *Knotenregel* und *Maschenregel* bezeichnet werden, von großer Bedeutung.

26.2.1 Knotenregel

Knotenpunkte sind Punkte, an denen sich der Stromkreis verzweigt. Das *1. Kirchhoff'sche Gesetz* bezieht sich auf diese Punkte, z. B. im Abb. 26.3 auf die Punkte A und B. In diesen Knotenpunkten muss die Summe der zufließenden Ströme gleich der Summe der abfließenden Ströme sein. Im Punkt A gilt

$$I_1 = I_2 + I_3 \tag{26.1}$$

$$I_1 - I_2 - I_3 = 0 \tag{26.2}$$

Werden alle in einen Knoten hineinfließenden Ströme mit positivem, alle abfließenden Ströme mit negativem Vorzeichen versehen, lässt sich die Knotenregel aufstellen.

$$\textit{Knotenregel} \quad \sum_{k=1}^{n} I_k = 0 \tag{26.3}$$

Diese Regel wird auch das *1. Kirchhoff'sche Gesetz* genannt.

26.2.2 Maschenregel

Das *2. Kirchhoff'sche Gesetz* bezieht sich auf *Maschen* von Netzwerken. Maschen sind geschlossene Teile von Netzwerken. In der Schaltung nach Abb. 26.4 gibt es drei Maschen. Masche I wird durch die Spannungsquelle, R_1 und R_2 gebildet, Masche II durch R_3, R_4 und R_2 und Masche III durch die Spannungsquelle, R_1, R_3 und R_4. Die *Maschenregel* ist eine Gleichung für die Spannungen innerhalb einer Masche. Bei einem frei wählbaren Umlaufsinn, hier gegen den Uhrzeigersinn, werden die Spannungen, die dem Umlaufpfeil entgegen gerichtet sind, negativ gezählt. Die Richtungen der Spannungspfeile an den Spannungsquellen ist von + nach −, an den einzelnen Widerständen werden die Spannungspfeile in den entsprechenden Stromrichtungen gezeichnet.

In der dargestellten Schaltung lassen sich z. B. folgende Maschenregeln aufstellen (Abb. 26.4):

$$\text{I)} \quad U - U_2 - U_1 = 0 \tag{26.4}$$

$$\text{II)} \quad U_2 - U_4 - U_3 = 0 \tag{26.5}$$

$$\text{III)} \quad U - U_4 - U_3 - U_1 = 0 \tag{26.6}$$

Innerhalb einer Masche gilt immer – unter Berücksichtigung der Vorzeichen – die Maschenregel.

$$\textit{Maschenregel} \quad \sum_{k=1}^{n} U_k = 0 \tag{26.7}$$

26.3 Schaltung von Widerständen

Für die Kombination von Widerständen gibt es im Prinzip zwei Möglichkeiten, *Reihen-* oder *Serienschaltung* und *Parallelschaltung*.

26.3.1 Reihenschaltung

In einer Reihenschaltung fließt durch jeden Widerstand derselbe Strom I. Für die Spannungsfälle an den einzelnen Widerständen gilt nach dem Ohm'schen Gesetz $U_k = I \cdot R_k$. Mit (26.7) gilt somit (Abb. 26.5)

$$U - I \cdot R_1 - I \cdot R_2 - I \cdot R_3 = 0 \tag{26.8}$$

$$U = I \cdot (R_1 + R_2 + R_3) = I \cdot R_{\text{ges}} \tag{26.9}$$

$$\textit{Reihen-Ersatzwiderstand} \quad R_{\text{ges}} = \sum_{k=1}^{n} R_k \tag{26.10}$$

Weiterhin folgt für die Spannungsfälle an den einzelnen Widerständen

$$U_1 : U_2 : U_3 = R_1 : R_2 : R_3 \tag{26.11}$$

Hieraus resultiert die *Spannungsteilerregel*. Sie besagt, dass sich die Teilspannungen zur Gesamtspannung verhalten, wie die Teilwiderstände zum Gesamtwiderstand.

$$\textit{Spannungsteiler} \quad \frac{U_k}{U} = \frac{R_k}{R_{\text{ges}}} \tag{26.12}$$

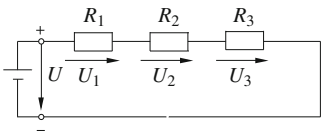


Abb. 26.5 Reihenschaltung

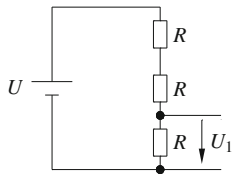


Abb. 26.6 Beispiel Spannungsteiler

Beispiel

In der Schaltung nach Abb. 26.6 soll die Teilspannung U_1 berechnet werden. Die Spannung U ist 300 V.

$$\frac{U_1}{U} = \frac{R}{3R} = \frac{1}{3}; \quad U_1 = 100\,\text{V} \qquad \blacktriangleleft$$

26.3.2 Parallelschaltung

In einer Parallelschaltung von Widerständen (Abb. 26.7) liegt an jedem Widerstand die gleiche Spannung U. Somit gilt mit (26.3) und

$$I_k = \frac{U}{R_k}; \quad \frac{U}{R} = \sum_{k=1}^{n} \frac{U}{R_k}$$

$$\textit{Parallel- Ersatzwiderstand} \quad \frac{1}{R} = \sum_{k=1}^{n} \frac{1}{R_k} \tag{26.13}$$

Der gesamte Leitwert errechnet sich als Summe der einzelnen Leitwerte.

$$\textit{Leitwert} \quad G = \sum_{k=1}^{n} G_k \tag{26.14}$$

Oft besteht eine Parallelschaltung nur aus zwei Widerständen. Dann errechnet sich der Ersatzwiderstand aus

$$\textit{Parallelwiderstand} \quad R = \frac{R_1 \cdot R_2}{R_1 + R_2} \tag{26.15}$$

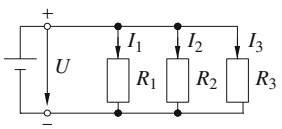


Abb. 26.7 Parallelschaltung

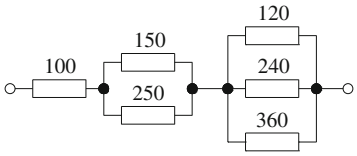


Abb. 26.8 Beispiel Ersatzwiderstand

Beispiel

Berechnung des Ersatzwiderstandes der Schaltung nach Abb. 26.8.

Die Widerstände werden zu Gruppen zusammengefasst.

$$R_{\mathrm{I}} = \frac{150\,\Omega \cdot 250\,\Omega}{150\,\Omega + 250\,\Omega} = 93{,}75\,\Omega$$

$$\frac{1}{R_{\mathrm{II}}} = \frac{1}{120\,\Omega} + \frac{1}{240\,\Omega} + \frac{1}{360\,\Omega} \qquad R_{\mathrm{II}} = 65{,}45\,\Omega$$

$$R = (100 + 93{,}75 + 65{,}45)\,\Omega = 259{,}2\,\Omega \quad \blacktriangleleft$$

26.3.3 Stern-Dreieck Umwandlung

Nicht immer lassen sich Widerstände zu Gruppen zusammenfassen. Als Beispiel wird die Schaltung nach Abb. 26.9 betrachtet. Zu einer Lösung gelangt man, wenn eine *Stern-Dreieckumwandlung* oder eine *Dreieck-Sternumwandlung* durchgeführt wird. Hierunter versteht man eine rechnerische Umwandlung einer Schaltung in eine elektrisch gleichwertige Schaltung. Die Schaltungen sind dann gleichwertig, wenn zwischen entsprechenden Punkten in den beiden Schaltungen gleiche Widerstände oder Leitwerte vorhanden sind. Wenn zwischen den Punkten 1 und 2 im Stern und im Dreieck derselbe Widerstand liegen soll, muss gelten:

$$R_1 + R_2 = \frac{R_{12}(R_{23} + R_{31})}{R_{12} + R_{23} + R_{31}} \tag{26.16}$$

$$R_2 + R_3 = \frac{R_{23}(R_{31} + R_{12})}{R_{12} + R_{23} + R_{31}} \tag{26.17}$$

$$R_3 + R_1 = \frac{R_{31}(R_{12} + R_{23})}{R_{12} + R_{23} + R_{31}} \tag{26.18}$$

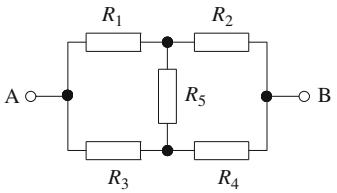


Abb. 26.9 Brückenschaltung

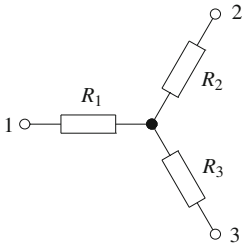


Abb. 26.10 Stern

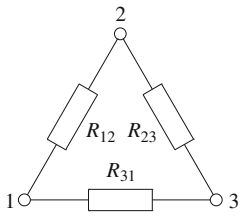


Abb. 26.11 Dreieck

Aus diesen drei Gleichungen lassen sich Bestimmungsgleichungen für die Umwandlungen ableiten. Sie lauten:

Dreieck in Stern:

$$R_1 = \frac{R_{12} \cdot R_{31}}{R_{12} + R_{23} + R_{31}} \tag{26.19}$$

$$R_2 = \frac{R_{23} \cdot R_{12}}{R_{12} + R_{23} + R_{31}} \tag{26.20}$$

$$R_3 = \frac{R_{31} \cdot R_{23}}{R_{12} + R_{23} + R_{31}} \tag{26.21}$$

Stern in Dreieck:

$$R_{12} = R_1 + R_2 + \frac{R_1 \cdot R_2}{R_3} \tag{26.22}$$

$$R_{23} = R_2 + R_3 + \frac{R_2 \cdot R_3}{R_1} \tag{26.23}$$

$$R_{31} = R_3 + R_1 + \frac{R_3 \cdot R_1}{R_2} \tag{26.24}$$

Beispiel

In der Schaltung Abb. 26.12 soll der Gesamtwiderstand zwischen den Punkten A und B berechnet werden. $R_1 = 10\,\Omega$, $R_2 = 15\,\Omega$, $R_3 = 25\,\Omega$, $R_4 = 5\,\Omega$, $R_5 = 15\,\Omega$.

Das Dreieck aus R_1, R_3 und R_5 wird in einen Stern umgerechnet.

$$R_A = \frac{R_1 R_3}{R_1 + R_3 + R_5} = \frac{10 \cdot 25}{10 + 25 + 15}\,\Omega = 5\,\Omega$$

$$R_C = \frac{R_1 R_5}{R_1 + R_3 + R_5} = \frac{10 \cdot 15}{10 + 25 + 15}\,\Omega = 3\,\Omega$$

$$R_D = \frac{R_3 R_5}{R_1 + R_3 + R_5} = \frac{25 \cdot 15}{10 + 25 + 15}\,\Omega = 7{,}5\,\Omega$$

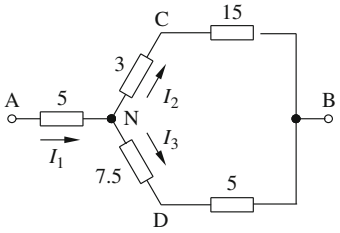


Abb. 26.12 Beispiel Stern-Dreieck-Umwandlung

Die Ersatzschaltung ergibt dann die in Abb. 26.9 dargestellte Schaltung. Der Gesamtwiderstand berechnet sich aus

$$R_{AB} = \left(5 + \frac{18 \cdot 12{,}5}{18 + 12{,}5}\right)\Omega = 12{,}38\,\Omega \qquad (26.25)$$

◀

26.3.4 Messbereichserweiterung

Durch geeignet gewählte Widerstände lassen sich die Messbereiche von Strom- und Spannungsgeräten anpassen. R_M ist jeweils der Innenwiderstand des verwendeten Messgerätes.

26.3.4.1 Voltmeter

Um den Messbereich eines Voltmeters zu erweitern, ist nach Abb. 26.13 ein Vorwiderstand zu schalten. Wenn mit U_1 der neu einzustellende Messbereich und mit U_M der vorhandene Messbereich des Gerätes bezeichnet wird, kann mit Hilfe der Maschen-regel der erforderliche Vorwiderstand berechnet werden.

$$U_1 = U_M + U_V \qquad (26.26)$$

$$\frac{U_1}{U_M} = 1 + \frac{U_V}{U_M} = 1 + \frac{R_V}{R_M} \qquad (26.27)$$

Messbereichsverhältnis $$n = \frac{U_1}{U_M} \qquad (26.28)$$

Vorwiderstand $$R_V = R_M(n-1) \qquad (26.29)$$

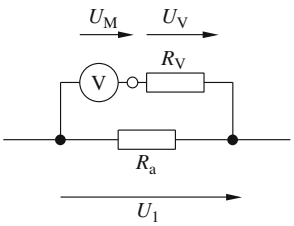


Abb. 26.13 Erweiterung des Spannungs-Messbereichs

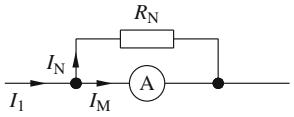


Abb. 26.14 Erweiterung des Strom-Messbereichs

26.3.4.2 Amperemeter

Für die Messbereichserweiterung eines Strommessgerätes nach Abb. 26.14 wird ein Widerstand parallel zum Messgerät geschaltet. Um die Größe dieses Nebenwiderstandes zu berechnen, muss man von der Knotenregel ausgehen.

$$I_1 = I_M + I_N \qquad (26.30)$$

$$\frac{I_1}{I_M} = 1 + \frac{I_N}{I_M} = 1 + \frac{R_M}{R_N} \qquad (26.31)$$

Messbereichsverhältnis $$n = \frac{I_1}{I_M} \qquad (26.32)$$

Nebenwiderstand $$R_N = \frac{R_M}{n-1} \qquad (26.33)$$

I_1 ist der zu messende Strom, also der neue Messbereich, I_M der vorhandene Messbereich.

Beispiel

Ein Strommessgerät mit einem Innenwiderstand von 0,45 Ω, welches den Skalenendwert von 2 A hat, soll für einen Verbraucherstrom von 20 A eingesetzt werden.

a) Welcher Nebenwiderstand ist zu schalten?
b) Welche Stellung nimmt der Zeiger an, wenn statt des unter a) errechneten Nebenwiderstandes ein Widerstand von 0,025 Ω genommen wird.

Lösung:

a)
$$n = \frac{20\,\text{A}}{2\,\text{A}} = 10; \quad R_N = \frac{0{,}45\,\Omega}{9} = 0{,}05\,\Omega$$

b)
$$n = 1 + \frac{R_M}{R_N} = 1 + \frac{0{,}45\,\Omega}{0{,}025\,\Omega} = 19;$$
$$I_1 = 19 \cdot 2\,\text{A} = 38\,\text{A}$$
$$\frac{20\,\text{A}}{38\,\text{A}} \cdot 100\,\% = 52{,}6\,\% \quad \text{des Skalenendwertes}$$

◀

26.4 Reale Spannungsquelle

Wird eine reale Spannungsquelle durch einen Widerstand belastet, sinkt die zwischen den Polen (Klemmen) vorhandene Spannung U_K, die als *Klemmspannung* bezeichnet wird. Ursache hierfür ist der immer vorhandene *innere Widerstand*

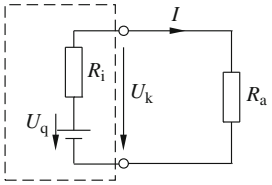


Abb. 26.15 Reale Spannungsquelle

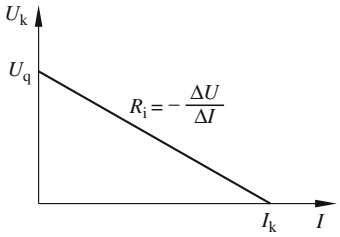


Abb. 26.16 Kennlinie einer Spannungsquelle

einer Spannungsquelle. Die Spannung U_q wird als *Quellspannung* oder *Leerlaufspannung* bezeichnet. Im Ersatzschaltbild Abb. 26.15 lässt sich ablesen

$$IR_i + IR_a - U_q = 0 \tag{26.34}$$

$$U_K = IR_a \tag{26.35}$$

$$\textit{Klemmspannung} \quad U_K = U_q - IR_i \tag{26.36}$$

Sind R_i und U_q unabhängig vom Belastungsstrom, liegt eine *lineare Quelle* vor, deren Kennlinie in Abb. 26.16 dargestellt ist. Wenn $I = 0$ ist, folgt aus (26.36) $U_K = U_q$. Ist dagegen der Widerstand $R_a = 0$ (Quelle kurzgeschlossen), nimmt der Strom seinen Maximalwert an.

$$\textit{Kurzschlussstrom} \quad I_K = \frac{U_q}{R_i} \tag{26.37}$$

Aus der Kennlinie lässt sich auch der Innenwiderstand R_i bestimmen.

$$\textit{Innenwiderstand} \quad R_i = -\frac{\Delta U}{\Delta I} \tag{26.38}$$

Eine Spannungsquelle mit einem Innenwiderstand $R_i = 0$ wird als *ideale Spannungsquelle* oder *Konstantspannungsquelle* bezeichnet.

26.4.1 Kombination von Spannungsquellen

Spannungsquellen können in Reihe oder parallel geschaltet werden. Von Interesse ist auch hier der resultierende Innenwiderstand und die Quellspannung der gesamten Anordnung.

26.4.1.1 Reihenschaltung

Die Kombination kann durch eine einzelne Quelle ersetzt werden, die durch *Quellspannung* und *Innenwiderstand* charakterisiert wird. Um diese Größen zu berechnen, nimmt man zunächst

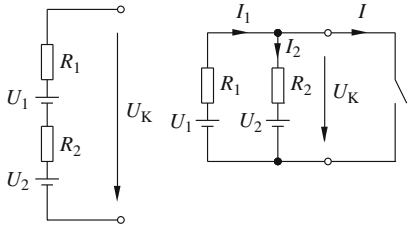


Abb. 26.17 Kombination von Spannungsquellen

an, dass die Quelle nicht belastet ist. Die Klemmspannung ist dann gleich der Quellspannung (Abb. 26.17).

$$\textit{Quellspannung} \quad U_q = U_1 + U_2 \tag{26.39}$$

$$\textit{Kurzschlussstrom} \quad I_k = \frac{U_1 + U_2}{R_1 + R_2} = \frac{U_q}{R_i} \tag{26.40}$$

$$\textit{Innenwiderstand} \quad R_i = R_1 + R_2 \tag{26.41}$$

26.4.1.2 Parallelschaltung

Um die Quellspannung und den Innenwiderstand zu bestimmen, wird zunächst die Quelle nicht belastet ($U_K = U_q$).

$$\textit{Ausgleichsstrom} \quad I_1 = I_2 = \frac{U_1 - U_2}{R_1 + R_2} \tag{26.42}$$

$$\textit{Quellspannung} \quad U_q = U_2 + I_2 R_2 \tag{26.43}$$

$$U_q = U_2 + R_2 \frac{U_1 - U_2}{R_1 + R_2} \tag{26.44}$$

$$U_q = \frac{U_1 R_2 + U_2 R_1}{R_1 + R_2} \tag{26.45}$$

Aus dem Kurzschlussstrom ($I = I_K$) lässt sich der Innenwiderstand berechnen.

$$\textit{Knoten} \quad I_1 = I_2 + I_K \tag{26.46}$$

$$\textit{rechte Masche} \quad I_2 R_2 + U_2 = 0 \tag{26.47}$$

$$\textit{linke Masche} \quad I_1 R_1 + I_2 R_2 + U_2 - U_1 = 0 \tag{26.48}$$

$$\textit{Kurzschlussstrom} \quad I_K = \frac{U_1}{R_1} + \frac{U_2}{R_2} \tag{26.49}$$

$$I_K = \frac{U_1 R_2 + U_2 R_1}{R_1 \cdot R_2} \tag{26.50}$$

$$\textit{Innenwiderstand} \quad R_i = \frac{U_q}{I_K} = \frac{R_1 R_2}{R_1 + R_2} \tag{26.51}$$

$$R_i = R_1 \text{ parallel } R_2 \tag{26.52}$$

26.5 Ersatzstromquelle

Jede Quelle kann auch entsprechend Abb. 26.18 durch eine Ersatzstromquelle mit einem Quellstrom I_q und einem parallel dazu liegenden Innenwiderstand betrachtet werden. Es gelten

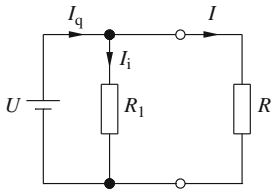


Abb. 26.18 Ersatzstromquelle

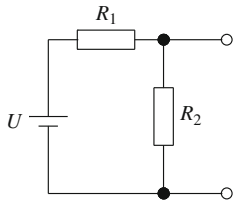


Abb. 26.19 Beispiel Ersatzquelle

die Beziehungen

Kurzschluss	$I = I_q$	(26.53)
	$U = 0$	(26.54)
	$I_i = 0$	(26.55)
Leerlauf	$I_i = I_q$	(26.56)
	$I = 0$	(26.57)
	$U = I_q R_i$	(26.58)

Im Beispiel Abb. 26.19 soll der aktive Zweipol einmal durch eine Ersatzspannungsquelle und einmal durch eine Ersatzstromquelle dargestellt werden.

Ersatzspannungsquelle	$U_q = U\dfrac{R_2}{R_1 + R_2}$	(26.59)
	$R_i = \dfrac{R_1 R_2}{R_1 + R_2}$	(26.60)
	$I_K = \dfrac{U}{R_1}$	(26.61)
Ersatzstromquelle	$I_q = \dfrac{U}{R_1} = I_K$	(26.62)
	$R_i = \dfrac{R_1 R_2}{R_1 + R_2}$	(26.63)

26.6 Netzwerkberechnung

Die Berechnung von Strömen, Spannungen oder Leistungen in einer beliebigen elektrischen Schaltung bezeichnet man als *Netzwerkberechnung*. Zur Berechnung der Ströme und Spannungen kann man im Prinzip immer Knoten- und Maschenregeln anwenden. Dies kann allerdings in komplizierten Netzwerken zu einer großen Zahl von Unbekannten und damit Gleichungen führen. Hier führen manchmal andere Verfahren schneller zum Ziel.

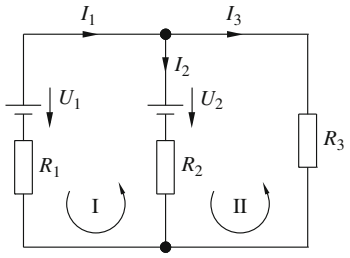


Abb. 26.20 Beispiel Netzwerk

26.6.1 Gemischte Schaltungen

Zunächst soll an Hand eines Beispiels die Berechnung mit Hilfe der Kirchhoff'schen Gesetze erfolgen. Die Spannungspfeile werden von + nach − gezeichnet. Die Strompfeile werden willkürlich angenommen. Sollte für einen der Ströme ein negativer Wert berechnet werden, ist der Pfeil falsch gewählt worden. Die Werte in der Schaltung Abb. 26.20 sind: $U_1 = 20\,\text{V}$, $U_2 = 30\,\text{V}$, $R_1 = R_2 = 100\,\Omega$, $R_3 = 50\,\Omega$.

Knotenregel	$I_1 - I_2 - I_3 = 0$
Masche I	$U_1 - I_1 R_1 - I_2 R_2 - U_2 = 0$
Masche II	$U_2 + I_2 R_2 - I_3 R_3 = 0$

Mit Zahlenwerten und nach Umstellung

Knotenregel	$I_1 - I_2 - I_3 = 0$
Masche I	$100\,\Omega \cdot I_1 + 100\,\Omega \cdot I_2 = -10\,\text{V}$
Masche II	$100\,\Omega \cdot I_2 - 50\,\Omega \cdot I_3 = -30\,\text{V}$

Durch Einsetzen kann dieses Gleichungssystem gelöst werden.

$$I_1 = 75\,\text{mA}; \quad I_2 = -175\,\text{mA}; \quad I_3 = 250\,\text{mA}$$

Aus dem negativen Wert für I_2 ist zu erkennen, dass die tatsächliche Stromrichtung der eingezeichneten Pfeilrichtung entgegengesetzt ist.

Beispiel

In der Schaltung nach Abb. 26.9 soll der Strom durch den Widerstand R_5 berechnet werden. Die angelegte Spannung ist 100 V.

Ausgehend vom Ergebnis in (26.25) ist der Gesamtstrom

$$I_{\text{ges}} = \frac{100\,\text{V}}{12{,}38\,\Omega} = 8{,}08\,\text{A}$$
$$U_{\text{AN}} = 8{,}08\,\text{A} \cdot 5\,\Omega = 40{,}4\,\text{V}$$
$$U_{\text{NB}} = 100\,\text{V} - 40{,}4\,\text{V} = 59{,}6\,\text{V}$$
$$I_2 = \frac{59{,}6\,\text{V}}{18\,\Omega} = 3{,}31\,\text{A}; \quad I_3 = \frac{59{,}6\,\text{V}}{12{,}5\,\Omega} = 4{,}77\,\text{A}$$

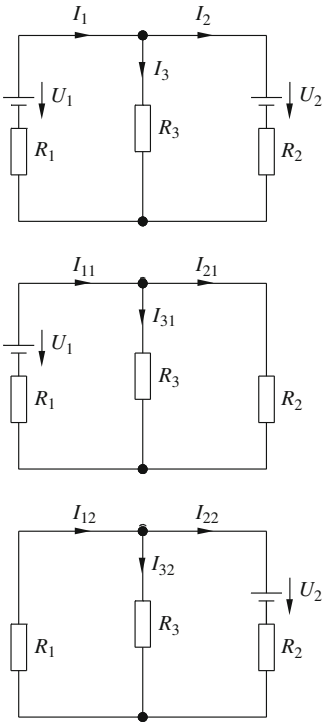


Abb. 26.21 Überlagerungsverfahren

Die Spannung über R_5 ist gleich der Spannung zwischen den Punkten C und D in Abb. 26.12:

$$U_{CD} = 4{,}77\,\text{A} \cdot 7{,}5\,\Omega - 3{,}31\,\text{A} \cdot 3\,\Omega = 25{,}85\,\text{V} \quad \blacktriangleleft$$

26.6.2 Überlagerungsverfahren

Ein weiteres Verfahren ist die Berechnung mit Hilfe des *Helmholtz'schen Überlagerungsverfahrens*. Bei der Berechnung von Zweigströmen in einem Netzwerk mit mehreren Spannungsquellen werden zunächst alle Spannungsquellen bis auf eine kurzgeschlossen. Für diesen Fall werden die Teilströme berechnet. Dann wird dasselbe Verfahren für eine weitere Spannungsquelle durchgeführt und wieder die Teilströme berechnet. Der gesamte Zweigstrom ergibt sich als Summe der Teilströme.

Das Verfahren soll an Hand der Schaltung in Abb. 26.21a dargestellt werden. Zunächst werden alle Spannungen bis auf U_1 kurzgeschlossen. Es entsteht die Ersatzschaltung nach Abb. 26.21b. Der Teilstrom I_{11} errechnet sich zu

$$I_{11} = \frac{U_1}{R_1 + \frac{R_2 \cdot R_3}{R_2 + R_3}} = U_1 \frac{R_2 + R_3}{R_1 R_2 + R_2 R_3 + R_3 R_1} \tag{26.64}$$

Werden im nächsten Schritt alle Spannungen bis auf U_2 kurzgeschlossen, entsteht die Ersatzschaltung Abb. 26.21c. Jetzt lässt

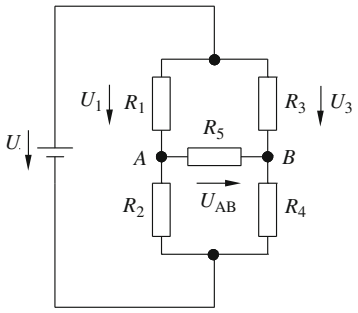


Abb. 26.22 Brücke

sich der Teilstrom I_{22} berechnen.

$$I_{22} = U_2 \frac{R_1 + R_3}{R_1 R_2 + R_2 R_3 + R_3 R_1} \tag{26.65}$$

und daraus weiter:

$$I_{12} = U_2 \frac{R_3}{R_1 R_2 + R_2 R_3 + R_3 R_1} \tag{26.66}$$

$$I_1 = I_{11} + I_{12} \tag{26.67}$$

Haben die Spannungsquellen einen nicht vernachlässigbaren Innenwiderstand, darf dieser nicht kurzgeschlossen werden.

26.6.3 Ersatzspannungsquelle

Oft ist in Netzwerken nur nach einer Teilspannung oder einem Teilstrom gefragt. In diesem Fall lässt sich die Methode der Ersatzspannungsquelle anwenden. Als Beispiel wird die in Abb. 26.22 dargestellte Brückenschaltung mit einer Last R_5 betrachtet. In der Messtechnik interessiert als Messgröße die Spannung U_{AB} zwischen den Punkten A und B bei einer Belastung durch ein Messgerät.

Bei nicht belasteter Brücke besteht zwischen den Punkten A und B die Spannung U_{AB0}. Diese Spannung wird als Quellspannung einer Ersatzspannungsquelle gesehen.

$$U_{AB0} = U_3 - U_1 \tag{26.68}$$

$$U_{AB0} = U \left(\frac{R_3}{R_3 + R_4} - \frac{R_1}{R_1 + R_2} \right) \tag{26.69}$$

Die Spannung U_{AB} ist dann als Klemmspannung einer Spannungsquelle mit der Quellspannung U_{AB0} anzusehen. Der Innenwiderstand dieser Ersatzspannungsquelle ist durch die Brückenwiderstände festgelegt. Er wird dadurch bestimmt, dass die äußere Spannungsquelle U kurzgeschlossen wird.

Innenwiderstand
$$R_i = \frac{R_1 R_2}{R_1 + R_2} + \frac{R_3 R_4}{R_3 + R_4} \tag{26.70}$$

Brückenspannung
$$U_{AB} = U_{AB0} \frac{R_5}{R_i + R_5} \tag{26.71}$$

Teil IV

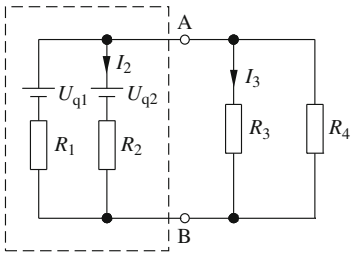


Abb. 26.23 Beispiel Methode der Ersatzspannungsquelle

Beispiel

In der Schaltung nach Abb. 26.22 gelten folgende Werte:

$$R_1, R_3, R_4 = 100\,\Omega, \quad R_2 = 200\,\Omega,$$
$$R_5 = 1\,\mathrm{k}\Omega, \quad U = 10\,\mathrm{V}.$$

Wie groß ist die Spannung U_{AB}?

$$U_{AB0} = 10\,\mathrm{V}\left(\frac{100}{200} - \frac{100}{300}\right) = 1{,}67\,\mathrm{V}$$
$$R_i = \frac{100 \cdot 200}{100 + 200}\,\Omega + \frac{100 \cdot 100}{100 + 100}\,\Omega = 116{,}67\,\Omega$$
$$U_{AB} = 1{,}67\,\mathrm{V}\frac{1000}{116{,}67 + 1000} = 1{,}49\,\mathrm{V}$$ ◀

Beispiel

Berechnung des Stromes I_3 durch den Widerstand R_3 (Abb. 26.23).

$$U_{q1} = 6\,\mathrm{V}, \quad U_{q2} = 4\,\mathrm{V}, \quad R_1 = 0{,}5\,\Omega,$$
$$R_2 = 0{,}8\,\Omega, \quad R_3 = R_4 = 6\,\Omega$$

Die im gestrichelten Rechteck liegenden Bauteile werden zu einer Ersatzspannungsquelle zusammengefasst. Die Quellspannung kann aus dem unbelasteten Fall bestimmt werden.

$$I_2 = \frac{U_{q1} - U_{q2}}{R_1 + R_2} = \frac{6\,\mathrm{V} - 4\,\mathrm{V}}{0{,}5\,\Omega + 0{,}8\,\Omega} = 1{,}54\,\mathrm{A}$$
$$U_q = U_{q2} + I_2 R_2 = 4\,\mathrm{V} + 1{,}54\,\mathrm{A} \cdot 0{,}8\,\Omega = 1{,}23\,\mathrm{V}$$
$$R_i = \frac{0{,}8 \cdot 0{,}5}{0{,}8 + 0{,}5}\,\Omega = 0{,}307\,\Omega$$

Die Klemmspannung sinkt bei Belastung mit

$$R_a = \frac{R_3 R_4}{R_3 + R_4} = 3\,\Omega \quad \text{auf}$$
$$U_{AB} = U_q \frac{R_a}{R_a + R_i} = 1{,}23\,\mathrm{V}\,\frac{3\,\Omega}{3\,\Omega + 0{,}307\,\Omega} = 1{,}11\,\mathrm{V}$$
$$I_3 = \frac{1{,}11\,\mathrm{V}}{3\,\Omega} = 0{,}37\,\mathrm{A}$$ ◀

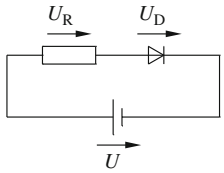


Abb. 26.24 Nichtlineare Kreise

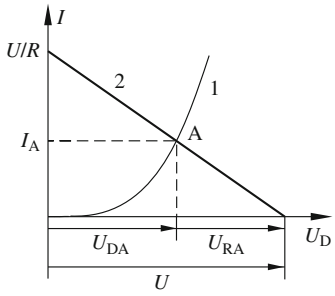


Abb. 26.25 Bestimmung des Arbeitspunktes

26.6.4 Nichtlineare Gleichstromkreise

Bisher wurde vorausgesetzt, dass die Widerstände in den Netzwerken nicht stromabhängig sind. Derartige Netzwerke werden als *lineare Netzwerke* bezeichnet. Ist der Widerstand eines Bauteils stromabhängig, wird diese Abhängigkeit normalerweise durch eine Kennlinie dargestellt. Die Bestimmung von Spannungen und Strömen im Kreis erfolgt meistens grafisch. Zur Erläuterung wird die Schaltung nach Abb. 26.24 betrachtet. In Abb. 26.25 sind auf der Ordinate der Strom und auf der Abszisse die Spannung U_D an der Diode aufgetragen. Die Kennlinie der Diode ist als Kurve 1 eingetragen. Weiter ist als Gerade 2 die Kennlinie des Ohm'schen Widerstandes eingetragen. Diese Gerade errechnet sich durch

$$I = \frac{U_R}{R} = \frac{U - U_D}{R} \tag{26.72}$$

Da durch Widerstand und Diode der gleiche Strom fließt, stellt sich als *Arbeitspunkt A* der Schnittpunkt beider Kennlinien ein. Es fließt der Strom I_A und an der Diode liegt die Spannung U_{DA}.

Beispiel

Die Diode in der Schaltung nach Abb. 26.24 hat eine Kennlinie gemäß Abb. 26.26. Der Widerstand R_1 ist 6 Ω, die Spannung U ist 1,8 V. Um wieviel Prozent ändert sich die Spannung U_R, wenn zu R_1 ein Widerstand R_2 von 4 Ω parallel geschaltet wird?

Die Kennlinie von R_1 ist die Gerade 2. Der Schnittpunkt mit der Spannungsachse liegt bei 1,8 V, der mit der Stromachse bei $I = \frac{U}{R_1} = \frac{1{,}8\,\mathrm{V}}{6\,\Omega} = 0{,}3\,\mathrm{A}$. Schnittpunkt mit Kennlinie 1: $U_D = o{,}9\,\mathrm{V}$;

$$U_R = U - U_D = 1{,}8\,\mathrm{V} - 0{,}9\,\mathrm{V} = 0{,}9\,\mathrm{V}.$$

Teil IV

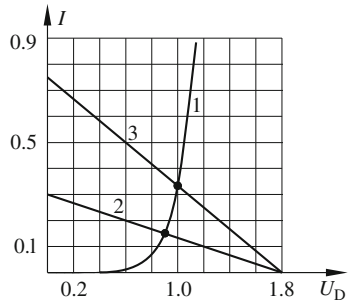


Abb. 26.26 Beispiel Nichtlinearer Kreis

Der Widerstand R ändert sich durch die Parallelschaltung auf den Wert $R = \frac{6 \cdot 4}{6+4}\,\Omega = 2{,}4\,\Omega$. Die zugehörige Kennlinie ist die Gerade 3, die bei $I = \frac{1{,}8\,\mathrm{V}}{2{,}4\,\Omega} = 0{,}75\,\mathrm{A}$ die Ordinate schneidet. Der Schnittpunkt mit der Kurve 1 liegt bei 1,0 V. Die Spannung am Widerstand ist jetzt $U_R' = 1{,}8\,\mathrm{V} - 1{,}0\,\mathrm{V} = 0{,}8\,\mathrm{V}$.

$$\frac{U_R' - U_R}{U_R} = \frac{0{,}8 - 0{,}9}{0{,}9}\,\mathrm{V} = -0{,}11$$

Die Spannung hat sich um 11 % verringert. ◄

26.7 Energie, Leistung, Wirkungsgrad

Fließt durch einen Widerstand R eine Ladung Q, wird die Arbeit W nach (25.3) bestimmt.

$$\textit{Arbeit} \quad W = U \cdot I \cdot t \qquad \begin{array}{c|c|c|c} W & U & I & t \\ \hline \mathrm{J = W\,s = V\,A\,s} & \mathrm{V} & \mathrm{A} & \mathrm{s} \end{array} \qquad (26.73)$$

Hieraus bestimmt sich die Leistung P.

$$\textit{Leistung} \quad P = \frac{W}{t} = U \cdot I \qquad \begin{array}{c|c|c} P & U & I \\ \hline \mathrm{W = V\,A} & \mathrm{V} & \mathrm{A} \end{array} \qquad (26.74)$$

Die Einheit der Leistung ist das *Watt*. Die Rückführung auf die SI-Grundeinheiten ergibt

$$1\,\mathrm{W} = 1\,\mathrm{V\,A} = 1\,\frac{\mathrm{J}}{\mathrm{s}} = 1\,\frac{\mathrm{N\,m}}{\mathrm{s}} = 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^3}} \qquad (26.75)$$

Für die Leistung lassen sich mit Hilfe des *Ohm'schen Gesetzes* nach (25.11) die folgenden Beziehungen ableiten:

$$P = \frac{U^2}{R} \qquad (26.76)$$

$$P = R \cdot I^2 \qquad (26.77)$$

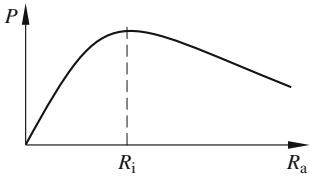


Abb. 26.27 Leistungsanpassung

26.7.1 Leistungsanpassung

Wird in einem Gleichstromkreis ein Ohm'scher Verbraucher an eine reale Spannungsquelle angeschlossen, ist die vom Verbraucher aufgenommene Leistung abhängig von dessen Widerstand. Betrachtet wird die Schaltung nach Abb. 26.15. U_q und R_i der Spannungsquelle sind bekannt. Die vom Verbraucher aufgenommene Leistung ist dann

$$P_V = U_K \cdot I = U_q \frac{R_a}{R_i + R_a} \cdot \frac{U_q}{R_i + R_a} \qquad (26.78)$$

$$P_V = U_q^2 \frac{R_a}{(R_i + R_a)^2} \qquad (26.79)$$

Für die Extremfälle $R_a = 0\,\Omega$ und $R_a = \infty$ verschwindet die aufgenommene Leistung. Dazwischen muss ein Maximum liegen. Die Lage dieses Maximums lässt sich mit Hilfe der Differentialrechnung bestimmen.

$$\frac{dP_V}{dR_a} = \frac{(R_i + R_a)^2 - 2R_a(R_i + R_a)}{(R_i + R_a)^4} \qquad (26.80)$$

Zur Bestimmung der Lage des Maximums muss dieser Ausdruck $= 0$ werden. Das ist dann der Fall, wenn der Zähler $= 0$ wird. Man erhält die Gleichung

$$(R_i + R_a)^2 - 2R_a(R_i + R_a) = 0 \qquad (26.81)$$

mit der Lösung

$$R_a = R_i \qquad (26.82)$$

In diesem Fall ist der Leistungsverlust am Innenwiderstand gleich der vom Verbraucher aufgenommenen Leistung. Bei großen Leistungen kann man keine Leistungsanpassung vornehmen; verwendet wird sie allerdings häufig in der Nachrichtentechnik.

Beispiel

In der Schaltung Abb. 26.28 sind $U_q = 10\,\mathrm{V}$ und $R_i = 0{,}5\,\Omega$.

a) Welchen Wert muss R_a haben, damit er eine Leistung von 10 W verbraucht?
b) Welche Leistung kann R_a maximal umsetzen?

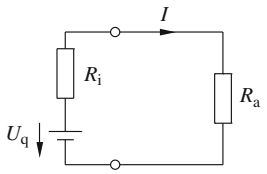


Abb. 26.28 Beispiel Leistungsanpassung

Lösung:

a)

$$P_a = I^2 \cdot R_a = \left(\frac{U_q}{R_i + R_a}\right)^2 \cdot R_a \tag{26.83}$$

$$P_a(R_i + R_a)^2 = U_q^2 \cdot R_a \tag{26.84}$$

$$R_a^2 + \left(2R_i - \frac{U_q^2}{P_a}\right) R_a + R_i^2 = 0 \tag{26.85}$$

$$R_a^2 - 9\,\Omega R_a + 0{,}25\,\Omega^2 = 0$$

$$R_{a1} = \left(4{,}5 + \sqrt{4{,}5^2 - 0{,}25}\right)\Omega = 8{,}97\,\Omega$$

$$R_{a2} = \left(4{,}5 - \sqrt{4{,}5^2 - 0{,}25}\right)\Omega = 0{,}28\,\Omega$$

b)

$$P_{a\max} = \left(\frac{U_q}{R_i + R_i}\right)^2 \cdot R_i = \left(\frac{10}{1}\right)^2 \cdot 0{,}5\,\mathrm{W} = 50\,\mathrm{W}$$

◀

26.7.2 Leistungsverlust auf Leitungen

Da jede Zuleitung zu einem Verbraucher einen Ohm'schen Widerstand R_L besitzt, wird immer elektrische Energie verbraucht. Der Widerstand der Zuleitung berechnet sich nach (25.14). Hierbei ist für l die gesamte Leiterlänge, also der doppelte Abstand zwischen Verbraucher und Spannungsquelle einzusetzen.

Beispiel

Ein Verbraucher ist 2,4 km von der Spannungsquelle entfernt. Die Leitung besteht aus Kupfer und hat einen Durchmesser von 8 mm. Wieviel Prozent der am Verbrauchsort zur Verfügung stehenden Leistung von 20 kW gehen verloren, wenn die Spannung beim Verbraucher 440 V beträgt ($\rho_{Cu} = 0{,}0173\,\frac{\Omega\,\mathrm{mm}^2}{\mathrm{m}}$)?

Leitungswiderstand

$$R_L = \frac{\rho \cdot 2l}{A} = \frac{0{,}0173 \cdot 4800 \cdot 4}{\pi \cdot 8^2}\,\Omega = 1{,}65\,\Omega$$

Strom

$$I = \frac{20\,\mathrm{kW}}{440\,\mathrm{V}} = 45{,}45\,\mathrm{A}$$

Verlust auf der Leitung

$$P_L = R_L \cdot I^2 = 1{,}65 \cdot 45{,}45^2\,\mathrm{W} = 3408\,\mathrm{W}$$

prozentualer Anteil

$$\frac{100 \cdot 3408}{20.000}\,\% = 17\,\%$$

◀

26.7.3 Wirkungsgrad

Der *Wirkungsgrad* η ist definiert als Verhältnis von zugeführter Energie W_{Zu} oder Leistung P_{Zu} zur Nutzenergie W_N oder Nutzleistung P_N. Da es bei allen Maschinen und Antrieben Verluste gibt, ist der Wirkungsgrad immer kleiner als 1 oder 100 %.

$$\textit{Wirkungsgrad} \quad \eta = \frac{W_N}{W_{Zu}} = \frac{P_N}{P_{Zu}} \leq 1 \tag{26.86}$$

Die Differenz zwischen zugeführter Leistung und Nutzleistung ist die Verlustleistung.

$$\textit{Verlustleistung} \quad P_V = P_{Zu} - P_N \tag{26.87}$$

26.7.4 Umwandlung elektrischer Energie

Elektrische Energie kann in andere Energieformen umgewandelt werden. Beispiele sind Umwandlung in Wärmeenergie (Kochplatte, Tauchsieder) oder mechanische Energie (Antriebe, Pumpen).

26.7.4.1 Wärme

Die für eine Temperaturerhöhung von ΔT oder ΔJ notwendige Wärmeenergie ΔQ ist gegeben durch

$$\Delta Q = mc\Delta T = mc\Delta\vartheta \tag{26.88}$$

Da die Kelvin-Skala und die Celsius-Skala dieselben Temperaturschritte haben, kann die Temperaturdifferenz sowohl in K als auch in °C angegeben werden. m ist die Masse und c die spezifische Wärmekapazität des zu erwärmenden Materials. Sie beträgt für Wasser

$$c_{\mathrm{Wasser}} = 4{,}19\,\frac{\mathrm{kJ}}{\mathrm{kg\,K}}$$

Beispiel

Wieviel Liter Wasser lassen sich mit einer elektrischen Energie von 15 kW h von $\vartheta_1 = 15\,°C$ auf $\vartheta_2 = 80\,°C$ erwärmen (1 Liter Wasser hat die Masse 1 kg)?

$$1\,\text{kW h} = 3600\,\text{kW s}$$

$$m = \frac{\Delta Q}{c \cdot (\vartheta_2 - \vartheta_1)}$$

$$= \frac{15 \cdot 3600\,\text{kJ}}{4{,}19\,\frac{\text{kJ}}{\text{kg K}} \cdot (80 - 15)\,\text{K}} = 198{,}3\,\text{kg} \quad \blacktriangleleft$$

Beispiel

Wie lange dauert es, um 5 l Wasser von $\vartheta_1 = 15\,°C$ mit einer Kochplatte, die eine Leistungsaufnahme von 800 W hat, zum Sieden zu bringen ($\vartheta_2 = 100\,°C$)? Der Wirkungsgrad beträgt 75 %.

$$P_N = \eta \cdot P_{Zu}; \quad P_N = \frac{\Delta Q}{t} = \frac{m \cdot c \cdot (\vartheta_2 - \vartheta_1)}{t}$$

$$t = \frac{m \cdot c \cdot (\vartheta_2 - \vartheta_1)}{P_{Zu} \cdot \eta}$$

$$t = \frac{5\,\text{kg} \cdot 4{,}19\,\frac{\text{kJ}}{\text{kg K}} \cdot (100 - 15)\,\text{K}}{0{,}8\,\frac{\text{kJ}}{\text{s}} \cdot 0{,}75}$$

$$= 2986\,\text{s} = 49{,}5\,\text{min} \quad \blacktriangleleft$$

26.7.4.2 Mechanische Energie

Elektrische Energie kann auch in potentielle oder kinetische Energie umgewandelt werden. Für diese Energieformen gelten die Gleichungen:

$$E_{pot} = mgh \quad \text{und} \quad E_{kin} = \frac{1}{2}mv^2$$

(siehe Abschn. 11.2.4).

Hierbei gilt natürlich ebenfalls der Energieerhaltungssatz.

Beispiel

Aus einer Grube sollen bei täglich 6 h Arbeitszeit innerhalb von 3 Tagen 12.000 m^3 Wasser über eine Förderhöhe von 10 m gepumpt werden. Welche Leistung muss der Motor aufnehmen, wenn ein Wirkungsgrad $\eta = 70\,\%$ vorhanden ist?

$$W_{pot} = mgh;$$

$$P_{Zu} = \frac{mgh}{t \cdot \eta} = \frac{12 \cdot 10^6\,\text{kg} \cdot 9{,}81\,\frac{\text{m}}{\text{s}^2} \cdot 10\,\text{m}}{0{,}7 \cdot (3 \cdot 6 \cdot 3600)\,\text{s}} = 25{,}9\,\text{kW} \quad \blacktriangleleft$$

Beispiel

Eine Güterzuglokomotive soll einen Güterzug von 500 t auf eine Geschwindigkeit von 100 km/h bringen. Die Lokomotive nimmt eine konstante Leistung von 2000 kW bei einem Wirkungsgrad von 40 % auf. Wie lange braucht der Zug bis zum Erreichen der Endgeschwindigkeit?

$$P_{Zu} \cdot \eta = \frac{W_{kin}}{t} = \frac{\frac{1}{2}mv^2}{t}$$

$$t = \frac{mv^2}{2 \cdot \eta \cdot P_{Zu}} = \frac{500 \cdot 10^3\,\text{kg} \cdot (\frac{100}{3{,}6}\,\frac{\text{m}}{\text{s}})^2}{2 \cdot 0{,}4 \cdot 2000 \cdot 10^3\,\frac{\text{kg m}^2}{\text{s}^3}}$$

$$= 241\,\text{s} = 4\,\text{min} \quad \blacktriangleleft$$

Beispiel

Ein Gleichstrommotor wird mit dem Drehmoment $M = 35\,\text{N m}$ belastet und hat dabei die Drehzahl $n = 1500\,\frac{1}{\text{min}}$. Die am Motor liegende Spannung beträgt $U = 400\,\text{V}$, der aufgenommene Strom $I = 17\,\text{A}$. Wie groß ist der Wirkungsgrad des Motors?

$$P_N = 2\pi nM; \quad P_N = 2\pi\,\frac{1500}{60}\,\frac{1}{\text{s}} \cdot 35\,\text{N m} = 5497\,\text{W};$$

$$P_{Zu} = UI = 400 \cdot 17\,\text{W} = 6800\,\text{W}$$

$$\eta = \frac{P_N}{P_{Zu}} = \frac{5497\,\text{W}}{6800\,\text{W}} = 0{,}81 \text{ oder } 81\,\% \quad \blacktriangleleft$$

Das Elektrische Feld

27

Reinhard von Liebenstein



27.1 Grundgrößen 324
27.2 Materie im elektrischen Feld 326
27.3 Kondensatoren 327

Teil IV

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_27

Wenn in jedem Punkt des Raumes eine physikalische oder technische Größe definiert ist, so bezeichnet man dies als *Feld*. So kann man z. B. in jedem Punkt eines Zimmers eine Temperatur messen; es ist ein *Temperaturfeld* definiert. An jedem Punkt über der Erdoberfläche herrscht eine Anziehungskraft, dies ergibt ein *Kraftfeld*, in diesem Fall das *Gravitationsfeld*. Im ersten Fall liegt, da die Temperatur eine skalare Größe ist, ein *Skalarfeld*, im zweiten Fall ein *Vektorfeld* vor.

27.1 Grundgrößen

27.1.1 Kräfte zwischen Ladungen

Zwischen elektrischen Ladungen bestehen immer Kräfte. Haben zwei Ladungen das gleiche Vorzeichen, stoßen sie sich ab, bei ungleichem Vorzeichen ziehen sie sich an. Die Größe der Kraft hängt von der Größe der Ladungen und von deren Abstand ab und wird durch das *Coulomb'sche Gesetz* beschrieben.

$$\textit{Coulomb'sches Gesetz} \quad F_{12} = \frac{1}{4\pi\varepsilon_0} \cdot \frac{Q_1 Q_2}{r_{12}^2} \tag{27.1}$$

Die Größe ε_0 heißt *elektrische Feldkonstante*

$$\varepsilon_0 = 8{,}854 \cdot 10^{-12} \, \frac{\mathrm{A\,s}}{\mathrm{V\,m}} \tag{27.2}$$

Der Abstand zwischen den beiden Ladungen ist r_{12}. Die Richtung der Kraft ist durch die Verbindungslinie zwischen beiden Ladungen gegeben. Die Ladung Q_1 übt auf die Ladung Q_2 die Kraft F_{12} aus. Im Abb. 27.1 oben haben die Ladungen unterschiedliches Vorzeichen, die Kraftpfeile sind aufeinander zu gerichtet. Im unteren Teil haben die Ladungen gleiches Vorzeichen. Die Ladung Q_2 übt auf Q_1 eine entgegengesetzt gleich große Kraft aus. Sind mehr als 2 Ladungen vorhanden, werden zunächst die Kräfte zwischen zwei Ladungen berechnet und dann aus allen Kräften die resultierende Kraft bestimmt.

Beispiel

Drei Ladungen ($Q_1 = -3\,\mathrm{A\,s}$, $Q_2 = Q_3 = 3\,\mathrm{A\,s}$, Abb. 27.2) befinden sich an den Eckpunkten eines gleichseitigen Dreiecks ($a = 5\,\mathrm{cm}$). Wie groß ist die Kraft auf die Ladung Q_1?

Da die Ladungen vom Betrag her gleich sind und gleiche Abstände haben, gilt für die Beträge $F_{21} = F_{31}$. Die Richtungen sind unterschiedlich und verlaufen in Verlängerung der Dreieckseiten.

$$F_{21} = \frac{1}{4\pi\varepsilon_0} \cdot \frac{3 \cdot 10^{-6}\,\mathrm{A\,s} \cdot 3 \cdot 10^{-6}\,\mathrm{A\,s}}{(0{,}05\,\mathrm{m})^2} = 32{,}35\,\mathrm{N}$$

Die Größe der resultierenden Kraft kann entweder aus der Zeichnung abgelesen werden oder mit Hilfe der Trigonometrie berechnet werden:

$$F_{\mathrm{res}} = 2 \cdot 32{,}35\,\mathrm{N} \cdot \cos 30^\circ = 56\,\mathrm{N} \quad \blacktriangleleft$$

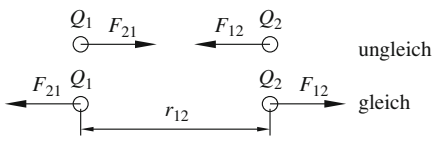


Abb. 27.1 Coulomb'sches Gesetz

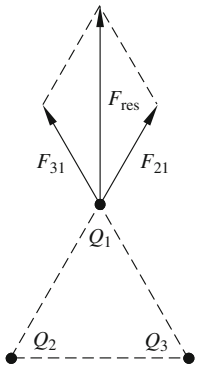


Abb. 27.2 Beispiel Kräfte zwischen Ladungen

27.1.2 Feldstärke

Kräfte zwischen Ladungen sind im ganzen Raum vorhanden, deshalb liegt ein *Feld* vor. Werden zwei Ladungen Q und q im Abstand r betrachtet, kann man willkürlich eine der beiden Ladungen als felderzeugende Ladung Q, die andere als Probeladung q annehmen. Das Coulomb'sche Gesetz lautet dann

$$F = \frac{1}{4\pi\varepsilon_0} \cdot \frac{Qq}{r^2} \tag{27.3}$$

Wird nun, um eine Definition unabhängig von der Probeladung q zu erhalten, der Quotient $\frac{F}{q}$ betrachtet, erhält man eine Größe, die allein von der felderzeugenden Ladung Q und dem Abstand r abhängig ist. Diese Größe nennt man die

elektrische Feldstärke

$$\vec{E} = \frac{\vec{F}}{q} \qquad \begin{array}{c|c|c} E & F & q \\ \hline \frac{\mathrm{N}}{\mathrm{A\,s}} = \frac{\mathrm{V}}{\mathrm{m}} & \mathrm{N} & \mathrm{A\,s} \end{array} \tag{27.4}$$

Da die Kraft ein Vektor ist, muss auch die elektrische Feldstärke ein Vektor sein. Die Richtung von $\vec{E}$ ist gleich der Richtung von $\vec{F}$ auf eine positive Probeladung. Die Feldstärke ist von der felderzeugenden Ladung und deren geometrischer Anordnung

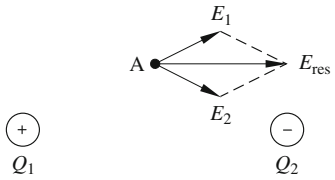


Abb. 27.3 Elektrisches Feld

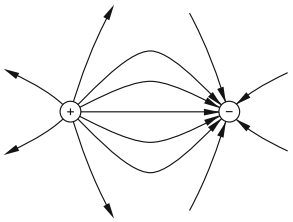


Abb. 27.4 Dipol

abhängig. Ist Q eine Punktladung, kann die elektrische Feldstärke nach (27.3) und (27.4) angegeben werden.

$$\textit{Feld einer Punktladung} \quad E = \frac{1}{4\pi\varepsilon_0} \cdot \frac{Q}{r^2} \tag{27.5}$$

Sind mehrere Ladungen beteiligt, können auch hier wieder die einzelnen Beiträge vektoriell addiert werden. Dies soll am Beispiel eines *Dipols* nach Abb. 27.3 dargestellt werden. Ein *Dipol* besteht aus zwei gleich großen, entgegengesetzten Ladungen. Die positive Ladung Q_1 erzeugt im Punkt A die Feldstärke $\vec{E}_1$. Da Q_1 positiv ist, zeigt $\vec{E}_1$ von Q_1 weg. Die negative Ladung Q_2 erzeugt die zur Ladung gerichtete Feldstärke $\vec{E}_2$. Hieraus ergibt sich die resultierende Feldstärke $\vec{E}_{\text{res}}$.

27.1.3 Feldlinien

Die Konstruktion nach Abb. 27.3 kann in beliebigen Punkten durchgeführt werden, somit kann in jedem Punkt der Vektorpfeil der resultierenden elektrischen Feldstärke konstruiert werden. Werden die Anfangspunkte der Pfeile durch tangentiale Linien verbunden, erhält man für diesen Dipol Linien wie in Abb. 27.4. Diese Linien sind die *elektrischen Feldlinien*.

Eigenschaften elektrischer Feldlinien: Sie verlaufen von der positiven zur negativen Ladung. Die Tangenten geben die Richtung der Kraft auf eine positive Probeladung an. An den Orten größerer Feldliniendichte ist die Kraftwirkung größer. Elektrische Feldlinien schneiden sich nie. Da in elektrischen Leitern die Ladungen frei beweglich sind, werden sie solange verschoben, bis keine tangentiale Kraft mehr vorhanden ist. Daher stehen Feldlinien immer senkrecht auf elektrischen Leitern.

Im Abb. 27.5 sind einige Feldlinienbilder angegeben. Links für eine positive Punktladung, in der Mitte für zwei entgegengesetzt

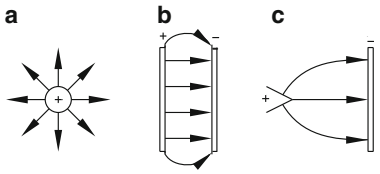


Abb. 27.5 Feldlinienbilder

geladene Platten, rechts für eine positiv geladene Spitze vor einer ebenen Leiterfläche.

27.1.4 Potential, Spannung

Wenn in einem elektrischen Feld eine positive Ladung von einem Punkt A nach einem Punkt B verschoben wird, muss gegen die herrschende Kraft *Verschiebungsarbeit* verrichtet werden.

Verschiebungsarbeit

$$W_{\text{AB}} = -\int_{\text{A}}^{\text{B}} \vec{F} \cdot d\vec{s} = -\int_{\text{A}}^{\text{B}} q\vec{E} \cdot d\vec{s} \tag{27.6}$$

Es kann gezeigt werden, dass diese Arbeit unabhängig vom Weg ist. Sie hängt nur von der Lage der beiden Punkte im elektrischen Feld ab. Das Vorzeichen in (27.6) ist so gewählt, dass zu leistende Arbeit positiv gezählt wird.

$$W_{\text{AB}} = -q\int_{\text{A}}^{\text{B}} \vec{E} d\vec{s} = qU_{\text{AB}} \tag{27.7}$$

$$\textit{Spannung} \quad U_{\text{AB}} = \frac{W_{\text{AB}}}{q} = -\int_{\text{A}}^{\text{B}} \vec{E} \cdot d\vec{s} \tag{27.8}$$

Wird der Ausgangspunkt A sehr weit (im Grenzfall ins Unendliche) entfernt, ist der Quotient aus der jetzt aufzubringenden Arbeit und der Probeladung q das

$$\textit{Potential} \quad \varphi_B = \frac{W_{\infty B}}{q} = -\int_{\infty}^{B} \vec{E} \cdot d\vec{s} \tag{27.9}$$

Die Spannung zwischen zwei Punkten A und B ist die Potentialdifferenz zwischen diesen Punkten.

$$\textit{Spannung} \quad U_{\text{AB}} = \varphi_B - \varphi_A = \Delta\varphi \tag{27.10}$$

$$\Delta W_{\text{AB}} = q \cdot \Delta\varphi \tag{27.11}$$

Ist das elektrische Feld nach Betrag und Richtung konstant, spricht man von einem *homogenen Feld*. Wird der Abstand zwischen den Punkten A und B mit d bezeichnet, folgt aus (27.8) im homogenen Feld

$$\textit{homogenes Feld} \quad U = E \cdot d \tag{27.12}$$

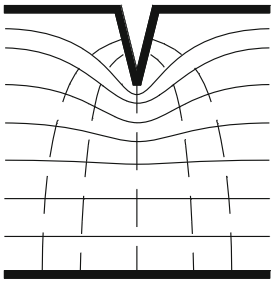


Abb. 27.6 Äquipotentiallinien

27.1.5 Äquipotentiallinien

Werden in einem elektrischen Feld Punkte mit gleichem Potential durch Linien verbunden, entstehen die *Äquipotentiallinien*. Zwischen zwei Punkten auf einer Äquipotentiallinie besteht keine Potentialdifferenz und auch keine Spannung. Hieraus folgt, dass bei Verschiebung einer Ladung auf einer Äquipotentiallinie keine Verschiebungsarbeit zu verrichten ist. Weiterhin folgt aus der Vektorrechnung, dass Äquipotentiallinien und Feldlinien immer senkrecht aufeinander stehen müssen. Da Feldlinien auch immer senkrecht auf Leiteroberflächen stehen, sind alle Leiteroberflächen auch Äquipotentiallinien. Messtechnisch einfacher ist es, Potentiallinien auszumessen. Sind sie bekannt, können die für die Kraftwirkungen wichtigen Feldlinien konstruiert werden. Im Abb. 27.6 sind für eine metallische Spitze und eine ebene metallische Gegenelektrode die Äquipotentiallinien (durchgezogen) und die elektrischen Feldlinien (gestrichelt) dargestellt.

27.1.6 Elektrischer Fluss

Die elektrischen Ladungen sind Ausgangspunkt der elektrischen Feldlinien. Je größer die Ladung, desto mehr Feldlinien gehen von ihr aus. Denkt man sich eine Kugel mit dem Radius r um eine Ladung, so bezeichnet man alle durch die Kugeloberfläche durchtretenden Feldlinien als den elektrischen Fluss ψ. Ladung Q und Fluss ψ sind *direkt proportional* zueinander. Die Konstante wird zu 1 festgelegt und es gilt

$$\psi = Q \tag{27.13}$$

Für eine Punktladung gilt nach (27.5)

$$Q = 4\pi\varepsilon_0 r^2 E \tag{27.14}$$

Oberfläche einer Kugel
$$A_{\text{Kugel}} = 4\pi r^2 \tag{27.15}$$

elektrischer Fluss
$$\psi = Q = A_{\text{Kugel}} \cdot \varepsilon_0 E \tag{27.16}$$

Der Fluss, dividiert durch die gesamte Oberfläche der Kugel, wird als *Flussdichte D* bezeichnet.

Flussdichte
$$D = \frac{\psi}{A} \qquad \begin{array}{c|c|c} D & \psi & A \\ \hline \frac{\text{A s}}{\text{m}^2} & \text{A s} & \text{m}^2 \end{array} \tag{27.17}$$

Flussdichte
$$D = \varepsilon_0 E \qquad \begin{array}{c|c|c} D & \varepsilon_0 & E \\ \hline \frac{\text{A s}}{\text{m}^2} & \frac{\text{A s}}{\text{V m}} & \frac{\text{V}}{\text{m}} \end{array} \tag{27.18}$$

Allgemeiner formuliert lautet (27.17) in Verbindung mit (27.13):

> Der durch eine geschlossene, beliebig geformte Oberfläche A gehende elektrische Fluss ψ ist gleich der Summe der von dieser Fläche eingeschlossenen Ladungen.
>
> $$\psi = \oint \vec{D} d\vec{A} = \sum_{i=1}^{n} Q_i \tag{27.19}$$

Daraus folgt für die elektrische Feldstärke eine der vier grundlegenden Maxwell'schen Gleichungen

$$\oint \vec{E} d\vec{A} = \frac{1}{\varepsilon_0 \varepsilon_r} \sum_{i=1}^{n} Q_i \tag{27.20}$$

Steht das elektrische Feld überall senkrecht zu A und ist es konstant, so ist

$$D = \frac{Q}{A} \tag{27.21}$$

27.1.7 Energie geladener Teilchen im elektrischen Feld

Da auf geladene Teilchen im elektrischen Feld immer eine Kraft wirkt, werden sie nach den Gesetzen der Mechanik immer beschleunigt.

$$F_{\text{el}} = qE = ma \tag{27.22}$$

Ist das elektrische Feld konstant, so ist auch die Beschleunigung konstant. Nach Durchlaufen einer Potentialdifferenz $\Delta\varphi$ sind die erreichte Bewegungsenergie und die elektrische Energie nach (27.11) gleich.

$$\frac{1}{2}mv^2 = q \cdot \Delta\varphi = qU \tag{27.23}$$

Wenn ein Elektron die Potentialdifferenz von 1 V durchläuft, hat es die Energie von $1{,}602 \cdot 10^{-19}\,\text{A s} \cdot 1\,\text{V} = 1{,}602 \cdot 10^{-19}\,\text{J}$. Hieraus leitet sich eine Energieeinheit ab:

1 Elektronenvolt
$$1\,\text{eV} = 1{,}602 \cdot 10^{-19}\,\text{J} \tag{27.24}$$

27.2 Materie im elektrischen Feld

Befindet sich Materie in einem elektrischen Feld, wirkt auf alle Ladungen in dieser Materie eine elektrische Kraft. Da in Leitern und Nichtleitern unterschiedliche Beweglichkeit der Ladungsträger vorliegt, sind auch die beobachteten Effekte von den Leitereigenschaften der Materialien abhängig.

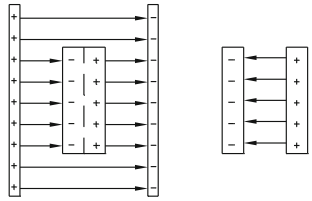


Abb. 27.7 Influenz

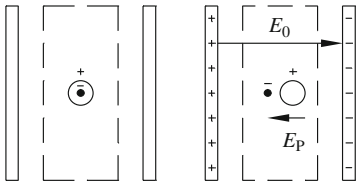


Abb. 27.8 Polarisation

27.2.1 Leiter

In einem Leiter sind die Ladungsträger (in Metallen sind das die Elektronen) frei beweglich. Bringt man zwei aufeinander liegende Metallplatten in ein elektrisches Feld, werden die Ladungen so verschoben (Abb. 27.7), dass sich eine gleiche Anzahl von Ladungen auf den entsprechenden Oberflächen gegenüber liegt. Im Leiter wird durch diese Ladungsverschiebung ein Feld E_{infl} erzeugt, das denselben Betrag wie das äußere Feld, aber entgegengesetzte Richtung hat. Das resultierende elektrische Feld im Inneren eines Leiters verschwindet daher immer.

$$E_i = E_a - E_{\text{infl}} = 0 \tag{27.25}$$

Diese Ladungsverschiebung heißt *Influenz*. Werden die Platten im elektrischen Feld getrennt und dann aus dem Feld genommen, bleiben die Platten geladen.

27.2.2 Nichtleiter

In Nichtleitern gibt es keine frei beweglichen Ladungen. Deshalb ist auch das Innere eines Nichtleiters im elektrischen Feld nicht feldfrei. Diese Stoffe werden *Dielektrika* genannt. Hier kann durch das äußere Feld ein inneres Feld erzeugt werden, indem die negative Elektronenhülle eines Atoms sich gegenüber dem positiven Kern verschieben kann. Es entsteht ein elektrischer Dipol. Auch können eventuell schon vorhandene Dipole im Feld gedreht werden. Es entsteht ein *Polarisationsfeld* E_p, das dem äußeren Feld E_0 entgegengesetzt gerichtet ist (Abb. 27.8). Im Inneren ist das resultierende Feld

$$E_i = E_0 - E_p \tag{27.26}$$

Das Feld im Inneren ist gegenüber dem ursprünglichen Feld E_0 geschwächt worden. Zahlenmäßig wird dies durch

$$E_i = \frac{E_0}{\varepsilon_r} \tag{27.27}$$

Tab. 27.1 Permittivitätszahlen bei 20 °C

Werkstoff	ε_r
Glas	4 … 12
Glimmer	5 … 8
Porzellan	6
Transformatorenöl	2,3
Wasser	81
Luft	1,0006

erfasst. ε_r ist die *Permittivitätszahl* des Materials (Tab. 27.1). Sind die äußeren Platten mit einer konstanten Spannungsquelle verbunden, fließen nach Einbringen eines Dielektrikums so viele Ladungen nach, bis sich das ursprüngliche Feld E_0 wieder eingestellt hat. Damit wird aber nach (27.13) der elektrische Fluss ψ und auch die Flussdichte D ebenfalls größer und es gilt

$$D = \varepsilon_r \varepsilon_0 E \tag{27.28}$$

27.3 Kondensatoren

Wenn zwischen zwei isolierten Leiteroberflächen eine Spannung liegt, laden sich die Leiteroberflächen auf. Ein solches Bauelement ist ein *Kondensator*. Er dient zur Speicherung elektrischer Ladungen und Energie. Die Ladungsmenge Q, die bei vorgegebener Spannung U gespeichert werden kann, hängt von der Geometrie und vom Material zwischen den Kondensatorplatten ab. Der Quotient aus Q und U wird als *Kapazität* bezeichnet.

27.3.1 Kapazität

Die Kapazität eines Kondensators ist definiert durch

$$\textit{Kapazität} \quad C = \frac{Q}{U} \qquad \begin{array}{c|c|c} C & Q & U \\ \hline \mathrm{F} = \frac{\mathrm{As}}{\mathrm{V}} & \mathrm{As} & \mathrm{V} \end{array} \tag{27.29}$$

Die Einheit der Kapazität ist 1 F (1 Farad).

27.3.1.1 Plattenkondensator

Ein Plattenkondensator besteht aus zwei parallelen Platten im Abstand d. Bis auf den Randbereich ist das elektrische Feld E homogen. Befindet sich zwischen den Platten Vakuum oder Luft, gilt

$$D = \frac{Q}{A} = \varepsilon_0 E = \varepsilon_0 \frac{U}{d} \tag{27.30}$$

$$C_{\text{Pl}} = \frac{Q}{U} = \varepsilon_0 \frac{A}{d} \tag{27.31}$$

Befindet sich zwischen den Platten ein Material mit der Permittivitätszahl ε_r, dann gilt

$$C_{\text{Pl}} = \varepsilon_0 \varepsilon_r \frac{A}{d} \tag{27.32}$$

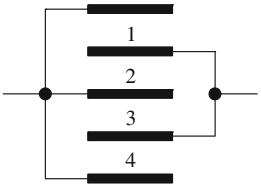


Abb. 27.9 Blockkondensator

27.3.1.2 Spezielle Kondensatoren

Ein *Blockkondensator* besteht aus n Platten eines Kondensators, die wie in Abb. 27.9 verbunden sind. Es ist eine Parallelschaltung mit $(n-1)$ wirksamen Kondensatoren.

$$\textit{Blockkondensator} \quad C_{\text{Block}} = (n-1)\varepsilon_0\varepsilon_r\frac{A}{d} \tag{27.33}$$

Ein *Kugelkondensator* besteht aus zwei konzentrisch angebrachten Kugeln. Die Vorteile eines Kugelkondensators liegen darin, dass keine Streufelder auftreten, da keine Ecken vorhanden sind. Die elektrischen Feldlinien verlaufen radial-symmetrisch und ändern ihren Abstand von innen nach außen. Das Feld ist also nicht homogen. Für die Kapazität eines Kugelkondensators ergibt sich mit der in Abb. 27.10 dargestellten Geometrie:

$$\textit{Kugelkondensator} \quad C_{\text{Kugel}} = 4\pi\varepsilon_0\varepsilon_r\frac{r_1 r_2}{r_2 - r_1} \tag{27.34}$$

Der *Zylinderkondensator* besteht aus zwei ineinander liegenden konzentrischen Zylindern. Wird der Radius des äußeren Zylinders mit r_2 und der des inneren Zylinders mit r_1 bezeichnet, gilt bei einer Länge L des Kondensators für die Kapazität:

$$\textit{Zylinderkondensator} \quad C_{\text{Zyl}} = 2\pi\varepsilon_0\varepsilon_r\frac{L}{\ln(\frac{r_2}{r_1})} \tag{27.35}$$

Jedes *Koaxialkabel* ist von seiner Bauform her ein Zylinderkondensator. In der Hochfrequenztechnik ist diese Kapazität daher zu berücksichtigen. Für die Kapazität einer *einzelnen zylindrischen Leitung* mit dem Radius r, die in einer Höhe h über der Erde liegt, gilt

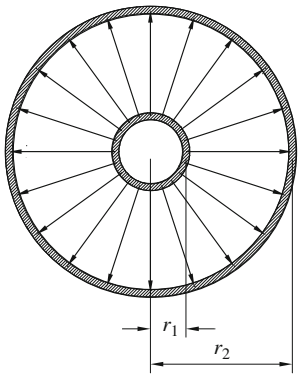


Abb. 27.10 Kugelkondensator

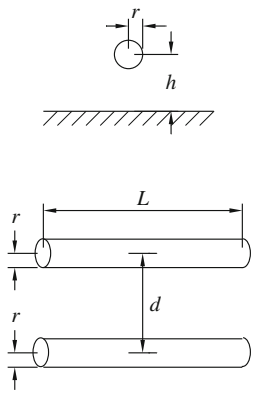


Abb. 27.11 Kapazität voin Leitungen

Kapazität einer Leitung

$$C_L = 2\pi\varepsilon_0\varepsilon_r\frac{L}{\ln(\frac{h}{r} + \sqrt{(\frac{h}{r})^2 - 1})} \tag{27.36}$$

für $h \gg r$

$$C_L = 2\pi\varepsilon_0\varepsilon_r\frac{L}{\ln(\frac{2h}{r})} \tag{27.37}$$

Auch zwei im Abstand d verlaufende Leitungen mit dem Radius r und der Länge L besitzen eine Kapazität (Abb. 27.11).

Kapazität einer Doppelleitung

$$C_{\text{DL}} = \pi\varepsilon_0\varepsilon_r\frac{L}{\ln(\frac{d}{2r} + \sqrt{(\frac{d}{2r})^2 - 1})} \tag{27.38}$$

für $d \gg r$

$$C_{\text{DL}} = \pi\varepsilon_0\varepsilon_r\frac{L}{\ln(\frac{d}{r})} \tag{27.39}$$

Beispiel

In einem Plattenkondensator nach Abb. 27.12 sind $d_1 = 0{,}3\,\text{mm}$, $d_2 = 0{,}5\,\text{mm}$ starke Isolierplatten untergebracht. Ihre Permittivitätszahlen sind $\varepsilon_{1r} = 3{,}8$ und $\varepsilon_{2r} = 4{,}7$. Die Fläche A einer Platte beträgt $900\,\text{cm}^2$. Die obere Platte trägt die positive Ladung $Q = 2{,}5 \cdot 10^{-6}\,\text{A s}$, die untere die negative Ladung $-Q$. Wie groß sind die Flussdichten und die elektrischen Feldstärken in den beiden Dielektrika und die Spannungen, mit denen die Isolierplatten beansprucht werden?

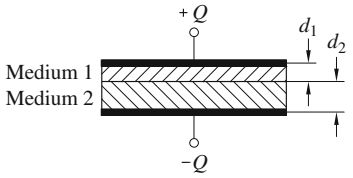


Abb. 27.12 Beispiel Plattenkondensator

Die Flussdichte muss in beiden Materialien gleich sein, da beide Platten des Kondensators die gleiche Ladung tragen.

$$D_1 = D_2 = \frac{Q}{A} = \frac{2{,}5 \cdot 10^{-6}\,\mathrm{A\,s}}{900 \cdot 10^{-4}\,\mathrm{m^2}} = 27{,}8 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{m^2}}$$

$$E_1 = \frac{D_1}{\varepsilon_0 \varepsilon_{r1}} = \frac{27{,}8 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{m^2}}}{8{,}854 \cdot 10^{-12}\,\frac{\mathrm{A\,s}}{\mathrm{V\,m}} \cdot 3{,}8} = 826 \cdot 10^3\,\frac{\mathrm{V}}{\mathrm{m}}$$

$$= 826\,\frac{\mathrm{V}}{\mathrm{mm}}$$

$$E_2 = \frac{D_2}{\varepsilon_0 \varepsilon_{r2}} = \frac{27{,}8 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{m^2}}}{8{,}854 \cdot 10^{-12}\,\frac{\mathrm{A\,s}}{\mathrm{V\,m}} \cdot 4{,}7} = 668 \cdot 10^3\,\frac{\mathrm{V}}{\mathrm{m}}$$

$$= 668\,\frac{\mathrm{V}}{\mathrm{mm}}$$

$$U_1 = E_1 \cdot d_1 = 826\,\frac{\mathrm{V}}{\mathrm{mm}} \cdot 0{,}3\,\mathrm{mm} = 248\,\mathrm{V}$$

$$U_2 = 668\,\frac{\mathrm{V}}{\mathrm{mm}} \cdot 0{,}5\,\mathrm{mm} = 334\,\mathrm{V}$$

◀

Beispiel

Ein *Koaxialkabel* enthält nach Abb. 27.13 zwei Isolationsschichten mit $\varepsilon_{r1} = 3$ und $\varepsilon_{r2} = 2$, einen Innenleiter mit dem Radius $r_1 = 5\,\mathrm{mm}$, einen Außenleiter mit $r_3 = 15\,\mathrm{mm}$. Der Radius r_2 ist 7,5 mm. Wie groß ist die Kapazität des Kabels je km Leitungslänge?

Mit (27.20) wird

$$E_1 = \frac{Q}{2\pi\varepsilon_0\varepsilon_{r1} r l}; \quad E_2 = \frac{Q}{2\pi\varepsilon_0\varepsilon_{r2} r l}$$

$$U = \int_{r_1}^{r_2} E_1 dr + \int_{r_2}^{r_3} E_2 dr \tag{27.40}$$

$$U = \frac{Q}{2\pi\varepsilon_0 l}\left(\frac{1}{\varepsilon_{r1}} \ln\frac{r_2}{r_1} + \frac{1}{\varepsilon_{r2}} \ln\frac{r_3}{r_2}\right) \tag{27.41}$$

Die Kapazität pro Länge l

$$C' = \frac{Q}{Ul} = \frac{2\pi\varepsilon_0}{\frac{1}{\varepsilon_{r1}} \ln\frac{r_2}{r_1} + \frac{1}{\varepsilon_{r2}} \ln\frac{r_3}{r_2}} \tag{27.42}$$

$$C' = \frac{2\pi 8{,}85 \cdot 10^{-12}\,\frac{\mathrm{F}}{\mathrm{m}}}{\frac{1}{3} \ln\frac{7{,}5}{5} + \frac{1}{2} \ln\frac{15}{7{,}5}} = 115{,}5 \cdot 10^{-12}\,\frac{\mathrm{F}}{\mathrm{m}} = 115\,\frac{\mathrm{nF}}{\mathrm{km}}$$

◀

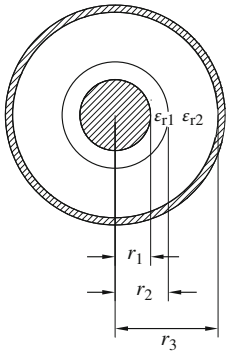


Abb. 27.13 Beispiel Koaxialkabel

27.3.2 Schaltungen mit Kondensatoren

27.3.2.1 Reihenschaltung

Werden zwei in Reihe geschaltete ungeladene Kondensatoren aufgeladen, muss auf jeden Kondensator dieselbe Ladungsmenge ΔQ geflossen sein (Abb. 27.14).

$$\Delta Q_1 = \Delta Q_2 = Q \tag{27.43}$$

$$U = U_1 + U_2 \tag{27.44}$$

$$\frac{1}{C_1} = \frac{U_1}{Q}, \quad \frac{1}{C_2} = \frac{U_2}{Q} \tag{27.45}$$

$$\frac{U}{Q} = \frac{U_1 + U_2}{Q} = \frac{U_1}{Q} + \frac{U_2}{Q} \tag{27.46}$$

$$\frac{1}{C_{\mathrm{ges}}} = \frac{1}{C_1} + \frac{1}{C_2} \tag{27.47}$$

Reihenschaltung

$$\frac{1}{C_{\mathrm{ges}}} = \sum_{i=1}^{n} \frac{1}{C_i} \tag{27.48}$$

27.3.2.2 Parallelschaltung

Sind zwei Kondensatoren parallel geschaltet, liegt an beiden die Spannung U. Für die während des Ladens aus der Spannungsquelle entnommene Ladung ΔQ gilt jetzt

$$\Delta Q = \Delta Q_1 + \Delta Q_2 \tag{27.49}$$

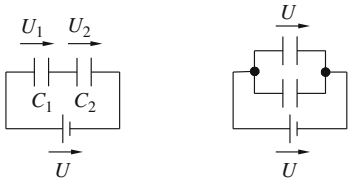


Abb. 27.14 Kondensatorschaltungen

Teil IV

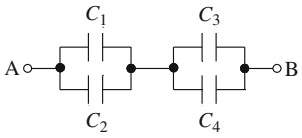


Abb. 27.15 Beispiel Berechnung der Gesamtkapazität

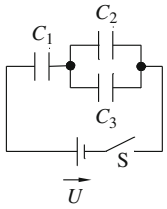


Abb. 27.16 Beispiel Umladungsvorgänge mit Spannungsquelle

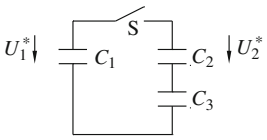


Abb. 27.17 Beispiel Umladungsvorgänge ohne Spannungsquelle

Da die Kondensatoren zu Beginn nicht geladen waren ($Q_v = 0$), gilt

$$\Delta Q_1 = Q_n - Q_v = Q \tag{27.50}$$

$$\frac{Q}{U} = \frac{Q_1 + Q_2}{U} = \frac{Q_1}{U} + \frac{Q_2}{U} \tag{27.51}$$

$$C_{\text{ges}} = C_1 + C_2 \tag{27.52}$$

Parallelschaltung

$$C_{\text{ges}} = \sum_{i=1}^{n} C_i \tag{27.53}$$

27.3.2.3 Gemischte Schaltungen

Beispiel

Die Gesamtkapazität der Schaltung nach Abb. 27.15 mit den Werten $C_1 = 80\,\text{pF}$, $C_2 = 20\,\text{pF}$, $C_3 = 200\,\text{pF}$, $C_4 = 100\,\text{pF}$ soll berechnet werden.

Die linke Gruppe ergibt 100 pF, die rechte Gruppe 300 pF. Die Gesamtkapazität ist dann $C_{\text{ges}} = \frac{100 \cdot 300}{100+300}\,\text{pF} = 75\,\text{pF}$. ◀

Beispiel

Die in der Schaltung nach Abb. 27.16 vorhandenen Kondensatoren sind zunächst nicht aufgeladen. Welche Ladungen und welche Teilspannungen haben die Kondensatoren nach Schließen des Schalters S ($C_1 = 300\,\text{nF}$, $C_2 = 200\,\text{nF}$, $C_3 = 400\,\text{nF}$, $U = 120\,\text{V}$)?

Die Ladungsmenge, die durch C_1 fließt, muss auch durch die Parallelkombination fließen. Es lassen sich folgende Gleichungen aufstellen:

$$U = U_1 + U_{23}; \quad \Delta Q_1 = \Delta Q_2 + \Delta Q_3$$

$$C_1 U_1 = C_2 U_{23} + C_3 U_{23} = U_{23}(C_2 + C_3)$$

$$U_{23} = U_1 \frac{C_1}{C_2 + C_3};$$

$$U = U_1 + U_1 \frac{C_1}{C_2 + C_3} = U_1 \left(1 + \frac{C_1}{C_2 + C_3}\right)$$

$$U_1 = \frac{U}{1 + \frac{C_1}{C_2+C_3}} = \frac{120\,\text{V}}{1 + \frac{300}{200+400}} = 80\,\text{V}$$

$$U_2 = U - U_1 = 120\,\text{V} - 80\,\text{V} = 40\,\text{V}$$

$$Q_1 = C_1 U_1 = 300 \cdot 10^{-9}\,\text{F} \cdot 80\,\text{V} = 24 \cdot 10^{-6}\,\text{A s}$$

$$Q_2 = C_2 U_{23} = 200 \cdot 10^{-9}\,\text{F} \cdot 40\,\text{V} = 8 \cdot 10^{-6}\,\text{A s}$$

$$Q_3 = C_3 U_{23} = 400 \cdot 10^{-9}\,\text{F} \cdot 40\,\text{V} = 16 \cdot 10^{-6}\,\text{A s}$$

◀

Beispiel

Der Kondensator $C_1 = 8\,\text{nF}$ ist zunächst mit einer Spannungsquelle von $U = 150\,\text{V}$ verbunden, C_2 ist in gleicher Richtung auf 50 V aufgeladen, C_3 ist ungeladen. Anschließend werden die Kondensatoren in geladenem Zustand von den Spannungsquellen getrennt und der Schalter S geschlossen ($C_2 = 3\,\text{nF}$ und $C_3 = 6\,\text{nF}$). Auf welche Spannungen werden die einzelnen Kondensatoren aufgeladen (Abb. 27.17)?

Die Spannungen an den einzelnen Kondensatoren vor Schließen des Schalters werden mit U^* bezeichnet, nach Umladen werden die Spannungen mit U benannt. Für die einzelnen Spannungen gilt $\Delta U_i = U_i - U_i^*$. Die Ladungsmenge, die vom Kondensator C_1 fließt, muss sowohl C_2 als auch C_3 aufladen. Außerdem gilt die Maschenregel.

$$\Delta U_i = U_i - U_i^*; \qquad \Delta Q_1 = \Delta Q_2 = \Delta Q_3$$

$$-C_1 \cdot \Delta U_1 = C_2 \cdot \Delta U_2; \quad -C_1 \cdot \Delta U_1 = C_3 \cdot \Delta U_3$$

Hieraus lassen sich für die drei gesuchten Spannungen für den allgemeinen Fall, dass die Kondensatoren 2 und 3 vor dem Umladungsprozess schon aufgeladen sind, die drei Gleichungen ableiten:

$$U_1 - U_2 - U_3 = 0;$$

$$C_1 \cdot U_1 + C_2 \cdot U_2 = C_1 \cdot U_1^* + C_2 \cdot U_2^*$$

$$C_1 \cdot U_1 + C_3 \cdot U_3 = C_1 \cdot U_1^* + C_3 \cdot U_3^*$$

Zahlenwerte:

$$U_1 - U_2 - U_3 = 0;$$

$$8U_1 + 3U_2 = 8 \cdot 150\,\text{V} + 3 \cdot 50\,\text{V};$$

$$8U_1 + 6U_3 = 8 \cdot 150\,\text{V}$$

Dieses Gleichungssystem hat die Lösungen

$$U_1 = 130\,\text{V}; \quad U_2 = 103{,}33\,\text{V}; \quad U_3 = 26{,}67\,\text{V}$$ ◀

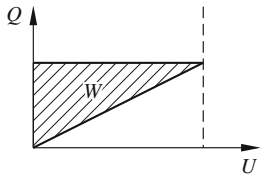


Abb. 27.18 Energie im elektrischen Feld

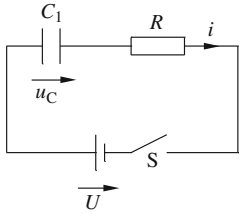


Abb. 27.19 Laden eines Kondensators

27.3.3 Energie des elektrostatischen Feldes

In einem aufgeladenen Kondensator ist eine bestimmte Energie gespeichert. Die Größe der gespeicherten Energie kann durch die Formel

$$\Delta W = u\Delta q \tag{27.54}$$

bestimmt werden.

Mit u ist die zeitlich veränderliche Spannung am Kondensator gemeint. Die Ladung q und die Spannung u sind aber nicht unabhängig voneinander, sondern über (27.29) verbunden.

Energie eines Kondensators

$$W = \int_0^Q u dq = \int_0^Q \frac{q}{C} dq = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}CU^2 \tag{27.55}$$

Grafisch ist die Energie, die bei einer Spannung U und der damit im Kondensator vorhandenen Ladung Q gespeichert ist, durch die schraffierte Dreiecksfläche in Abb. 27.18 gegeben.

$$W = \frac{1}{2}QU = \frac{1}{2}CU^2 \tag{27.56}$$

Die im Kondensator gespeicherte Energie befindet sich nicht auf den Elektroden, sondern ist im elektrischen Feld zwischen den Kondensatorplatten gespeichert. Ein wichtiger Begriff ist die Energie pro Volumen, die *Energiedichte*. Zur Berechnung wird die in einem bestimmten Volumen vorhandene Energie durch das Volumenelement dividiert.

$$\textit{Energiedichte} \quad w = \frac{\Delta W}{\Delta V} \tag{27.57}$$

Für einen Plattenkondensator lässt sich diese Größe unter Verwendung oben abgeleiteter Formeln einfach berechnen.

$$W = \frac{1}{2}CU^2 = \frac{1}{2}\frac{\varepsilon_0\varepsilon_r A}{d}(Ed)^2 \tag{27.58}$$

$$W = \frac{1}{2}\varepsilon_0\varepsilon_r E^2 Ad = \frac{1}{2}\varepsilon_0\varepsilon_r E^2 V \tag{27.59}$$

Das Volumen des elektrischen Feldes im Kondensator ist $V = Ad$. Mit $D = e_0 e_r E$ gilt

$$\textit{Energiedichte} \quad w = \frac{1}{2}ED \tag{27.60}$$

Gleichung (27.60) gilt in dieser Form nicht nur im homogenen Feld, sondern für konstantes ε_r allgemein.

Energiedichte

$$w = \int EdD = \int \varepsilon_0\varepsilon_r EdE = \frac{1}{2}\varepsilon_0\varepsilon_r E^2 = \frac{1}{2}ED \tag{27.61}$$

Soll in einem beliebigen inhomogenen Feld die Energiedichte berechnet werden, muss das Integral bestimmt werden.

$$W = \int_V w dV = \frac{1}{2}\int_V EDdV \tag{27.62}$$

27.3.4 Laden und Entladen eines Kondensators

Wenn ein Kondensator nach Abb. 27.19 über einen Widerstand aufgeladen wird, vergrößert sich die Ladungsmenge auf den Kondensatorplatten und damit die Kondensatorspannung u_C als Funktion der Zeit. Aus der

$$\textit{Maschenregel} \quad U = u_C + iR \tag{27.63}$$

folgt, dass der Strom dann kleiner werden muss. Wegen

$$i = \frac{dQ}{dt} = C\frac{du_C}{dt} \tag{27.64}$$

wird die zeitliche Änderung von u_C ebenfalls kleiner.

$$U = u_c + RC\frac{du}{dt} \tag{27.65}$$

Wird zum Zeitpunkt $t = 0$ der Schalter S geschlossen, hat diese *Differentialgleichung* die Lösungsfunktion

Laden eines Kondensators

$$u_C = U(1 - e^{-t/RC}) \tag{27.66}$$

Der Kondensator lädt sich auf die Endspannung U auf. Der Strom nimmt ab.

$$\textit{Ladestrom} \quad i = I_0 \cdot e^{-t/RC} \tag{27.67}$$

Der Strom I_0 ist der Strom zu Beginn des Ladevorganges, also zur Zeit $t = 0$. Er ist bestimmt durch

$$I_0 = \frac{U}{R} \tag{27.68}$$

Das Produkt RC hat die Dimension einer Zeit und wird Zeitkonstante genannt.

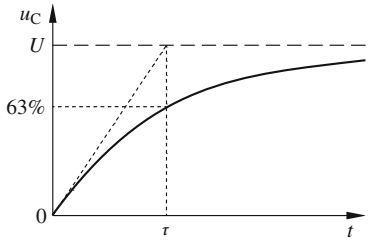


Abb. 27.20 Spannung am Kondensator

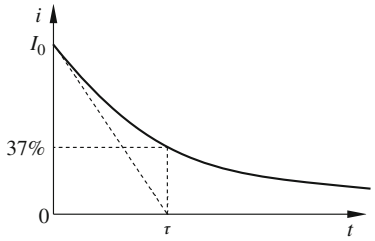


Abb. 27.21 Ladestrom

Zeitkonstante

$$\tau = RC \qquad \begin{array}{c|c|c} \tau & R & C \\ \hline \mathrm{s} = \Omega\,\mathrm{F} & \Omega = \frac{\mathrm{V}}{\mathrm{A}} & \mathrm{F} = \frac{\mathrm{As}}{\mathrm{V}} \end{array} \qquad (27.69)$$

Diese Größe gibt an, nach welcher Zeit die Spannung auf den Wert $1 - e^{-1} = 63\,\%$ angestiegen ist, bzw. nach welcher Zeit der Strom auf den Wert $e^{-1} = 37\,\%$ gefallen ist. Der Wert für t kann aus der grafischen Darstellung der Kondensatorspannung ermittelt werden. Er ergibt sich als Schnittpunkt der Tangente an die Spannungskurve zur Zeit $t = 0$ mit der gestrichelt gezeichneten horizontalen Geraden U (Abb. 27.20).

Beim Entladen eines Kondensators über einen ohmschen Widerstand R nimmt die Kondensatorspannung im Lauf der Zeit ab; der fließende Strom ebenfalls. Ist der Kondensator zu Beginn auf die Spannung U aufgeladen, nimmt diese Spannung ab.

Entladen eines Kondensators

$$u_C = Ue^{-t/RC} = Ue^{-t/\tau} \qquad (27.70)$$

$$\textit{Strom} \quad i = I_0 \cdot e^{-t/RC} \qquad (27.71)$$

Die in Abb. 27.21 dargestellte Kurve gilt in dieser Form für den Strom beim Laden und Entladen und für die Kondensatorspannung beim Entladen eines Kondensators.

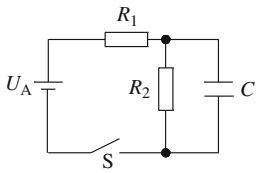


Abb. 27.22 Beispiel Aufladen eines Kondensators

Beispiel

In der Schaltung nach Abb. 27.22 wird der Kondensator $C = 100\,\mathrm{mF}$ über den Widerstand $R_1 = 2\,\mathrm{k\Omega}$ aufgeladen ($R_2 = 1\,\mathrm{k\Omega}$, $U_A = 100\,\mathrm{V}$).

a) Auf welche Endspannung lädt sich der Kondensator auf?
b) Wie groß ist die Spannung am Kondensator nach einer Zeit von 30 ms?

Die Spannungsquelle und die beiden Widerstände können als Ersatzspannungsquelle mit Innenwiderstand aufgefasst werden. Die Spannung, die den Kondensator auflädt, ergibt sich aus

$$U = U_A \frac{R_2}{R_1 + R_2} \qquad (27.72)$$

Der Innenwiderstand der Ersatzspannungsquelle ist

$$R_i = \frac{R_1 R_2}{R_1 + R_2} \qquad (27.73)$$

Mit (27.66) gilt

$$u_C = U_A \frac{R_2}{R_1 + R_2}(1 - e^{\frac{-t}{\tau}}) \qquad (27.74)$$

mit

$$\tau = R_i C \qquad (27.75)$$

a) Endspannung

$$U = 100\,\mathrm{V}\frac{1\,\mathrm{k\Omega}}{3\,\mathrm{k\Omega}} = 33{,}33\,\mathrm{V}$$

b)

$$R_i = \frac{1 \cdot 2}{1 + 2}\,\mathrm{k\Omega} = \frac{2}{3}\,\mathrm{k\Omega}$$

$$\tau = \frac{2}{3} \cdot 10^3\,\Omega \cdot 100 \cdot 10^{-6}\,\mathrm{F} = 66{,}7\,\mathrm{ms}$$

$$u_{30\,\mathrm{ms}} = 33{,}33(1 - e^{-30\,\mathrm{ms}/66{,}7\,\mathrm{ms}}) = 12{,}07\,\mathrm{V} \quad \blacktriangleleft$$

Teil IV

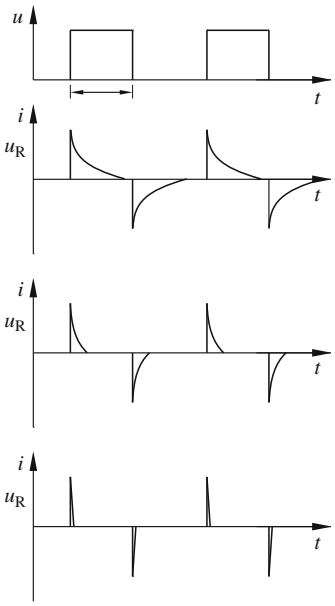


Abb. 27.23 Differenzieren

27.3.5 RC-Reihenschaltung

Legt man an eine RC-Reihenschaltung einen Rechteck-Impuls an, wird der Kondensator im gleichen Takt aufgeladen und entladen. Ist die Zeitkonstante τ ungefähr so groß wie die Impulsdauer t_0, kann der Kondensator nicht vollständig aufgeladen werden. Der Ladestrom, und damit auch die Spannung am ohmschen Widerstand, hat einen Verlauf wie in Abb. 27.23 oben. Die Zeitkonstante nimmt von oben nach unten ab. Je kleiner die Zeitkonstante τ gegenüber der Impulsdauer t_0 wird, desto schneller fällt die Spannung u_R auf 0 ab. Im Grenzfall entstehen Nadelimpulse. Wird das Rechtecksignal mathematisch differenziert, entsteht das gleiche Funktionsbild. Deshalb spricht man in diesem Fall auch von einem *Differenzierglied*.

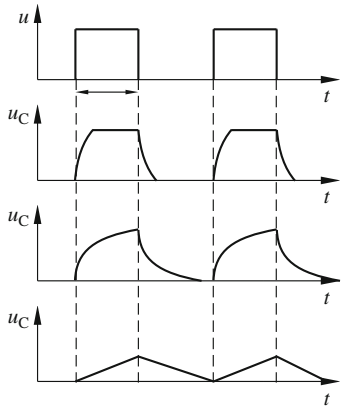


Abb. 27.24 Integrieren

Die Spannung u_C am Kondensator ist im Abb. 27.24 aufgetragen. Die Zeitkonstante τ wird hier von oben nach unten immer größer. Bei sehr großem Wert von τ wird der Kondensator nahezu linear aufgeladen. Dies entspricht dem mathematischen Vorgang des Integrierens. Die Schaltung stellt jetzt ein *Integrierglied* dar.

Das Magnetische Feld

28

Reinhard von Liebenstein



28.1 Feldlinien 336
28.2 Magnetische Grundgrößen 337
28.3 Kräfte im Magnetfeld 341
28.4 Energie des Magnetfeldes 344
28.5 Materie im Magnetfeld 344
28.6 Magnetische Kreise 348

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_28

Neben den elektrischen Kräften und Feldern gibt es in der Natur auch magnetische Kräfte und Felder.

Jede sich bewegende elektrische Ladung erzeugt ein magnetisches Feld.

Jedes magnetische Feld wird durch bewegte Ladungen erzeugt.

Diese Aussage gilt für *Elektromagnete* und *Dauermagnete*. Fließt durch Elektromagnete ein Strom, ist damit eine Bewegung von Ladungen verbunden. In Dauermagneten sind die bewegten Ladungen die Elektronen, die immer um die Atomkerne kreisen und somit *Kreisströme* hervorrufen. Außerdem rotieren alle Elektronen um die eigene Achse. Diese Bewegung wird *Spin* genannt und erzeugt ebenfalls magnetische Felder. In vielen Stoffen heben sich die von den einzelnen Elektronen erzeugten Magnetfelder gegenseitig auf; in Dauermagneten nicht.

28.1 Feldlinien

28.1.1 Dauermagnet

Wie beim elektrischen Feld wird auch das Magnetfeld durch Feldlinien dargestellt. Wenn eine drehbar gelagerte Magnetnadel um einen Dauermagneten herumgeführt wird, stellt sie sich parallel zu den Feldlinien ein. Verbindet man alle Stellungen tangential, erhält man das Feldlinienbild eines Dauermagneten (Abb. 28.1).

Wird ein Dauermagnet geteilt, haben beide Teile dieselben magnetischen Eigenschaften wie der Ausgangsmagnet. Es gibt also, im Gegensatz zu den elektrischen Ladungen, keine einzelne magnetische Pole, sondern nur Dipole. Magnetische Feldlinien sind daher grundsätzlich geschlossene Linien. Die magnetischen Pole werden aus historischen Gründen *Nord-* und *Südpol* genannt. Auch die Erde besitzt ein Magnetfeld. Eine Magnetnadel wird sich auch hier parallel zu den Feldlinien einstellen. Der Pol der Nadel, der in die geographische Nordrichtung zeigt, heißt Nordpol.

Sind zwei Magnete vorhanden, ziehen sich ungleiche Pole an. Die magnetischen Feldlinien werden mit Richtungspfeilen versehen. Ihre Richtung ist willkürlich festgelegt. Im Außenraum eines Dauermagneten verlaufen sie vom Nordpol zum Südpol. Da die Feldlinien immer geschlossen sind, verlaufen sie im Inneren eines Dauermagneten vom Südpol zum Nordpol.

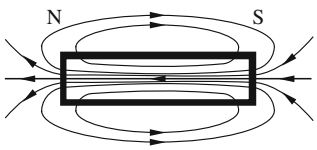


Abb. 28.1 Dauermagnet

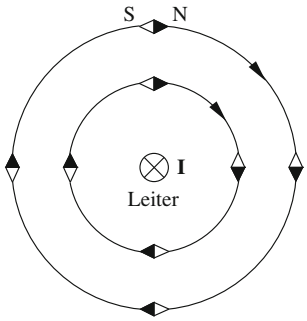


Abb. 28.2 Magnetfeld um einen Leiter

28.1.2 Stromdurchflossene Leiter

Bringt man eine Magnetnadel in die Nähe eines stromdurchflossenen Leiters und führt sie um den Leiter herum, dann ergibt die tangentiale Verbindung aller Einstellungen Kreise, in deren Zentrum der Leiter liegt. Die technische Stromrichtung ist durch das Kreuz angegeben, sie zeigt in die Zeichenebene hinein. Die Richtung der Feldlinien kann durch die *Rechtsschrauben-Regel* beschrieben werden (Abb. 28.2).

Dreht man eine rechtsgängige Schraube so, dass sie sich in der technischen Stromrichtung weiterbewegt, ist die Richtung der magnetischen Feldlinien durch die Drehrichtung gegeben.

Diese Richtung lässt sich auch mit Hilfe der rechten Hand bestimmen.

Der Daumen der rechten Hand wird in Richtung der technischen Stromrichtung gehalten. Die gekrümmten Finger zeigen dann die Richtung der Magnetfeldlinien an.

28.1.3 Stromdurchflossene Spule

Wenn stärkere Magnetfelder erzeugt werden sollen, werden *Spulen* verwendet. Wird ein Leiter auf einen zylindrischen Spulenkörper gewickelt, erhält man eine *Zylinderspule* (Abb. 28.3). Jeder Leiter erzeugt ein Magnetfeld. Zwischen den Leitern heben sich die von den einzelnen Leitern erzeugten Felder gegenseitig auf. Im Außenraum und im Inneren der Spule addieren sich die einzelnen Beiträge zum Gesamtfeld. Das in der Mitte der Spule entstehende Feld ist dann die Vektorsumme aller einzelnen Felder. In der Mitte der *Zylinderspule* kann das Feld bei großer Spulenlänge als homogen angesehen werden. Im Außenraum ist das Feldlinienbild gleich dem eines Dauermagneten. Man kann einer *Zylinderspule* somit auch einen Nordpol und einen Südpol zuordnen.

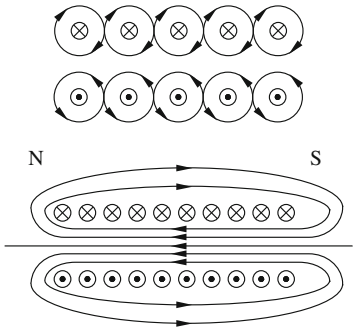


Abb. 28.3 Zylinderspule

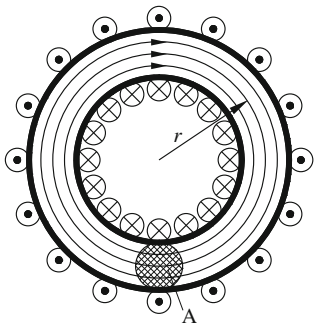


Abb. 28.4 Ringspule

Wird der zylindrische Spulenkörper zu einem Kreis gebogen, entsteht eine *Ringspule* (Abb. 28.4). Das hier aus der Überlagerung aller einzelnen Magnetfelder entstehende resultierende Feld hat die Eigenschaft, nur im Inneren des Spulenkerns vorhanden zu sein. Es gibt keine Stelle, an der die Feldlinien, die ja geschlossene Linien darstellen, austreten können.

28.1.4 Magnetfeld der Erde

Die Erde besitzt ein Magnetfeld. Eine Magnetnadel zeigt in die Richtung zum magnetischen Pol. Da der Nordpol der Magnetnadel in die geographische Nordrichtung zeigt, andererseits aber zum magnetischen Südpol weist, liegt der magnetische Südpol auf der Nordhalbkugel. Er befindet sich bei ungefähr 74° nördlicher Breite und 100° westlicher Länge nordwestlich von Nordgrönland. Der magnetische Nordpol auf der Südhalbkugel liegt nicht genau gegenüber. Er liegt bei ungefähr 72° südlicher Breite und 155° östlicher Länge in der Antarktis (Abb. 28.5). Die Abweichung des Erdmagnetfeldes von der geographischen Nord-Süd-Richtung ist die *Deklination*; sie ist abhängig vom Ort auf der Erde. In Deutschland beträgt diese Abweichung etwa 2° in westlicher Richtung. Die Feldlinien verlaufen – bis auf die Äquatorgegend – nicht parallel zur Erdoberfläche. Die Neigung zur Horizontalen wird *Inklination* genannt.

Das Magnetfeld der Erde ist nicht so symmetrisch, wie in Abb. 28.5 dargestellt. Durch Satellitenmessungen ist bekannt, dass das Magnetfeld durch den *Sonnenwind* stark deformiert

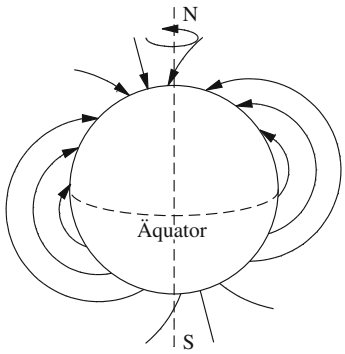


Abb. 28.5 Magnetfeld der Erde

wird. Der *Sonnenwind* besteht aus Wasserstoff- und Heliumionen sowie aus Elektronen. Diese Teilchen treffen mit einigen 100 km/s die Erde. Dadurch wird das Magnetfeld auf der Tagseite stark zusammengedrückt, auf der Nachtseite der Erde wird das Magnetfeld zu einem langen Schweif deformiert, der bis über die Mondbahn hinausgeht.

Die Ursache für das Erdmagnetfeld ist nicht genau bekannt. Wahrscheinlich rührt es von der Bewegung des ionisierten flüssigen Erdkerns her. Das Erdmagnetfeld ist in seiner Stärke und seiner Richtung nicht konstant. Vor etwa 4 Millionen Jahren war die Polarität entgegengesetzt. Dies kann aus der Richtung der Magnetisierung in alten Gesteinen festgestellt werden.

28.2 Magnetische Grundgrößen

28.2.1 Feldstärke

Bringt man in die Nähe eines stromdurchflossenen Leiters eine Magnetnadel und stellt sie senkrecht zur Richtung der Feldlinien, die wie konzentrische Kreise um den Leiter liegen, so wird die Nadel durch die Wechselwirkung mit dem Magnetfeld in Richtung der Magnetlinien gedreht. Es wird also ein Drehmoment M ausgeübt. Die Größe von M ist ein Maß für die Stärke des Magnetfeldes. Dieses wird mit H bezeichnet

$$M \sim H \tag{28.1}$$

Erhöht man den Strom I im Leiter, erhöht sich auch das Drehmoment und somit die Feldstärke

$$H \sim I \tag{28.2}$$

Werden n Leiter verwendet, hängt die Stärke des Drehmomentes von allen Strömen ab:

$$H \sim \sum_{i=1}^{n} I_i \tag{28.3}$$

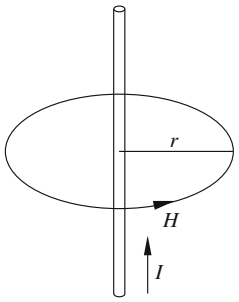


Abb. 28.6 Magnetfeld um einen Leiter

Die Summe aller Ströme, die von einer geschlossenen Feldlinie umgeben sind, ist die *Durchflutung*.

$$\textit{Durchflutung} \quad \Theta = \sum_{i=1}^{n} I_i \qquad \begin{array}{c|c} \Theta & I \\ \hline \mathrm{A} & \mathrm{A} \end{array} \qquad (28.4)$$

Zur Berechnung von Feldstärken kann das *Durchflutungsgesetz* benutzt werden.

Ist die Feldstärke nicht konstant, gilt für die Durchflutung:

$$\Theta = \oint \vec{H} d\vec{s} \qquad (28.5)$$

$$\textit{Durchflutungsgesetz} \quad \oint \vec{H} d\vec{s} = \sum_{i=1}^{n} I_i \qquad (28.6)$$

Für den Sonderfall einer konstanten Feldstärke ist

$$\Theta = H \cdot l. \qquad (28.7)$$

Der Weg l muss dabei geschlossen sein.

Zur Berechnung des Feldes in einer Ringspule nach Abb. 28.4 wird der mittlere Radius r gewählt. Der geschlossene Weg ist dann $l = 2\pi r$ und es gilt:

$$\textit{Feld in einer Ringspule} \quad H = \frac{IN}{2\pi r} \qquad (28.8)$$

Dieses Feld ist nicht konstant, sondern fällt nach außen mit $1/r$ ab. Ist jedoch der mittlere Radius groß gegenüber dem Windungsdurchmesser d, kann H innerhalb des Ringquerschnittes als konstant angenommen werden.

Ein weiteres Beispiel für die Anwendung von (28.6) ist die Berechnung des Magnetfeldes um einen geraden stromdurchflossenen Leiter. Das Magnetfeld ist auf Kreisbahnen um den Leiter konstant. Der nach dem Durchflutungsgesetz geforderte geschlossene Weg mit konstanter Feldstärke muss gleich dem Umfang eines Kreises sein. Es fließt nur ein Strom I. Nach (28.6) gilt

$$H \cdot 2\pi r = I \qquad (28.9)$$

$$\textit{Feld um einen Leiter} \quad H = \frac{I}{2\pi r} \qquad (28.10)$$

Beispiel

Ein massiver Kupferleiter ist nach Abb. 28.7 von einem Kupferrohr umschlossen. Diese Anordnung dient als Hin- und Rückleitung. In beiden Leitern fließt der gleiche Strom $I = 100\,\mathrm{A}$. Wie groß ist die Feldstärke als Funktion des Abstandes vom Mittelpunkt?

Zunächst wird der Bereich $0 < r < r_1$ betrachtet. Die Stromdichte S ist überall konstant.

$$\textit{Stromdichte} \qquad S_1 = \frac{I}{\pi \cdot r_1^2}$$

$$\textit{Durchflutung} \qquad \Theta = S_1 \cdot \pi r^2$$

Das Feld in einem beliebigen Abstand r wird mit dem Durchflutungsgesetz berechnet.

$$H = \frac{\Theta}{2\pi r} = \frac{S_1 \cdot \pi r^2}{2\pi r} = \frac{I \cdot \pi r^2}{\pi r_1^2 2\pi r} = \frac{Ir}{2\pi r_1^2} \qquad (28.11)$$

H nimmt linear mit r zu. Für $r = r_1$ ist die Grenze dieses Bereiches erreicht.

$$H_1 = \frac{I}{2\pi r_1} = \frac{100\,\mathrm{A}}{2\pi \cdot 0{,}01\,\mathrm{m}} = 1591\,\frac{\mathrm{A}}{\mathrm{m}}$$

Im Bereich $r_1 < r < r_2$ kommt kein weiterer Strom hinzu, die Durchflutung bleibt konstant ($Q = I$). H nimmt nach (28.10) mit $1/r$ ab bis zum Wert

$$H_2 = \frac{I}{2\pi r_2} = \frac{100\,\mathrm{A}}{2\pi \cdot 0{,}02\,\mathrm{m}} = 795\,\frac{\mathrm{A}}{\mathrm{m}}$$

Im Bereich $r_2 < r < r_3$ nimmt die Durchflutung wieder ab.

$$\Theta = \Theta_1 - I\frac{\pi(r^2 - r_2^2)}{\pi(r_3^2 - r_2^2)} \qquad (28.12)$$

$$H = \frac{1}{2\pi r} \cdot \left(I - I\frac{(r^2 - r_2^2)}{(r_3^2 - r_2^2)} \right) \qquad (28.13)$$

Für Abstände $r > r_3$ ist die Feldstärke $H = 0$. Die Abhängigkeit der Feldstärke vom Abstand ist in Abb. 28.7 unten aufgetragen. ◀

Soll das Feld in einer Zylinderspule berechnet werden, ist kein geschlossener Weg zu finden, auf dem die Feldstärke konstant ist. Der für die Berechnung notwendige geschlossene Weg wird in einzelne Stücke aufgeteilt. Auf diesen ist entweder H konstant, oder Beiträge von verschiedenen Wegelementen kompensieren sich. Wird der Umlaufweg nach Abb. 28.8 gelegt, gilt für die einzelnen Wegabschnitte:

Auf dem Teilstück 1 ist das Feld konstant und gleich dem gesuchten Feld, $H = \text{const}$.

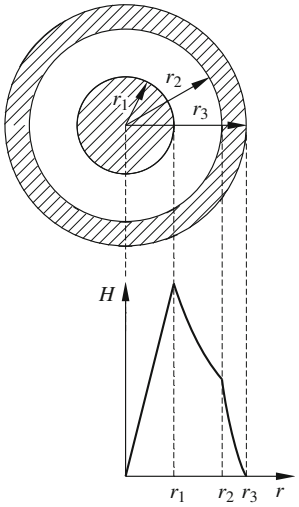


Abb. 28.7 Beispiel Magnetfeld um konzentrische Doppelleiter

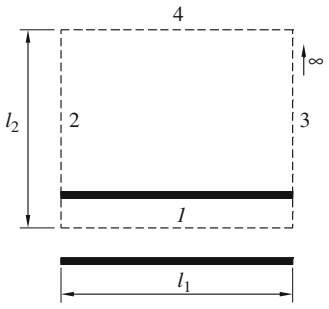


Abb. 28.8 Feld in einer Spule

Alle Beiträge, die auf dem Teilstück 2 entstehen, werden durch entgegengesetzte Beiträge auf dem Teilstück 3 kompensiert. Das Teilstück 4 ist so weit entfernt gedacht, dass dort $H = 0$ gilt. Das Wegintegral aus (28.6) ergibt dann

$$H_1 \cdot l_1 + H_2 \cdot l_2 + H_3 \cdot l_3 + H_4 \cdot l_4 = H_1 \cdot l_1 \tag{28.14}$$

Wird die Windungszahl mit N bezeichnet, so ist die Durchflutung

$$\Theta = NI \tag{28.15}$$

Feld in einer Zylinderspule

$$H = \frac{\Theta}{l} = \frac{NI}{l} \tag{28.16}$$

In den Randbereichen der Spule ist allerdings das Feld H_1 nicht konstant. Damit (28.16) trotzdem angewendet werden kann, muss die Spulenlänge groß sein gegenüber dem Durchmesser der Spule. Ist diese Bedingung nicht erfüllt, liegt also eine kurze Spule vor, ist das Feld nicht homogen und muss auf andere Weise ausgerechnet werden. Der Durchmesser der Spulenwicklung ist d, die Länge l.

kurze Zylinderspule

$$H_{\text{Mitte}} = \frac{NI}{\sqrt{l^2 + d^2}} \tag{28.17}$$

$$H_{\text{Rand}} = \frac{NI}{2\sqrt{l^2 + d^2}} \tag{28.18}$$

Wenn $l \gg d$ ist, kann d im Nenner von (28.17) vernachlässigt werden und es folgt (28.16).

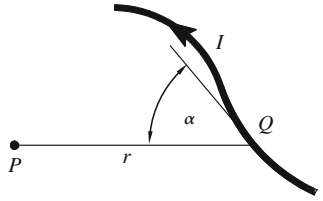


Abb. 28.9 Biot-Savart'sches Gesetz

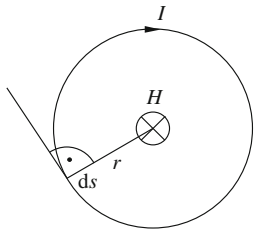


Abb. 28.10 Feld im Zentrum eines Kreisstromes

Das Gesetz, nach dem Magnetfelder bei beliebiger Leitergeometrie zu berechnen sind, ist das *Biot-Savart'sche Gesetz*. Jedes Teilstück ds eines stromdurchflossenen Leiters, durch den der Strom I fließt, trägt zum Feld in einem beliebigen Punkt P bei. Der Beitrag eines Teilstückes ds im Punkt Q ist (Abb. 28.9)

Biot-Savart'sches Gesetz

$$dH = \frac{I \cdot ds}{4\pi r^2} \sin\alpha \tag{28.19}$$

Um das gesamte Feld im Punkt P zu bestimmen, müssen alle einzelnen Beiträge zusammengefasst werden. Ist ein Leiter zum Beispiel zu einem Kreis gebogen, lässt sich das Magnetfeld genau im Mittelpunkt mit (28.19) berechnen (Abb. 28.10). Für jedes Leiterelement ds gilt $\alpha = 90°$ und $\sin\alpha = 1$. Der Abstand zum Punkt P (Zentrum) ist überall gleich. Die Gesamtlänge der Strombahn ist $2\pi r$.

Zentrum eines Kreisstromes

$$H = \oint dH = \frac{I \cdot 2\pi r}{4\pi r^2} = \frac{I}{2r} \tag{28.20}$$

Für einen Punkt P auf der Mittelpunktsachse ist die Berechnung etwas komplizierter (Abb. 28.11).

$$H = \frac{I}{2R} \sin^3\beta \tag{28.21}$$

mit $\sin\beta = \frac{R}{r}$, $r = \sqrt{R^2 + l^2}$.

Achse eines Kreisstromes

$$H = \frac{IR^2}{2(\sqrt{R^2 + l^2})^3} \tag{28.22}$$

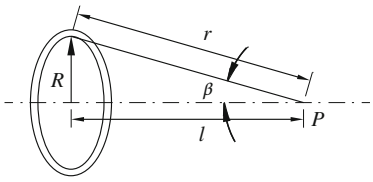


Abb. 28.11 Feld auf der Achse eines Kreisstromes

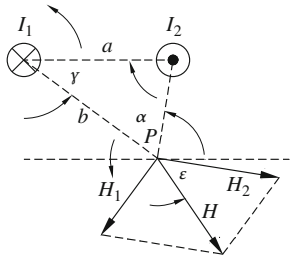


Abb. 28.12 Beispiel Magnetfeld von 2 Leitern

Beispiel

Zwei parallel stromdurchflossene Leiter mit einem Abstand von $a = 45\,\text{mm}$ führen nach Abb. 28.12 die Ströme $I_1 = 90\,\text{A}$, $I_2 = 70\,\text{A}$. Der angegebene Punkt P ist von den Leitern $b = 50\,\text{mm}$, $c = 30\,\text{mm}$ entfernt. Wie groß ist die Feldstärke im Punkt P? Welche Richtung hat sie zur Verbindungslinie der beiden Leiter?

Die von den einzelnen Strömen erzeugten Magnetfelder stehen senkrecht auf den jeweiligen Verbindungslinien zum Punkt P. Die Beträge berechnen sich nach (28.10).

$$H_1 = \frac{90\,\text{A}}{2\pi \cdot 0{,}05\,\text{m}} = 286{,}5\,\frac{\text{A}}{\text{m}};$$
$$H_2 = \frac{70\,\text{A}}{2\pi \cdot 0{,}03\,\text{m}} = 371{,}4\,\frac{\text{A}}{\text{m}}$$

Diese Komponenten können im entsprechenden Maßstab unter dem richtigen Winkel gezeichnet werden. Das resultierende Feld kann aus der Zeichnung abgelesen werden zu $H = 350\,\frac{\text{A}}{\text{m}}$. Der gesuchte Winkel kann ebenfalls abgelesen werden: $\varepsilon = 56°$.

Soll die Lösung rechnerisch gefunden werden, müssen zunächst die Winkel im Dreieck PI_1I_2 berechnet werden. Dies kann durch die Anwendung von Sinus- und Kosinussatz erfolgen.

$$\cos\gamma = \frac{a^2 + b^2 - c^2}{2ab} = 0{,}805$$
$$\gamma = 36{,}37°; \quad \alpha = 62{,}73°, \quad \beta = 80{,}9°$$

Der Winkel zwischen H_1 und H_2 ist $180 - \alpha = 117{,}3°$, für die resultierende Feldstärke H gilt

$$H = \sqrt{286{,}5^2 + 371{,}4^2 - 2 \cdot 286{,}5 \cdot 371{,}4 \cdot \cos 62{,}73°}\,\frac{\text{A}}{\text{m}}$$
$$= 350\,\frac{\text{A}}{\text{m}}$$

Für den Winkel zwischen H und H_2 gilt

$$\sin\varphi = \frac{286{,}5}{350} \cdot \sin 62{,}73°; \quad \varphi = 46{,}7°$$
$$\varepsilon = (90° - 80{,}94°) + 46{,}7° = 55{,}76° \quad \blacktriangleleft$$

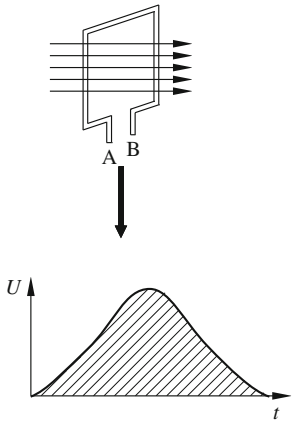


Abb. 28.13 Spannungsstoß

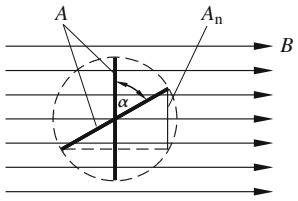


Abb. 28.14 Magnetischer Fluss

28.2.2 Fluss, Flussdichte

Die Gesamtheit aller Feldlinien nennt man den *magnetischen Fluss* Φ. Diese Größe kann gemessen werden, indem eine geschlossene Leiterschleife aus dem Magnetfeld herausgezogen wird. Während der Bewegung der Leiterschleife entsteht zwischen den Punkten A und B eine zeitlich veränderliche Spannung. Die Fläche unter der Spannungskurve wird als *Spannungsstoß* bezeichnet (Abb. 28.13).

Dieser Spannungsstoß wird proportional zur Flussänderung $\Delta\Phi$ durch die Leiterschleife gesetzt. Der magnetische Fluss wird in *Weber* (Wb) gemessen.

$$\textit{Fluss} \quad \phi = \int u\,dt \qquad \begin{array}{c|c|c} \phi & U & t \\ \hline \text{Wb} = \text{V s} & \text{V} & \text{s} \end{array} \qquad (28.23)$$

Wird der Fluss auf eine Flächeneinheit bezogen, ergibt sich die *Flussdichte* B.

$$\textit{Flussdichte} \quad B = \frac{\phi}{A_n} \qquad \begin{array}{c|c|c} B & \phi & A \\ \hline \text{T} = \frac{\text{V s}}{\text{m}^2} & \text{V s} & \text{m}^2 \end{array} \qquad (28.24)$$

Die Einheit ist 1 *Tesla*.

Die Fläche A_n ist dabei die vom Magnetfeld durchsetzte Fläche. Steht das Magnetfeld nicht senkrecht auf der Fläche, gilt nach Abb. 28.14

$$A_n = A \cdot \cos\alpha \qquad (28.25)$$

Zwischen der magnetischen Flussdichte und der magnetischen Feldstärke gilt im Vakuum die Beziehung

$$\vec{B} = \mu_0 \cdot \vec{H} \tag{28.26}$$

Die Konstante m_0 ist die magnetische Feldkonstante oder auch Permeabilität des Vakuums.

magnetische Feldkonstante

$$\mu_0 = 4\pi \cdot 10^{-7} \frac{\mathrm{V\,s}}{\mathrm{A\,m}} \tag{28.27}$$

Beispiel

Eine lange Zylinderspule wird von einem Strom I durchflossen. Wie groß sind das magnetische Feld und die Flussdichte im Inneren der Spule ($N = 1000$, $I = 5\,\mathrm{A}$, $l = 1\,\mathrm{m}$)?

$$H = \frac{NI}{l} = \frac{1000 \cdot 5\,\mathrm{A}}{1\,\mathrm{m}} = 5000\,\frac{\mathrm{A}}{\mathrm{m}}$$

$$B = \mu_0 H = 4\pi \cdot 10^{-7}\,\frac{\mathrm{V\,s}}{\mathrm{A\,m}} \cdot 5000\,\frac{\mathrm{A}}{\mathrm{m}} = 6{,}28 \cdot 10^{-3}\,\mathrm{T}$$

◀

Die elektrische Feldkonstante und die magnetische Feldkonstante bestimmen die Ausbreitung von elektromagnetischen Wellen im Vakuum.

$$\textit{Ausbreitungsgeschwindigkeit} \quad c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{28.28}$$

Für die Lichtgeschwindigkeit im Vakuum gilt entsprechend der SI Definition (s. Abschn. 11.2) $c = 299.792.458\,\frac{\mathrm{m}}{\mathrm{s}}$. Somit lässt sich hieraus die elektrische Feldkonstante bestimmen zu

$$\varepsilon_0 = \frac{1}{c^2 \cdot \mu_0} = 8{,}8541878 \cdot 10^{-12}\,\frac{\mathrm{A\,s}}{\mathrm{V\,m}} \tag{28.29}$$

28.3 Kräfte im Magnetfeld

28.3.1 Kräfte auf bewegliche Ladungsträger

Im Gegensatz zu elektrischen Feldern, in denen auf Ladungen immer eine Kraft ausgeübt wird, werden im Magnetfeld Kräfte nur auf bewegte Ladungen ausgeübt. Sie dürfen sich dabei nicht parallel zu den magnetischen Feldlinien bewegen (Abb. 28.15).

$$\textit{Lorentzkraft} \quad F_L = qvB \cdot \sin\varphi \tag{28.30}$$

Der Winkel φ ergibt sich aus der Richtung der Geschwindigkeit v und der Richtung des Magnetfeldes B. Mit dieser Gleichung lässt sich nur der Betrag der Lorentzkraft berechnen. Die Richtung kann mit Hilfe der *Drei-Finger-Regel* der rechten Hand bestimmt werden. Die drei Finger werden rechtwinklig zueinander gestellt.

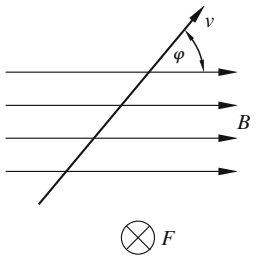


Abb. 28.15 Lorentzkraft

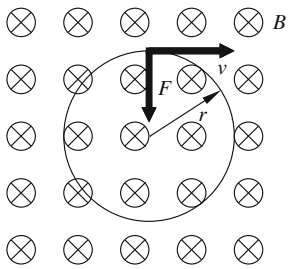


Abb. 28.16 Kreisbahn unter dem Einfluss der Lorentzkraft

Der Daumen zeigt in Richtung der Geschwindigkeit, der Zeigefinger in Richtung des Magnetfeldes und der Mittelfinger bei positiver Ladung in Richtung der Kraft; bei negativer Ladung entgegengesetzt zur Kraft.

Die drei Vektoren $\vec{v}$, $\vec{B}$ und $\vec{F}$ stehen alle senkrecht aufeinander. Bewegt sich ein Elektron (negative Ladung) in einem Magnetfeld nach Abb. 28.16, stehen die Feldlinien immer senkrecht auf dem Vektor der Geschwindigkeit des Elektrons. Damit muss auch die Kraft immer senkrecht auf der Geschwindigkeit stehen. Der Kraftvektor liegt immer in der Zeichenebene. Die Bewegung des Elektrons unter dem Einfluss einer zu jedem Zeitpunkt senkrecht stehenden Kraft ist eine Kreisbahn. Die Kraft ist dann zum Mittelpunkt gerichtet, die Richtung der momentanen Geschwindigkeit ist die Tangente an den Kreis. Der Radius der Kreisbahn ergibt sich aus der Überlegung, dass die Zentrifugalkraft gleich der Lorentzkraft ist.

$$\frac{mv^2}{r} = e_0 vB \tag{28.31}$$

$$r = \frac{mv}{e_0 B} \tag{28.32}$$

28.3.2 Stromdurchflossener Leiter

Auch in einem stromdurchflossenen Leiter werden Ladungen bewegt. Bringt man diesen Leiter in ein äußeres Magnetfeld, wird auch hier eine Kraft auf die Ladungen und damit auf den Leiter ausgeübt (Abb. 28.17). Da die Ladungen negativ sind und die technische Stromrichtung der tatsächlichen Richtung der Elektronenbewegung entgegengesetzt ist, heben sich beide

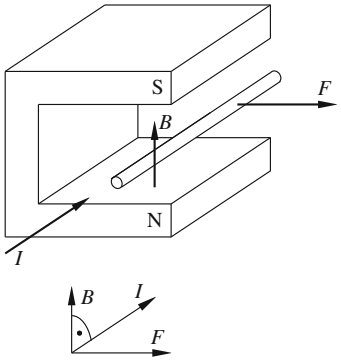


Abb. 28.17 Kraft auf Leiter im Magnetfeld

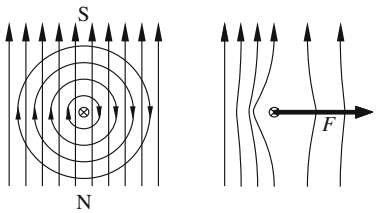


Abb. 28.18 Resultierendes Feld um einen Leiter

negativen Vorzeichen auf. Die Richtung der Kraft kann mit der Drei-Finger-Regel bestimmt werden.

> Zeigt der Daumen in Stromrichtung und der Zeigefinger in Richtung des Magnetfeldes, dann gibt der Mittelfinger die Kraftrichtung an.

Die Ermittlung der Kraftrichtung ist auch unter dem Namen *Motorregel* bekannt. Hier wird allerdings die *linke Hand* benutzt. Sie wird so in das Magnetfeld gehalten, dass die magnetischen Feldlinien in die Innenfläche der linken Hand eintreten und die Fingerspitzen in Stromrichtung zeigen. Der abgespreizte Daumen zeigt dann die Kraftrichtung an.

Das Magnetfeld des Dauermagneten und das durch den Strom im Leiter erzeugte Magnetfeld überlagern sich. Auf der linken Seite von Abb. 28.18 sind das durch den stromführenden Leiter und das durch den Dauermagneten erzeugte Feld gezeichnet, rechts das daraus entstehende resultierende Feld. Die Richtung der Kraft auf den Leiter ist aus diesem Feldlinienbild zu entnehmen. Die Kraft ist immer zu kleineren Feldstärken, also zu weiter auseinander liegenden Feldlinien, gerichtet. Die Größe der Kraft hängt neben der Flussdichte B und der Stromstärke I auch noch von der Länge des Leiters l im Magnetfeld und vom Winkel, den der Leiter mit den Magnetfeldlinien einschließt, ab. Dies folgt direkt aus (28.30).

Kraft im Magnetfeld
$$F = lIB \sin\varphi \tag{28.33}$$

Die Kraft wirkt immer senkrecht zu der Fläche, die durch das Magnetfeld und die Stromrichtung definiert wird. Sind Leiter und Magnetfeld senkrecht zueinander, dann ist die Kraft

$$F = lIB \tag{28.34}$$

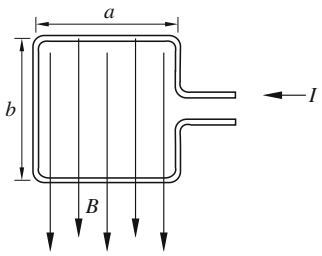


Abb. 28.19 Leiterschleife im Magnetfeld

28.3.3 Magnetisches Moment

In einem äußeren Magnetfeld wird auf einen weiteren Magneten ein Drehmoment ausgeübt. Eine Magnetnadel wird sich in Richtung der Feldlinien einstellen. Der Quotient aus Drehmoment und Flussdichte wird *magnetisches Moment* genannt.

magnetisches Moment

$$|m| = \frac{|M|}{|B|} \qquad \begin{array}{c|c|c} m & M & B \\ \hline \mathrm{A\,m^2} & \mathrm{N\,m = V\,A\,s} & \frac{\mathrm{V\,s}}{\mathrm{m^2}} \end{array} \tag{28.35}$$

Das magnetische Moment ist ein Vektor, der vom Süd- zum Nordpol zeigt. Eine vom Strom durchflossene Spule erfährt in einem Magnetfeld ein Drehmoment, das versucht, die Schleifenfläche senkrecht zu den Magnetfeldlinien zu stellen. Die rechteckige Leiterschleife und das Magnetfeld sind in Abb. 28.19 dargestellt. Die Kantenlänge b liegt parallel zum Feld B, daher wird auf diesen Teil des Leiters keine Kraft ausgeübt. Auf die Rechteckseiten der Länge a wirken gleichgroße, entgegengesetzte Kräfte $F = IBa$. Die Schleife dreht sich um die Achse bei $b/2$.

Drehmoment
$$M = 2 \cdot I \cdot B \cdot a \cdot \frac{b}{2} = I \cdot B \cdot a \cdot b = I \cdot B \cdot A \tag{28.36}$$

magnetisches Moment
$$m = \frac{M}{B} = I \cdot A \tag{28.37}$$

28.3.4 Kräfte zwischen zwei parallelen Leitern

Befindet sich ein stromführender Leiter in einem Magnetfeld, welches durch einen anderen Leiter hervorgerufen wird, muss auf den Leiter ebenfalls eine Kraft ausgeübt werden. Das Magnetfeld am Ort von Leiter 2 ist nach (28.10) und Abb. 28.20:

$$H_1 = \frac{I_1}{2\pi d} \tag{28.38}$$

Mit (28.34) gilt für die Kraft auf den Leiter 2, die durch das Magnetfeld des Leiters 1 hervorgerufen wird:

$$F_{12} = \mu_0 H_1 I_2 l = \frac{\mu_0 I_1 I_2 l}{2\pi d} \tag{28.39}$$

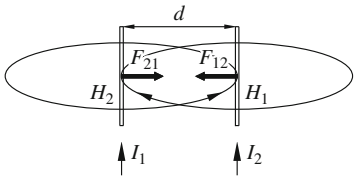


Abb. 28.20 Kräfte zwischen zwei Leitern

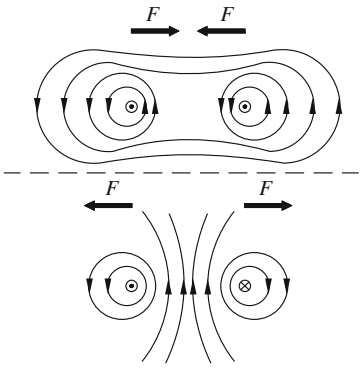


Abb. 28.21 Feldlinien um zwei Leiter

Die Richtung dieser Kraft kann nach der *Rechte-Hand-Regel* bestimmt werden und ist in Abb. 28.20 eingezeichnet. Durch das Magnetfeld des Leiters 2 wird am Ort des Leiters 1 ein Magnetfeld

$$H_2 = \frac{I_2}{2\pi d} \tag{28.40}$$

erzeugt. Hierdurch wird nach der gleichen Überlegung die Kraft

$$F_{21} = \mu_0 H_2 I_1 l = \frac{\mu_0 I_2 I_1 l}{2\pi d} \tag{28.41}$$

ausgeübt.

Die Beträge sind gleich, die Richtungen entgegengesetzt. Es gilt die Aussage:

> Zwei parallele vom Strom in gleichen Richtungen durchflossene Leiter ziehen sich an.

Die von beiden Strömen hervorgerufenen Magnetfelder überlagern sich.

Zwischen den beiden Leitern sind die Richtungen der Feldlinien entgegengesetzt, das resultierende Feld wird hier geschwächt. Sind beide Ströme gleich groß, verschwindet das Feld genau in der Mitte. Die Kraft wirkt in Richtung des schwächeren Magnetfeldes. Die beiden Leiter ziehen sich also an. Sind die Stromrichtungen in beiden Leitern entgegengesetzt, wird das resultierende Feld zwischen den Leitern verstärkt. Die Kräfte sind jetzt so gerichtet, dass sich beide Leiter abstoßen (Abb. 28.21).

> Zwei parallele vom Strom in entgegengesetzten Richtungen durchflossene Leiter stoßen sich ab.

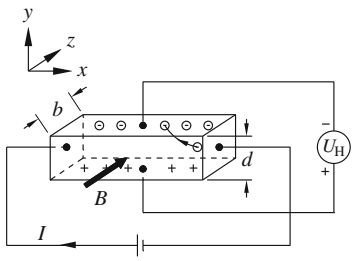


Abb. 28.22 Hall-Effekt

28.3.5 Hall-Effekt

Wird ein leitendes Plättchen in ein senkrecht zu seiner Oberfläche stehendes Magnetfeld gebracht und eine Spannungsquelle nach Abb. 28.22 angeschlossen, fließt im Plättchen ein Strom in der angegebenen Richtung. Auf diese bewegten Ladungsträger wird durch das senkrecht zur Bewegungsrichtung stehende Magnetfeld die Lorentzkraft nach (28.30) ausgeübt. Bei der angegebenen Anordnung wird auf die Ladungsträger eine solche Kraft ausgeübt, dass die negativen Ladungsträger zur Oberseite des Plättchens abgelenkt werden und somit an der Unterseite ein Überschuss an positiven Ladungen entsteht. Durch die Ladungstrennung wird ein elektrisches Feld zwischen Ober- und Unterseite aufgebaut. Es entsteht die *Hallspannung* U_H. Im Gleichgewicht sind die Lorentzkraft und die durch das elektrische Feld auf die Ladungen wirkende Kraft entgegengesetzt gleich groß.

$$F_{Ly} = -e v_x B_z \tag{28.42}$$

$$F_{\text{el}} = -eE = -e\frac{U_H}{d} \tag{28.43}$$

$$U_H = d\, v_x B_z \tag{28.44}$$

Mit (25.9) ist

$$v_x = \frac{I}{neA} = \frac{I}{nebd} \tag{28.45}$$

Hallspannung

$$U_H = \frac{1}{ne} \cdot \frac{dIB}{bd} = R_H \cdot \frac{IB}{b} \tag{28.46}$$

Hall-Koeffizient

$$R_H = \frac{1}{ne} \quad \begin{array}{c|c|c} R_H & n & e \\ \hline \frac{\text{m}^3}{\text{A s}} & \frac{1}{\text{m}^3} & \text{A s} \end{array} \tag{28.47}$$

Durch Messung dieser Größe können die Ladungsträgerdichte und die Art des Leitungsmechanismus (Elektronen oder Löcherleitung) bestimmt werden. Sind die Ladungsträger Elektronen, ist der Hall-Koeffizient negativ; bei Löcherleitung ist er positiv (Tab. 28.1). Da die Hallspannung direkt proportional zur Flussdichte B ist, werden *Hall-Sonden* zur Messung von Magnetfeldern verwendet. Bei bekannter Geometrie und bekanntem und konstantem Strom durch die Hallsonde ist die Hallspannung proportional zum Magnetfeld.

Eine weitere Anwendung findet der *Hallgenerator* in der Messtechnik. Er wird dann eingesetzt, wenn das Produkt zweier

Tab. 28.1 Hall-Koeffizienten

Werkstoff	R_H [10^{-11} m^3/A s]
Cu	−5,5
Al	−3,6
Ag	−9,0
Au	−7,5
Cd	+6,0
Sn	+14,0
InAs	-10^{-7}

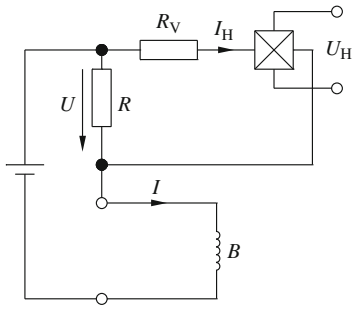


Abb. 28.23 Hallgenerator

Größen gemessen werden soll. In Abb. 28.23 ist die Prinzip-Schaltung angegeben. Der Strom I_H wird durch die Verbraucherspannung U festgelegt. Dabei muss dieser Strom klein gegen den Strom durch den Verbraucher gehalten werden. Dies wird durch den geeigneten Vorwiderstand R_V erreicht. Der Strom durch den Verbraucher wird durch eine in Reihe geschaltete Spule geschickt, die das für den Halleffekt erforderliche Magnetfeld erzeugt. Die Hallspannung ist proportional zum Produkt aus Strom und Spannung des Verbrauchers, also auch zu seiner Leistung

$$U_H \sim I_H \cdot B \sim U \cdot I \sim P. \tag{28.48}$$

Der Hallgenerator ist im Prinzip eine Multiplizierschaltung für zwei Ströme. Zur Leistungsmessung muss ein Strom als Messgröße an Stelle der Verbraucherspannung genommen werden.

28.4 Energie des Magnetfeldes

Analog zu (27.61) gilt im Falle eines Magnetfeldes für die Energiedichte

$$\textit{Energiedichte} \quad w = \int H dB \tag{28.49}$$

Wenn der Zusammenhang zwischen Flussdichte B und Feldstärke H nicht linear ist, kann dieses Integral nicht in einfacher Weise berechnet werden.

28.5 Materie im Magnetfeld

Wird Materie in ein magnetisches Feld gebracht, ändert sich, analog zur Materie im elektrischen Feld, die Flussdichte B. Diese Änderung wird relativ zur Flussdichte im Vakuum angegeben.

$$\textit{relative Permeabilität} \quad \mu_r = \frac{B_M}{B_0} = \frac{B_M}{\mu_0 H_0} \tag{28.50}$$

$$\textit{Induktionsflussdichte} \quad B_M = \mu_r B_0 = \mu_r \mu_0 H_0 \tag{28.51}$$

Die dimensionslose Zahl m_r ist die *relative magnetische Feldkonstante* oder *relative Permeabilität*. Die durch die Materie zusätzlich hervorgerufene Flussdichte ist die

magnetische Polarisation

$$J = B_M - B_0 \tag{28.52}$$

$$J = (\mu_r - 1)B_0 = (\mu_r - 1)\mu_0 H_0 \tag{28.53}$$

magnetische Suszeptibilität

$$\chi_M = (\mu_r - 1) \tag{28.54}$$

$$J = \chi_M B_0 = \chi_M \mu_0 H_0 \tag{28.55}$$

Die *magnetische Suszeptibilität* kann gemessen werden, indem der zu untersuchende Stoff in ein inhomogenes Magnetfeld gebracht wird. Stoffe mit negativer Suszeptibilität werden zu Orten geringerer Feldstärke herausgedrückt, Stoffe mit positiver Suszeptibilität zu Orten größerer Feldstärke gezogen (Abb. 28.24).

Die magnetische Polarisation gibt die Änderung der magnetischen Flussdichte an. Die Änderung der magnetischen Feldstärke wird durch die *Magnetisierung* M beschrieben.

$$\textit{Magnetisierung} \quad M = H_M - H_0 \tag{28.56}$$

$$M = \frac{B_M}{\mu_0} - H_0 \tag{28.57}$$

$$M = \mu_r H_0 - H_0 = (\mu_r - 1)H_0 \tag{28.58}$$

$$M = \chi_M H_0 \tag{28.59}$$

$$\textit{Polarisation} \quad J = \mu_0 M \tag{28.60}$$

Die Magnetisierung ist bei vielen Stoffen proportional zur magnetischen Feldstärke H und somit ist für diese Stoffe χ_M

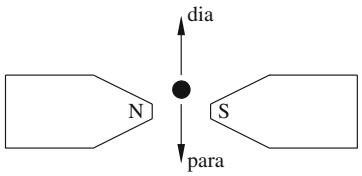


Abb. 28.24 Messung der Suszeptibilität

Tab. 28.2 Einteilung magnetischer Werkstoffe

Diamagnetisch	$\mu_r < 1,\ \chi_M < 0$	$-10^{-4} < \chi_M < -10^{-9}$
Paramagnetisch	$\mu_r > 1,\ \chi_M > 0$	$10^{-6} < \chi_M < 10^{-2}$
Ferromagnetisch	$\mu_r \gg 1,\ \chi_M \gg 1$	$\mu_r > 500$

konstant. Ausnahme hiervon bilden die nichtlinearen magnetischen Werkstoffe, wie z. B. die ferromagnetischen Werkstoffe. Nach ihrem Verhalten in Magnetfeldern werden Werkstoffe in die in der Tab. 28.2 aufgeführten Hauptklassen eingeteilt.

Das unterschiedliche Verhalten der Materie im Magnetfeld ist auf die Elektronenstruktur und die damit verbundenen magnetischen Momente zurückzuführen. Jedes Elektron erzeugt durch seine Bewegung magnetische Momente. Es gibt zwei unterschiedliche Bewegungen der Elektronen, die magnetische Momente erzeugen. Zum einen ist es die Bewegung um den Atomkern. Hierdurch wird das *magnetische Bahnmoment* erzeugt. Zum anderen rotiert das Elektron um seine eigene Achse. Diese Rotation wird *Spin* genannt. Durch diese Rotation entsteht das *magnetische Spinmoment*. In jedem Stoff liegt eine Überlagerung von Bahnmoment und Spinmoment vor. Die Magnetisierung M ergibt sich als Summe aller magnetischen Momente.

Ein auf einer Kreisbahn um den Atomkern kreisendes Elektron kann als kreisförmiger Strom betrachtet werden, da Ladung transportiert wird. Bei einer Umlaufzeit T ist dieser

$$\textit{Kreisstrom} \quad I = \frac{e_0}{T} = \frac{\omega e_0}{2\pi} \tag{28.61}$$

Das magnetische Moment dieses Kreisstromes ist nach (28.37):

magnetisches Moment

$$m_B = IA = \frac{\omega e_0 \pi r^2}{2\pi} = \frac{e_0 \omega r^2}{2} \tag{28.62}$$

In der Quantenmechanik der Physik wird gezeigt, dass die Beziehung

$$m_{\mathrm{el}} \omega r^2 = \frac{h}{2\pi} \tag{28.63}$$

gilt, mit der Naturkonstanten h, dem Planck'schen Wirkungsquant und der Elektronenmasse m_{el}. Gleichung (28.62) kann hiermit umgeformt werden zu

Bohr'sches Magneton

$$m_B = \frac{e_0 h}{4\pi m_{\mathrm{el}}} \tag{28.64}$$

$$m_B = \frac{1{,}602 \cdot 10^{-19}\,\mathrm{A\,s} \cdot 6{,}626 \cdot 10^{-34}\,\mathrm{J\,s}}{4\,\pi \cdot 9{,}109 \cdot 10^{-31}\,\mathrm{kg}} = 9{,}274 \cdot 10^{-24}\,\mathrm{A\,m^2} \tag{28.65}$$

Je nach Anordnung und Eigenschaften der magnetischen Momente in der Materie reagiert diese in unterschiedlicher Weise auf äußere Magnetfelder.

Tab. 28.3 Suszeptibilität von diamagnetischen Stoffen

Material	χ_M
N_2 (gasförmig)	$-6{,}75 \cdot 10^{-9}$
Bi	$-1{,}5 \cdot 10^{-4}$
Au	$-2{,}9 \cdot 10^{-5}$
Cu	$-1 \cdot 10^{-5}$
Wasser	$-7 \cdot 10^{-6}$

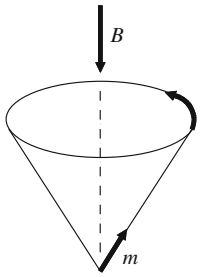


Abb. 28.25 Ursache des Diamagnetismus

28.5.1 Diamagnetismus

Der Diamagnetismus ist bei allen Stoffen vorhanden, aber bei bestimmten Konfigurationen der magnetischen Momente der einzelnen Atome wird er von anderen magnetischen Eigenschaften überdeckt. Zu beobachten ist der Diamagnetismus bei allen Stoffen mit abgeschlossenen Elektronenschalen. Ohne äußeres Magnetfeld kompensieren sich die magnetischen Momente aller Elektronen eines Atoms. Unter dem Einfluss eines äußeren Feldes führen die Momente, analog zur Bewegung eines Kreisels, *Präzessionsbewegungen* aus. Dies ist eine Bewegung, die ein rotierender Körper durchführt, an dem eine äußere Kraft angreift. Der Körper weicht senkrecht zur Kraftrichtung aus. Unter dem Einfluss des Magnetfeldes B laufen die Spitzen der einzelnen magnetischen Momente auf Kreisbahnen um die Richtung von B. Hierdurch entsteht ein zusätzliches Moment, welches der Richtung des äußeren Feldes entgegengesetzt ist und es schwächt. Somit muss die Suszeptibilität negativ sein. Der Diamagnetismus wird also durch das äußere Feld induziert. Die Suszeptibilität diamagnetischer Stoffe ist nicht temperaturabhängig (Abb. 28.25 und Tab. 28.3).

28.5.2 Paramagnetismus

Diese Stoffe haben normalerweise eine ungerade Anzahl von Elektronen, oder die Elektronenschalen sind nicht gefüllt. Somit verbleibt für jedes Atom ein resultierendes magnetisches Moment. Ohne äußeres Magnetfeld sind alle diese einzelnen Momente statistisch in alle Richtungen verteilt und das resultierende Moment verschwindet. Unter dem Einfluss eines äußeren Magnetfeldes richten sich die atomaren Momente in Feldrichtung aus und verstärken das Magnetfeld. Die Suszeptibilität ist daher positiv. Da die thermische Bewegung der Atome, und damit auch die der atomaren Momente, mit steigender Temperatur

Tab. 28.4 Suszeptibilität von paramagnetischen Stoffen

Material	c_M
O_2 (flüssig)	$3{,}6 \cdot 10^{-3}$
O_2 (gasförmig)	$1{,}5 \cdot 10^{-6}$
Pt	$2{,}5 \cdot 10^{-4}$
Al	$2{,}4 \cdot 10^{-5}$

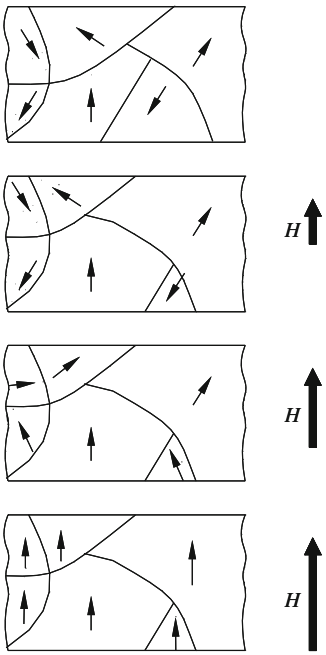


Abb. 28.26 Weiß'sche Bezirke

zunimmt, wird dieser Ordnungszustand der Momente im äußeren Feld bei höherer Temperatur schlechter. Daher muss die Suszeptibilität von der Temperatur abhängig sein (Tab. 28.4).

$$\textit{Curie-Gesetz} \quad \chi_M = \frac{C}{T} \tag{28.66}$$

Die Größe C heißt *Curie-Konstante* und ist materialabhängig.

28.5.3 Ferromagnetismus

Ferromagnetische Stoffe haben, wie paramagnetische Materialien, ein resultierendes magnetisches Moment pro Atom. Durch gegenseitige Kräfte stellt sich ein Zustand ein, bei dem in größeren Bereichen ohne äußeres Magnetfeld alle Momente in einer Richtung ausgerichtet sind. Diese Bereiche, die nach ihrem Entdecker benannten *Weiß'schen Bezirke*, können Durchmesser bis zu 1 mm haben. Die Wände, die diese Bezirke trennen, sind die *Bloch-Wände*. Ohne äußeres Feld sind die Momente der einzelnen Bezirke statistisch verteilt, der Werkstoff ist dann nach außen unmagnetisch (Abb. 28.26).

Wird ein solches Material in ein Magnetfeld gebracht und dieses langsam vergrößert, geschieht die Magnetisierung des Materials

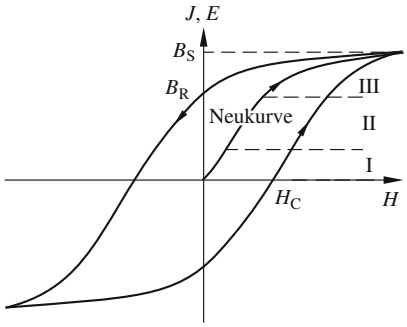


Abb. 28.27 Hysterese

in mehreren Schritten. Zunächst werden sich die Bezirke, deren Momente in etwa in Richtung des Magnetfeldes liegen, vergrößern. Es werden also die Wände zwischen den Weiß'schen Bezirken verschoben. Bei kleinen Feldstärken ist dieser Vorgang *reversibel* (Bereich I in Abb. 28.27), bei Abschalten des Feldes werden die Verschiebungen rückgängig gemacht. Wird die Feldstärke weiter vergrößert, werden diese Wandverschiebungen *irreversibel* (Bereich II). Werden die Feldstärken noch größer, so werden schließlich alle Momente der einzelnen Bezirke gedreht, bis sie in der Endstellung parallel zum angelegten Magnetfeld liegen (Bereich III). Die Magnetisierung hat jetzt ihren *Sättigungswert* M_s erreicht.

28.5.3.1 Magnetisierungskurve

Diese komplizierten Vorgänge verlaufen nicht linear mit steigender Feldstärke, daher ist auch die Magnetisierungskurve keine Gerade. Üblicherweise wird nicht die Magnetisierung M aufgetragen, sondern die Polarisation J oder die Flussdichte B. Da für diese Stoffe

$$\mu_r \gg 1 \text{ ist:} \quad \chi_M \approx \mu_r \text{ und } J \approx B \tag{28.67}$$

Wird die Feldstärke nach Erreichen der *Sättigungsmagnetisierung* verringert, verläuft die Magnetisierungskurve auf einem anderen Weg, da die Vorgänge teilweise irreversibel sind. Wird das Magnetfeld umgepolt und dann wieder bis zur Sättigung magnetisiert, erhält man die in Abb. 28.27 aufgezeigte Kurve. Aus dieser *Hysteresekurve* lassen sich folgende Größen ablesen:

Die verbleibende Flussdichte bei abgeschaltetem Magnetfeld wird als *Remanenzflussdichte* oder *Remanenz* B_r bezeichnet. Sie ist ein Maß für die Stärke von Dauermagneten. Die magnetische Feldstärke, die notwendig ist, um den Stoff zu entmagnetisieren (Wert $B = 0$ einzustellen), wird als *Koerzitivfeldstärke* H_c bezeichnet.

Wird ein Stoff magnetisiert, muss hierzu Energie aufgewendet werden. Diese Energie wird für die Verschiebung der Wände und die Drehung der Momente aufgewendet.

$$w = \int H dB = \int H \mu_0 \mu_r(H) dH \tag{28.68}$$

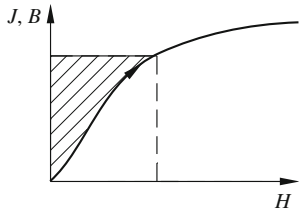


Abb. 28.28 Magnetisierungsenergie

Da die relative Permeabilität keine Konstante und der mathematische Zusammenhang nicht genau bekannt ist, kann dieses Integral normalerweise nicht berechnet werden. Die Bestimmung der aufzuwendenden Energie pro Volumeneinheit bis zu einer bestimmten Magnetisierung kann grafisch geschehen, indem die schraffierte Fläche in Abb. 28.28 bestimmt wird.

Da bei einem vollen Durchlauf der Hysteresekurve in Bereichen, in denen der Werkstoff entmagnetisiert wird, Energie frei wird, ist die gesamte pro Volumeneinheit benötigte Energie bei einem vollen Durchlauf der Kurve aus der Fläche der Hysteresekurve zu entnehmen (Abb. 28.29).

$$w = \oint H dB \mathrel{\widehat{=}} A_{\text{Hysterese}} \tag{28.69}$$

Magnetische Werkstoffe werden in *magnetisch harte* und *magnetisch weiche* Materialien unterteilt. Magnetisch harte Werkstoffe besitzen eine große Fläche, und damit bei gleicher Remanenz eine große *Koerzitivfeldstärke*, *magnetisch weiche* Stoffe eine kleine Hystereseфläche (siehe Abb. 28.30).

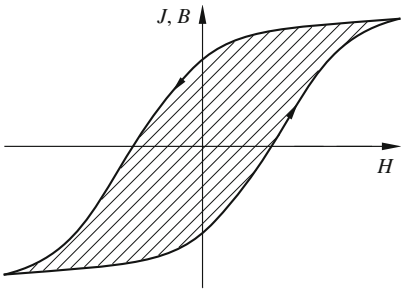


Abb. 28.29 Ummagnetisierungsverluste

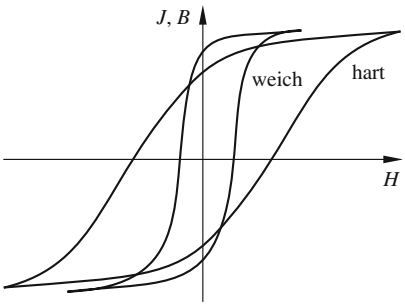


Abb. 28.30 Harte und weiche Werkstoffe

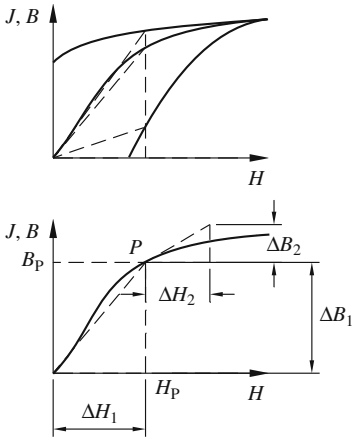


Abb. 28.31 Bestimmung der Permeabilität

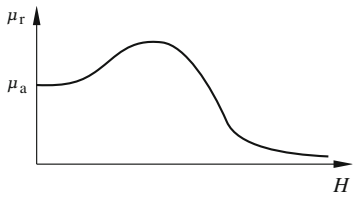


Abb. 28.32 Verlauf der Permeabilität

28.5.3.2 Verlauf der Permeabilität

Die relative Permeabilität m_r ist durch (28.50) definiert. Ihr Wert kann aus der Hysteresekurve bestimmt werden. Hierzu wird der 1. Quadrant der Magnetisierungskurve betrachtet. Bei vorgegebener Feldstärke H wird die zugehörige Flussdichte B abgelesen und dann mit (28.50) m_r berechnet.

Wie aus Abb. 28.31 oben zu entnehmen ist, sind bei einer Feldstärke bis zu drei verschiedene Flussdichten möglich. Daher kann es auch drei verschiedene Werte für die relative Permeabilität geben. Die relative Permeabilität ist von der magnetischen Vorgeschichte des Werkstoffes abhängig. Wird die Bestimmung an der Neukurve vorgenommen (Abb. 28.31 unten), erhält man für jede Feldstärke nur einen Wert. Dieser Wert ist nicht konstant, sondern von H abhängig. Bestimmt wird μ_r für die Feldstärke H_P, indem vom Nullpunkt eine Gerade zum Punkt P auf der Neukurve gezeichnet wird und dann ΔB_1 und ΔH_1 abgelesen werden. Hieraus wird m_r berechnet.

$$\textit{relative Permeabilität} \quad \mu_r = \frac{1}{\mu_0}\frac{\Delta B_1}{\Delta H_1} \tag{28.70}$$

Der prinzipielle Verlauf ist in Abb. 28.32 dargestellt. Ausgehend von einem Anfangswert μ_a durchläuft die Kurve ein Maximum. Die Werte für μ_a liegen im Bereich von einigen Hundert bis zu einigen Tausend.

Für manche Überlegungen ist es notwendig, nicht die relative Permeabilität, sondern eine Größe zu verwenden, die aus der Steigung der Kurve bestimmt wird. Dies wird *differentielle Permeabilität* μ_d genannt. Im Punkt P wird die Tangente an die

Tab. 28.5 Curie-Temperatur und Sättigungspolarisation

Werkstoff	T_C [K]	$J_s(0)$ [V s/m^2]
Gd	289	2,54
Ni	631	0,64
Fe	1042	2,18
Co	1400	1,81

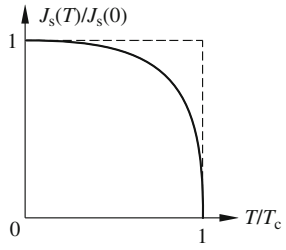


Abb. 28.33 Sättigungsmagnetisierung

Hysteresekurve gezeichnet und deren Steigung nach Abb. 28.31 unten durch ΔB_2 und ΔH_2 bestimmt.

$$\textit{differentielle Permeabilität} \quad \mu_d = \frac{1}{\mu_0}\frac{\Delta B_2}{\Delta H_2} \tag{28.71}$$

28.5.3.3 Temperaturabhängigkeit

Grundlage für die ferromagnetischen Eigenschaften ist die Ausrichtung der magnetischen Momente in den Weiß'schen Bezirken. Diese Ausrichtung wird als *spontane Magnetisierung* bezeichnet. Eine solche Ausrichtung wird durch die thermische Energie beeinflusst. Je höher die Temperatur, desto größer ist die thermische Bewegung, und desto geringer wird der Ordnungsgrad sein. Bei Überschreitung einer kritischen Temperatur werden die Weiß'schen Bezirke zerstört und der Werkstoff wird paramagnetisch. Als Maß für diesen Ordnungsgrad dient der Sättigungswert der Magnetisierung oder der Polarisation. Die Temperaturabhängigkeit von J_s ist in Abb. 28.33 gezeigt.

Aufgetragen sind relative Sättigungswerte, bezogen auf den Wert bei 0 K als Funktion der Temperatur, bezogen auf die kritische Temperatur T_C. Diese kritische Temperatur ist die *Curie-Temperatur*. Für Temperaturen oberhalb der *Curie-Temperatur* gilt für die Suszeptibilität das

$$\textit{Curie-Weiß'sche Gesetz} \quad \chi_M = \frac{C}{T - T_C} \tag{28.72}$$

Die Curie-Temperaturen einiger Stoffe sind in Tab. 28.5 aufgeführt.

28.5.3.4 Magnetostriktion

Durch die Verschiebung der Bloch-Wände bei der Magnetisierung von ferromagnetischen Stoffen kann eine Längenänderung des Werkstückes auftreten. Es kommen sowohl Verlängerungen, *positive Magnetostriktion*, als auch Verkürzungen, *negative Magnetostriktion*, vor. Die relativen Längenänderungen sind sehr klein $\frac{\Delta l}{l} \approx 4 \cdot 10^{-5}$. Technische Anwendung findet dieser Effekt bei der Erzeugung von Ultraschall bis zu einer Frequenz bis zu ungefähr 60 kHz.

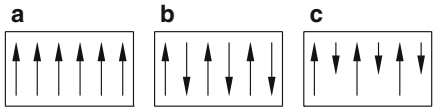


Abb. 28.34 Magnetische Stoffe. **a** Ferro, **b** antiferro, **c** ferri

28.5.4 Antiferromagnetismus

Voraussetzung für die ferromagnetischen Eigenschaften sind die Weiß'schen Bezirke mit den parallel eingestellten magnetischen Momenten (Abb. 28.34a). Unter bestimmten Bedingungen stellen sich in einigen Werkstoffen die Momente nicht parallel, sondern antiparallel ein (Abb. 28.34b). Für diese Stoffe gilt wie für die Ferromagnetika, dass sie sich oberhalb einer kritischen Temperatur, der *Néel-Temperatur* T_N, paramagnetisch verhalten, da die Ordnung der Momente durch die höhere Temperatur zerstört wird. Unterhalb der Néel-Temperatur sind diese Stoffe schwach magnetisch, da sich die Momente gegenseitig kompensieren. Typische Substanzen sind MnO, NiO, CoO, CrF_3, FeO.

28.5.5 Ferrimagnetismus

In ferrimagnetischen Stoffen sind die Momente ebenfalls antiparallel eingestellt. Im Gegensatz zu den antiferromagnetischen Stoffen sind sie hier nicht gleich groß (Abb. 28.34c). Die Ferrite haben große technische Bedeutung, sowohl als weichmagnetische, als auch als dauermagnetische Werkstoffe. Sie sind keine Metalle, sondern Ionenkristalle. Im Vergleich zu Metallen ($\rho \approx 10^{-7}\,\Omega\,\mathrm{m}$) haben sie einen wesentlich größeren spezifischen Widerstand ($1 < \rho < 10^3\,\Omega\,\mathrm{m}$). Daher treten kaum Wirbelströme auf. Ferrite werden als Spulenkerne für Anwendungen bei hohen Frequenzen eingesetzt. Die Abhängigkeit der spontanen Magnetisierung von der Temperatur ist sehr kompliziert und kann durchaus für die einzelnen Werkstoffe unterschiedlich sein. Auch diese Stoffe werden oberhalb der ferrimagnetischen Curie-Temperatur paramagnetisch. Typische Werkstoffe sind $\mathrm{MeO} \cdot \mathrm{Fe_2O_3}$, (Me = Fe, Co, Ni).

28.6 Magnetische Kreise

Manchmal ist es zur Berechnung der Kombinationen von magnetischen Bauteilen hilfreich, formale Analogien zu elektrischen Schaltungen herzustellen. Dazu werden eine magnetische Spannung und ein magnetischer Widerstand eingeführt.

28.6.1 Magnetische Spannung

In Analogie zur elektrischen Spannung nach (27.8) kann man auch in magnetischen Kreisen eine magnetische Spannung definieren. Sie hat dann die gleiche formale Struktur wie die

elektrische Spannung. Die Spannung zwischen den Punkten 1 und 2 ist die magnetische Spannung.

$$\textit{magnetische Spannung} \quad V_{12} = \int_1^2 \vec{H} d\vec{s} \tag{28.73}$$

Die Addition aller Teilspannungen ergibt die

magnetische Umlaufspannung

$$V = \oint \vec{H} \cdot d\vec{l} = NI \tag{28.74}$$

Dieser Ausdruck ist aber im Durchflutungsgesetz nach (28.6) aufgetreten.

$$\Theta = V \tag{28.75}$$

28.6.2 Magnetischer Widerstand

Der magnetische Fluss in einem konstanten Magnetfeld einer Ringspule ist

$$\phi = BA = \mu_0 \mu_r HA = \mu_0 \mu_r \frac{IN}{l} A \tag{28.76}$$

$$NI = V = \phi \frac{l}{\mu_0 \mu_r A} \tag{28.77}$$

In Analogie zu (25.14) wird ein magnetischer Widerstand definiert.

Magnetischer Widerstand

$$R_m = \frac{l}{\mu_0 \mu_r A} \tag{28.78}$$

R_m	l	μ_0	A
$\frac{\text{A}}{\text{Vs}}$	m	$\frac{\text{Vs}}{\text{Am}}$	m^2

magnetischer Leitwert

$$\Lambda = \frac{1}{R_m} = \frac{\mu_0 \mu_r A}{l} \tag{28.79}$$

Das Produkt $\mu_0 \mu_r$ kann als magnetische Leitfähigkeit gedeutet werden. Die relative Permeabilität ist ein Maß dafür, magnetische Feldlinien zu leiten. Der formal gleiche Aufbau der Formeln für elektrische und magnetische Größen hat aber nur den Zweck, den Umgang mit magnetischen Größen zu vereinfachen. Es werden völlig andere Größen miteinander verglichen.

28.6.3 Unverzweigte Kreise

Ein unverzweigter magnetischer Kreis kann formal behandelt werden wie eine elektrische Masche

$$\sum V = \Theta = NI \tag{28.80}$$

Ein oft benutzter magnetischer Kreis ist der geschlossenen Eisenkern ohne Luftspalt, wie er in Abb. 28.35 gezeichnet ist.

Der Eisenkern hat längs des Umlaufweges l überall den gleichen Querschnitt A. Abweichungen in den Ecken werden vernachlässigt. Die Neukurve ist in Abb. 28.36 gegeben. Die Windungszahl der verwendeten Spule ist N. Da die relative Permeabilität von Eisen sehr viel größer ist als die von Luft ($\mu_{r\,\text{Fe}} \approx 2000$), verläuft der in der Spule erzeugte magnetische Fluss fast vollständig im Eisenkern, obwohl die felderzeugende Spule nur an einer Stelle um den Eisenkern liegt. Der Fluss im Eisenkern ist daher überall gleich groß. Der Strom I durch die Spule erzeugt das Magnetfeld $H_E = \frac{IN}{l}$. Zu diesem Magnetfeld kann aus der Magnetisierungskurve die zugehörige Flussdichte B_E abgelesen werden.

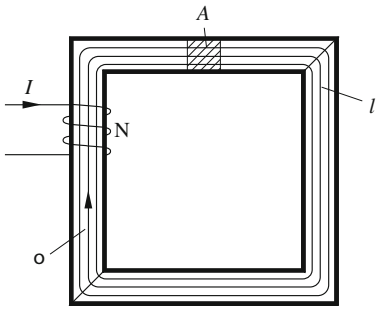


Abb. 28.35 Spule mit Eisenkern

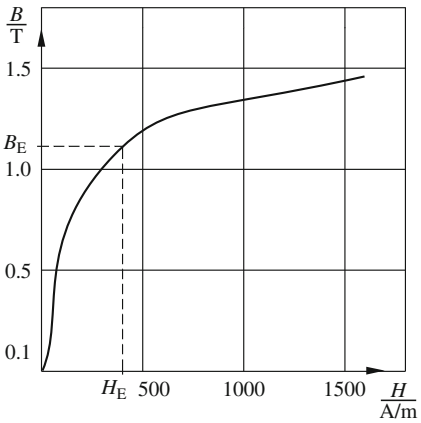


Abb. 28.36 Magnetisierungskurve von Eisen

Wird nun in den Eisenkern ein Luftspalt der Breite d geschnitten (Abb. 28.37), sind analog zur Maschenregel zwei magnetische Spannungen zu addieren. Der magnetische Gesamtwiderstand ist gleich der Summe der Einzelwiderstände. Da der Luftspalt als sehr klein angenommen wird, ist der magnetische Fluss im Eisen und im Spalt gleich.

Nach (28.77) ist

$$NI = \frac{\Phi}{\mu_0 A} \left(\frac{l_E}{\mu_r} + \frac{l_L}{1} \right) \tag{28.81}$$

$$NI = H_E l_E + H_L l_L \tag{28.82}$$

$$H_L = \frac{B_E}{\mu_0} \tag{28.83}$$

$$NI = H_E l_E + \frac{B_E}{\mu_0} l_L \tag{28.84}$$

Teil IV

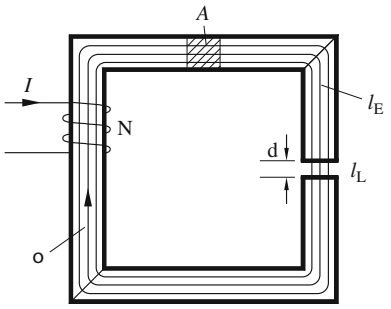


Abb. 28.37 Spule mit Luftspalt

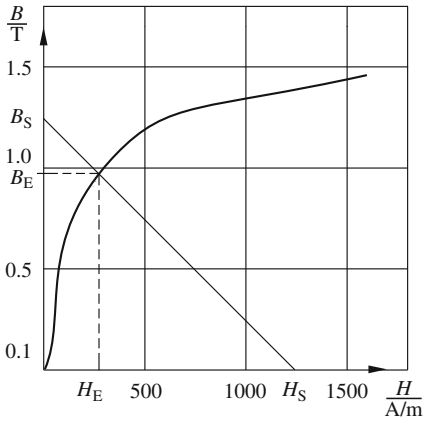


Abb. 28.38 Einfluss des Luftspaltes

Beispiel

Ein Ringkern mit der in Abb. 28.36 angegebenen Magnetisierungskurve hat eine mittlere Länge von $l = 35\,\text{cm}$ und eine Querschnittsfläche von $A = 4\,\text{cm}^2$. Die verwendete Spule hat eine Windungszahl von $N = 200$.

a) Welche Stromstärke I wird benötigt, um einen Fluss von $f = 4{,}4 \cdot 10^{-4}\,\text{Wb}$ zu erzeugen?
b) in den Kern wird ein Luftspalt von 0,3 mm geschnitten. Welche Stromstärke wird jetzt benötigt?

Lösung:

a)

$$B_E = \frac{\Phi}{A} = \frac{4{,}4 \cdot 10^{-4}\,\text{Wb}}{4 \cdot 10^{-4}\,\text{m}^2} = 1{,}1\,\text{T}$$

Aus der Magnetisierungskurve wird zu diesem B-Wert ein Feld $H_E = 400\,\frac{\text{A}}{\text{m}}$ abgelesen. Hierzu wird ein Strom von

$$I = \frac{H_E l}{N} = \frac{400\,\frac{\text{A}}{\text{m}} \cdot 0{,}35\,\text{m}}{200} = 0{,}7\,\text{A}$$

benötigt.

b) B_E und H_E werden wie unter a) bestimmt. Das Feld im Luftspalt ist

$$H_L = \frac{B_E}{\mu_0} = \frac{1{,}1\,\text{T}}{4\pi \cdot 10^{-7}\,\frac{\text{Vs}}{\text{Am}}} = 8{,}75 \cdot 10^5\,\frac{\text{A}}{\text{m}}$$

Nach (28.82) ist

$$I_L = \frac{H_E l_E + H_L l_L}{N}$$

$$I_L = \frac{(400 \cdot 0{,}35 + 8{,}75 \cdot 10^5 \cdot 3 \cdot 10^{-4})\,\text{A}}{200} = 2{,}01\,\text{A}$$

◀

Ist hingegen bei einer Spule mit Luftspalt die Durchflutung gegeben und nach dem magnetischen Feld gefragt, kann dies Problem nur grafisch gelöst werden, da die mathematische Form der Magnetisierungskurve nicht bekannt ist. Ausgehend von (28.84) ist zu sehen, dass das Feld im Eisen linear mit zunehmender Breite des Luftspaltes abnimmt.

$$H_E l_E = NI - \frac{B_E}{\mu_0} l_L \tag{28.85}$$

Die Schnittpunkte dieser Geraden mit den Koordinatenachsen berechnen sich nach

$$\textit{Schnittpunkte} \quad H_S = \frac{NI}{l_E} \tag{28.86}$$

$$B_S = \frac{\mu_0 NI}{l_L} \tag{28.87}$$

Die Schnittpunkte dieser „Luftspaltgeraden“ mit der Magnetisierungskurve ergeben die gesuchten Werte für H_E und B_E.

Beispiel

Wie groß ist die Feldstärke und die Flussdichte in einem Eisenkern mit 0,3 mm Luftspalt und einer mittleren Länge von 24 cm, wenn durch eine Spule mit 150 Windungen ein Strom von 2 A fließt?

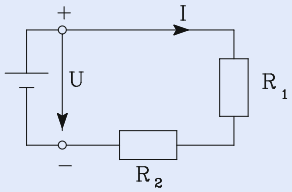
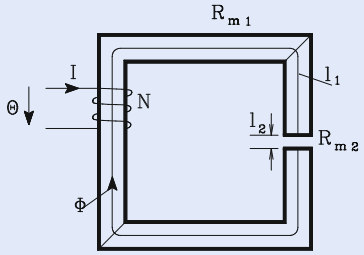
$$H_S = \frac{2\,\text{A} \cdot 150}{0{,}24\,\text{m}} = 1250\,\frac{\text{A}}{\text{m}}$$

$$B_S = \frac{4\pi \cdot 10^{-7}\,\frac{\text{Vs}}{\text{Am}} \cdot 2\,\text{A} \cdot 150}{3 \cdot 10^{-4}\,\text{m}} = 1{,}256\,\text{T}$$

Diese Werte und die Gerade werden in Abb. 28.38 eingetragen. Die Koordinaten des Schnittpunktes mit der Magnetisierungskurve sind $H_E = 280\,\frac{\text{A}}{\text{m}}$ und $B_E = 0{,}97\,\text{T}$.

◀

Tab. 28.6 Analogie elektrischer – magnetischer Kreis

	Elektrischer Kreis	Magnetischer Kreis
		
Ursache	Elektrische Spannung U; $[U] = 1\,\mathrm{V}$	Magnetische Spannung $\Theta = \oint Hds = NI$; $[\Theta] = 1\,\mathrm{A}$
Wirkung	Strom I; $[I] = 1\,\mathrm{A}$	Fluss ϕ; $[\phi] = 1\,\mathrm{Wb} = 1\,\mathrm{V\,s}$
Ohm'sches Gesetz	$R = \frac{U}{I}$	$R_m = \frac{\Theta}{\phi}$
Widerstand	$R = \frac{l}{\chi A}$; $[R] = 1\,\Omega$	$R = \frac{l}{\mu_0 \mu_r A}$; $[R_m] = 1\,\frac{\mathrm{A}}{\mathrm{Wb}}$
Leitfähigkeit	χ; $[\chi] = 1\,\frac{\mathrm{A}}{\mathrm{V\,m}}$	$\mu_0\mu_r$; $[\mu_0\mu_r] = 1\,\frac{\mathrm{Wb}}{\mathrm{A\,m}} = 1\,\frac{\mathrm{V\,s}}{\mathrm{A\,m}}$

28.6.4 Verzweigte Kreise

Die Berechnung kann analog zu den Überlegungen in elektrischen Kreisen durchgeführt werden (Tab. 28.6).

$$\textit{Knotenregel} \qquad \sum_{i=1}^{n} \phi_i = 0 \tag{28.88}$$

$$\textit{Maschenregel} \qquad \Theta = \sum_{i=1}^{n} H_i l_i \tag{28.89}$$

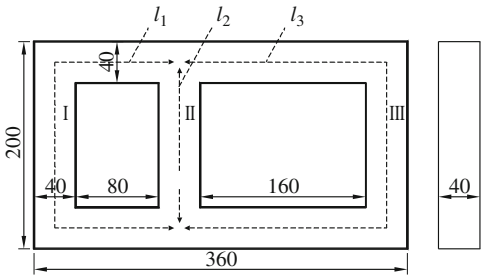


Abb. 28.39 verzweigter Kreis

Beispiel

Im mittleren Steg II des Transformatorkernes nach Abb. 28.39 soll ein Fluss von $\phi_2 = 1{,}6 \cdot 10^{-3}\,\mathrm{Wb}$ erzeugt werden. Auf Steg I befindet sich eine Spule mit 150 Windungen. Welcher Strom muss durch die Spule fließen? Die Magnetisierungskurve ist durch Abb. 28.38 gegeben.

Aus der Zeichnung lassen sich folgende geometrische Größen bestimmen:

$$l_1 = 40\,\mathrm{cm}, \quad l_2 = 16\,\mathrm{cm}, \quad l_3 = 56\,\mathrm{cm},$$
$$A = 1{,}6 \cdot 10^{-3}\,\mathrm{m}^2$$

Da sich die Spule nur auf dem Steg I befindet, sind die magnetischen Spannungen der beiden Zweige II und III gleich

$$H_2 l_2 = H_3 l_3; \quad B_2 = \frac{\phi_2}{A} = \frac{1{,}6 \cdot 10^{-3}\,\mathrm{Wb}}{1{,}6 \cdot 10^{-3}\,\mathrm{m}^2} = 1\,\mathrm{T}$$

Hierzu gehört nach Abb. 28.38

$$H_2 = 300\,\frac{\mathrm{A}}{\mathrm{m}}; \quad H_3 = \frac{H_2 l_2}{l_3} = 85{,}7\,\frac{\mathrm{A}}{\mathrm{m}};$$
$$B_3 = 0{,}4\,\mathrm{T}$$
$$\phi_3 = B_3 A = 6{,}4 \cdot 10^{-4}\,\mathrm{Wb};$$
$$\phi_1 = \phi_2 + \phi_3 = 2{,}24 \cdot 10^{-3}\,\mathrm{Wb}$$
$$B_1 = \frac{\phi_1}{A} = 1{,}4\,\mathrm{T}; \quad H_1 = 1100\,\frac{\mathrm{A}}{\mathrm{m}}$$
$$NI = H_1 l_1 + H_2 l_2 = 488\,\mathrm{A};$$
$$I_1 = 3{,}25\,\mathrm{A}$$

◀

Teil IV

Induktion

29

Reinhard von Liebenstein



29.1 Induktion bei Änderung der Fläche 354
29.2 Induktion bei Änderung des Magnetfeldes 355
29.3 Die Induktivität einer Spule 356

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_29

Eine der wichtigsten Erscheinungen für die Elektrotechnik ist die *Induktion*. *Faraday* erkannte im Jahr 1831, dass jede zeitliche Änderung des magnetischen Flusses in einer Leiterschleife eine Spannung hervorruft, die Spannung *induziert*. Ursache der Induktion ist die Lorentzkraft.

29.1 Induktion bei Änderung der Fläche

Bei der Bewegung eines Leiters in einem Magnetfeld senkrecht zu der Richtung der Feldlinien wird auf die Elektronen eine Kraft ausgeübt. Die Kraft ist bei den in Abb. 29.1 angegebenen Richtungen der Geschwindigkeit v und des Magnetfeldes B so, dass die Elektronen zum hinteren Teil des Leiters bewegt werden.

Im Abb. 29.1 ist die Lorentzkraft auf negative Ladungen eingezeichnet. Diese Trennung erfolgt solange, bis das elektrostatische Gegenfeld und die dadurch erzeugte Kraft F_{ind} sich aufheben. Für die Beträge gilt

$$|F_{\text{ind}}| = |F_L| \tag{29.1}$$

$$|e_0 vB| = |e_0 E_{\text{ind}}| = \frac{|u_{\text{ind}}|}{l} \tag{29.2}$$

$$|vB| = |E_{\text{ind}}| \tag{29.3}$$

$$|u_{\text{ind}}| = |vBl| \tag{29.4}$$

Wenn auf die Größen u_{ind}, v, B die rechte Handregel angewendet wird, (v in Daumen, B in Zeigefinger und u_{ind} in Mittelfingerrichtung nach Abb. 29.2), ist:

$$\textit{Induzierte Spannung} \quad u_{\text{ind}} = -vBl \tag{29.5}$$

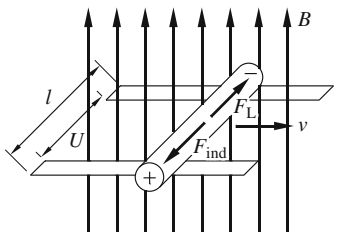


Abb. 29.1 Änderung der Fläche durch Bewegung des Leiters

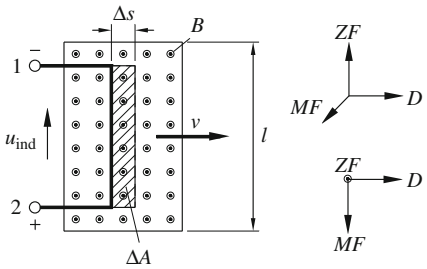


Abb. 29.2 Bewegter Leiter im Magnetfeld

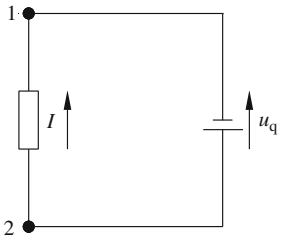


Abb. 29.3 Induktion als Ersatzspannungsquelle

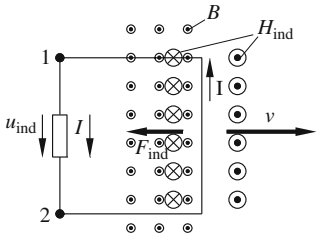


Abb. 29.4 Lenz'sche Regel

Bei der betrachteten Bewegung des Leiters wird die Fläche um das Stück vergrößert

$$\Delta A = l \cdot \Delta s = l \cdot v \cdot \Delta t \tag{29.6}$$

$$\textit{Flussänderung} \quad \Delta\phi = B \cdot \Delta A = B \cdot l \cdot v \cdot \Delta t \tag{29.7}$$

$$\textit{induzierte Spannung} \quad u_{\text{ind}} = -\frac{\Delta\phi}{\Delta t} \tag{29.8}$$

Wird die Leiterschleife in Abb. 29.2 zwischen den Punkten 1 und 2 durch einen Widerstand verbunden, fließt ein Strom. Hierfür kann das in Abb. 29.3 dargestellte Ersatzschaltbild verwendet werden.

Der Strom, der in Abb. 29.2 durch den Leiter fließt, erzeugt selbst wieder ein Magnetfeld. Die Richtung dieses induzierten Feldes ist so, dass auf den Leiter eine Kraft ausgeübt wird, die die Bewegung des Leiters hemmt (Abb. 29.4). Das wird als *Lenz'sche Regel* bezeichnet.

> Die durch Bewegung von Magneten oder Leitern erzeugten Induktionsströme haben eine solche Richtung, dass ihr Magnetfeld der Bewegung entgegengesetzt wirkt.

Es muss also bei der Bewegung Arbeit geleistet werden.

Das Magnetfeld stand bei der bisherigen Betrachtung senkrecht zur Fläche. Ist dies nicht der Fall, wird der Fluss durch den Flächenanteil, der senkrecht zum Magnetfeld steht, bestimmt. Eine Flussänderung kann auch durch Rotation einer Leiterschleife nach Abb. 29.5 bewirkt werden. Die Schleife dreht sich mit der Winkelgeschwindigkeit ω.

$$\textit{Momentaner Fluss } t \quad \phi = B \cdot A \cdot \cos\omega t \tag{29.9}$$

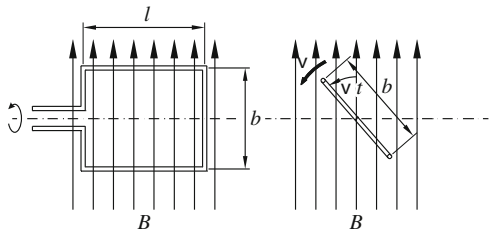


Abb. 29.5 Rotation einer Leiterschleife

Daraus resultiert eine induzierte Spannung

$$u_{\text{ind}} = -\frac{d\phi}{dt} = B \cdot A \cdot \omega \cdot \sin \omega t \quad (29.10)$$

Wechselspannung $u = \hat{u} \cdot \sin \omega t$ (29.11)

Scheitelspannung $\hat{u} = B \cdot A \cdot \omega$ (29.12)

Wird nicht nur eine Schleife, sondern eine Spule mit N Windungen gedreht, ist die

Scheitelspannung $\hat{u} = N \cdot B \cdot A \cdot \omega$ (29.13)

Beispiel

Ein in Richtung Ost-West liegender Metallstab von 2 m Länge fällt aus 15 m Höhe zu Boden. Welche Spannung wird durch das Erdmagnetfeld induziert, wenn der Stab den Boden erreicht ($B_{\text{Erde}} = 20\,\mu\text{T}$)?

$$v = \sqrt{2gh} = \sqrt{2 \cdot 9{,}81 \cdot 15}\,\frac{\text{m}}{\text{s}} = 17{,}16\,\frac{\text{m}}{\text{s}}$$

$$|u_{\text{ind}}| = |vBl| = 17{,}16\,\frac{\text{m}}{\text{s}} \cdot 20 \cdot 10^{-6}\,\text{T} \cdot 2\,\text{m}$$

$$= 0{,}68\,\text{mV}$$ ◀

Beispiel

Eine quadratische Spule mit der Kantenlänge $a = 5\,\text{cm}$ und $N_1 = 50$ Windungen rotiert mit $n = 250\,\text{min}^{-1}$ um ihre Symmetrieachse im Inneren einer langen Zylinderspule, die je cm 8 Windungen hat und von 6 A durchflossen wird. Welchen Scheitelwert erreicht die induzierte Spannung?

$$B_{\text{Sp}} = \mu_0 \frac{N_{\text{Sp}} I}{l_{\text{Sp}}}; \quad B_{\text{Sp}} = 4\pi \cdot 10^{-7}\,\frac{\text{V s}}{\text{A m}}\,\frac{8 \cdot 6\,\text{A}}{10^{-2}\,\text{m}} = 6\,\text{mT}$$

$$\hat{u} = N_1 \cdot B \cdot A \cdot 2\pi \cdot n$$

$$\hat{u} = 50 \cdot 6 \cdot 10^{-3}\,\frac{\text{V s}}{\text{m}^2} \cdot (5 \cdot 10^{-2}\,\text{m})^2 \cdot 2\pi \cdot \frac{250}{60}\,\frac{1}{\text{s}}$$

$$= 20\,\text{mV}$$ ◀

29.2 Induktion bei Änderung des Magnetfeldes

Bei konstanter Fläche einer Leiterschleife oder einer Spule kann durch Magnetfeldänderungen eine Flussänderung hervorgerufen werden. Dies kann z. B. durch Stromänderung in einer Erregerspule erreicht werden. Die Flussänderung ist dann

Flussänderung $\frac{d\phi}{dt} = \frac{\mu_0 N_{\text{err}}}{l_{\text{err}}}\frac{dI_{\text{err}}}{dt}$ (29.14)

Dieser Effekt wird bei Transformatoren verwendet. Das Induktionsgesetz lautet bei N Windungen einer Feldspule

Induktionsgesetz $u_{\text{ind}} = -N\frac{d\phi}{dt}$ (29.15)

$$u_{\text{ind}} = -N\left(A_n\frac{dB}{dt} + B\frac{dA_n}{dt}\right) \quad (29.16)$$

A_n ist der Flächenanteil, der senkrecht zu den Feldlinien steht.

Aus (29.16) geht hervor, dass es gleichgültig ist, ob sich die magnetische Flussdichte bei gleichbleibender Fläche (*Transformatorenprinzip*) oder die Fläche bei gleichbleibender Flussdichte (*Generatorenprinzip*) ändert.

Entscheidend ist, dass sich der Fluss in einer Leiterschleife ändert. Wird eine Leiterschleife in einem homogenen Magnetfeld bewegt, ändert sich der Fluss nicht. Die durch die auch hier vorhandene Lorentzkraft verursachte Verschiebung der Elektronen hebt sich in gegenüberliegenden Teilen der Schleife gegenseitig auf.

Wird eine Schleife im inhomogenen Feld bewegt, entsteht eine Induktionsspannung, da sich bei der Bewegung der Fluss ändert und in den gegenüberliegenden Teilen der Schleife unterschiedliche Spannungen induziert werden. In den in Abb. 29.6 parallel zur Bewegungsrichtung liegenden Teilen der Schleife werden keine Spannungen induziert.

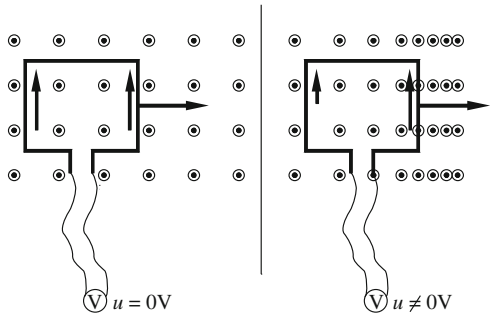


Abb. 29.6 Bewegung einer Schleife im homogenen und inhomogenen Feld

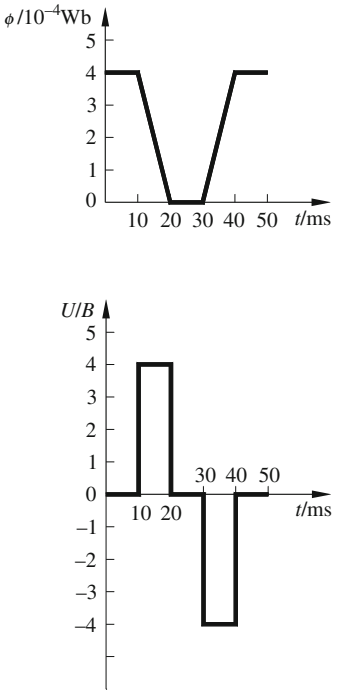


Abb. 29.7 Beispiel Induktion

Beispiel

Eine Spule mit $N = 100$ Windungen wird von einem Fluss Φ durchsetzt, der den in Abb. 29.7 oben dargestellten Verlauf hat. Welchen zeitlichen Verlauf hat die induzierte Spannung?

In den Zeitintervallen 0 … 10 ms, 20 … 30 ms, > 40 ms ändert sich der Fluss nicht, somit ist in diesen Bereichen $u_{\text{ind}} = 0\,\text{V}$.

Im Intervall 10 … 20 ms gilt

$$\frac{d\phi}{dt} = -\frac{4 \cdot 10^{-4}\,\text{V s}}{10 \cdot 10^{-3}\,\text{s}} = -40\,\text{mV}$$

$$u_{\text{ind}} = -N\frac{d\phi}{dt} = -100 \cdot (-40\,\text{mV}) = 4\,\text{V}$$

Im Intervall 30 … 40 ms hat die induzierte Spannung denselben Wert mit umgekehrtem Vorzeichen. Die induzierte Spannung ist in Abb. 29.7 unten aufgetragen. ◀

29.3 Die Induktivität einer Spule

29.3.1 Selbstinduktion

In der Ringspule in Abb. 28.4 mit der Windungszahl N soll der Strom i beliebig veränderbar sein. Wird der eingespeiste Strom geändert, so ändert sich auch der von diesem Strom erzeugte magnetische Fluss ϕ. Wegen dieser Flussänderung wird eine Spannung in der Spule induziert. Dieser Vorgang wird *Selbstinduktion* genannt. Die induzierte Spannung ist nach der Lenz'schen-Regel dem erzeugenden Strom entgegen gesetzt.

$$u = -N\frac{d\phi}{dt} \tag{29.17}$$

Da die Flussänderung durch eine Stromänderung verursacht wird, gilt auch

$$u = -L\frac{di}{dt} \tag{29.18}$$

Induktivität $L = \left|u\frac{dt}{di}\right|$

L	u	i	t
$\text{H} = \frac{\text{V s}}{\text{A}}$	V	A	s

(29.19)

Die Einheit ist 1 *Henry*, 1 H.

Aus (29.17) und (29.18) folgt

$$L = N \cdot \frac{d\phi}{di} = N \cdot \frac{A \cdot dB}{di} \tag{29.20}$$

$$L = NA\frac{dB}{dH} \cdot \frac{dH}{di} \tag{29.21}$$

Befindet sich kein ferromagnetischer Stoff in der Spule, so ist

$$\frac{dB}{dH} = \mu_0 \tag{29.22}$$

$$\frac{dH}{di} = \frac{N}{l} \tag{29.23}$$

Induktivität $L = \mu_0 N^2 \frac{A}{l}$ (29.24)

Ist der Spulenkörper ferromagnetisch, gilt

$$\frac{dB}{dH} = \mu_0\mu_d \tag{29.25}$$

$$L = \mu_0\mu_d N^2 \frac{A}{l} \tag{29.26}$$

μ_d ist die *differentielle Permeabilität*.

Die Induktivität spielt in Wechselstromkreisen eine große Rolle. Sie ist für beliebige Leiteranordnungen und Leitergeometrien schwierig zu berechnen, lässt sich aber in Wechselstromkreisen gut durch Messungen bestimmen (Abb. 29.8).

Einfachleitung $L = \frac{\mu_0\mu_r l}{2\pi}\left(\ln\left(\frac{2l}{r}\right) - \frac{3}{4}\right)$ (29.27)

Doppelleitung $L = \frac{\mu_0\mu_r l}{\pi}\left(\ln\left(\frac{a}{r}\right) + \frac{1}{4}\right)$ (29.28)

Bei koaxialen Leitern ist die Gesamtinduktivität nach folgenden Gleichungen zu berechnen:

Innenleiter $L_i = \frac{\mu_0\mu_r l}{8\pi}$ (29.29)

Zwischenraum $L_{\text{ZW}} = \frac{\mu_0\mu_r l}{2\pi}\ln\frac{r_1}{r}$ (29.30)

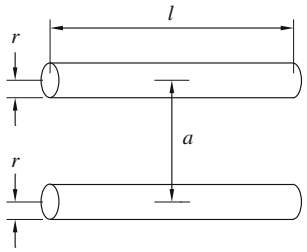


Abb. 29.8 Doppel-Leitung

Außenleiter

$$L_a = \frac{\mu_0 \mu_r l r_2^4}{2\pi(r_2^2 - r_1^2)} \left(\ln \frac{r_2}{r_1} - \frac{(3r_2^2 - r_1^2)(r_2^2 - r_1^2)}{4r_2^4} \right) \tag{29.31}$$

gesamte Induktivität $L_{\text{ges}} = L_i + L_{\text{ZW}} + L_a$ (29.32)

29.3.2 Gegeninduktion

Befindet sich in der Nähe einer stromdurchflossenen Spule Sp_1 eine zweite Spule Sp_2, so verläuft ein Teil des von der Spule 1 erzeugten Magnetfeldes auch durch die Spule 2. Die Spulen sind *magnetisch gekoppelt*. Der vom Strom I_1 in der Spule 1 erzeugte magnetische Fluss durchsetzt die Spule 2 nur zum Teil. Dieser Anteil ist ϕ_{12}. Der Gesamtfluss in der Spule 2 ist dann

$$\phi_{G\,12} = N_2 \phi_{12} \tag{29.33}$$

$$N_2 \phi_{12} \sim I_1 \tag{29.34}$$

$$N_2 \phi_{12} = M_{12} I_1 \tag{29.35}$$

Der Faktor M_{12} wird als *Gegeninduktivität* bezeichnet. Bei Abwesenheit von ferromagnetischen Materialien ist die Gegeninduktivität eine Konstante. Wird der Strom in der Spule 1 geändert, wird in der Spule 2 eine Spannung induziert.

$$u_2 = -N_2 \frac{d\phi_{12}}{dt} = -M_{12} \frac{di_1}{dt} \tag{29.36}$$

Wird die Spule 2 vom Strom durchflossen, so wird in der Spule 1 eine Spannung

$$u_1 = -M_{21} \frac{di_2}{dt} \tag{29.37}$$

induziert. Die Gegeninduktivitäten sind gleich, wenn die Permeabilitäten nicht vom Magnetfeld abhängen, wenn also keine

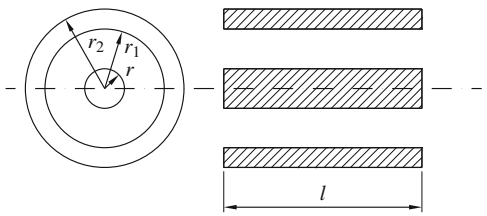


Abb. 29.9 Koaxialleiter

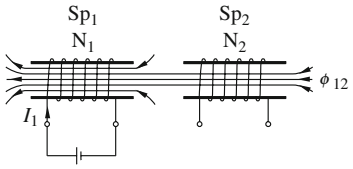


Abb. 29.10 Gegeninduktion

ferromagnetischen Stoffe beteiligt sind. Wie aus Abb. 29.10 ersichtlich, wird Spule 2 nicht vom gesamten in Spule 1 erzeugten Fluss durchsetzt.

$$\phi_1 = \phi_{12} + \phi_{\sigma 1} \tag{29.38}$$

$$\phi_2 = \phi_{21} + \phi_{\sigma 2} \tag{29.39}$$

ϕ_{12} wird als *Hauptfluss*, $\phi_{\sigma 1}$ als *Streufluss* bezeichnet. Es lassen ich folgende Ausdrücke definieren:

Streufaktoren

$$\sigma_1 = \frac{\phi_{\sigma 1}}{\phi_1}; \quad \sigma_2 = \frac{\phi_{\sigma 2}}{\phi_2} \tag{29.40}$$

Streuinduktivitäten

$$L_{\sigma 1} = \sigma_1 L_1; \quad L_{\sigma 2} = \sigma_2 L_2 \tag{29.41}$$

Koppelungsfaktoren

$$k_1 = 1 - \sigma_1; \quad k_2 = 1 - \sigma_2 \tag{29.42}$$

totaler Koppelungsfaktor

$$k = \sqrt{k_1 k_2} \tag{29.43}$$

Hauptflüsse

$$\phi_{12} = k_1 \Phi_1; \quad \phi_{21} = k_2 \phi_2 \tag{29.44}$$

Gegeninduktivität

$$M = \sqrt{k_1 k_2 L_1 L_2} = k\sqrt{L_1 L_2} \tag{29.45}$$

Beispiel

Wie groß sind die Streufaktoren für die in Abb. 29.11 dargestellte Anordnung zweier konzentrisch ineinander liegender Zylinderspulen? Die Abmessungen der beiden Spulen sind

$$d_1 = 18\,\text{mm}, \quad l_1 = 60\,\text{mm}, \quad N_1 = 250$$
$$d_2 = 10\,\text{mm}, \quad l_2 = 30\,\text{mm}, \quad N_2 = 24$$

Fluss

$$\phi_1 = \mu_0 H_1 A_1 = \mu_0 \frac{i_1 N_1}{l_1} \frac{\pi d_1^2}{4}$$

$$\phi_{12} = \mu_0 H_1 A_2 = \mu_0 \frac{i_1 N_1}{l_1} \frac{\pi d_2^2}{4}$$

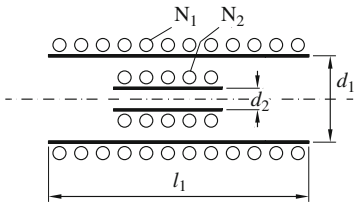


Abb. 29.11 Beispiel Gegeninduktion

induzierte Spannung

$$u_2 = -N_2 \frac{d\phi_{12}}{dt} = -\mu_0 \frac{N_1 N_2 \pi d_2^2}{4 l_1} \frac{di_1}{dt}$$

Der Faktor vor $\frac{di_1}{dt}$ ist die Gegeninduktivität M.

Gegeninduktivität

$$M = \mu_0 \frac{N_1 N_2 \pi d_2^2}{4 l_1}$$

$$M = \frac{4\pi \cdot 10^{-7} \frac{\mathrm{Vs}}{\mathrm{Am}} \cdot 250 \cdot 24 \cdot \pi \cdot (10 \cdot 10^{-3}\,\mathrm{m})^2}{4 \cdot 60 \cdot 10^{-3}\,\mathrm{m}}$$

Streufaktor

$$\phi_{\sigma 1} = \phi_1 - \phi_{12}$$

$$\sigma_1 = \frac{\phi_{\sigma 1}}{\phi_1} = 1 - \frac{\phi_{12}}{\phi_1};$$

$$\sigma_1 = 1 - \frac{\mu_0 \frac{i_1 N_1}{l_1} \frac{\pi d_2^2}{4}}{\mu_0 \frac{i_1 N_1}{l_1} \frac{\pi d_1^2}{4}} = 1 - \left(\frac{d_2}{d_1}\right)^2$$

$$\sigma_1 = 1 - \left(\frac{10\,\mathrm{mm}}{18\,\mathrm{mm}}\right)^2 = 0{,}691 \quad \blacktriangleleft$$

29.3.3 Energie im Magnetfeld einer Spule

Zum Aufbau des Magnetfeldes in einer Spule ist Energie erforderlich. Ist die Induktivität L bekannt, kann die benötigte Energie berechnet werden.

$$u = L \frac{di}{dt} \tag{29.46}$$

$$dW = u \cdot i \cdot dt \tag{29.47}$$

$$dW = L \frac{di}{dt} \cdot i \cdot dt = L \cdot i \cdot di \tag{29.48}$$

Wird die Spule nach Abb. 29.12 geschaltet, steigt der Strom vom Wert $i = 0\,\mathrm{A}$ bis zum Maximalwert $i = \frac{U}{R} = I$ an. Damit beträgt die der Spule zugeführte und im Magnetfeld gespeicherte Energie:

$$\textit{Energie des Magnetfeldes} \quad W = \int_0^I Li \cdot di \tag{29.49}$$

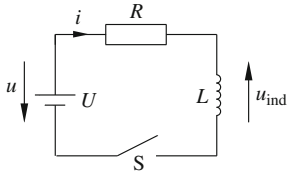


Abb. 29.12 Spule

Ist die Induktivität konstant, enthält die Spule kein ferromagnetisches Material, kann dieses Integral berechnet werden.

$$\textit{Energie in einer Spule} \quad W = \frac{1}{2} L I^2 \tag{29.50}$$

Bei Anwesenheit von ferromagnetischen Materialien ist zur Bestimmung der Energie das grafische Verfahren nach Abb. 28.28 anzuwenden.

29.3.4 Ein- und Ausschaltvorgänge

Wird im Stromkreis von Abb. 29.12 zur Zeit $t = 0$ der Schalter S geschlossen, ändert sich der Strom i. Damit wird in der Spule eine Spannung induziert. Es gilt nach der Maschenregel

$$u + u_{\mathrm{ind}} - iR = 0 \tag{29.51}$$

$$u - L \frac{di}{dt} = iR \tag{29.52}$$

Zu Beginn des Vorganges ist der Strom $i = 0$, am Ende hat der Strom seinen Maximalwert $I = \frac{U}{R}$ erreicht. Die mathematische Form dieser Differentialgleichung ist wie die beim Aufladen eines Kondensators. Die Lösungsfunktion ist wie in (27.67) durch eine e-Funktion darzustellen.

$$\textit{Einschalten} \quad i = I(1 - e^{-R \cdot t/L}) \tag{29.53}$$

$$i = I(1 - e^{-t/\tau}) \tag{29.54}$$

$$\textit{Zeitkonstante} \quad \tau = \frac{L}{R} \tag{29.55}$$

Wird eine Spule, durch die ein konstanter Strom I fließt und in der ein konstantes Magnetfeld besteht, kurzgeschlossen, wird eine Spannung induziert, die nach der Lenz'schen Regel versucht, das Magnetfeld aufrecht zu halten. Der Strom wird seinen Endwert ebenfalls nach einer e-Funktion erreichen.

$$\textit{Kurzschließen} \quad i = I e^{-R \cdot t/L} \tag{29.56}$$

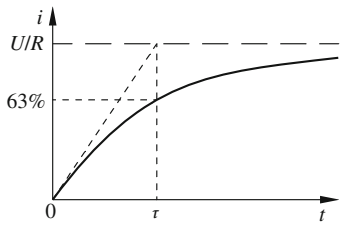


Abb. 29.13 Einschaltvorgang einer Spule

Teil IV

Beispiel

Eine Konstantstromquelle nach Abb. 29.14 speist die Reihenschaltung aus $R = 50\,\Omega$ und $L = 200\,\text{mH}$ mit einem Strom $I = 2\,\text{A}$. Zum Zeitpunkt $t = 0$ wird der Schalter geschlossen.

a) Welcher Strom fließt durch den Widerstand nach 2 ms?
b) Welche Spannung liegt dann am Widerstand?
c) Welche Spannung liegt dann an der Spule?

Lösung:

$$\tau = \frac{0{,}2\,\text{H}}{50\,\Omega} = \frac{0{,}2\,\frac{\text{Vs}}{\text{A}}}{50\,\frac{\text{V}}{\text{A}}} = 4\,\text{ms}$$

a) $$i = Ie^{-t/\tau} = 2\,\text{A} \cdot e^{-2\,\text{ms}/4\,\text{ms}} = 1{,}21\,\text{A}$$

b) $$u_R = i \cdot R = 1{,}21\,\text{A} \cdot 50\,\Omega = 60{,}65\,\text{V}$$

c) $$u_L = L \cdot \frac{di}{dt} = L \cdot I \cdot \left(-\frac{R}{L}\right) \cdot e^{-Rt/L}$$

$$u_L = 0{,}2\,\text{H} \cdot 2\,\text{A} \cdot \left(-\frac{50\,\Omega}{0{,}2\,\text{H}}\right) e^{-2\text{ms}/4\text{ms}}$$

$$= -60{,}65\,\text{V}$$

$$u_R + u_L = 0$$

◀

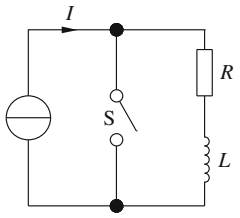


Abb. 29.14 Beispiel Kurzschließen einer Spule

29.3.5 Zusammenschalten von Induktivitäten

Für die Zusammenschaltung von Induktivitäten gelten dieselben Gesetze wie bei Ohm'schen Widerständen.

Reihenschaltung
$$L_{\text{ges}} = \sum_{i=1}^{n} L_i \tag{29.57}$$

Parallelschaltung
$$\frac{1}{L_{\text{ges}}} = \sum_{i=1}^{n} \frac{1}{L_i} \tag{29.58}$$

Wechselstrom

30

Horst Steffen



30.1 Grundbegriffe des Wechselstroms 362
30.2 Grundschaltelemente im Wechselstromkreis 364
30.3 Schaltungen von Wechselstromwiderständen 367
30.4 Passive Filter 376
30.5 Schwingkreise 378
30.6 Leistung und Arbeit im Wechselstromkreis 379

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_30

30.1 Grundbegriffe des Wechselstroms

Wechselstrom oder Wechselspannung sind Größen, bei denen sich die Werte zeitabhängig periodisch wiederholen. Der arithmetische Mittelwert der sich periodisch mit der Zeit ändernden Augenblickswerte ist gleich Null. Die Zeitspanne, nach der sich der periodische Verlauf gleichartig wiederholt, ist die *Periodendauer* T. Die Zahl der Perioden pro Zeit nennt man die *Frequenz* f.

$$\textit{Frequenz} \quad f = \frac{1}{T} \qquad \begin{array}{c|c} f & T \\ \hline \frac{1}{\mathrm{s}} & \mathrm{s} \end{array} \tag{30.1}$$

Die Einheit der Frequenz ist $[f] = \mathrm{s}^{-1} = \mathrm{Hz}$ (Hertz).

Nach DIN 5488 unterscheidet man periodische Vorgänge, deren Mittelwert nicht Null wird und solche, deren Mittelwert Null wird. Ist der Mittelwert Null, spricht man von einer Wechselgröße. Bei den Wechselgrößen unterscheidet man die nichtsinusförmigen und die sinusförmigen Vorgänge (Sinusgrößen). Im Abb. 30.1 ist der Verlauf einer Sinusgröße dargestellt.

30.1.1 Erzeugung einer sinusförmigen Wechselspannung

Wird eine Leiterschleife mit konstanter Winkelgeschwindigkeit ω in einem homogenen Magnetfeld gedreht (Abb. 30.2), so hängt die Größe der induzierten Spannung von der Winkelgeschwindigkeit und der vom Magnetfeld durchsetzten Fläche ab.

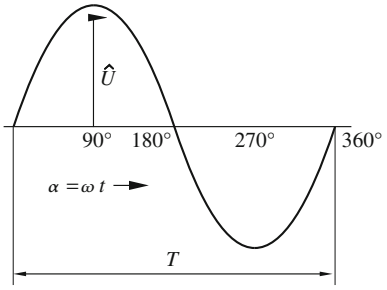


Abb. 30.1 Verlauf einer Sinusgröße

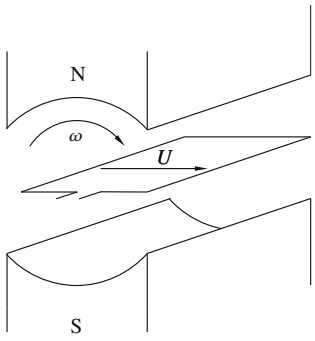


Abb. 30.2 Induktionsvorgang

induzierte Spannung

$$u_{\mathrm{ind}} = -\frac{d\Phi}{dt} = B \cdot A \cdot \omega \cdot \sin \omega t \qquad \begin{array}{c|c|c|c} u_{\mathrm{ind}} & B & A & \omega \\ \hline \mathrm{V} & \frac{\mathrm{V\,s}}{\mathrm{m}^2} & \mathrm{m}^2 & \frac{1}{\mathrm{s}} \end{array} \tag{30.2}$$

Maximalwert der Spannung (Scheitelwert $\hat{u}$):

$$\textit{Scheitelwert} \quad \hat{u} = B \cdot A \cdot \omega \tag{30.3}$$

Rotiert die Leiterschleife einmal im Magnetfeld, hat sie einen Winkel von 360° überstrichen. Dabei wurde in die Leiterschleife eine Wechselspannung induziert, deren zeitlicher Verlauf einer Sinusschwingung entspricht. Die Leiterschleife hat mit der Winkelgeschwindigkeit ω in der Periodendauer T einen Winkel $\alpha = 2\pi$ (Bogenmaß) überstrichen.

In Anlehnung an die gleichförmige Rotation mit der Winkelgeschwindigkeit ω hat man für Wechselgrößen die *Kreisfrequenz* definiert.

$$\textit{Kreisfrequenz} \quad \omega = 2\pi f \qquad \begin{array}{c|c} \omega & f \\ \hline \mathrm{s}^{-1} & \mathrm{s}^{-1} \end{array} \tag{30.4}$$

Die Kreisfrequenz ω gibt an, welcher Bogen in einer Sekunde überstrichen wird. Hieraus ergibt sich die *Periodendauer* T.

$$\textit{Periodendauer} \quad T = \frac{2\pi}{\omega} \tag{30.5}$$

Beispiel

Eine Leiterschleife dreht sich mit einer Umdrehung pro Sekunde. Welchen Wert hat die Kreisfrequenz und die Periodendauer?

Mit $f = 1\,\mathrm{s}^{-1}$ folgt für die Kreisfrequenz $\omega = 2\pi f = 2 \cdot 3{,}14 \cdot 1\,\mathrm{s}^{-1} = 6{,}28\,\mathrm{s}^{-1}$. Die Leiterschleife hat also einen Bogen von 2π in einer Sekunde zurückgelegt.

Für die Periodendauer ergibt sich

$$T = \frac{2\pi}{\omega} = \frac{2 \cdot 3{,}14}{6{,}28\,\mathrm{s}^{-1}} = 1\,\mathrm{s}^{-1}. \quad \blacktriangleleft$$

Beispiel

Eine Leiterschleife dreht sich mit 100 Umdrehungen pro Sekunde. Welchen Wert hat die Kreisfrequenz und die Periodendauer?

Mit $f = 100\,s^{-1}$ folgt für die Kreisfrequenz $\omega = 2\pi f = 2 \cdot 3{,}14 \cdot 100\,\mathrm{s}^{-1} = 6{,}28 \cdot 10^2\,\mathrm{s}^{-1}$. Die Leiterschleife hat also einen Bogen von $2\pi \cdot 100$ in einer Sekunde zurückgelegt.

Für die Periodendauer ergibt sich

$$T = \frac{2\pi}{\omega} = \frac{2 \cdot 3{,}14}{6{,}28 \cdot 100\,\mathrm{s}^{-1}} = 10^{-2}\,\mathrm{s}^{-1}. \quad \blacktriangleleft$$

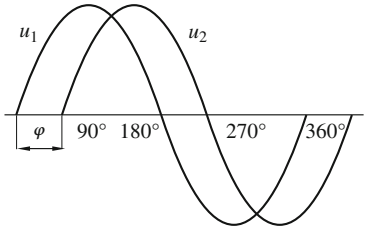


Abb. 30.3 Liniendiagramm zweier phasenverschobener Spannungen

30.1.2 Phasenverschiebung

Werden in einem Magnetfeld zwei Leiterschleifen, die in einem Winkel φ zueinander versetzt sind, mit gleicher Winkelgeschwindigkeit gedreht, so werden in ihnen zwei sinusförmige Spannungen induziert, die sich nicht decken. Nach Abb. 30.3 erreichen beide Spannungen ihr Maximum um den *Phasenwinkel* φ verschoben.

Die Spannungen sind *phasenverschoben*. Nach Abb. 30.3 erreicht die Spannung u_2 den Maximalwert später als die Spannung u_1. Die Spannung u_2 eilt der Spannung u_1 nach.

30.1.3 Effektivwert

Mit dem Scheitelwert $\hat{i}$ ergibt sich der *Effektivwert* I_{eff} eines sinusförmigen Wechselstroms:

$$\textit{Effektivwert} \quad I_{\text{eff}} = \frac{\hat{i}}{\sqrt{2}} = 0{,}707 \cdot \hat{i} \qquad (30.6)$$

Der *Effektivwert* I_{eff} eines sinusförmigen Stroms verrichtet an einem Widerstand R die gleiche elektrische Arbeit wie ein gleich großer Gleichstrom.

Gleiche Überlegungen für die Spannung ergeben den *Effektivwert* U_{eff} einer sinusförmigen Wechselspannung:

$$\textit{Effektivwert} \quad U_{\text{eff}} = \frac{\hat{u}}{\sqrt{2}} = 0{,}707 \cdot \hat{u} \qquad (30.7)$$

Effektivwerte werden wie Gleichstromwerte durch Großbuchstaben bezeichnet, eventuell zusätzlich durch den Index „eff“.

Beispiel

Welchen Effektivwert weist ein sinusförmiger Wechselstrom mit Scheitelwert $\hat{i} = 32\,\text{A}$ auf?

$$I = \frac{\hat{i}}{\sqrt{2}} = 0{,}707 \cdot \hat{i} = 0{,}707 \cdot 32\,\text{A} = 22{,}624\,\text{A} \quad \blacktriangleleft$$

Beispiel

Welchen Scheitelwert weist eine sinusförmige Spannung mit dem Effektivwert $U = 220\,\text{V}$ auf?

$$\hat{u} = \sqrt{2} \cdot U = 1{,}414 \cdot 220\,\text{V} = 311{,}08\,\text{V} \quad \blacktriangleleft$$

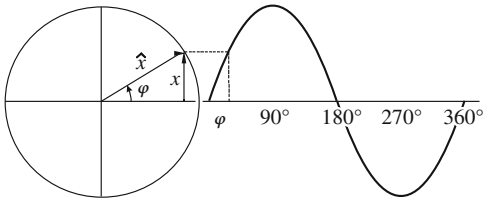


Abb. 30.4 Konstruktion eines Zeigers aus einer Sinusgröße

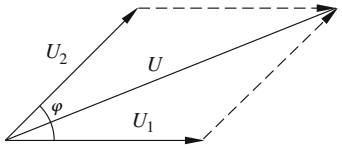


Abb. 30.5 Addition zweier Zeiger

30.1.4 Darstellungsarten

Eine Sinusgröße ist eindeutig bestimmt durch die Angabe von Amplitude, Frequenz und Phasenlage. Handelt es sich um Sinusgrößen mit gleicher Frequenz, genügen zwei Angaben, um die Sinusgröße eindeutig zu bestimmen. Aus diesem Grund kann bei der Berechnung von sinusförmigen Vorgängen auf die Darstellung in Abhängigkeit von der Zeit (Frequenz) verzichtet werden. Für die Berechnung von Sinusgrößen in der Elektrotechnik haben sich zwei Verfahren durchgesetzt: die Zeigerdarstellung und die Darstellung der Größen in der komplexen Zahlenebene.

30.1.4.1 Zeigerdarstellung von Sinusgrößen

Der zeitliche Verlauf einer sinusförmigen Wechselgröße kann über ein Linien- oder ein *Zeigerdiagramm* dargestellt werden.

Aus der Geometrie ist bekannt, dass sich eine Sinuskurve aus der Drehung eines Zeigers im Einheitskreis konstruieren lässt. Umgekehrt kann jede Sinuskurve durch einen Zeiger, der nach Abb. 30.4 rotiert, dargestellt werden.

Wird die Länge des Zeigers als Maximalwert der Sinusgröße festgelegt, kann der Augenblickswert in Abhängigkeit vom zurückgelegten Winkel bestimmt werden.

$$\textit{Augenblickswert} \quad x = \hat{x} \cdot \sin\varphi \qquad (30.8)$$

Zeichnerisch lässt sich der Augenblickswert ermitteln, indem von der Zeigerspitze das Lot auf die durch den Kreismittelpunkt gehende Gerade gefällt wird. Das Zeigerdiagramm stellt eine „Momentaufnahme“ der sich drehenden Zeiger dar.

Die Addition zweier sinusförmiger Wechselgrößen in der Zeigerdarstellung ergibt sich, wenn ein Zeiger parallelverschoben wird, bis sein Ursprung auf der Spitze des zweiten Zeigers liegt (Abb. 30.5).

Die Subtraktion zweier Zeiger ergibt sich, wenn ein Zeiger parallelverschoben wird, bis seine Spitze auf der Spitze des zweiten Zeigers liegt (Abb. 30.6).

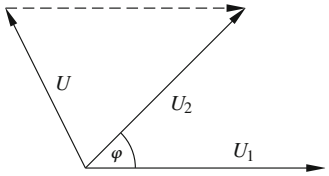


Abb. 30.6 Subtraktion zweier Zeiger

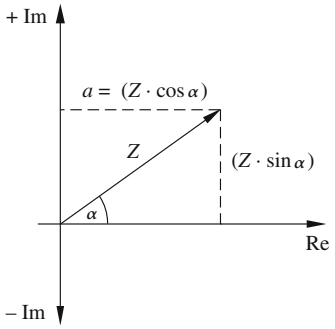


Abb. 30.7 Zeiger in der komplexen Zahlenebene

30.1.4.2 Darstellung von Sinusgrößen in der komplexen Zahlenebene

Eine Sinusgröße lässt sich durch die Zeigerdarstellung relativ einfach darstellen. Durch das graphische Verfahren dieser Darstellungsart ist eine genaue Auswertung oft unmöglich oder sehr zeitaufwendig. Aus diesem Grund wurden die Zeiger in der Gauß'schen Zahlenebene als komplexe Größen dargestellt und können somit berechnet werden.

In der Gauß'schen Zahlenebene wird jeder Punkt dieser Ebene durch eine komplexe Zahl beschrieben.

$$\textit{komplexe Zahl} \quad \underline{Z} = a + jb \tag{30.9}$$

Summand a stellt den reellen Anteil dar und Summand jb den imaginären Anteil. Beschreibt man den Endpunkt eines Zeigers $\underline{Z}$, der vom Nullpunkt der Gauß'schen Zahlenebene ausgeht, durch eine komplexe Zahl (Abb. 30.7), so ist damit auch der komplexe Zeiger eindeutig bestimmt.

$$\textit{kartesische Form} \quad \underline{Z} = a + jb = \underline{Z}_1 + \underline{Z}_2 \tag{30.10}$$

(mit $\underline{Z}_1 = a$ und $\underline{Z}_2 = jb$).

Der Zeiger $\underline{Z}$ kann also durch die Addition der beiden Zeiger $\underline{Z}_1$ und $\underline{Z}_2$ gebildet werden.

Vereinbarungsgemäß wird für die komplexe Darstellung von Zeigern das gleiche Symbol und die gleiche Definition benutzt wie für den Zeiger selbst. Der komplexe Zeiger wird mit dem unterstrichenen lateinischen Buchstaben der Größe bezeichnet. Der Betrag wird nur durch den lateinischen Buchstaben angegeben.

$$\textit{Betrag des Zeigers} \quad |\underline{Z}| = Z = \sqrt{Z_1^2 + Z_2^2} \tag{30.11}$$

Der Winkel mit der reellen Achse kann folgendermaßen berechnet werden:

$$\textit{Winkel des Zeigers} \quad \alpha = \arctan \frac{Z_2}{Z_1} \tag{30.12}$$

Der Betrag des Zeigers lässt sich über die Phasenlage in den Realteil und den Imaginärteil zerlegen.

$$\textit{Realteil} \quad \underline{Z}_1 = |\underline{Z}| \cdot \cos\alpha \tag{30.13}$$

$$\textit{Imaginärteil} \quad \underline{Z}_2 = |\underline{Z}| \cdot \sin\alpha \tag{30.14}$$

Mit Hilfe dieser Zerlegung kann der Zeiger in der polaren Form beschrieben werden.

$$\textit{polare Form} \quad \underline{Z} = |\underline{Z}| \cdot (\cos\alpha + j \sin\alpha) \tag{30.15}$$

Mit Hilfe der Euler'schen Gleichung ist die Exponentialform des Zeigers möglich:

$$\textit{Euler'sche Gleichung} \quad \cos\alpha + j \sin\alpha = e^{j\alpha} \tag{30.16}$$

$$\textit{Exponentialform} \quad \underline{Z} = |\underline{Z}| \cdot e^{j\alpha} = Z \cdot e^{j\alpha} \tag{30.17}$$

Die bisher aufgeführten Möglichkeiten der Darstellung eines Zeigers in der komplexen Ebene gelten allgemein, also unabhängig davon, ob der Winkel α konstant oder eine Funktion der Zeit ist. Bei der Behandlung von Wechselstromaufgaben muss jedoch zwischen den Operatoren und den Zeigern unterschieden werden.

Operatoren sind zeitlich konstante Größen. Sie symbolisieren meistens Widerstände. Ihr Winkel α zu der reellen Achse ist zeitlich konstant.

Zeiger symbolisieren sinusförmige Wechselgrößen. Ihr Winkel α ist eine Funktion der Zeit (Summe $\omega \cdot t + \varphi$), obwohl diese Tatsache bei der Darstellung nicht berücksichtigt wird.

Berücksichtig wird der Unterschied zwischen Operatoren und Zeigern bei der mathematischen Behandlung dieser Größen und zwar beim Integrieren und Differenzieren (Abb. 30.11, 30.12 und 30.13).

30.2 Grundschaltelemente im Wechselstromkreis

30.2.1 Ohm'scher Widerstand

Wird ein ohmscher Widerstand an eine Wechselspannung angeschlossen, ergibt der 2. Kirchhoff'sche Satz:

$$\textit{2. Kirchhoff'sche Satz} \quad u - i \cdot R = 0 \tag{30.18}$$

$$i = \frac{u}{R} \tag{30.19}$$

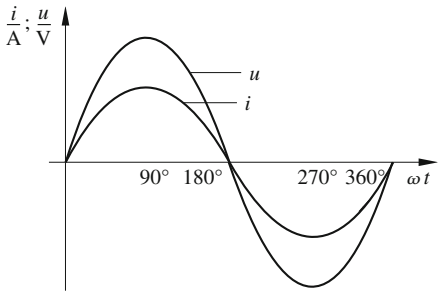


Abb. 30.8 Verlauf von Strom und Spannung am Ohm'schen Widerstand

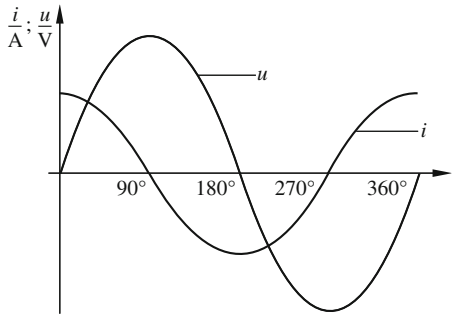


Abb. 30.9 Verlauf von Strom und Spannung an einer Kapazität

Der Wechselstrom hat also zu jeder Zeit einen Wert abhängig von u und R. Zwischen Strom und Spannung tritt keine Phasenverschiebung auf (Abb. 30.8). Der Scheitelwert des Stroms und der ohmsche Wechselstromwiderstand berechnet sich zu

$$\textit{Scheitelwert} \quad \hat{i} = \frac{\hat{u}}{R} \tag{30.20}$$

ohmscher Wechselstromwiderstand

$$R = \frac{\hat{u}}{\hat{i}} = \frac{U}{I} \tag{30.21}$$

Wechselstromleitwert

$$G = \frac{1}{R} = \frac{\hat{i}}{\hat{u}} = \frac{I}{U} \tag{30.22}$$

Die Darstellung in der komplexen Schreibweise ergibt folgende Gleichungen:

$$\textit{ohmsches Gesetz} \quad \underline{I} = \frac{\underline{U}}{\underline{Z}} \tag{30.23}$$

$$\textit{Widerstandsoperator} \quad \underline{Z} = R = \frac{\underline{U}}{\underline{I}} \tag{30.24}$$

$$\textit{Leitwertoperator} \quad \underline{Y} = \frac{1}{\underline{Z}} = G \tag{30.25}$$

30.2.2 Kapazität

Wird eine Kapazität an eine Wechselspannung angeschlossen, so ergibt der 2. Kirchhoff'sche Satz

$$\textit{2. Kirchhoff'sche Satz} \quad u - \frac{1}{C}\int i dt = 0 \tag{30.26}$$

$$i = C \cdot \frac{du}{dt} = C \cdot \hat{u} \cdot \omega \cdot \cos(\omega \cdot t) \tag{30.27}$$

(mit $u = \hat{u} \cdot \sin(\omega \cdot t)$).

Zwischen dem Strom i und der Spannung u besteht eine Phasenverschiebung von $\varphi = 90°$ (Abb. 30.9). Teilt man die Spannung durch den Strom, ergibt sich der *kapazitive Blindwiderstand* X_C

kapazitiver Blindwiderstand

$$|X_C| = X_C = \frac{\hat{u}}{\hat{i}} = \frac{U}{I} = \frac{1}{\omega \cdot C} \quad \begin{array}{c|c} \omega & C \\ \hline \frac{1}{\mathrm{s}} & \frac{\mathrm{As}}{\mathrm{V}} \end{array} \tag{30.28}$$

$$\textit{kapazitiver Leitwert} \quad B_C = \omega \cdot C \tag{30.29}$$

Der Kondensator im Wechselstromkreis ist also für den Wechselstrom kein unüberwindliches Hindernis, er wirkt aber trotzdem wie ein Widerstand. Dieses ist leicht zu verstehen, wenn man sich vorstellt, dass bei jeder Periode Ladungsträger zwischen den Kondensatorplatten und der Spannungsquelle bewegt werden.

Die Darstellung in der komplexen Schreibweise ergibt (30.30) bis (30.32).

$$\textit{ohmsches Gesetz} \quad \underline{I} = \frac{\underline{U}}{\underline{Z}} \tag{30.30}$$

Widerstandsoperator

$$\underline{Z} = \frac{1}{j\omega \cdot C} = -jX_C = \frac{\underline{U}}{\underline{I}} \tag{30.31}$$

Leitwertoperator

$$\underline{Y} = \frac{1}{\underline{Z}} = j\omega \cdot C = jB_C \tag{30.32}$$

30.2.3 Spule

Wird eine Spule an eine Wechselspannung angeschlossen, so ergibt der 2. Kirchhoff'sche Satz:

$$\textit{2. Kirchhoff'sche Satz} \quad u - L\frac{di}{dt} = 0 \tag{30.33}$$

$$u = L\frac{di}{dt} \tag{30.34}$$

$$i = \frac{1}{L} \cdot \int u dt = -\frac{1}{\omega \cdot L} \cdot \hat{u} \cdot \cos \omega \cdot t \tag{30.35}$$

(mit $u = \hat{u} \cdot \sin(\omega \cdot t)$).

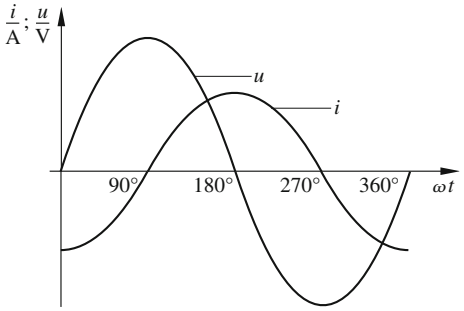


Abb. 30.10 Verlauf von Strom und Spannung an einer Induktivität

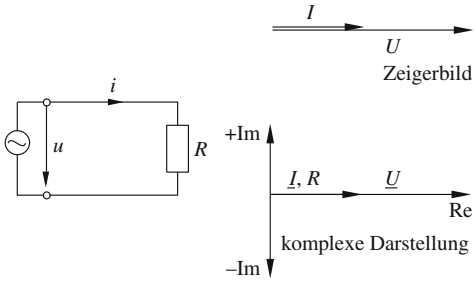


Abb. 30.11 Darstellung von Strom und Spannung am ohmschen Widerstand in der Zeiger- und komplexen Form

Zwischen dem Strom i und der Spannung u besteht eine Phasenverschiebung von $\varphi = -90°$ (Abb. 30.10). Teilt man die Spannung durch den Strom, ergibt sich der *induktive Blindwiderstand* X_L.

induktiver Blindwiderstand

$$|X_L| = X_L = \frac{\hat{u}}{\hat{i}} = \frac{U}{I} = \omega \cdot L \qquad \begin{array}{c|c} \omega & L \\ \hline \frac{1}{\mathrm{s}} & \frac{\mathrm{V\,s}}{\mathrm{A}} \end{array} \qquad (30.36)$$

$$\text{induktiver Leitwert} \quad B_L = \frac{1}{\omega \cdot L} \qquad (30.37)$$

Die Spule im Wechselstromkreis ist also nicht widerstandslos, obwohl der ohmsche Anteil vernachlässigt wird. Der induktive Blindwiderstand der Spule ist durch die auftretende Selbstinduktionsspannung an der Spule zu erklären. Diese Spannung der Spule bewirkt, dass der Spulenstrom der Wechselspannung um 90° nacheilt.

Die Darstellung in der komplexen Schreibweise ergibt (30.38) bis (30.40).

$$\text{ohmsches Gesetz} \qquad \underline{I} = \frac{\underline{U}}{\underline{Z}} \qquad (30.38)$$

$$\text{Widerstandsoperator} \qquad \underline{Z} = j\omega \cdot L = jX_L = \frac{\underline{U}}{\underline{I}} \qquad (30.39)$$

$$\text{Leitwertoperator} \qquad \underline{Y} = \frac{1}{\underline{Z}} = \frac{1}{j\omega \cdot L} = -jB_L \qquad (30.40)$$

Beispiel

Ein Widerstand von $R = 100\,\Omega$ wird an einem Wechselstromnetz mit $U = 220\,\mathrm{V}$ und $f = 50\,\mathrm{Hz}$ betrieben. Welchen Wert hat der sich einstellende Strom? Darstellung des Zeigerbildes.

Zeigerdarstellung:

$$\hat{i} = I = \frac{\hat{u}}{R} = \frac{U}{R} = \frac{220\,\mathrm{V}}{100\,\Omega} = 2{,}2\,\mathrm{A}$$

Winkel ist nicht darstellbar.

Komplexe Darstellung:

$$\underline{I} = \frac{\underline{U}}{\underline{Z}} = \frac{\underline{U}}{R} = \frac{220\,\mathrm{V} \cdot \mathrm{e}^{\mathrm{j}0}}{100\,\Omega \cdot e^{j0}} = 2{,}2\,\mathrm{A} \cdot e^{j0} \quad \blacktriangleleft$$

Beispiel

Eine Induktivität von $L = 100\,\mathrm{m}\,\frac{\mathrm{V\,s}}{\mathrm{A}}$ wird an einem Wechselstromnetz mit $U = 220\,\mathrm{V}$ und $f = 50\,\mathrm{Hz}$ betrieben. Welchen Wert hat der sich einstellende Strom? Darstellung des Zeigerbildes.

Zeigerdarstellung:

$$i = \frac{1}{L} \cdot \int u dt = -\frac{1}{\omega \cdot L} \cdot \hat{u} \cdot \cos\omega \cdot t$$

$$\Rightarrow \quad I = \frac{U}{\omega \cdot L} = \frac{220\,\mathrm{V}}{2 \cdot \pi \cdot 50\,\mathrm{s}^{-1} \cdot 100 \cdot 10^{-3}\,\frac{\mathrm{V\,s}}{\mathrm{A}}} = 7\,\mathrm{A}$$

Komplexe Darstellung:

$$\underline{I} = \frac{\underline{U}}{\underline{Z}} = \frac{U \cdot e^{j0}}{j\omega \cdot L} = \frac{220\,\mathrm{V} \cdot e^{j0}}{j \cdot 2 \cdot \pi \cdot 50\,\mathrm{s}^{-1} \cdot 100 \cdot 10^{-3}\,\frac{\mathrm{V\,s}}{\mathrm{A}}}$$

$$= \frac{220\,\mathrm{V} \cdot e^{j0}}{31{,}42 \cdot \frac{\mathrm{V\,s}}{\mathrm{A}} e^{j90}} = 7 \cdot \mathrm{A} e^{-j90} \quad \blacktriangleleft$$

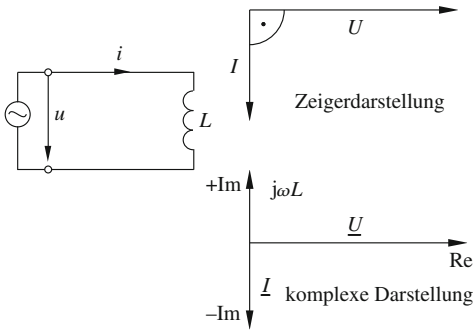


Abb. 30.12 Darstellung von Strom und Spannung am induktiven Widerstand in der Zeiger- und komplexen Form

Teil IV

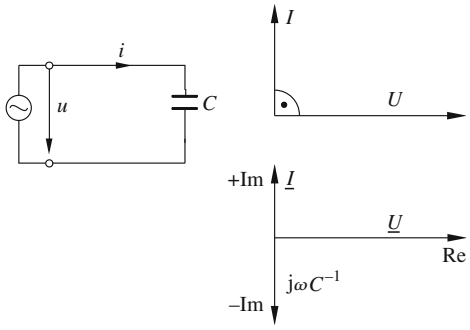


Abb. 30.13 Darstellung von Strom und Spannung am kapazitiven Widerstand in der Zeiger- und komplexen Form

Beispiel

Eine Kapazität von $C = 100\,\mu\frac{\mathrm{A\,s}}{\mathrm{V}}$ wird an einem Wechselstromnetz mit $U = 220\,\mathrm{V}$ und $f = 50\,\mathrm{Hz}$ betrieben. Welchen Wert hat der sich einstellende Strom? Darstellung des Zeigerbildes.

Zeigerdarstellung:

$$i = C \cdot \frac{du}{dt} = \omega \cdot C \cdot \hat{u} \cdot \cos\omega \cdot t$$

$$\Rightarrow \quad I = \omega \cdot C \cdot U$$

$$= 2 \cdot \pi \cdot 50\,\mathrm{s}^{-1} \cdot 100 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}} \cdot 220\,\mathrm{V} = 6{,}9\,\mathrm{A}$$

Komplexe Darstellung:

$$\underline{I} = \frac{\underline{U}}{\underline{Z}} = j \cdot \omega \cdot C \cdot U \cdot e^{j0}$$

$$= j \cdot 2 \cdot \pi \cdot 50\,\mathrm{s}^{-1} \cdot 100 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}} \cdot 220\,\mathrm{V}e^{j0}$$

$$= 0{,}03142\,\frac{\mathrm{A\,s}}{\mathrm{V}}e^{j90} \cdot 220 \cdot e^{j0}\,\mathrm{V} = 6{,}9\,\mathrm{A}e^{j90}$$ ◀

30.3 Schaltungen von Wechselstromwiderständen

Bei der weiteren Betrachtung von Wechselstromwiderständen werden nur noch die Effektivwerte der Wechselstromgrößen verwendet. Es wird nicht mehr ausdrücklich darauf hingewiesen, ob es sich bei dem benutzten Wert um einen Operator oder um einen Zeiger handelt.

30.3.1 Reihenschaltung von Wechselstromwiderständen

Im Zeigerbild für die Reihenschaltung drehen sich die Zeiger der Spannungen gleichmäßig um einen gemeinsamen Drehpunkt. Die Lage der Spannungen zueinander bleibt immer erhalten. Die Zeiger der Spannungen können also addiert oder subtrahiert werden. Da Wirk- und Blindanteil immer senkrecht aufeinander stehen, ist das *Spannungsdreieck* stets rechtwinklig. Für die Größen im rechtwinkligen Dreieck gilt:

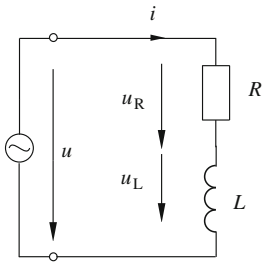


Abb. 30.14 Induktivität und ohmscher Widerstand im Wechselstromkreis

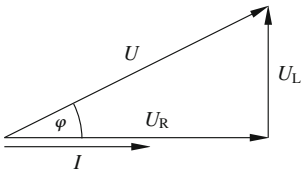


Abb. 30.15 Zeigerbild der Reihenschaltung von ohmschen Widerstand und Induktivität

$$\textit{Gesamtspannung} \quad U^2 = U_R^2 + U_B^2 \tag{30.41}$$

$$U = \sqrt{U_R^2 + U_B^2} \tag{30.42}$$

Wirkanteil der Spannung

$$U_R = \sqrt{U^2 - U_B^2} = U \cdot \cos\varphi \tag{30.43}$$

Blindanteil der Spannung

$$U_B = \sqrt{U^2 - U_R^2} = U \cdot \sin\varphi \tag{30.44}$$

30.3.1.1 Wirkwiderstand und Induktivität

Im allgemeinen enthält jede Induktionsspule neben dem induktiven Widerstand X_L auch einen ohmschen Widerstand R. Im Ersatzschaltbild werden beide Größen in Reihe geschaltet, da sie vom selben Strom durchflossen werden (Abb. 30.14).

Der Strom erzeugt am Widerstand den Spannungsfall $U_R = I \cdot R$ und an der Induktivität den Spannungsfall $U_L = \omega \cdot L \cdot I$. Beide Spannungen haben im Zeiger-bild verschiedene Richtungen. Addiert ergeben die beiden Spannungszeiger den Zeiger der Gesamtspannung U. Nach Abb. 30.15 liegt der Zeiger U_R in Richtung des Stroms I, der Zeiger U_L eilt dem Strom um 90° vor.

Das durch die Zeigeraddition entstehende Spannungsdreieck kann nach dem Lehrsatz von Pythagoras berechnet werden (Abb. 30.16).

Gesamtspannung

$$U^2 = U_R^2 + U_L^2 = I^2 \cdot R^2 + I^2 \cdot X_L^2 = I^2(R^2 + X_L^2) \quad (30.45)$$

Hieraus folgt das ohmsche Gesetz für Wirkwiderstand und Induktivität im Wechselstromkreis.

$$\textit{Gesamtspannung} \qquad U = I \cdot \sqrt{R^2 + X_L^2} \quad (30.46)$$

$$\textit{Scheinwiderstand} \qquad Z = \sqrt{R^2 + X_L^2} \quad (30.47)$$

Der Scheinwiderstand entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Wirkwiderstand R und dem induktiven Blindwiderstand X_L entsprechen.

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.15.

Phasenverschiebung

$$\varphi = \arccos \frac{U_R}{U} = \arccos \frac{I \cdot R}{I \cdot Z} = \arccos \frac{R}{Z} \quad (30.48)$$

$$\varphi = \arctan \frac{U_L}{U_R} = \arctan \frac{I \cdot X_L}{I \cdot R} = \arctan \frac{X_L}{R} \quad (30.49)$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge:

Gesamtspannung

$$\underline{U} = \underline{U}_R + \underline{U}_L = \underline{I} \cdot R + \underline{I} \cdot jX_L = \underline{I} \cdot (R + jX_L) = \underline{I} \cdot \underline{Z} \quad (30.50)$$

Scheinwiderstand

$$\underline{Z} = R + jX_L \quad (30.51)$$

$$Z = \sqrt{R^2 + X_L^2} \quad (30.52)$$

Phasenwinkel

$$\varphi = \arctan \frac{X_L}{R} \quad (30.53)$$

Beispiel

Eine Spule mit einem ohmschen Widerstand von $R = 100\,\Omega$ und einem induktiven Blindwiderstand von $X_L = 200\,\Omega$ wird an eine Spannung von $U = 200\,\text{V}$ angeschlossen.

Wie groß sind Scheinwiderstand und Stromstärke? Welche Spannungsfälle treten an den Bauelementen des Ersatzschaltbildes auf? Welcher Phasenwinkel herrscht zwischen Spannung und Strom?

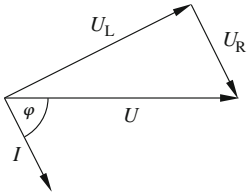


Abb. 30.16 Beispiel: Reihenschaltung von ohmschem Widerstand und Induktivität

Zeigerdarstellung:

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{100^2\,\Omega^2 + 200^2\,\Omega^2} = 223{,}6\,\Omega$$

$$I = \frac{U}{Z} = \frac{200\,\text{V}}{223{,}6\,\Omega} = 0{,}89\,\text{A}$$

$$U_R = I \cdot R = 0{,}89\,\text{A} \cdot 100\,\Omega = 89\,\text{V}$$

$$U_L = I \cdot X_L = 0{,}89\,\text{A} \cdot 200\,\Omega = 178\,\text{V}$$

$$\varphi = \arctan \frac{X_L}{R} = \arctan \frac{200}{100} = 63{,}4^\circ$$

Komplexe Darstellung:

$$\varphi = \arctan \frac{X_L}{R} = \arctan \frac{200}{100} = 63{,}4^\circ$$

$$\underline{Z} = R + jX_L = 100\,\Omega + j200\,\Omega = 223{,}6\,\Omega \cdot e^{j63{,}4}$$

$$\underline{I} = \frac{\underline{U}}{\underline{Z}} = \frac{220\,\text{V} \cdot e^{j0}}{223{,}6\,\Omega \cdot e^{j63{,}4}} = 0{,}89\,\text{A} \cdot e^{-j63{,}4}$$

$$\underline{U_R} = \underline{I} \cdot R$$

$$= 0{,}89\,\text{A} \cdot e^{-j63{,}4} \cdot 100\,\Omega \cdot e^{j0} = 89\,\text{V} \cdot e^{-j63{,}4}$$

$$\underline{U_L} = \underline{I} \cdot X_L$$

$$= 0{,}89\,\text{A} \cdot e^{-j63{,}4} \cdot 200\,\Omega \cdot e^{j90} = 178\,\text{V} \cdot e^{j26{,}6} \quad \blacktriangleleft$$

30.3.1.2 Wirkwiderstand und Kapazität

Sind in einem Wechselstromkreis ein ohmscher Widerstand R und eine Kapazität mit dem kapazitiven Blindwiderstand X_C nach Ersatzschaltbild (Abb. 30.17) in Reihe geschaltet, erzeugt der Strom am Widerstand den Spannungsfall $U_R = I \cdot R$ und an der Kapazität den Spannungsfall $U_C = \frac{1}{\omega \cdot C} \cdot I$.

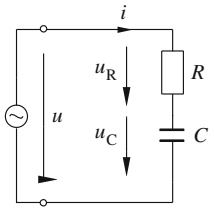


Abb. 30.17 Kapazität und ohmscher Widerstand im Wechselstromkreis

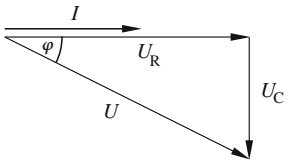


Abb. 30.18 Zeigerbild der Reihenschaltung von ohmschem Widerstand und Kapazität

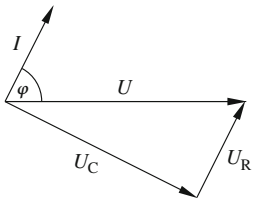


Abb. 30.19 Beispiel: Reihenschaltung von ohmschem Widerstand und Kapazität

Beide Spannungen haben im Zeigerbild verschiedene Richtungen. Addiert ergeben die beiden Spannungszeiger den Zeiger der Gesamtspannung U. Nach Abb. 30.18 liegt der Zeiger U_R in Richtung des Stroms I, der Zeiger U_C eilt dem Strom um 90° nach.

Das durch die Zeigeraddition entstehende Spannungsdreieck kann nach dem Lehrsatz von Pythagoras berechnet werden (Abb. 30.19).

Gesamtspannung

$$U^2 = U_R^2 + U_L^2 = I^2 \cdot R^2 + I^2 \cdot X_C^2 = I^2(R^2 + X_C^2) \quad (30.54)$$

Hieraus folgt das ohmsche Gesetz für Wirkwiderstand und Induktivität im Wechselstromkreis.

Gesamtspannung $$U = I \cdot \sqrt{R^2 + X_C^2} \quad (30.55)$$

Scheinwiderstand $$Z = \sqrt{R^2 + X_C^2} \quad (30.56)$$

Der Scheinwiderstand entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Wirkwiderstand R und dem kapazitiven Blindwiderstand X_C entsprechen.

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.18.

Phasenverschiebung

$$\varphi = \arccos \frac{U_R}{U} = \arccos \frac{I \cdot R}{I \cdot Z} = \arccos \frac{R}{Z} \quad (30.57)$$

$$\varphi = \arctan \frac{U_C}{U_R} = \arctan \frac{I \cdot X_C}{I \cdot R} = \arctan \frac{X_C}{R} \quad (30.58)$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge.

Gesamtspannung

$$\underline{U} = \underline{U}_R + \underline{U}_C = \underline{I} \cdot R + \underline{I} \cdot jX_C = \underline{I} \cdot (R + jX_C) = \underline{I} \cdot \underline{Z} \quad (30.59)$$

Scheinwiderstand

$$\underline{Z} = R + jX_C \quad (30.60)$$

$$Z = \sqrt{R^2 + X_C^2} \quad (30.61)$$

Phasenwinkel

$$\varphi = \arctan \frac{X_C}{R} \quad (30.62)$$

Beispiel

Ein ohmscher Widerstand von $R = 100\,\Omega$ und eine Kapazität mit einem kapazitiven Blindwiderstand von $X_C = 200\,\Omega$ werden an eine Spannung von $U = 200\,\text{V}$ angeschlossen.

Wie groß sind Scheinwiderstand und Stromstärke? Welche Spannungsfälle treten an den Bauelementen auf? Welcher Phasenwinkel herrscht zwischen Spannung und Strom?

Zeigerdarstellung:

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{100^2\,\Omega^2 + 200^2\,\Omega^2} = 223{,}6\,\Omega$$

$$I = \frac{U}{Z} = \frac{200\,\text{V}}{223{,}6\,\Omega} = 0{,}89\,\text{A}$$

$$U_R = I \cdot R = 0{,}89\,\text{A} \cdot 100\,\Omega = 89\,\text{V}$$

$$U_C = I \cdot X_C = 0{,}89\,\text{A} \cdot 200\,\Omega = 178\,\text{V}$$

$$\varphi = \arctan \frac{X_C}{R} = \arctan \frac{200}{100} = 63{,}4°$$

Komplexe Darstellung:

$$\varphi = \arctan \frac{X_C}{R} = \arctan \frac{200}{100} = 63{,}4°$$

$$\underline{Z} = R - jX_C = 100\Omega - j200\,\Omega = 223{,}6\,\Omega \cdot e^{-j63{,}4}$$

$$\underline{I} = \frac{\underline{U}}{\underline{Z}} = \frac{220\,\text{V} \cdot e^{j0}}{223{,}6\,\Omega \cdot e^{-j63{,}4}} = 0{,}89\,\text{A} \cdot e^{j63{,}4}$$

$$\underline{U}_R = \underline{I} \cdot R = 0{,}89\,\text{A} \cdot e^{j63{,}4} \cdot 100\,\Omega \cdot e^{j0} = 89\,\text{V} \cdot e^{j63{,}4}$$

$$\underline{U}_C = \underline{I} \cdot X_C = 0{,}89\,\text{A} \cdot e^{j63{,}4} \cdot 200\,\Omega \cdot e^{-j90} = 178\,\text{V} \cdot e^{-j26{,}6}$$ ◀

30.3.1.3 Wirkwiderstand, Induktivität und Kapazität

Sind in einem Wechselstromkreis ein ohmscher Widerstand R, eine Induktivität mit dem induktiven Blindwiderstand X_L und eine Kapazität mit dem kapazitiven Blindwiderstand X_C nach Abb. 30.20 in Reihe geschaltet, erzeugt der Strom am Widerstand den Spannungsfall $U_R = I \cdot R$, an der Induktivität den Spannungsfall $U_L = \omega \cdot L \cdot I$ und an der Kapazität den Spannungsfall $U_C = \frac{1}{\omega \cdot C} \cdot I$.

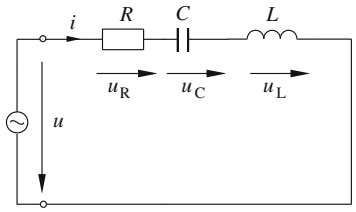


Abb. 30.20 Kapazität, Induktivität und ohmscher Widerstand im Wechselstromkreis

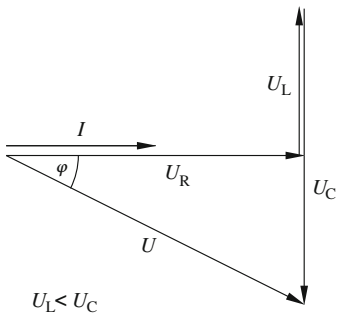


Abb. 30.21 Zeigerbild der Reihenschaltung vom ohmschem Widerstand, Induktivität und Kapazität

Die Spannungen haben im Zeigerbild verschiedenen Richtungen. Addiert ergeben die Spannungszeiger den Zeiger der Gesamtspannung U. Nach Abb. 30.21 liegt der Zeiger U_R in Richtung des Stroms I, der Zeiger U_C eilt dem Strom um 90° nach und der Zeiger U_L eilt dem Strom um 90° vor.

Das durch die Zeigeraddition entstehende Zeigerbild kann nach dem Lehrsatz von Pythagoras berechnet werden.

Gesamtspannung

$$U^2 = U_R^2 + U'^2 = U_R^2 + (U_L - U_C)^2 = I^2 \cdot R^2 + I^2 \cdot X'^2 = I^2(R^2 + (X_L - X_C)^2) \tag{30.63}$$

Hieraus folgt das ohmsche Gesetz für die Reihenschaltung von Wirkwiderstand, Induktivität und Kapazität im Wechselstromkreis.

$$\textit{Gesamtspannung} \qquad U = I \cdot \sqrt{R^2 + (X_L - X_C)^2} \tag{30.64}$$

$$\textit{Scheinwiderstand} \qquad Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{30.65}$$

Der Scheinwiderstand entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Wirkwiderstand R und der Differenz des induktiven Blindwiderstandes X_L und des kapazitiven Blindwiderstand X_C entsprechen.

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.21.

Phasenverschiebung

$$\varphi = \arccos \frac{U_R}{U} = \arccos \frac{I \cdot R}{I \cdot Z} = \arccos \frac{R}{Z} \tag{30.66}$$

$$\varphi = \arctan \frac{U'}{U_R} = \arctan \frac{I \cdot X'}{I \cdot R} = \arctan \frac{X_L - X_C}{R} \tag{30.67}$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge.

Gesamtspannung

$$\underline{U} = \underline{U_R} + \underline{U'} = \underline{I} \cdot R + \underline{I} \cdot j \cdot (X_L - X_C) = \underline{I} \cdot (R + j \cdot (X_L - X_C)) = \underline{I} \cdot \underline{Z} \tag{30.68}$$

Scheinwiderstand

$$\underline{Z} = R + j \cdot (X_L - X_C) \tag{30.69}$$

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{30.70}$$

Phasenwinkel

$$\varphi = \arctan \frac{(X_L - X_C)}{R} \tag{30.71}$$

Bei der Reihenschaltung von R, L und C stellt der ohmsche Widerstand den Realteil dar, während der Imaginärteil eine Zusammensetzung aus induktiven und kapazitiven Blindwiderstand ist. Der Imaginärteil kann verschiedene elektrische Eigenschaften aufweisen. Ist der induktive Anteil größer als der kapazitive, ergibt die Zusammenfassung einen induktiv wirkenden Blindwiderstand.

$$X_L > X_C \quad \Rightarrow \quad X' = X_{L\,\text{ers}} = X_L - X_C \tag{30.72}$$

Ist der kapazitive Anteil größer als der induktive, ergibt die Zusammenfassung einen kapazitiv wirkenden Blindwiderstand.

$$X_C > X_L \quad \Rightarrow \quad X' = X_{C\,\text{ers}} = X_C - X_L \tag{30.73}$$

Sind kapazitiver und induktiver Anteil gleich groß, wird der Imaginärteil Null; die Schaltung hat einen rein ohmschen Charakter. In diesem Fall spricht man von einer Reihenresonanz.

Reihenresonanz

$$X_L = X_C \quad \Rightarrow \quad X' = 0 = \omega \cdot L - \frac{1}{\omega \cdot C} \tag{30.74}$$

Über die Resonanzbedingung von (30.74) kann die *Resonanzfrequenz* f_0 bzw. ω_0 berechnet werden.

Resonanzbedingung

$$\omega_0 \cdot L = \frac{1}{\omega_0 \cdot C} \quad \Rightarrow \quad \omega_0^2 = \frac{1}{L \cdot C} \quad \Rightarrow \quad \omega_0 = \frac{1}{\sqrt{L \cdot C}} \tag{30.75}$$

Resonanzfrequenz

$$f_0 = \frac{\omega_0}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}} \tag{30.76}$$

Beispiel

Ein ohmscher Widerstand von $R = 100\,\Omega$, eine Induktivität mit einem induktiven Blindwiderstand von $X_L = 300\,\Omega$ und eine Kapazität mit einem kapazitiven Blindwiderstand von $X_C = 200\,\Omega$ werden an eine Spannung von $U = 200\,\mathrm{V}$ angeschlossen. Die Netzfrequenz beträgt $f = 50\,\mathrm{Hz}$.

Wie groß sind Scheinwiderstand und Stromstärke? Welche Spannungsfälle treten an den Bauelementen auf? Welcher Phasenwinkel herrscht zwischen Spannung und Strom? Welche Resonanzfrequenz hat diese Schaltung?

Zeigerdarstellung:

$$\begin{aligned}
Z &= \sqrt{R^2 + (X_L - X_C)^2} \\
&= \sqrt{100^2\,\Omega^2 + (300\,\Omega - 200\,\Omega)^2} = 141{,}4\,\Omega \\
I &= \frac{U}{Z} = \frac{200\,\mathrm{V}}{141{,}4\,\Omega} = 1{,}41\,\mathrm{A} \\
U_R &= I \cdot R = 1{,}41\,\mathrm{A} \cdot 100\,\Omega = 141\,\mathrm{V} \\
U_L &= I \cdot X_L = 1{,}41\,\mathrm{A} \cdot 300\,\Omega = 423\,\mathrm{V} \\
U_C &= I \cdot X_C = 1{,}41\,\mathrm{A} \cdot 200\,\Omega = 282\,\mathrm{V} \\
\varphi &= \arctan \frac{(X_L - X_C)}{R} \\
&= \arctan \frac{(300\,\Omega - 200\,\Omega)}{100\,\Omega} = 45\,^\circ \\
L &= \frac{X_L}{\omega} = \frac{300\,\frac{\mathrm{V}}{\mathrm{A}}}{2 \cdot \pi \cdot 50\,\frac{1}{\mathrm{s}}} = 0{,}955\,\frac{\mathrm{V\,s}}{\mathrm{A}} \\
C &= \frac{1}{\omega \cdot X_C} = \frac{1}{2 \cdot \pi \cdot 50\,\frac{1}{\mathrm{s}} \cdot 200\,\frac{\mathrm{V}}{\mathrm{A}}} = 15{,}9 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}} \\
f_0 &= \frac{\omega_0}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}} \\
&= \frac{1}{2 \cdot \pi \cdot \sqrt{0{,}955\,\frac{\mathrm{V\,s}}{\mathrm{A}} \cdot 15{,}9 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}}}} = 40{,}84\,\frac{1}{\mathrm{s}}
\end{aligned}$$

Komplexe Darstellung:

$$\begin{aligned}
\varphi &= \arctan \frac{(X_L - X_C)}{R} = \arctan \frac{300\,\Omega - 200\,\Omega}{100\,\Omega} = 45^\circ \\
\underline{Z} &= R - j(X_L - X_C) \\
&= 100\,\Omega - j(300 - 200)\,\Omega = 141{,}4\,\Omega \cdot e^{j45} \\
\underline{I} &= \frac{\underline{U}}{\underline{Z}} = \frac{200\,\mathrm{V} \cdot e^{j0}}{141{,}4\,\Omega \cdot e^{j45}} = 1{,}41\,\mathrm{A} \cdot e^{-j45} \\
\underline{U}_R &= \underline{I} \cdot R \\
&= 1{,}41\,\mathrm{A} \cdot e^{-j45} \cdot 100\,\Omega \cdot e^{j0} = 141\,\mathrm{V} \cdot e^{-j45} \\
\underline{U}_L &= \underline{I} \cdot X_L \\
&= 1{,}41\,\mathrm{A} \cdot e^{-j45} \cdot 300\,\Omega \cdot e^{j90} = 423\,\mathrm{V} \cdot e^{j45} \\
\underline{U}_C &= \underline{I} \cdot X_C \\
&= 1{,}41\,\mathrm{A} \cdot e^{-j45} \cdot 200\,\Omega \cdot e^{-j90} = 282\,\mathrm{V} \cdot e^{-j135}
\end{aligned}$$

◀

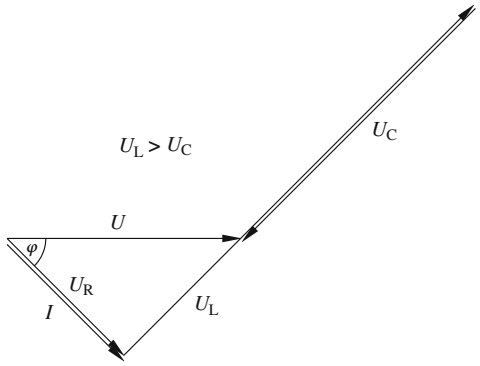


Abb. 30.22 Beispiel: Reihenschaltung von ohmschen Widerstand, Induktivität und Kapazität

30.3.2 Parallelschaltung von Wechselstromwiderständen

Im Zeigerbild für die Parallelschaltung drehen sich die Zeiger der Ströme gleichmäßig um einen gemeinsamen Drehpunkt. Die Lage der Ströme zueinander bleibt immer erhalten. Die Zeiger der Ströme können also addiert oder subtrahiert werden. Da Wirk- und Blindanteil immer senkrecht aufeinander stehen, ist das *Stromdreieck* stets rechtwinklig.

Für die Größen im rechtwinkligen Dreieck gilt:

$$\text{Gesamtstrom} \qquad I^2 = I_R^2 + I_B^2 \tag{30.77}$$

$$I = \sqrt{I_R^2 + I_B^2} \tag{30.78}$$

$$\text{Wirkstrom} \qquad I_R = \sqrt{I^2 - I_B^2} = I \cdot \cos\varphi \tag{30.79}$$

$$\text{Blindstrom} \qquad I_B = \sqrt{I^2 - I_R^2} = I \cdot \sin\varphi \tag{30.80}$$

30.3.2.1 Wirkwiderstand und Induktivität

Für jede Parallelschaltung gilt, dass der Gesamtstrom die Summe der Teilströme in den Parallelzweigen ist. In der Parallelschaltung gilt als wesentlicher Grundsatz, dass Leitwerte addiert werden können.

Sind ein Wirkwiderstand R und ein induktiver Blindwiderstand X_L parallelgeschaltet, treibt die anliegende Wechselspannung nach Abb. 30.23 einen Wirkstrom $I_W = \frac{U}{R} = U \cdot G$ durch den Wirkwiderstand und einen Blindstrom $I_B = \frac{U}{X_L} = U \cdot B_L$ durch den Blindwiderstand.

Der Wirkstrom zeigt in Richtung der Spannung U (Abb. 30.24), der Blindstrom eilt um 90° nach.

Das durch die Zeigeraddition entstehende Zeigerbild kann nach dem Lehrsatz von Pythagoras berechnet werden.

Gesamtstrom

$$I^2 = I_R^2 + I_B^2 = U^2 \cdot G^2 + U^2 \cdot B_L^2 = U^2 \cdot (G^2 + B_L^2) \tag{30.81}$$

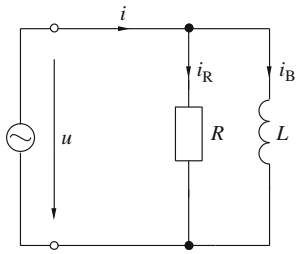


Abb. 30.23 Parallelschaltung von Induktivität und ohmschem Widerstand im Wechselstromkreis

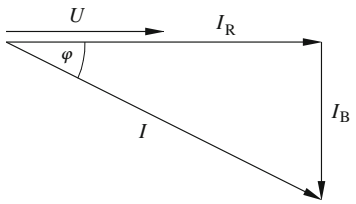


Abb. 30.24 Zeigerbild der Parallelschaltung von ohmschem Widerstand und Induktivität

Hieraus folgt das ohmsche Gesetz für die Parallelschaltung von Wirkwiderstand und Induktivität im Wechselstromkreis.

Gesamtstrom
$$I = U \cdot \sqrt{G^2 + B_L^2} \tag{30.82}$$

Scheinleitwert
$$Y = \sqrt{G^2 + B_L^2} \tag{30.83}$$

Scheinwiderstand
$$Z = \frac{1}{Y} = \frac{1}{\sqrt{G^2 + B_L^2}} \tag{30.84}$$

Der Scheinleitwert entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Leitwert des Wirkwiderstandes R und dem Leitwert des induktiven Blindwiderstandes X_L entsprechen.

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.24.

Phasenverschiebung
$$\varphi = \arccos \frac{I_R}{I} = \arccos \frac{U \cdot G}{U \cdot Y} = \arccos \frac{G}{Y} \tag{30.85}$$

$$\varphi = \arctan \frac{I_L}{I_R} = \arctan \frac{U \cdot B_L}{U \cdot G} = \arctan \frac{B_L}{G} \tag{30.86}$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge:

Gesamtstrom
$$\underline{I} = \underline{I}_R + \underline{I}_L = \underline{U} \cdot G + \underline{U} \cdot (-\mathrm{j}B_L) = \underline{U} \cdot (G - \mathrm{j}B_L) = \underline{U} \cdot \underline{Y} \tag{30.87}$$

Scheinleitwert
$$\underline{Y} = G - jB_L \tag{30.88}$$

$$Y = \sqrt{G^2 + B_L^2} \tag{30.89}$$

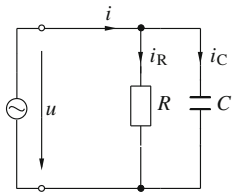


Abb. 30.25 Parallelschaltung von Kapazität und ohmschem Widerstand im Wechselstromkreis

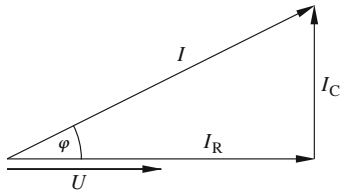


Abb. 30.26 Zeigerbild der Parallelschaltung von ohmschem Widerstand und Kapazität

Scheinwiderstand
$$\underline{Z} = \frac{1}{\underline{Y}} = \frac{1}{G - jB_l} \tag{30.90}$$

Phasenwinkel
$$\varphi = \arctan \frac{B_L}{G} \tag{30.91}$$

30.3.2.2 Wirkwiderstand und Kapazität

Sind in einem Wechselstromkreis ein ohmscher Widerstand Rund eine Kapazität mit dem kapazitiven Blindwiderstand X_C nach Ersatzschaltbild (Abb. 30.25) parallel geschaltet, erzeugt die Spannung U durch den Widerstand den Strom $I_R = U \cdot G$ und durch die Kapazität den Strom $I_C = B_C \cdot U$. Beide Ströme haben im Zeigerbild verschiedene Richtungen. Addiert ergeben die beiden Stromzeiger den Zeiger des Gesamtstromes I. Nach Abb. 30.26 liegt der Zeiger I_R in Richtung der Spannung U; der Zeiger I_C eilt der Spannung um 90° vor.

Das durch die Zeigeraddition entstehende Stromdreieck kann nach dem Lehrsatz von Pythagoras berechnet werden.

Gesamtstrom
$$I^2 = I_R^2 + I_C^2 = U^2 \cdot G^2 + U^2 \cdot B_C^2 = U^2(G^2 + B_C^2) \tag{30.92}$$

Hieraus folgt das ohmsche Gesetz für die Parallelschaltung von Wirkwiderstand und Kapazität im Wechselstromkreis.

Gesamtstrom
$$I = U \cdot \sqrt{G^2 + B_C^2} \tag{30.93}$$

Scheinleitwert
$$Y = \sqrt{G^2 + B_C^2} \tag{30.94}$$

Der Scheinleitwert entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Leitwert des Wirkwiderstandes R und dem Leitwert des kapazitiven Blindwiderstandes X_C entsprechen.

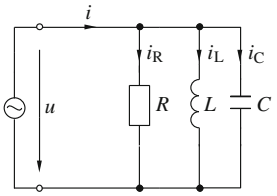


Abb. 30.27 Parallelschaltung von Induktivität, Kapazität und ohmschen Widerstand im Wechselstromkreis

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.26.

Phasenverschiebung

$$\varphi = \arccos \frac{I_R}{I} = \arccos \frac{U \cdot G}{U \cdot Y} = \arccos \frac{G}{Y} \tag{30.95}$$

$$\varphi = \arctan \frac{I_C}{I_R} = \arctan \frac{U \cdot B_C}{U \cdot G} = \arctan \frac{B_C}{G} \tag{30.96}$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge.

Gesamtstrom

$$\underline{I} = \underline{I}_R + \underline{I}_C = \underline{U} \cdot G + \underline{U} \cdot jB_C = \underline{U} \cdot (G + jB_C) = \underline{U} \cdot \underline{Y} \tag{30.97}$$

Scheinleitwert

$$\underline{Y} = G + jB_C \tag{30.98}$$

$$Y = \sqrt{G^2 + B_C^2} \tag{30.99}$$

Phasenwinkel

$$\varphi = \arctan \frac{B_C}{G} \tag{30.100}$$

30.3.2.3 Wirkwiderstand, Induktivität und Kapazität

Sind in einem Wechselstromkreis ein ohmscher Widerstand R, eine Induktivität mit dem induktiven Blindwiderstand X_L und eine Kapazität mit dem kapazitiven Blindwiderstand X_C nach Ersatzschaltbild (Abb. 30.27) parallel geschaltet, erzeugt die Spannung durch den Widerstand den Strom $I_R = U \cdot G$, durch die Induktivität den Strom $I_L = B_L \cdot U$ und durch die Kapazität den Strom $I_C = B_C \cdot U$.

Die Ströme haben im Zeigerbild verschiedene Richtungen. Addiert ergeben die Stromzeiger den Zeiger des Gesamtstroms I. Nach Abb. 30.28 liegt der Zeiger I_R in Richtung der Spannung U, der Zeiger I_C eilt der Spannung um 90° vor und der Zeiger I_L eilt der Spannung um 90° nach.

Das durch die Zeigeraddition entstehende Zeigerbild kann nach dem Lehrsatz von Pythagoras berechnet werden.

Gesamtstrom

$$I^2 = I_R^2 + I'^2 = I_R^2 + (I_C - I_L)^2 = U^2 \cdot G^2 + U^2 \cdot B'^2 = U^2(G^2 + (B_C - B_L)^2) \tag{30.101}$$

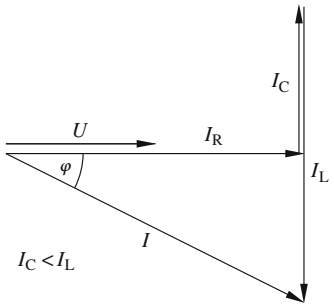


Abb. 30.28 Zeigerbild der Parallelschaltung von ohmschem Widerstand, Induktivität und Kapazität

Hieraus folgt das ohmsche Gesetz für die Parallelschaltung von Wirkwiderstand, Induktivität und Kapazität im Wechselstromkreis.

Gesamtstrom $$I = U \cdot \sqrt{G^2 + (B_C - B_L)^2} \tag{30.102}$$

Scheinleitwert $$Y = \sqrt{G^2 + (B_C - B_L)^2} \tag{30.103}$$

Der Scheinleitwert entspricht der Hypotenuse eines rechtwinkligen Dreiecks, dessen Katheten dem Leitwert des Wirkwiderstand R und der Differenz des Leitwertes des kapazitiven Blindwiderstand X_C und des Leitwertes des induktiven Blindwiderstandes X_L entsprechen.

Die Phasenverschiebung zwischen der Gesamtspannung U und dem Strom I erhält man nach Abb. 30.28.

Phasenverschiebung

$$\varphi = \arccos \frac{I_R}{I} = \arccos \frac{U \cdot G}{U \cdot Y} = \arccos \frac{G}{Y} \tag{30.104}$$

$$\varphi = \arctan \frac{I'}{I_R} = \arctan \frac{U \cdot B'}{U \cdot G} = \arctan \frac{B_C - B_L}{G} \tag{30.105}$$

In der komplexen Darstellung ergeben sich folgende Zusammenhänge:

Gesamtstrom

$$\underline{I} = \underline{I}_R + \underline{I}' = \underline{U} \cdot G + \underline{U} \cdot j \cdot (B_C - B_L) = \underline{U} \cdot (G + j \cdot (B_C - B_L)) = \underline{U} \cdot \underline{Y} \tag{30.106}$$

Scheinleitwert

$$\underline{Y} = G + j \cdot (B_C - B_L) \tag{30.107}$$

$$Y = \sqrt{G^2 + (B_C - B_L)^2} \tag{30.108}$$

Phasenwinkel

$$\varphi = \arctan \frac{(B_C - B_L)}{G} \tag{30.109}$$

Bei der Parallelschaltung von R, L und C stellt der ohmsche Widerstand den Realteil dar, während der Imaginärteil eine Zusammensetzung aus induktivem und kapazitivem Blindwiderstand ist. Der Imaginärteil kann verschiedene elektrische

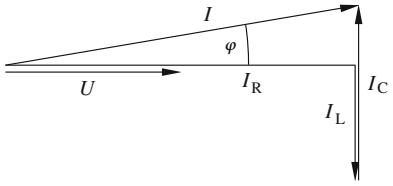


Abb. 30.29 Beispiel: Parallelschaltung von ohmschem Widerstand, Induktivität und Kapazität

Eigenschaften aufweisen. Ist der induktive Anteil größer als der kapazitive, ergibt die Zusammenfassung einen induktiv wirkenden Leitwert.

$$B_L > B_C \quad \Rightarrow \quad B' = B_{L\,\mathrm{ers}} = B_L - B_C \tag{30.110}$$

Ist der kapazitive Anteil größer als der induktive, ergibt die Zusammenfassung einen kapazitiv wirkenden Leitwert.

$$B_C > B_L \quad \Rightarrow \quad B' = B_{C\,\mathrm{ers}} = B_C - B_L \tag{30.111}$$

Sind der kapazitive und der induktive Anteil gleich groß, wird der Imaginärteil Null; die Schaltung hat einen rein ohmschen Charakter. In diesem Fall spricht man von einer Parallelresonanz.

Parallelresonanz

$$B_L = B_C \quad \Rightarrow \quad B' = 0 = \omega \cdot C - \frac{1}{\omega \cdot L} \tag{30.112}$$

Über die Resonanzbedingung von (30.112) kann die *Resonanzfrequenz* f_0 bzw. ω_0 berechnet werden.

Resonanzbedingung

$$\omega_0 \cdot C = \frac{1}{\omega_0 \cdot L} \quad \Rightarrow \quad \omega_0^2 = \frac{1}{L \cdot C} \quad \Rightarrow \quad \omega_0 = \frac{1}{\sqrt{L \cdot C}} \tag{30.113}$$

Resonanzfrequenz

$$f_0 = \frac{\omega_0}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}} \tag{30.114}$$

Beispiel

Ein ohmscher Widerstand von $R = 100\,\Omega$, eine Induktivität mit einem induktiven Blindwiderstand von $X_L = 300\,\Omega$ und eine Kapazität mit einem kapazitiven Blindwiderstand von $X_C = 200\,\Omega$ werden an eine Spannung von $U = 200\,\mathrm{V}$ angeschlossen. Die Netzfrequenz beträgt $f = 50\,\mathrm{Hz}$ (Abb. 30.29).

Wie groß sind Scheinleitwert und Gesamtstromstärke? Welche Ströme fließen durch die Bauelemente, welcher Phasenwinkel herrscht zwischen Spannung und Strom? Welche Resonanzfrequenz hat diese Schaltung?

Zeigerdarstellung:

$$G = \frac{1}{R} = \frac{1}{100\,\Omega} = 10 \cdot 10^{-3}\,\mathrm{S};$$

$$B_L = \frac{1}{X_L} = \frac{1}{300\,\Omega} = 3{,}33 \cdot 10^{-3}\,\mathrm{S}$$

$$B_C = \frac{1}{X_C} = \frac{1}{200\,\Omega} = 5 \cdot 10^{-3}\,\mathrm{S}$$

$$Y = \sqrt{G^2 + (B_C - B_L)^2}$$

$$= \sqrt{(10 \cdot 10^{-3})^2\,\mathrm{S}^2 + (5 \cdot 10^{-3}\,\mathrm{S} - 3{,}3 \cdot 10^{-3}\,\mathrm{S})^2}$$

$$= 10{,}14 \cdot 10^{-3}\,\mathrm{S}$$

$$I = U \cdot Y = 200\,\mathrm{V} \cdot 10{,}14 \cdot 10^{-3}\,\mathrm{S} = 2{,}03\,\mathrm{A}$$

$$I_R = U \cdot G = 200\,\mathrm{V} \cdot 10 \cdot 10^{-3}\,\mathrm{S} = 2\,\mathrm{A}$$

$$I_L = U \cdot B_L = 200\,\mathrm{V} \cdot 3{,}3 \cdot 10^{-3}\,\mathrm{S} = 0{,}66\,\mathrm{A}$$

$$I_C = U \cdot B_C = 200\,\mathrm{V} \cdot 5 \cdot 10^{-3}\,\mathrm{S} = 1\,\mathrm{A}$$

$$\varphi = \arccos\frac{G}{Y} = \arccos\frac{10 \cdot 10^{-3}\,\mathrm{S}}{10{,}14 \cdot 10^{-3}\,\mathrm{S}} = 9{,}5\,^\circ$$

$$L = \frac{X_L}{\omega} = \frac{300\,\frac{\mathrm{V}}{\mathrm{A}}}{2 \cdot \pi \cdot 50\,\frac{1}{\mathrm{s}}} = 0{,}955\,\frac{\mathrm{V\,s}}{\mathrm{A}}$$

$$C = \frac{1}{\omega \cdot X_C} = \frac{1}{2 \cdot \pi \cdot 50\,\frac{1}{\mathrm{s}} \cdot 200\,\frac{\mathrm{V}}{\mathrm{A}}} = 15{,}9 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}}$$

$$f_0 = \frac{\omega_0}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}}$$

$$= \frac{1}{2 \cdot \pi \cdot \sqrt{0{,}955\,\frac{\mathrm{V\,s}}{\mathrm{A}} \cdot 15{,}9 \cdot 10^{-6}\,\frac{\mathrm{A\,s}}{\mathrm{V}}}} = 40{,}84\,\frac{1}{\mathrm{s}}$$

Komplexe Darstellung:

$$\underline{Y} = G - j(B_C - B_L)$$

$$= 10 \cdot 10^{-3}\,\mathrm{S} - j(5 \cdot 10^{-3} - 3{,}3 \cdot 10^{-3})\,\mathrm{S}$$

$$= 10{,}14 \cdot 10^{-3}\,\mathrm{S} \cdot e^{j9{,}6}$$

$$\underline{I} = \underline{U} \cdot \underline{Y}$$

$$= 200\,\mathrm{V} \cdot e^{j0} \cdot 10{,}14 \cdot 10^{-3}\,\mathrm{S} \cdot e^{j9{,}6} = 2{,}03\,\mathrm{A} \cdot e^{j9{,}6}$$

$$\underline{I_R} = \underline{U} \cdot G$$

$$= 200\,\mathrm{V} \cdot e^{j0} \cdot 10 \cdot 10^{-3}\,\mathrm{S} \cdot e^{j0} = 2\,\mathrm{A} \cdot e^{j0}$$

$$\underline{I_L} = \underline{U} \cdot B_L$$

$$= 200\,\mathrm{V} \cdot e^{j0} \cdot 3{,}3 \cdot 10^{-3}\,\mathrm{S} \cdot e^{-j90} = 0{,}66\,\mathrm{A\,V} \cdot e^{-j90}$$

$$\underline{I_C} = \underline{U} \cdot B_C$$

$$= 200\,\mathrm{V} \cdot e^{j0} \cdot 5 \cdot 10^{-3}\,\mathrm{S} \cdot e^{j90} = 1\,\mathrm{A} \cdot e^{j90}$$ ◀

30.3.3 Gemischte Schaltungen

Die Lösung von gemischten Schaltungen im Wechselstromkreis kann ebenfalls mit Hilfe der Zeigerdarstellung und der komplexen Rechnung erfolgen. Der Lösungsweg soll an einem Beispiel (Abb. 30.30) dargestellt werden.

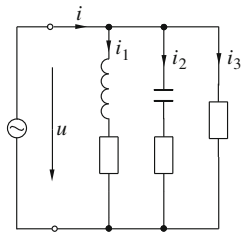


Abb. 30.30 Gemischte Schaltung im Wechselstromkreis

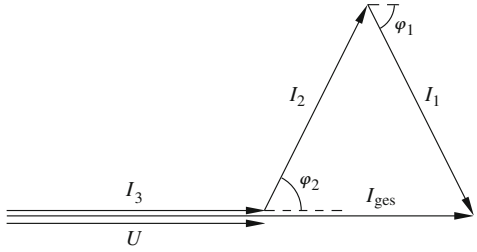


Abb. 30.31 Beispiel: Gemischte Schaltung im Wechselstromkreis in der Zeigerdarstellung

Beispiel

Lösung mit Hilfe der Zeigerdarstellung:

$$U = 200\,\text{V}; \quad f = 50\,\text{Hz}$$

Parallelzweig 1 mit Widerstand $R_1 = 100\,\Omega$ und Spule $X_L = 200\,\Omega$

$$Z_1 = \sqrt{R_1^2 + X_L^2} = \sqrt{100^2\,\Omega^2 + 200^2\,\Omega^2} = 223{,}6\,\Omega$$

$$I_1 = \frac{U}{Z_1} = \frac{200\,\text{V}}{223{,}6\,\Omega} = 0{,}89\,\text{A}$$

$$\varphi_1 = \arctan\frac{X_L}{R_1} = \arctan\frac{200}{100} = 63{,}4^\circ$$

Strom eilt der Spannung nach!

Parallelzweig 2 mit Widerstand $R_2 = 100\,\Omega$ und Kapazität $X_C = 200\,\Omega$

$$Z_2 = \sqrt{R_2^2 + X_C^2} = \sqrt{100^2\,\Omega^2 + 200^2\,\Omega^2} = 223{,}6\,\Omega$$

$$I_2 = \frac{U}{Z_2} = \frac{200\,\text{V}}{223{,}6\,\Omega} = 0{,}89\,\text{A}$$

$$\varphi_2 = \arctan\frac{X_C}{R_2} = \arctan\frac{200}{100} = 63{,}4^\circ$$

Strom eilt der Spannung vor!

Parallelzweig 3 mit Widerstand $R_3 = 200\,\Omega$

$$I_3 = \frac{U}{R_3} = \frac{200\,\text{V}}{200\,\Omega} = 1\,\text{A}$$

$$\varphi_3 = \arctan\frac{0}{R_3} = 0^\circ$$

Die rechnerische Addition der Ströme kann erfolgen, wenn die Zweigströme in ihre Real- und Blindstromanteile aufgelöst werden.

$$\begin{aligned}
I_{1R} &= I_1 \cdot \cos\varphi_1 = 0{,}89\,\text{A} \cdot \cos(-63{,}4^\circ)\\
&= 0{,}89\,\text{A} \cdot 0{,}448 = 0{,}399\,\text{A}\\
I_{1B} &= I_1 \cdot \sin\varphi_1 = 0{,}89\,\text{A} \cdot \sin(-63{,}4^\circ)\\
&= 0{,}89\,\text{A} \cdot (-0{,}894) = -0{,}796\,\text{A}\\
I_{2R} &= I_2 \cdot \cos\varphi_2 = 0{,}89\,\text{A} \cdot \cos(63{,}4^\circ)\\
&= 0{,}89\,\text{A} \cdot 0{,}448 = 0{,}399\,\text{A}\\
I_{2B} &= I_2 \cdot \sin\varphi_2 = 0{,}89\,\text{A} \cdot \sin(63{,}4^\circ)\\
&= 0{,}89\,\text{A} \cdot 0{,}894 = 0{,}796\,\text{A}
\end{aligned}$$

$$\begin{aligned}
I_{3R} &= I_3 \cdot \cos\varphi_3 = 0{,}89\,\text{A} \cdot \cos(0^\circ) = 1\,\text{A} \cdot 1 = 1\,\text{A}\\
I_{3B} &= I_3 \cdot \sin\varphi_3 = 0{,}89\,\text{A} \cdot \sin(0^\circ) = 1\,\text{A} \cdot 0 = 0\,\text{A}\\
I_{\text{ges}\,R} &= I_{1R} + I_{2R} + I_{3R}\\
&= 0{,}399\,\text{A} + 0{,}399\,\text{A} + 1\,\text{A} = 1{,}798\,\text{A}\\
I_{\text{ges}\,B} &= I_{1B} + I_{2B} + I_{3B}\\
&= -0{,}796\,\text{A} + 0{,}796\,\text{A} + 0\,\text{A} = 0\,\text{A}\\
I_{\text{ges}} &= \sqrt{I_{\text{ges}\,R}^2 + I_{\text{ges}\,B}^2} = \sqrt{1{,}798^2 + 0^2}\,\text{A} = 1{,}798\,\text{A}
\end{aligned}$$

Die zeichnerische Lösung erfolgt durch Addition der Stromzeiger nach Abb. 30.31

Lösung in der komplexen Darstellung:

In der komplexen Darstellung wird zuerst die *Gesamtimpedanz* Z_{ges} ermittelt.

$$\begin{aligned}
\underline{Z}_1 &= R_1 + jX_L = 100\,\Omega + j200\,\Omega = 223{,}6\,\Omega \cdot e^{j63{,}43}\\
\underline{Z}_2 &= R_2 - jX_C = 100\,\Omega - j200\,\Omega = 223{,}6\,\Omega \cdot e^{-j63{,}43}\\
\underline{Z}_3 &= R_3 + j0 = 200\,\Omega + j0 = 200\,\Omega \cdot e^{j0}
\end{aligned}$$

In der Parallelschaltung können Leitwerte addiert werden, die Impedanzen müssen also in Leitwerte umgerechnet werden.

$$\begin{aligned}
\underline{Y}_1 &= \frac{1}{\underline{Z}_1} = \frac{1}{223{,}6\,\Omega \cdot e^{j63{,}4}} = 4{,}47 \cdot 10^{-3}\,\text{S} \cdot e^{-j63{,}4}\\
&= (2{,}0 - j4{,}0) \cdot 10^{-3}\,\text{S}\\
\underline{Y}_2 &= \frac{1}{\underline{Z}_2} = \frac{1}{223{,}6\,\Omega \cdot e^{-j63{,}4}} = 4{,}47 \cdot 10^{-3}\,\text{S} \cdot e^{j63{,}4}\\
&= (2{,}0 + j4{,}0) \cdot 10^{-3}\,\text{S}\\
\underline{Y}_3 &= \frac{1}{\underline{Z}_3} = \frac{1}{200\,\Omega \cdot e^{j0}} = 5 \cdot 10^{-3}\,\text{S} \cdot e^{j0}\\
&= (5{,}0 - j0) \cdot 10^{-3}\,\text{S}\\
\underline{Y}_{\text{ges}} &= \underline{Y}_1 + \underline{Y}_2 + \underline{Y}_3\\
&= (2{,}0 + j4{,}0) \cdot 10^{-3}\,\text{S} + (2{,}0 - j4{,}0) \cdot 10^{-3}\,\text{S}\\
&\quad + (5{,}0 - j0) \cdot 10^{-3}\text{S}\\
&= (9{,}0 + j0) \cdot 10^{-3}\,\text{S} = 9{,}0 \cdot 10^{-3}\,\text{S} \cdot e^{j0}\\
\underline{I}_{\text{ges}} &= \underline{U} \cdot \underline{Y}_{\text{ges}}\\
&= (200\,\text{V} \cdot e^{j0}) \cdot (9 \cdot 10^{-3}\,\text{S} \cdot e^{j0}) = 1{,}8\,\text{A} \cdot e^{j0}
\end{aligned}$$

◀

30.4 Passive Filter

Passive Filter sind Schaltungen mit einem frequenzabhängigen Übertragungsverhältnis der Eingangsgröße zur Ausgangsgröße. Sie bestehen aus einer Zusammenschaltung von ohmschen Widerständen R, Kapazitäten C und Induktivitäten L zu komplexen Netzwerken.

Ein Filter ist ein Netzwerk, das aus einem Frequenzgemisch einen bestimmten Frequenzbereich hervorhebt und einen anderen unterdrückt.

Im Durchlassbereich der Filter werden die zu übertragenden Größen kaum, im Sperrbereich dagegen stark gedämpft. Der Übergang von Sperr- in den Durchlassbereich wird durch die *Grenzfrequenz* f_G festgelegt. Nach der Lage der Sperr- bzw. Durchlassbereiche unterscheidet man zwischen *Tief-*, *Hoch-* und *Bandpassschaltungen*.

30.4.1 Hochpassschaltung mit RC- und RL-Glied

Das in Abb. 30.32 dargestellte RC-Glied stellt einen Hochpass dar.

Für die Filter sind Grenzfrequenz und Phasengang von besonderem Interesse. Die Grenzfrequenz ist die Frequenz, bei der das Verhältnis Ausgangsspannung U_A zu Eingangsspannung U_E den Wert $\frac{1}{\sqrt{2}}$ beträgt.

Das Verhältnis von Ausgangsspannung zu Eingangsspannung wird *Amplitudengang* genannt.

Amplitudengang (RC)

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{R}{\sqrt{R^2 + X_C^2}} = \frac{1}{\sqrt{1 + \frac{1}{(\omega \cdot C \cdot R)^2}}} \tag{30.115}$$

Der Amplitudengang hat für den Fall $R = \frac{1}{\omega \cdot C}$ den Wert

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{1}{\sqrt{1 + \frac{1}{(\omega \cdot C \cdot \frac{1}{\omega \cdot C})^2}}} = \frac{1}{\sqrt{2}} \tag{30.116}$$

Aus (30.116) ergibt sich die Grenzfrequenz:

$$\textit{Grenzfrequenz (RC)} \quad f_g = \frac{1}{2 \cdot \pi \cdot R \cdot C} \tag{30.117}$$

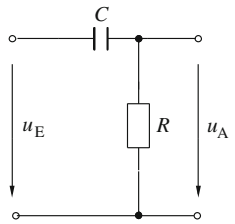


Abb. 30.32 Hochpassschaltung mit Kapazität und ohmschem Widerstand

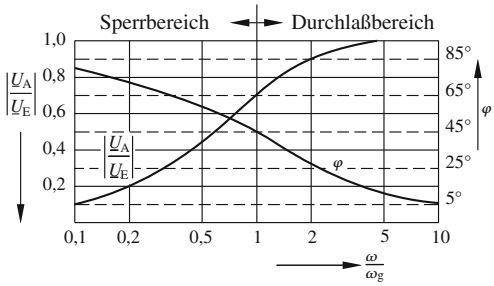


Abb. 30.33 Amplituden- und Phasengang der Hochpassschaltung mit Kapazität und Ohm'schem Widerstand

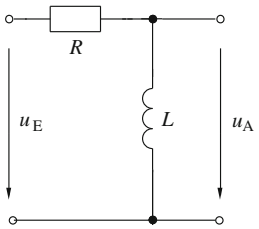


Abb. 30.34 Hochpassschaltung mit Induktivität und ohmschem Widerstand

Der *Phasengang*, die Abhängigkeit des Winkels φ zwischen Ausgangs- und Eingangsspannung, ist stark von der Frequenz abhängig.

$$\textit{Phasengang} \quad \arctan \varphi = f(\omega) \tag{30.118}$$

Der Winkel φ nimmt für verschiedene Frequenzen unterschiedliche Werte an. Häufig gebrauchte Grenzwerte:

$$\omega = 0 \quad \varphi = \arctan \frac{1}{\omega \cdot C \cdot R} = \arctan(\infty) = 90°$$

$$\omega = \frac{1}{R \cdot C}$$

$$\varphi = \arctan \frac{1}{\omega \cdot C \cdot R}$$

$$= \arctan \frac{C \cdot R}{C \cdot R} = \arctan(1) = 45° \quad \text{(Grenzfrequenz)}$$

$$\omega = \infty \quad \varphi = \arctan \frac{1}{\omega \cdot C \cdot R} = \arctan(0) = 0°$$

Im Abb. 30.33 sind Amplituden- und Phasengang einer Hochpassschaltung dargestellt.

Wird anstatt des RC-Gliedes ein RL-Glied (Abb. 30.34) als Hochpass eingesetzt, ergeben sich (30.119) und (30.120):

Amplitudengang (RL)

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{X_L}{\sqrt{R^2 + X_L^2}} = \frac{1}{\sqrt{1 + (\frac{R}{\omega \cdot L})^2}} \tag{30.119}$$

Grenzfrequenz (RL)

$$f_g = \frac{R}{2 \cdot \pi \cdot L} \tag{30.120}$$

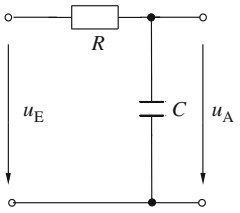


Abb. 30.35 Tiefpassschaltung mit Kapazität und ohmschem Widerstand

30.4.2 Tiefpassschaltung mit RC- und RL-Glied

Das in Abb. 30.35 dargestellte RC-Glied stellt einen Tiefpass dar.

Für einen RC-Tiefpass ergeben sich die charakterisierenden Gleichungen (30.121) bis (30.123):

Amplitudengang (RC)

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{X_C}{\sqrt{R^2 + X_C^2}} = \frac{1}{\sqrt{1 + (\omega \cdot C \cdot R)^2}} \tag{30.121}$$

Der Amplitudengang hat für den Fall $R = \frac{1}{\omega \cdot C}$ den Wert

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{1}{\sqrt{1 + \frac{1}{(\omega \cdot C \cdot \frac{1}{\omega \cdot C})^2}}} = \frac{1}{\sqrt{2}} \tag{30.122}$$

Daraus folgt die Grenzfrequenz

$$\textit{Grenzfrequenz (RC)} \quad f_g = \frac{1}{2 \cdot \pi \cdot R \cdot C} \tag{30.123}$$

Häufig gebrauchte Grenzwerte:

$$\omega = 0 \quad \varphi = \arctan \frac{1}{\omega \cdot C \cdot R} = \arctan(0) = 0°$$

$$\omega = \frac{1}{R \cdot C}$$

$$\varphi = \arctan \omega \cdot C \cdot R$$

$$= \arctan \frac{C \cdot R}{C \cdot R} = \arctan(1) = 45° \quad \text{(Grenzfrequenz)}$$

$$\omega = \infty \quad \varphi = \arctan \omega \cdot C \cdot R = \arctan(\infty) = 90°$$

Im Abb. 30.36 sind Amplituden- und Phasengang einer Tiefpassschaltung dargestellt.

Wird anstatt des RC-Gliedes ein RL-Glied (Abb. 30.37) als Tiefpass eingesetzt, ergeben sich (30.124) und (30.125):

Amplitudengang (RL)

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{X_L}{\sqrt{R^2 + X_L^2}} = \frac{1}{\sqrt{1 + (\frac{R}{\omega \cdot L})^2}} \tag{30.124}$$

Grenzfrequenz (RL)

$$f_g = \frac{R}{2 \cdot \pi \cdot L} \tag{30.125}$$

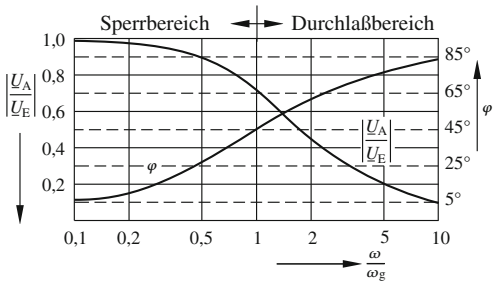


Abb. 30.36 Amplituden- und Phasengang der Tiefpassschaltung mit Kapazität und ohmschem Widerstand

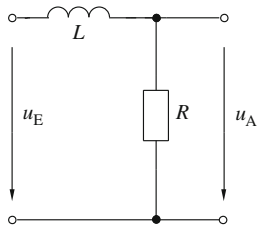


Abb. 30.37 Tiefpassschaltung mit Induktivität und ohmschem Widerstand

30.4.3 Bandpassschaltung

Bei einem Bandpass sind zwei verschiedene Filter (Hoch- und Tiefpass) hintereinander geschaltet (Abb. 30.38).

Ein Bandpass überträgt zwischen zwei Grenzfrequenzen eine Eingangsspannung an einen Ausgang mit einem Verhältnis

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} \geq \frac{1}{\sqrt{2}} \cdot \frac{|\underline{U}_A|_{\max}}{|\underline{U}_E|}$$

Die Differenz der beiden Grenzfrequenzen wird als *Bandbreite B* bezeichnet.

$$\textit{Bandbreite} \quad B = f_{go} - f_{gu} \tag{30.126}$$

Für den Bandpass nach Abb. 30.38 folgt für die Ausgangsspannung:

Ausgangsspannung Bandpass

$$\underline{U}_A = \frac{\frac{1}{\frac{1}{R} + j\omega C}}{\frac{1}{\frac{1}{R} + j\omega C} + R + \frac{1}{j\omega C}} \cdot \underline{U}_E \tag{30.127}$$

$$\underline{U}_A = \frac{j\omega RC}{(j\omega RC + 1)^2 + j\omega RC} \cdot \underline{U}_E \tag{30.128}$$

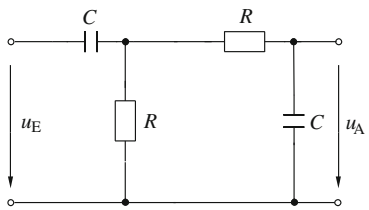


Abb. 30.38 Bandpassschaltung mit Kapazität und ohmschem Widerstand

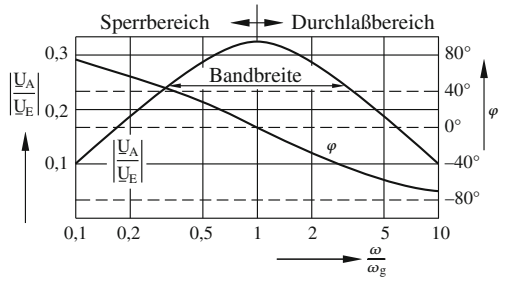


Abb. 30.39 Amplituden- und Phasengang der Bandpassschaltung mit Kapazität und ohmschem Widerstand

Amplitudengang (BP)

$$\frac{|\underline{U}_A|}{|\underline{U}_E|} = \frac{1}{\sqrt{9 + \frac{[(\omega RC)^2 - 1]^2}{(\omega RC)^2}}} \tag{30.129}$$

Die Ausgangsspannung für den Bandpass wird maximal, wenn der Faktor $\omega RC = 1$ wird.

maximale Ausgangsspannung

$$|\underline{U}_A|_{\max} = \frac{1}{3} \cdot |\underline{U}_E| \tag{30.130}$$

Die Grenzfrequenzen stellen sich bei einem Bandpass für folgende Frequenz ein:

untere Grenzfrequenz

$$f_{\text{gu}} = 0{,}303 \cdot \frac{1}{2 \cdot \pi \cdot R \cdot C} \tag{30.131}$$

obere Grenzfrequenz

$$f_{\text{go}} = 3{,}303 \cdot \frac{1}{2 \cdot \pi \cdot R \cdot C} \tag{30.132}$$

Häufig gebrauchte Grenzwerte:

$$\omega = 0 \qquad \varphi = \arctan \frac{1 - (\omega \cdot R \cdot C)^2}{3 \cdot \omega \cdot C \cdot R} = \arctan(\infty) = 90°$$

$$\omega = \frac{1}{R \cdot C} \qquad \varphi = \arctan \frac{1 - (\omega \cdot R \cdot C)^2}{3 \cdot \omega \cdot C \cdot R} = \arctan \frac{1 - 1}{3} = \arctan(0) = 0°$$

$$\omega = \infty \qquad \varphi = \arctan \frac{1 - (\omega \cdot R \cdot C)^2}{3 \cdot \omega \cdot C \cdot R} = \frac{-\infty^2}{3 \cdot \infty} = \arctan(-\infty) = -90°$$

Im Abb. 30.39 sind Amplituden- und Phasengang einer Bandpassschaltung dargestellt.

30.5 Schwingkreise

Anstelle der Bandpassschaltungen können bestimmte Frequenzbereiche auch durch eine Zusammenschaltung von ohmschem Widerstand, Induktivität und Kapazität hervorgehoben oder bedämpft werden. Diese Schaltungen werden als Schwingkreise bezeichnet.

30.5.1 Reihenresonanz

Bei der Reihenschaltung stellt sich für die Resonanzfrequenz f_r der maximale Strom $I_r = \frac{U}{R}$ ein, da sich die Blindwiderstände gegenseitig kompensieren.

Resonanzfrequenz

$$f_r = \frac{\omega_r}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}} \tag{30.133}$$

Die Reihenschaltung verhält sich nach außen wie ein rein ohmscher Widerstand mit dem Phasenwinkel $\varphi = 0°$. Innerhalb der Reihenschaltung erreichen die Spannungen an Induktivität und Kapazität wesentlich höhere Werte als die Klemmenspannung.

Resonanzspannung an der Induktivität

$$U_{rL} = I_r \cdot \omega_r \cdot L \tag{30.134}$$

Resonanzspannung an der Kapazität

$$U_{rC} = \frac{I_r}{\omega_r \cdot C} \tag{30.135}$$

Die beiden Resonanzspannungen sind gleich groß, haben ihr Maximum jedoch um 180° verschoben.

Die im Kreis auftretenden Verluste werden über den *Dämpfungsfaktor d* angegeben. Der Dämpfungsfaktor beinhaltet die an der Spule, am Kondensator und am Widerstand auftretenden ohmschen Verluste im Verhältnis zum Blindwiderstand.

$$\textit{Dämpfungsfaktor} \quad d = \frac{R}{\omega_r \cdot L} \tag{30.136}$$

Der Kehrwert des Dämpfungsfaktors wird als *Güte Q* des Schwingkreises bezeichnet.

$$\textit{Kreisgüte} \quad Q = \frac{1}{d} = \frac{\omega_r \cdot L}{R} = \frac{1}{R} \cdot \sqrt{\frac{L}{C}} \tag{30.137}$$

Die Güte des Reihenschwingkreises wird größer, wenn die Induktivität größer und die Kapazität kleiner wird. Mit Hilfe der Güte kann die Bandbreite des Reihenschwingkreises bestimmt werden.

$$\textit{Bandbreite} \quad \Delta f = f_{\text{go}} - f_{\text{gu}} = \frac{f_r}{Q} = f_r \cdot d \tag{30.138}$$

Bei den Grenzfrequenzen erreicht die Gesamtimpedanz Z des Kreises den $\sqrt{2}$-fachen Wert der Resonanzimpedanz.

$$Z_{\text{go}} = Z_{\text{gu}} = \sqrt{2} \cdot Z_r \tag{30.139}$$

30.5.2 Parallelresonanz

Ein Parallelschwingkreis ist die Zusammenschaltung von ohmschem Widerstand, Induktivität und Kapazität. Bei der Parallelschaltung stellt sich für die Resonanzfrequenz f_r der minimale Strom $I_r = \frac{U}{R}$ ein, da sich die Blindwiderstände gegenseitig kompensieren.

Resonanzfrequenz

$$f_r = \frac{\omega_r}{2 \cdot \pi} = \frac{1}{2 \cdot \pi \cdot \sqrt{L \cdot C}} \tag{30.140}$$

Die Parallelschaltung verhält sich nach außen wie ein rein ohmscher Widerstand mit dem Phasenwinkel $j = 0°$. Innerhalb der Parallelschaltung erreichen die Ströme durch die Induktivität und die Kapazität wesentlich höhere Werte als der Strom in der Zuleitung.

Resonanzstrom durch die Induktivität

$$I_{rL} = \frac{U}{\omega_r \cdot L} \tag{30.141}$$

Resonanzstrom durch die Kapazität

$$I_{rC} = U \cdot \omega_r \cdot C \tag{30.142}$$

Die beiden Resonanzströme sind gleich groß, haben ihr Maximum jedoch um 180° verschoben.

Die im Kreis auftretenden Verluste werden wie bei der Reihenresonanz über den *Dämpfungsfaktor d* angegeben.

$$\textit{Dämpfungsfaktor} \quad d = \frac{\omega_r \cdot L}{R} \tag{30.143}$$

$$\textit{Kreisgüte} \quad Q = \frac{1}{d} = \frac{R}{\omega_r \cdot L} = R \cdot \sqrt{\frac{C}{L}} \tag{30.144}$$

Die Güte des Reihenschwingkreises wird größer, wenn die Induktivität kleiner und die Kapazität größer wird.

$$\textit{Bandbreite} \quad \Delta f = f_{go} - f_{gu} = \frac{f_r}{Q} = f_r \cdot d \tag{30.145}$$

Bei den Grenzfrequenzen erreicht die Gesamtimpedanz Z des Kreises den $\frac{1}{\sqrt{2}}$-fachen Wert der Resonanzimpedanz.

$$Z_{go} = Z_{gu} = \frac{1}{\sqrt{2}} \cdot Z_r \tag{30.146}$$

30.6 Leistung und Arbeit im Wechselstromkreis

30.6.1 Leistung und Arbeit bei Phasengleichheit von Spannung und Strom

Die Berechnung der Leistung im Gleichstromkreis erfolgt nach (30.147).

$$\textit{Leistung im Gleichstromkreis} \quad P = U \cdot I \tag{30.147}$$

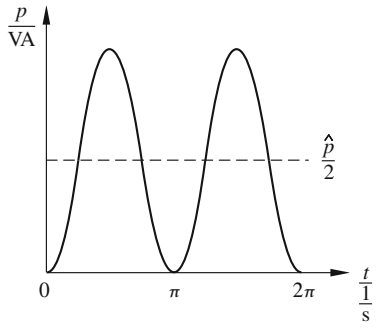


Abb. 30.40 Liniendiagramm der Leistung

Die Leistung im Wechselstromkreis kann analog zur Berechnung der Leistung im Gleichstromkreis für Werte erfolgen, die zeitlich zusammenfallen. Die Augenblicks- (Momentan-) Leistung ergibt sich zu:

$$\textit{Momentanleistung} \quad p = u \cdot i \tag{30.148}$$

Da am Wirkwiderstand Spannung und Strom gleiche Phasenlage haben, folgt

$$p = \hat{u} \cdot \hat{i} \cdot \sin^2 \omega \cdot t. \tag{30.149}$$

Das Liniendiagramm (Abb. 30.40) zeigt, dass die Leistung mit doppelter Frequenz oberhalb der Nullachse um den Wert $p_m = \frac{\hat{p}}{2}$ schwingt.

Der Mittelwert der Leistung berechnet sich zu:

Mittelwert der Leistung

$$p_m = \frac{\hat{u} \cdot \hat{i}}{2 \cdot \pi} \int_0^{2\pi} \sin^2 \omega t d(t) = \frac{\hat{u} \cdot \hat{i}}{2} = \frac{\hat{u} \cdot \hat{i}}{\sqrt{2} \cdot \sqrt{2}} = U \cdot I \tag{30.150}$$

Aus (30.150) ergibt sich, dass die Wirkleistung eines Wechselstroms in einem Stromkreis, der nur mit einem Wirkwiderstand belastet ist, gleich dem Produkt aus den Effektivwerten von Spannung und Stromstärke ist

$$\textit{Wirkleistung} \quad P = U \cdot I$$

P	U	I
$VA = W$	V	A

(30.151)

Die Stromarbeit ist das Produkt der Wirkleistung P und der Zeit t.

$$\textit{Stromarbeit} \quad W = P \cdot t$$

W	P	t
Wh	W	h

(30.152)

30.6.2 Leistung und Arbeit bei Phasenverschiebung von Spannung und Strom

Sind die Spannungs- und Stromwellen phasenverschoben (Abb. 30.41), wird die Augenblicksleistung in den Bereichen

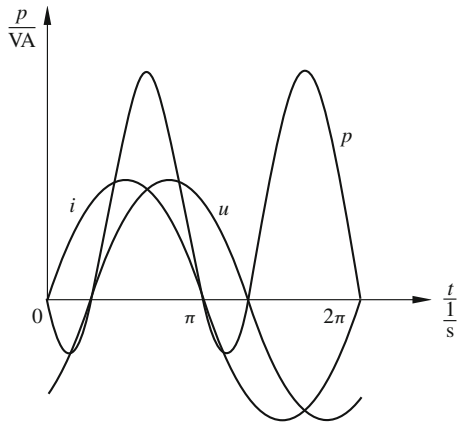


Abb. 30.41 Liniendiagramm der Leistung bei einer Phasenverschiebung zwischen Strom und Spannung

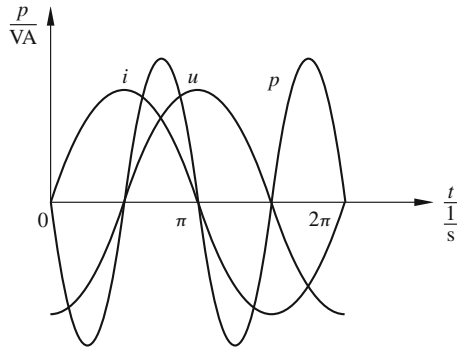


Abb. 30.42 Liniendiagramm der Leistung bei einer Phasenverschiebung von 90° zwischen Strom und Spannung

negativ, in denen Spannung und Strom entgegengesetzte Vorzeichen haben. Der Mittelwert der Leistung verschiebt sich in Richtung Zeitachse. Die Wirkleistung wird also durch die Phasenverschiebung kleiner und erreicht ein Minimum, wenn die Phasenverschiebung zwischen Spannung und Strom 90° beträgt (Abb. 30.42).

Der Mittelwert der Leistung liegt im Abb. 30.42 auf der Zeitachse und hat den Wert Null. Die Leistungskurve besteht aus positiven und negativen Halbwellen gleicher Größe, die Wirkleistung hat den Wert $P = 0$.

Beträgt die Phasenverschiebung zwischen Strom und Spannung 90°, hat der Stromkreis eine rein induktive oder kapazitive Last. Der Effektivwert der Leistungskurve kann wiederum durch die Effektivwerte von Strom und Spannung bestimmt werden. Es handelt sich bei dieser Leistung jedoch um keine Wirkleistung, sondern um *Blindleistung* Q.

$$\textit{Blindleistung}\ (\varphi = 90°) \quad Q = U \cdot I \tag{30.153}$$

Von *Wirkleistung* spricht man, wenn Strom und Spannung keine Phasenverschiebung ($\varphi = 0°$) aufweisen, die Zeiger beider Größen also gleiche Richtung haben.

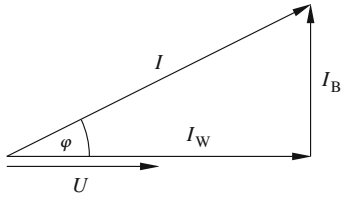


Abb. 30.43 Zeigerdiagramm von Strom und Spannung bei beliebigem Winkel

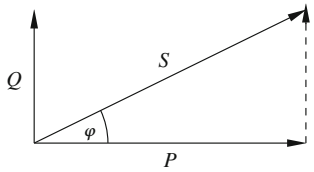


Abb. 30.44 Zeigerdiagramm der Leistung bei beliebigem Winkel (Operatoren)

Von *Blindleistung* spricht man, wenn Strom und Spannung eine Phasenverschiebung von 90° haben, die Zeiger beider Größen einen rechten Winkel bilden.

Liegt der Phasenwinkel zwischen 0° und 90°, kann die Wirkleistung ermittelt werden, indem eine Zeigergröße in den Winkel 0° gelegt und die andere Zeigergröße in ihre Komponenten zerlegt wird, die in Richtung 0° und 90° zeigen (Abb. 30.43).

Im Abb. 30.43 wird die Spannung nicht verschoben und der Strom in den *Wirkstrom* I_W und den *Blindstrom* I_B aufgeteilt.

$$\textit{Wirkstrom} \qquad I_W = I \cdot \cos\varphi \tag{30.154}$$

$$\textit{Blindstrom} \qquad I_B = I \cdot \sin\varphi \tag{30.155}$$

Für die entsprechenden Leistungen folgt:

$$\textit{Wirkleistung} \qquad P = U \cdot I \cdot \cos\varphi \tag{30.156}$$

$$\textit{Blindleistung} \qquad Q = U \cdot I \cdot \sin\varphi \tag{30.157}$$

Wirkleistung und Blindleistung haben ihre Maxima im Liniendiagramm um 90° verschoben, das heißt, man kann die beiden Leistungsanteile als Zeiger darstellen, die einen Winkel von 90° einschließen. Zwischen resultierendem Zeiger und Wirkleistung ist der Winkel φ (Abb. 30.44).

Der resultierende Zeiger stellt die *Scheinleistung* S dar.

$$\textit{Scheinleistung} \qquad S = \sqrt{P^2 + Q^2} \tag{30.158}$$

$$P = S \cdot \cos\varphi \tag{30.159}$$

$$Q = S \cdot \sin\varphi \tag{30.160}$$

$$S = U \cdot I \tag{30.161}$$

Die bei effektiver Spannung U und Stromstärke I abgegebene nutzbare Leistung beträgt also nur den durch den Faktor $\cos\varphi$ gegebenen Teil der Scheinleistung. Der andere Teil pendelt als Blindleistung zwischen der Spannungsquelle und dem Verbraucher hin und her.

Zur Unterscheidung der verschiedenen Leistungsarten, hat man unterschiedliche Maßbezeichnungen eingeführt, und zwar misst man die Wirkleistung in $[P] = \mathrm{W}$, die Scheinleistung in $[S] = \mathrm{V\,A}$ und die Blindleistung in $[Q] = \mathrm{var}$.

30.6.3 Leistung in komplexer Schreibweise

Die Leistung ist definiert als $S = U \cdot I$. Hieraus ergibt sich für die komplexe Schreibweise in der Exponentialform:

komplexe Leistung

$$\underline{S} = \underline{U} \cdot \underline{I}^* = U \cdot I \cdot e^{j\varphi_U} \cdot e^{-j\varphi_I} = U \cdot I \cdot e^{j(\varphi_U - \varphi_I)} \qquad (30.162)$$

Die komplexe Leistung ist das Produkt aus der komplexen Spannung $\underline{U}$ und dem konjugiert komplexen Strom $\underline{I}^*$.

Da $\varphi = \varphi_U - \varphi_I$ die Verschiebung zwischen Spannung $\underline{U}$ und Strom $\underline{I}$ darstellt, gilt:

$$\underline{U} \cdot \underline{I}^* = U \cdot I \cdot (\cos\varphi + j \sin\varphi) \qquad (30.163)$$

Aus der vorhergehenden Definition ergeben sich (30.164) bis (30.166).

Wirkleistung P	$\mathrm{Re}(\underline{U} \cdot \underline{I}^*) = U \cdot I \cdot \cos\varphi$	(30.164)
Blindleistung Q	$\mathrm{Im}(\underline{U} \cdot \underline{I}^*) = U \cdot I \cdot \sin\varphi$	(30.165)
Scheinleistung S	$\lvert\underline{U} \cdot \underline{I}^*\rvert = U \cdot I$	(30.166)

Beispiel

Ein induktiver Verbraucher nimmt bei $U_n = 220\,\mathrm{V}$ einen Strom $I = 2\,\mathrm{A}$ auf. Der Winkel zwischen Strom und Spannung beträgt $\varphi = 40°$.

Welchen Wert haben a) die Scheinleistung, b) die Wirkleistung und c) die Blindleistung?

Lösung:

a) $S = U \cdot I = 220\,\mathrm{V} \cdot 2\,\mathrm{A} = 440\,\mathrm{V\,A}$
b) $P = S \cdot \cos\varphi = 440\,\mathrm{V\,A} \cdot 0{,}766 = 337\,\mathrm{W}$
c) $Q = S \cdot \sin\varphi = 440\,\mathrm{V\,A} \cdot 0{,}643 = 283\,\mathrm{var}$ ◀

30.6.4 Leistungsfaktor

Die von einem Verbraucher aufgenommene Wirkleistung ist über den Faktor $\cos\varphi$ mit der Scheinleistung verknüpft. Der Faktor $\cos\varphi$ wird daher als *Leistungsfaktor* bezeichnet.

$$\textit{Leistungsfaktor} \quad \cos\varphi = \frac{P}{S} \qquad (30.167)$$

Der Leistungsfaktor gibt an, wieviel Prozent der *Scheinleistung S* als *Wirkleistung P* im Verbraucher umgesetzt werden.

In der Praxis soll der Leistungsfaktor möglichst nahe bei Eins liegen, damit die Zuleitungen nicht unnötig mit Blindstrom belastet werden, da dieser Blindstrom auf der Zuleitung Verluste erzeugt.

Drehstrom

31

Horst Steffen



31.1 Erzeugung von mehrphasigem Wechselstrom 384
31.2 Phasenverkettung 384
31.3 Leistung des Dreiphasenstroms 386
31.4 Das unsymmetrische Dreiphasensystem 387
Literatur 389

Teil IV

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_31

31.1 Erzeugung von mehrphasigem Wechselstrom

Dreht man in einem homogenen Magnetfeld drei in einem Winkel von 120° versetzte gleiche Spulen (Abb. 31.1) mit der Winkelgeschwindigkeit ω, so werden folgende Spannungen in den Spulen induziert:

$$\textit{Spannung Spule 1} \quad u_1 = \hat{u} \cdot \sin(\omega \cdot t) \tag{31.1}$$

$$\textit{Spannung Spule 2} \quad u_2 = \hat{u} \cdot \sin(\omega \cdot t + 120°) \tag{31.2}$$

$$\textit{Spannung Spule 3} \quad u_3 = \hat{u} \cdot \sin(\omega \cdot t + 240°) \tag{31.3}$$

Das Zeigerdiagramm der erzeugten Spannung ist in Abb. 31.2 dargestellt.

Schließt man an jede der Spulen einen ohmschen Verbraucher mit gleichen Werten an, so treiben die Spannungen einen Strom durch die Verbraucher, der gleiche Amplitude hat, aber jeweils um den Winkel von 120° versetzt ist. Solche voneinander unabhängigen Dreiphasensysteme werden in der Praxis nicht verwendet, da sie einen hohen Aufwand an Leitungsmaterial erfordern und keinen Vorteil gegenüber Wechselstromsystemen bieten.

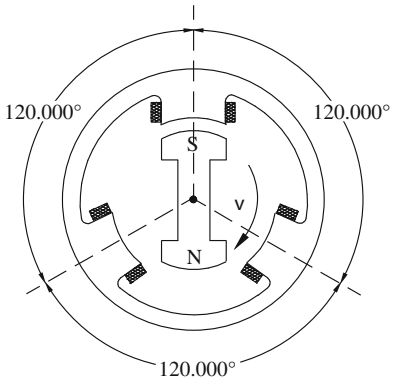


Abb. 31.1 Erzeugen eines Dreiphasensystems

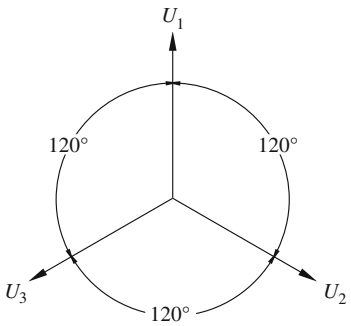


Abb. 31.2 Zeigerdiagramm der Spannungen gemäß Abb. 31.1

31.2 Phasenverkettung

Betrachtet man die im Dreiphasensystem erzeugten Spannungen in einem beliebigen Zeitpunkt, so ergeben sich beispielsweise folgende Werte:

Zeitpunkt $\varphi = 0°$

$$u_1 = 0; \quad u_2 = 0{,}866 \cdot \hat{u}; \quad u_3 = -0{,}866 \cdot \hat{u}$$

Zeitpunkt $\varphi = 90°$

$$u_1 = \hat{u}; \quad u_2 = -0{,}5 \cdot \hat{u}; \quad u_3 = -0{,}5 \cdot \hat{u}$$

Zeitpunkt $\varphi = 240°$

$$u_1 = -0{,}866 \cdot \hat{u}; \quad u_2 = 0; \quad u_3 = 0{,}866 \cdot \hat{u}$$

Bei der Bildung der Summen der Augenblickswerte erkennt man, dass zu jedem beliebigem Zeitpunkt die Gesamtspannung der drei Stromkreise den Wert Null ergibt. Da die induzierten Spannungen stets in ihrer Addition den Wert Null ergeben, können die Spulen an einem Spulenende zusammengefasst werden, ohne dass ein Kurzschluss entsteht (Abb. 31.3). Werden die Spulen derart zusammengefasst, spricht man von einer *Verkettung*. Man unterscheidet zwei Arten der Verkettung: die *Sternschaltung* und die *Dreieckschaltung*.

31.2.1 Sternschaltung

Eine Sternschaltung erhält man, wenn die Enden der drei Generatorspulen in einem Punkt, dem *Sternpunkt*, zusammengeschaltet werden (Abb. 31.4). Die in den Spulen induzierte Spannung wird *Strangspannung* U_{Str} genannt; die zwischen den Abgangsklemmen des Generators gemessene Spannung ist die *Leiterspannung* U_L.

Ein Zusammenhang zwischen den Strangwerten und den Leiterwerten des symmetrischen Drehstromsystems in Sternschaltung ist im Abb. 31.5 zu erkennen.

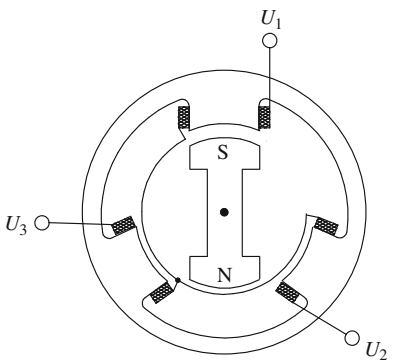


Abb. 31.3 Verkettung eines Dreiphasensystems

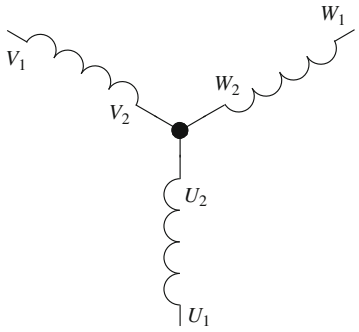


Abb. 31.4 Spulenanordnung bei einer Verkettung im Stern

Nach Abb. 31.5 ergeben sich folgende mathematischen Zusammenhänge.

Strangspannung

$$U_{\mathrm{Str}} = U_{1N} = U_{2N} = U_{3N} \tag{31.4}$$

Leiterspannung 1–2:

$$\begin{aligned}\underline{U}_{L1L2} &= \underline{U}_{1N} - \underline{U}_{2N} = U_{1N} \cdot e^{-j90} - U_{2N} \cdot e^{j150} \\ &= U_{\mathrm{Str}} \cdot (0{,}866 - j1{,}5) = 1{,}732 \cdot e^{-j60} \cdot U_{\mathrm{Str}}\end{aligned} \tag{31.5}$$

Leiterspannung 2–3:

$$\begin{aligned}\underline{U}_{L2L3} &= \underline{U}_{2N} - \underline{U}_{3N} = U_{2N} \cdot e^{j150} - U_{3N} \cdot e^{j30} \\ &= U_{\mathrm{Str}} \cdot (1{,}732 - j0) = 1{,}732 \cdot e^{-j180} \cdot U_{\mathrm{Str}}\end{aligned} \tag{31.6}$$

Leiterspannung 3–1:

$$\begin{aligned}\underline{U}_{L3L1} &= \underline{U}_{3N} - \underline{U}_{1N} = U_{3N} \cdot e^{j30} - U_{1N} \cdot e^{-j90} \\ &= U_{\mathrm{Str}} \cdot (0{,}866 + j1{,}5) = 1{,}732 \cdot e^{+j60} \cdot U_{\mathrm{Str}}\end{aligned} \tag{31.7}$$

Strangstrom

$$I_{\mathrm{Str}} = I_{1N} = I_{2N} = I_{3N} = I_{L1} = I_{L2} = I_{L3} \tag{31.8}$$

Die Leiterströme entsprechen in der symmetrischen Sternschaltung den Strangströmen, die Leiterspannungen einer symmetrischen Sternschaltung betragen das $1{,}732 = \sqrt{3}$-fache der Strangspannungen.

Für einen symmetrischen Drehstromverbraucher in Sternschaltung gelten gleiche Bedingungen.

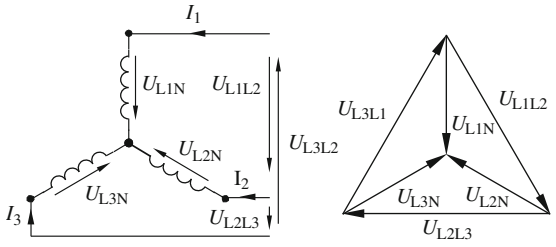


Abb. 31.5 Sternschaltung mit Zeigerbild der Spannungen

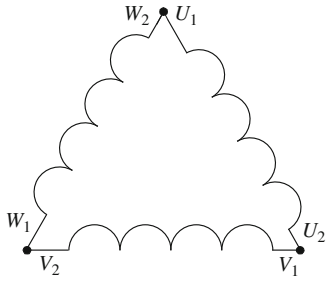


Abb. 31.6 Spulenanordnung bei einer Verkettung im Dreieck

31.2.2 Dreieckschaltung

Eine Dreiecksschaltung erhält man, wenn die drei Generatorspulen in Reihe geschaltet werden (Abb. 31.6). Die in den Spulen induzierte Spannung ist die zwischen den Abgangsklemmen des Generators gemessene Spannung.

Ein Zusammenhang zwischen den Strangwerten und den Leiterwerten des symmetrischen Drehstromsystems in Dreieckschaltung ist im Abb. 31.7 dargestellt.

Nach Abb. 31.7 ergeben sich folgende mathematischen Zusammenhänge:

Strangstrom

$$I_{\mathrm{Str}} = I_{12} = I_{23} = I_{31} \tag{31.9}$$

Leiterstrom 1:

$$\begin{aligned}\underline{I}_1 &= \underline{I}_{12} - \underline{I}_{31} = I_{12} \cdot e^{-j60} - I_{31} \cdot e^{j60} \\ &= I_{\mathrm{Str}} \cdot (0 - j1{,}732) = 1{,}732 \cdot e^{-j90} \cdot I_{\mathrm{Str}}\end{aligned} \tag{31.10}$$

Leiterstrom 2:

$$\begin{aligned}\underline{I}_2 &= \underline{I}_{23} - \underline{I}_{12} = I_{23} \cdot e^{-j90} - I_{12} \cdot e^{-j60} \\ &= I_{\mathrm{Str}} \cdot (-1{,}5 + j0{,}866) = 1{,}732 \cdot e^{j150} \cdot I_{\mathrm{Str}}\end{aligned} \tag{31.11}$$

Leiterstrom 3:

$$\begin{aligned}\underline{I}_3 &= \underline{I}_{31} - \underline{I}_{23} = I_{31} \cdot e^{j60} - I_{23} \cdot e^{j180} \\ &= I_{\mathrm{Str}} \cdot (1{,}5 + j0{,}866) = 1{,}732 \cdot e^{j30} \cdot I_{\mathrm{Str}}\end{aligned} \tag{31.12}$$

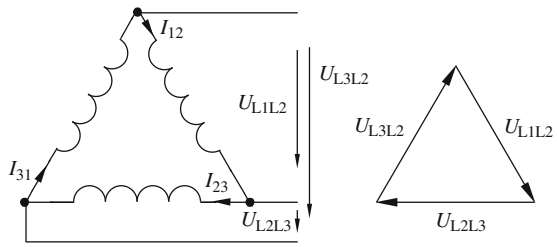


Abb. 31.7 Dreieckschaltung mit Zeigerbild der Spannungen

Teil IV

Strangspannung

$$U_{\mathrm{Str}} = U_{12} = U_{23} = U_{31} \tag{31.13}$$

Die Leiterspannungen entsprechen in der symmetrischen Dreieckschaltung den Strangspannungen; die Leiterströme einer symmetrischen Dreieckschaltung betragen das 1,732 $= \sqrt{3}$-fache der Strangströme.

Für einen symmetrischen Drehstromverbraucher in Dreieckschaltung gelten gleiche Bedingungen.

31.3 Leistung des Dreiphasenstroms

In symmetrisch belasteten Dreiphasensystemen sind die Werte für U, I und φ für die drei Stromkreise (Phasen) gleich. Die Leistung in jedem Strang ist somit identisch und kann nach (30.156) bis (30.161) berechnet werden.

Strangscheinleistung	$S_{\mathrm{Str}} = U_{\mathrm{Str}} \cdot I_{\mathrm{Str}}$	(31.14)
Strangwirkleistung	$P_{\mathrm{Str}} = U_{\mathrm{Str}} \cdot I_{\mathrm{Str}} \cdot \cos\varphi$	(31.15)
Strangblindleistung	$Q_{\mathrm{Str}} = U_{\mathrm{Str}} \cdot I_{\mathrm{Str}} \cdot \sin\varphi$	(31.16)

Die Gesamtleistung des Dreiphasensystems ist die Summe der Strangleistungen.

Scheinleistung	$S = 3 \cdot S_{\mathrm{Str}}$	(31.17)
Wirkleistung	$P = 3 \cdot P_{\mathrm{Str}}$	(31.18)
Blindleistung	$Q = 3 \cdot Q_{\mathrm{Str}}$	(31.19)

Berechnung der Leistung über die Leiterspannung und den Leiterstrom:

Sternschaltung:

$$S = 3 \cdot U_{\mathrm{Str}} \cdot I_{\mathrm{Str}} = 3 \cdot \frac{U_L}{\sqrt{3}} \cdot I_L = \sqrt{3} \cdot U_L \cdot I_L \tag{31.20}$$

Dreieckschaltung:

$$S = 3 \cdot U_{\mathrm{Str}} \cdot I_{\mathrm{Str}} = 3 \cdot U_L \cdot \frac{I_L}{\sqrt{3}} = \sqrt{3} \cdot U_L \cdot I_L \tag{31.21}$$

Aus der Übereinstimmung der beiden Gleichungen ergibt sich, dass die Leistung eines Dreiphasensystems unabhängig davon berechnet wird, ob eine Stern- oder Dreieckschaltung vorliegt.

Die Wirk- und Blindleistungen im Dreiphasensystem bei symmetrischer Belastung werden entsprechend berechnet.

Aus Gründen der Vereinfachung werden bei Berechnungen die Leiterwerte des Dreiphasensystems ohne Indizes geschrieben.

Beispiel

Ein symmetrischer Drehstromverbraucher ist im Stern geschaltet. Bekannt sind $U = 100\,\mathrm{V}$ und $R = 100\,\Omega$.

Wie groß sind die Strangströme und die Strangspannungen und welche Phasenlage haben sie? Welchen Wert hat die Scheinleistung?

$$U_{\mathrm{Str}} = \frac{U}{\sqrt{3}} = \frac{100\,\mathrm{V}}{\sqrt{3}} = 57{,}7\,\mathrm{V};$$
$$I = \frac{U_{\mathrm{Str}}}{R} = \frac{57{,}7\,\mathrm{V}}{100\,\Omega} = 0{,}577\,\mathrm{A}$$

Mit den Vorgaben im symmetrischen Drehstromnetz folgt für die Ströme und Spannungen:

$$\underline{I}_1 = 0{,}577\,\mathrm{A}\, e^{-j90}; \quad \underline{U}_{1N} = 57{,}7\,\mathrm{V}\, e^{-j90}$$
$$\underline{I}_2 = 0{,}577\,\mathrm{A}\, e^{j150}; \quad \underline{U}_{2N} = 57{,}7\,\mathrm{V}\, e^{j150}$$
$$\underline{I}_3 = 0{,}577\,\mathrm{A}\, e^{j60}; \quad \underline{U}_{3N} = 57{,}7\,\mathrm{V}\, e^{j60}$$

Scheinleistung:

$$S = \sqrt{3} \cdot U \cdot I = \sqrt{3} \cdot 100\,\mathrm{V} \cdot 0{,}577\,\mathrm{A} = 100\,\mathrm{V\,A}$$

oder in der komplexen Form

$$\underline{S} = 3 \cdot (\underline{U}_{\mathrm{Str}} \cdot \underline{I}^*_{\mathrm{Str}}) = 3 \cdot 57{,}7\,\mathrm{V} \cdot 0{,}577\,\mathrm{A} \cdot e^{j((-90)-(-90))}$$
$$= 100\,\mathrm{V\,A} \cdot e^{j0} \quad \blacktriangleleft$$

Beispiel

Gegeben ist ein symmetrischer Verbraucher mit induktiver Last in Dreieckschaltung mit folgenden Werten: $U = 100\,\mathrm{V}$; $Z = 300\,\Omega\, e^{j20}$. Welchen Wert und Phasenlage haben die Ströme in den Zuleitungen? Welchen Wert hat die Scheinleistung?

Strangströme:

$$\underline{I}_{12} = \frac{\underline{U}_{L1L2}}{\underline{Z}_{12}} = \frac{100\,\mathrm{V} \cdot e^{-j60}}{300\,\Omega \cdot e^{j20}}$$
$$= 0{,}33\,\mathrm{A} \cdot e^{-j80} = (0{,}057 - j0{,}325)\,\mathrm{A}$$
$$\underline{I}_{23} = \frac{\underline{U}_{L2L3}}{\underline{Z}_{23}} = \frac{100\,\mathrm{V} \cdot e^{j180}}{300\,\Omega \cdot e^{j20}}$$
$$= 0{,}33\,\mathrm{A} \cdot e^{j160} = (-0{,}31 + j0{,}113)\,\mathrm{A}$$
$$\underline{I}_{31} = \frac{\underline{U}_{L3L1}}{\underline{Z}_{31}} = \frac{100\,\mathrm{V} \cdot e^{j60}}{300\,\Omega \cdot e^{j20}}$$
$$= 0{,}33\,\mathrm{A} \cdot e^{j40} = (0{,}25 + j0{,}212)\,\mathrm{A}$$

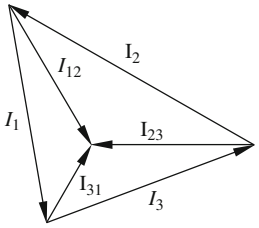


Abb. 31.8 Zeigerdiagramm der Ströme gemäß Beispiel

Ströme in der Zuleitung (Abb. 31.8):

$$\begin{aligned}
\underline{I}_1 &= \underline{I}_{12} - \underline{I}_{31} \\
&= 0{,}057\,\text{A} - j0{,}325\,\text{A} - (0{,}25\,\text{A} + j0{,}212\,\text{A}) \\
&= -0{,}193\,\text{A} - j0{,}537\,\text{A} = 0{,}57\,\text{A}\,e^{-j110} \\
\underline{I}_2 &= \underline{I}_{23} - \underline{I}_{12} \\
&= -0{,}31\,\text{A} + j0{,}113\,\text{A} - (0{,}057\,\text{A} - j0{,}325\,\text{A}) \\
&= -0{,}367\,\text{A} + j0{,}438\,\text{A} = 0{,}57\,\text{A}\,e^{j130} \\
\underline{I}_3 &= \underline{I}_{31} - \underline{I}_{23} \\
&= 0{,}25\,\text{A} + j0{,}212\,\text{A} - (-0{,}31\,\text{A} + j0{,}113\,\text{A}) \\
&= 0{,}56\,\text{A} + j0{,}1\,\text{A} = 0{,}57\,\text{A}\,e^{j10}
\end{aligned}$$

Scheinleistung:

$$\begin{aligned}
S &= \sqrt{3} \cdot U \cdot I = \sqrt{3} \cdot 100\,\text{V} \cdot 0{,}57\,\text{A} \\
&= 98{,}7\,\text{V A} \approx 100\,\text{V A}
\end{aligned}$$

◀

31.4 Das unsymmetrische Dreiphasensystem

Nicht bei jedem Verbraucher oder bei jedem Belastungsfall kann gewährleistet werden, dass die Impedanzen der drei Stränge gleichen Wert oder gleiche Phasenlage haben. Tritt ein Ungleichgewicht im Dreiphasensystem auf, hat das Verschiebungen im Netz zur Folge. Damit diese Verschiebungen, besonders im Niederspannungsnetz, nicht zu große Auswirkungen auf die Phasenlage haben, wird das Netz, wenn ungleichmäßige Belastungen zu erwarten sind, nicht als Dreileiternetz, sondern als *Vierleiternetz* ausgeführt.

31.4.1 Das unsymmetrische Dreileiternetz

Im Dreiphasennetz sind die Leiterspannungen in Größe und Phasenlage vorgegeben. Treten am Verbraucher Unsymmetrien auf, so verschieben sich die Ströme in den Zuleitungen, ergeben als Summe aber immer noch Null!

Beim Verbraucher in der Sternschaltung nehmen die Strangspannungen einen anderen Winkel und eine andere Größe ein.

Beispiel

Gegeben ist ein unsymmetrischer, im Stern geschalteter Drehstromverbraucher mit folgenden Werten: $U = 100\,\text{V}$; $R_1 = 100\,\Omega$; $R_2 = 50\,\Omega$; $R_3 = 100\,\Omega$.

Für die Berechnung muss eine Stern-Dreieck-Umwandlung durchgeführt werden.

$$\begin{aligned}
R_{12} &= \frac{R_1 \cdot R_3 + R_3 \cdot R_2 + R_2 \cdot R_1}{R_3} \\
R_{23} &= \frac{R_1 \cdot R_3 + R_3 \cdot R_2 + R_2 \cdot R_1}{R_1} \\
R_{31} &= \frac{R_1 \cdot R_3 + R_3 \cdot R_2 + R_2 \cdot R_1}{R_2}
\end{aligned}$$

Mit Hilfe der Umwandlung können die Dreieckwiderstände bestimmt werden, die den gleichen Leiterstrom erzeugen, wie die Strangwiderstände der Sternschaltung.

$$\begin{aligned}
R_{12} &= \frac{100\,\Omega \cdot 100\,\Omega + 100\,\Omega \cdot 50\,\Omega + 100\,\Omega \cdot 50\,\Omega}{100\,\Omega} \\
&= 200\,\Omega \\
R_{23} &= \frac{100\,\Omega \cdot 100\,\Omega + 100\,\Omega \cdot 50\,\Omega + 100\,\Omega \cdot 50\,\Omega}{100\,\Omega} \\
&= 200\,\Omega \\
R_{31} &= \frac{100\,\Omega \cdot 100\,\Omega + 100\,\Omega \cdot 50\,\Omega + 100\,\Omega \cdot 50\,\Omega}{50\,\Omega} \\
&= 400\,\Omega
\end{aligned}$$

Nun können die Ströme in der Dreieckschaltung berechnet werden.

$$\begin{aligned}
\underline{I}_{12} &= \frac{\underline{U}_{L1L2}}{R_{12}} = \frac{100\,\text{V}\,e^{-j60}}{200\,\frac{\text{V}}{\text{A}}\,e^{j0}} \\
&= 0{,}5\,\text{A}\,e^{-j60} = 0{,}25\,\text{A} - j0{,}433\,\text{A} \\
\underline{I}_{23} &= \frac{\underline{U}_{L2L3}}{R_{23}} = \frac{100\,\text{V}\,e^{j180}}{200\,\frac{\text{V}}{\text{A}}\,e^{j0}} \\
&= 0{,}5\,\text{A}\,e^{j180} = -0{,}5\,\text{A} + j0\,\text{A} \\
\underline{I}_{31} &= \frac{\underline{U}_{L3L1}}{R_{31}} = \frac{100\,\text{V}\,e^{j60}}{400\,\frac{\text{V}}{\text{A}}\,e^{j0}} \\
&= 0{,}25\,\text{A}\,e^{j60} = +0{,}125\,\text{A} + j0{,}216\,\text{A}
\end{aligned}$$

Ströme in der Zuleitung (Abb. 31.8):

$$\begin{aligned}
\underline{I}_1 &= \underline{I}_{12} - \underline{I}_{31} \\
&= 0{,}25\,\text{A} - j0{,}433\,\text{A} - (0{,}125\,\text{A} + j0{,}216\,\text{A}) \\
&= 0{,}125\,\text{A} + j0{,}649\,\text{A} = 0{,}66\,\text{A}\,e^{j79} \\
\underline{I}_2 &= \underline{I}_{23} - \underline{I}_{12} \\
&= -0{,}5\,\text{A} + j0\,\text{A} - (0{,}25\,\text{A} - j0{,}433\,\text{A}) \\
&= -0{,}75\,\text{A} - j0{,}433\,\text{A} = 0{,}866\,\text{A}\,e^{j150} \\
\underline{I}_3 &= \underline{I}_{31} - \underline{I}_{23} \\
&= 0{,}125\,\text{A} + j0{,}216\,\text{A} - (-0{,}5\,\text{A} + j0\,\text{A}) \\
&= 0{,}625\,\text{A} + j0{,}216\,\text{A} = 0{,}66\,\text{A}\,e^{j19}
\end{aligned}$$

Sind die Ströme in den Zuleitungen bekannt, können die Spannungen an den Widerständen der Sternschaltung berechnet werden.

$$\underline{U}_{1N} = \underline{I}_1 \cdot R_1 = 0{,}66\,\text{A} \cdot e^{j79} \cdot 100\,\Omega = 66\,\text{V} \cdot e^{j79}$$
$$\underline{U}_{2N} = \underline{I}_2 \cdot R_2 = 0{,}866\,\text{A} \cdot e^{-j150} \cdot 50\,\Omega = 43{,}3\,\text{V} \cdot e^{-j150}$$
$$\underline{U}_{3N} = \underline{I}_3 \cdot R_3 = 0{,}66\,\text{A} \cdot e^{j19} \cdot 100\,\Omega = 66\,\text{V} \cdot e^{j19}$$

Die Strangspannungen haben bei einer unsymmetrischen Belastung also nicht mehr den Wert $U_{\text{Str}} = \frac{U}{\sqrt{3}} = \frac{100\,\text{V}}{\sqrt{3}} = 57{,}7\,\text{V}$, den sie bei einer symmetrischen Belastung aufweisen. Der Sternpunkt verschiebt sich um einen Wert U_0 aus der Ursprungslage.

Scheinleistung:

$$\begin{aligned} S_{\text{ges}} &= S_1 + S_2 + S_3 = U_{1N} \cdot I_1 + U_{2N} \cdot I_2 + U_{3N} \cdot I_3 \\ &= 66\,\text{V} \cdot 0{,}66\,\text{A} + 43{,}3\,\text{V} \cdot 0{,}866\,\text{A} + 66\,\text{V} \cdot 0{,}66\,\text{A} \\ &= 43{,}56\,\text{V A} + 37{,}5\,\text{V A} + 43{,}56\,\text{V A} \\ &= 124{,}62\,\text{V A} \end{aligned}$$ ◀

Beispiel

Gegeben ist ein unsymmetrischer, im Dreieck geschalteter Drehstromverbraucher mit folgenden Werten: $U = 100\,\text{V}$; $R_{12} = 200\,\Omega$; $R_{23} = 200\,\Omega$; $R_{31} = 400\,\Omega$.

Ströme in der Dreieckschaltung:

$$\begin{aligned} \underline{I}_{12} &= \frac{\underline{U}_{L1L2}}{R_{12}} = \frac{100\,\text{V}\, e^{-j60}}{200\,\frac{\text{V}}{\text{A}}\, e^{j0}} \\ &= 0{,}5\,\text{A}\, e^{-j60} = 0{,}25\,\text{A} - j0{,}433\,\text{A} \\ \underline{I}_{23} &= \frac{\underline{U}_{L2L3}}{R_{23}} = \frac{100\,\text{V}\, e^{j180}}{200\,\frac{\text{V}}{\text{A}}\, e^{j0}} \\ &= 0{,}5\,\text{A}\, e^{j180} = -0{,}5\,\text{A} + j0\,\text{A} \\ \underline{I}_{31} &= \frac{\underline{U}_{L3L1}}{R_{31}} = \frac{100\,\text{V}\, e^{j60}}{400\,\frac{\text{V}}{\text{A}}\, e^{j0}} \\ &= 0{,}25\,\text{A}\, e^{j60} = 0{,}125\,\text{A} + j0{,}216\,\text{A} \end{aligned}$$

Ströme in der Zuleitung:

$$\begin{aligned} \underline{I}_1 &= \underline{I}_{12} - \underline{I}_{31} \\ &= 0{,}25\,\text{A} - j0{,}433\,\text{A} - (0{,}125\,\text{A} + j0{,}216\,\text{A}) \\ &= 0{,}125\,\text{A} + j0{,}649\,\text{A} = 0{,}66\,\text{A}\, e^{j79} \\ \underline{I}_2 &= \underline{I}_{23} - \underline{I}_{12} \\ &= -0{,}5\,\text{A} + j0\,\text{A} - (0{,}25\,\text{A} - j0{,}433\,\text{A}) \\ &= -0{,}75\,\text{A} - j0{,}433\,\text{A} = 0{,}866\,\text{A}\, e^{j150} \\ \underline{I}_3 &= \underline{I}_{31} - \underline{I}_{23} \\ &= 0{,}125\,\text{A} + j0{,}216\,\text{A} - (-0{,}5\,\text{A} + j0\,\text{A}) \\ &= 0{,}625\,\text{A} + j0{,}216\,\text{A} = 0{,}66\,\text{A}\, e^{j19} \end{aligned}$$

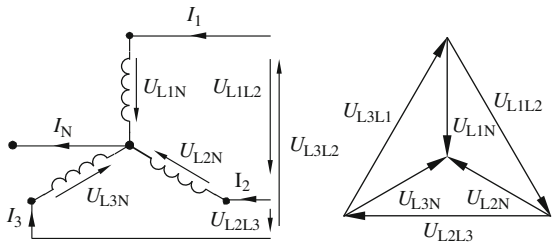


Abb. 31.9 Sternschaltung mit Zeigerbild der Spannungen im Vierleiternetz

Scheinleistung:

$$\begin{aligned} S_{\text{ges}} &= S_1 + S_2 + S_3 \\ &= U_{L1L2} \cdot I_{12} + U_{L2L3} \cdot I_{23} + U_{L3L1} \cdot I_{31} \\ &= 100\,\text{V} \cdot 0{,}5\,\text{A} + 100\,\text{V} \cdot 0{,}5\,\text{A} + 100\,\text{V} \cdot 0{,}25\,\text{A} \\ &= 50\,\text{V A} + 50\,\text{V A} + 25\,\text{V A} = 125\,\text{V A} \end{aligned}$$ ◀

31.4.2 Das unsymmetrische Vierleiternetz

Muss in einem Dreiphasensystem mit starken Unsymmetrien gerechnet werden, wird das Vierleiternetz eingesetzt. Beim Vierleiternetz ist der Sternpunkt des Verbrauchers mit dem Sternpunkt der Spannungsquelle über einen separaten Leiter (Neutralleiter) verbunden (Abb. 31.9). Treten am Verbraucher ungleiche Belastungen in den verschiedenen Phasen auf, wird der Differenzstrom des Systems über den Neutralleiter zur Spannungsquelle zurückgeleitet.

$$\textit{Neutralleiterstrom} \quad \underline{I}_N = \underline{I}_1 + \underline{I}_2 + \underline{I}_3 \qquad (31.22)$$

Ist der Widerstand des Neutalleiters sehr klein gegenüber dem Widerstand des Verbrauchers, verschiebt sich beim Vierleiternetz die Sternpunktspannung nicht; die Strangspannungen bleiben in Phasenlage und Größe erhalten.

Beispiel

Gegeben ist ein unsymmetrischer, im Stern geschalteter Drehstromverbraucher mit folgenden Werten: $U = 100\,\text{V}$; $R_1 = 100\,\Omega$; $R_2 = 50\,\Omega$; $R_3 = 100\,\Omega$.

Unter der Voraussetzung, dass der Widerstand des Neutralleiters gegen Null geht, folgt:

$$U_{\text{Str}} = \frac{U}{\sqrt{3}} = \frac{100\,\text{V}}{\sqrt{3}} = 57{,}7\,\text{V};$$

Strangströme:

$$\underline{I}_1 = \frac{\underline{U}_{1N}}{R} = \frac{57{,}7\,\text{V} \cdot e^{-j90}}{100\,\Omega} = 0{,}577\,\text{A} \cdot e^{-j90}$$
$$\underline{I}_2 = \frac{\underline{U}_{2N}}{R} = \frac{57{,}7\,\text{V} \cdot e^{j150}}{50\,\Omega} = 1{,}154\,\text{A} \cdot e^{j150}$$
$$\underline{I}_3 = \frac{\underline{U}_{3N}}{R} = \frac{57{,}7\,\text{V} \cdot e^{j30}}{100\,\Omega} = 0{,}577\,\text{A} \cdot e^{j30}$$

Der über den Neutralleiter abfließende Strom ergibt sich zu:

$$\begin{aligned}\underline{I}_N &= \underline{I}_1 + \underline{I}_2 + \underline{I}_3 \\ &= 0{,}577\,\mathrm{A} \cdot e^{-j90} + 1{,}154\,\mathrm{A} \cdot e^{j150} \\ &\quad + 0{,}577\,\mathrm{A} \cdot e^{j30} \\ &= (0 - j0{,}577)\,\mathrm{A} + (-1{,}0 + j0{,}577)\,\mathrm{A} \\ &\quad + (0{,}5 + j0{,}289)\,\mathrm{A} \\ &= (-0{,}5 + j0{,}289)\,\mathrm{A} = 0{,}577\,\mathrm{A} \cdot e^{j150}\end{aligned}$$

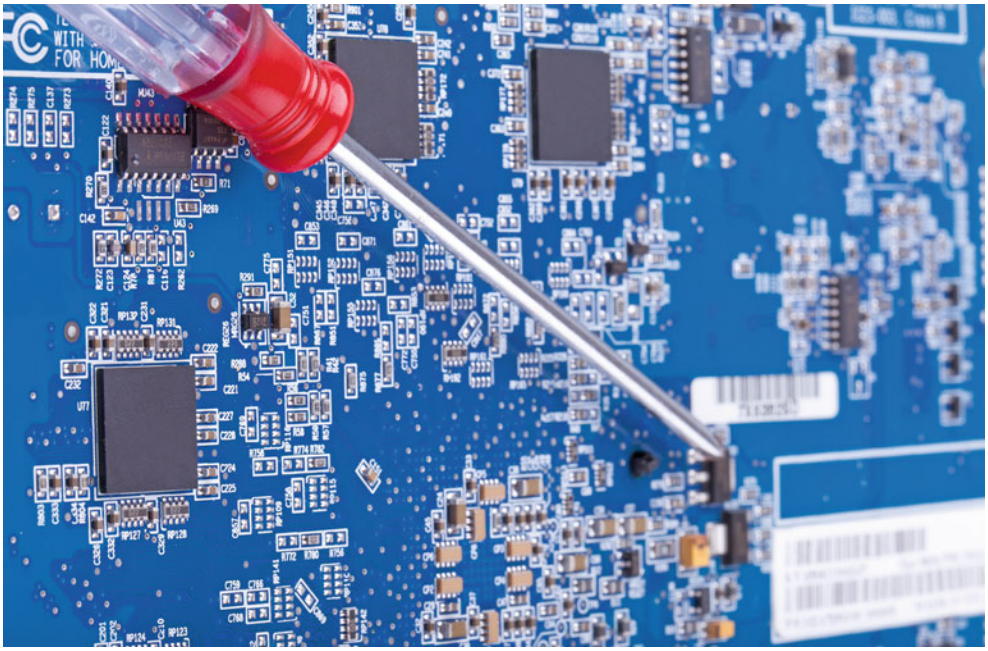
◀

Literatur

1. *Hagmann, G.:* Grundlagen der Elektrotechnik: Aula-Verlag, 2006
2. *Hering, E.; Martin, R.; Stohrer, M.:* Physik für Ingenieure: Springer, 2002
3. *Frohne, H.:* Moeller Grundlagen der Elektrotechnik: Teubner, 2005
4. *Zastrow, D.:* Elektrotechnik: Vieweg, 2006
5. *Lindner, H.:* Physik für Ingenieure: Fachbuchverlag Leipzig im Carl-Hanser-Verlag, 2006

Elektronik

V

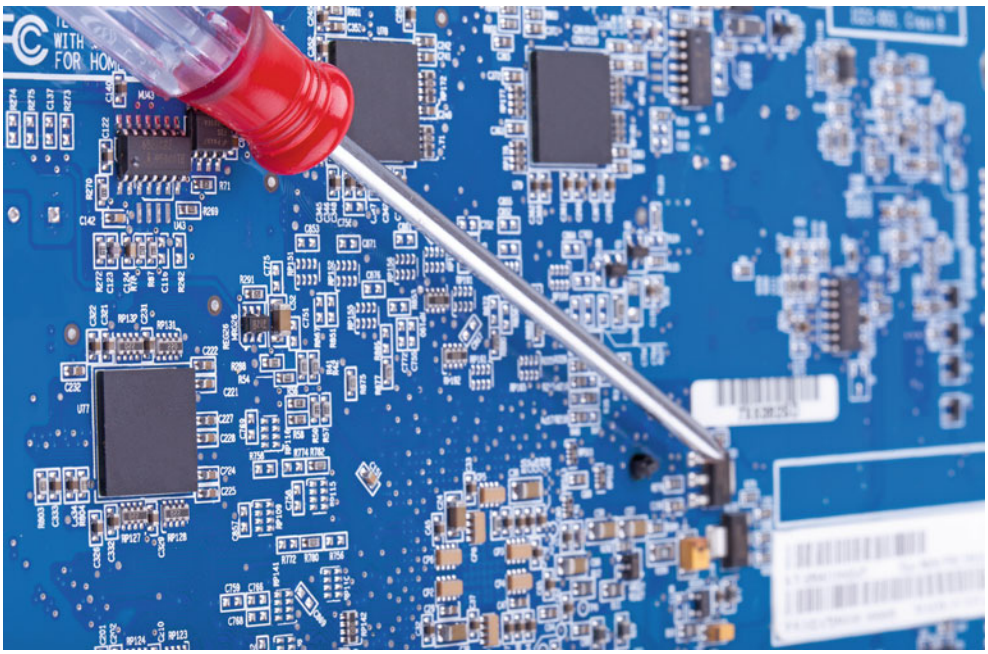


32 Leitungsmechanismen bei Halbleitern, pn-Übergang 393
33 Dioden 397
34 Mehrschichtdioden und -trioden 415
35 Transistoren 425
36 Besondere Halbleiter-Bauelemente 441
37 Analoge Verstärker 449
38 Endstufen 477
39 Operationsverstärker 485
40 Elektronische Schalter, Kippstufen 499
41 Oszillatoren 513
42 Schaltungstechniken 519
43 Optoelektronik 525
44 Analog-Digital-Wandler 537
45 Digital-Analog-Wandler 541

Leitungsmechanismen bei Halbleitern, pn-Übergang

32

Peter Döring



32.1 Einführung in die Halbleiterphysik 394
32.2 Der pn-Übergang 396

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_32

32.1 Einführung in die Halbleiterphysik

Im Periodischen System der Elemente findet man zwischen den Metallen und den Nichtmetallen Elemente, die als *Halbleiter* bezeichnet werden. Es handelt sich hierbei um Materialien mit einer *spezifischen Leitfähigkeit*, die in dem Bereich zwischen der spezifischen Leitfähigkeit von metallischen Leitern und der von Isolatoren liegt.

Die wichtigsten Halbleiterwerkstoffe sind das *Silizium* (Si) und das *Germanium* (Ge). Selen (Se) wird dagegen nur noch zur Herstellung kleinerer Netzgleichrichter verwendet.

Hinzu kommen jedoch noch die intermetallischen Verbindungen wie Gallium-Arsenid (GaAs), Gallium-Phosphid (GaP) und Indium-Arsenid (InAs). Sie werden hauptsächlich zur Herstellung von Fotohalbleitern und Hallgeneratoren verwendet.

Silizium und *Germanium* sind 4wertige Elemente; also bilden Atome mit vier Valenzelektronen das Kristallgitter (Tetraeder-Gefüge) von reinem Silizium und Germanium.

Bei diesen Elementen werden die Atome durch die Elektronenpaarbindung zusammengehalten. In den Atomkoordinationsgittern lagern sich die Atome mit ihren äußeren Elektronenschalen so aneinander, dass ihre Bindung über zwei Elektronen erfolgt. Die Atome teilen sich die Elektronen der äußersten Schale. Diese erhält so nach Abb. 32.1 den Bau einer Edelgasschale (Teil III).

Bei einem störungsfreien Kristallgitteraufbau und absolutem Temperaturnullpunkt ($\vartheta_0 = -273\,°\mathrm{C} = 0\,\mathrm{K(elvin)}$ befinden sich alle Atome im Ruhezustand. Die *Leitfähigkeit* des Halbleitermaterials ist unter den genannten Bedingungen gleich Null, und damit ist das Material ein absoluter Nichtleiter.

In reinem Silizium gibt nun jedes Si-Atom an vier Nachbaratome je ein Elektron ab oder nimmt von jedem der vier ein Elektron zur Aufrechterhaltung der Elektronenpaarbindung an. In einem reinen Siliziumkristall sind alle Valenzelektronen fest gebunden, so dass keine freien Elektronen zur Verfügung stehen.

Wird ein solcher Kristall der Einwirkung von Energie in Form von Licht und/oder Wärme ausgesetzt, so beginnen die Atome zu schwingen (thermische Eigenbewegung). Unter diesen Bedingungen können Valenzelektronen aus ihren Bindungen herausspringen und werden damit zu freien Elektronen, so dass die Leitfähigkeit des Materials größer wird.

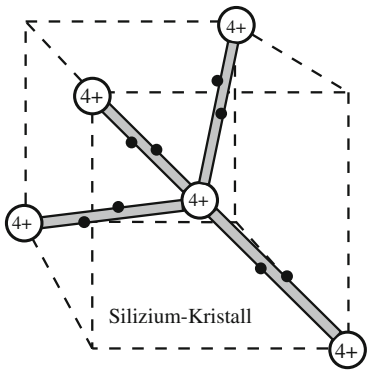


Abb. 32.1 Elektronenpaarbindung bei Silizium

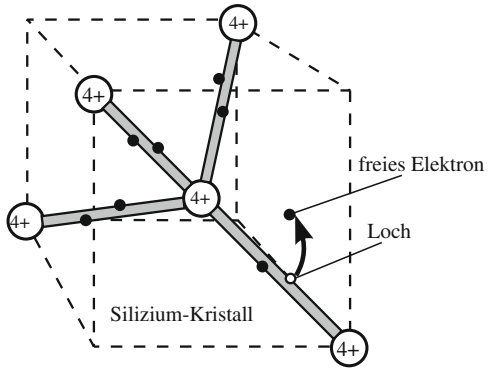


Abb. 32.2 Thermische Paarbildung beziehungsweise Generation im Teilchen-Modell

An der Stelle, an der ein Valenzelektron aus seiner Bindung gerissen wurde, fehlt nun jedoch eine negative Ladung. Infolge der positiven Ladung der Protonen im Atom verbleibt dem Atom eine positive Ladung, die als „Defektelektron“ oder „Loch“ bezeichnet wird.

Das Entstehen eines freien Elektrons und eines Loches nach Abb. 32.2 wird als „thermische Paarbildung“ oder „Generation“ bezeichnet. Bei andauernder Energiezuführung werden fortlaufend Elektronen frei, die scheinbar ziellos durch den Kristall wandern, bis sie auf ein Loch treffen und dort wieder in einen festen Atomverband zurückspringen. Dieser Vorgang wird „Rekombination“ genannt. Paarbildung und Rekombination sind stets im Gleichgewicht. Die Anzahl der verfügbaren Ladungsträger hängt aber von der Temperatur ab und wird mit steigender Temperatur größer.

Im Bereich der Raumtemperatur verdoppelt sich in etwa die Anzahl der Ladungsträgerpaare, wenn eine Temperaturerhöhung um 10 K vorgenommen wird. Diese Erhöhung der Leitfähigkeit mit steigender Temperatur wird allgemein „Eigenleitung“ genannt.

Beim Anlegen einer Spannung entsteht im Kristall ein elektrisches Feld, dass die frei gewordenen Elektronen vom Minuspol zum Pluspol der Spannungsquelle zieht. Sie springen auf ihrem Weg dahin immer von einem Loch zum anderen und „fallen hinein“ (rekombinieren). Damit wandern die Löcher aber scheinbar vom Pluspol zum Minuspol der Spannungsquelle.

In jedem Halbleitermaterial stehen also zum *Ladungstransport* sowohl *Elektronen* als auch *Löcher* zur Verfügung. Diesen bipolaren Leitungsmechanismus gilt es zu nutzen.

Neben dem *Teilchen-Modell* gibt es noch das *Bändermodell*, das auf Überlegungen der Quantentheorie zurückgeführt wird (siehe Teil III).

Danach kann Ladung (Energie) nur in Portionen, also bestimmten ganzzahligen Vielfachen der Grundladung e (Elektronenenergie) existieren. In einem Kristall können nach *Pauli* nur

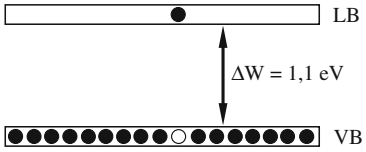


Abb. 32.3 Generation im Bändermodell

zwei Elektronen dasselbe Energieniveau haben. Dadurch liegen die Energieniveaulinien der Elektronen des Kristalls sehr dicht in Form von Bändern beieinander. Bei Festkörpern geben also *Bänder* die möglichen Energiezustände der Elektronen an:

1. Die Elektronen können keine beliebigen *Energiewerte* im Kristall einnehmen, sondern nur solche, die *innerhalb* eines Bandes liegen.
2. Die Elektronen können im Kristall keine *Energiezustände* einnehmen, deren Werte zwischen den *Bändern*, das heißt, in der verbotenen Zone, liegen.
3. Eine *Energiezufuhr* kann ein Anheben der Elektronen auf höhere Energiebänder bewirken, das heißt, die verbotene Zone kann von den Elektronen übersprungen werden.
4. Nur in teilweise besetzten Bändern kann eine Elektronenleitung vonstattengehen. In absolut leeren Bändern sind keine Elektronen vorhanden, und in völlig besetzten sind keine erlaubten Energieniveaus mehr vorhanden.
5. Bei Energieabgabe springen alle vorher angeregten Elektronen auf noch freie Plätze in den niederen Energiebändern zurück.

Werden Elektronen aus dem Valenzband VB in das Leitungsband LB angehoben, so ist die Elektronenbesetzung des VB nicht mehr komplett. Dieses VB mit relativ wenig unbesetzten Elektronenzuständen verhält sich nun hinsichtlich der Leitfähigkeit so, als wären die entstandenen Lücken positiv geladene Teilchen in einem sonst leeren Band.

Bei den Halbleitern ist deshalb eine *n-Leitung* durch die negativ geladenen Elektronen im Leitungsband und eine *p-Leitung* durch die positiv geladenen Defektelektronen des Valenzbandes zu unterscheiden.

Der „thermischen Paarbildung“ im Teilchenmodell steht also hier nach Abb. 32.3 die Anhebung eines Valenzelektrons durch Temperaturanstieg vom oberen Rand des Valenzbandes zum unteren Rand des Leitungsbandes adäquat gegenüber. Die zur Paarbildung notwendige Energie wird *Bindungsenergie* genannt und entspricht der Breite der verbotenen Zone.

Bei der bisher besprochenen „Eigenleitung“ hängt die Leitfähigkeit, also die Anzahl der freien Ladungsträger (Elektronen und Löcher) nur von der Temperatur ab.

Wie macht man die Leitfähigkeit möglichst temperaturunabhängig?

Die Leitfähigkeit wird fast temperaturunabhängig, wenn man *Störstellen* im Gitter (Abb. 32.4) durch das gezielte Einbringen von Fremdatomen in das Kristallgitter des reinen Halbleitermaterials einbaut.

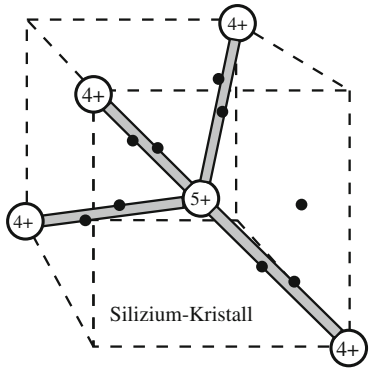


Abb. 32.4 n-Dotierung von Silizium

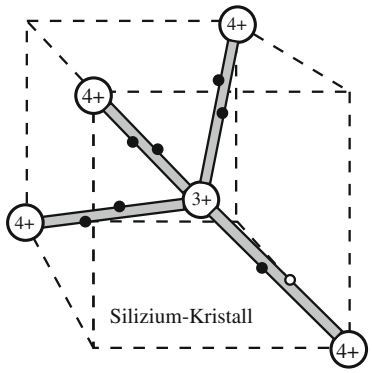


Abb. 32.5 p-Dotierung von Silizium

Ein derartiger Einbau von Fremdatomen wird als *Dotieren* bezeichet. Zum Dotieren sind Stoffe geeignet, die entweder ein Valenzelektron mehr oder ein Valenzelektron weniger besitzen als zum Beispiel Silizium.

Die Elemente *Phosphor* (P), *Arsen* (As) und *Antimon* (Sb) haben fünf Valenzelektronen und damit ein Elektron mehr als Silizium hat.

Bringt man Atome dieser Elemente in das Kristallgitter der Halbleiter, so wird ein Elektron für die Elektronenpaarbindung im Kristall nicht gebraucht. Es entsteht ein Elektronenüberschuss im Kristallgitter (Elektronenspender = Donatoren).

Wird dagegen mit *Bor* (B), *Aluminium* (Al), *Gallium* (Ga) oder *Indium* (In) dotiert, entsteht im Kristall ein Elektronenmangel, weil diese Atome nur über drei Valenzelektronen verfügen. Eine Paarbindung ist nicht komplett, das heißt es fehlt eine negative Ladung, was bedeutet, dass wir nun nach Abb. 32.5 Löcher (Defektelektronen) eingebracht haben (Elektronenempfänger = Akzeptoren).

Diese durch gezielte Verunreinigung bewirkte Leitfähigkeit wird *Störstellenleitung* genannt. Sie ist maßgebend für die Leitfähigkeit aller dotierten Halbleitermaterialien und ist fast unabhängig von Temperaturänderungen.

Mit Hilfe des Bändermodells lässt sich dieses Phänomen auch erklären: Durch die Dotierung entstehen zusätzliche Energieniveaus, so dass ein Ladungsträgertransport bei äußerst kleiner

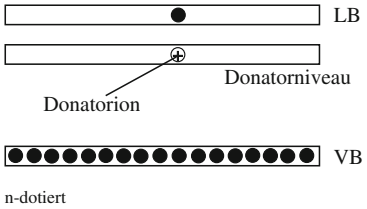


Abb. 32.6 Energieniveau bei n-dotiertem Silizium

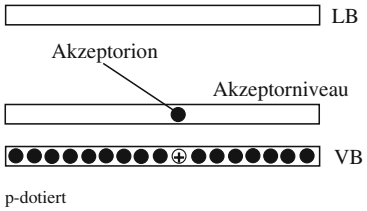


Abb. 32.7 Energieniveau bei p-dotiertem Silizium

Energiezufuhr möglich ist. Die Lage der neuen Energieniveaus ist im Abb. 32.6 für n-dotiertes Material und Abb. 32.7 für p-dotiertes Material erkennbar.

Wir haben nun n-dotiertes und p-dotiertes Halbleitermaterial zur Verfügung und kennen zwei Leitungsmechanismen, nämlich n-Leitung und p-Leitung.

32.2 Der pn-Übergang

Fügt man n- und p-dotiertes Halbleitermaterial aneinander, liegt ein *pn-Übergang* nach Abb. 32.8 vor. Infolge *Diffusion* (Verschieben von Ladungsträgern zwischen den beiden unterschiedlich dotierten Siliziumschichten infolge ihrer Wärmebewegung) verlagern sich Elektronen aus dem n-Bereich, in dem sie die Majoritätsträger (Mehrheit) darstellen, in den p-Bereich und rekombinieren mit den Löchern. Ebenso diffundieren Defektelektronen (Löcher) aus dem p-Bereich in den n-Bereich und rekombinieren mit den Elektronen.

Diese Ladungsträger sind nunmehr für den Stromtransport verloren. Es bildet sich eine Grenzschicht aus, die keine oder nur sehr wenige freie Ladungsträger enthält, die sogenannte *Sperrschicht*. Infolge der fehlenden Ladungsträger ist die Leitfähigkeit dieser Zone sehr gering.

Da im p-Bereich aufgrund der Dotierung negative Atomrümpfe (Ionen) und im n-Bereich positive Atomrümpfe im Kristallgitter eingebunden sind, entsteht eine *Raumladung*, also eine

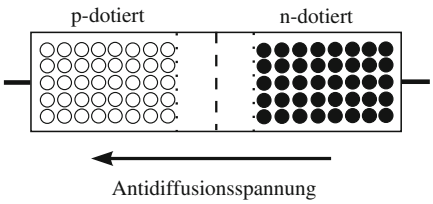


Abb. 32.8 pn-Übergang mit Antidiffusionsspannung

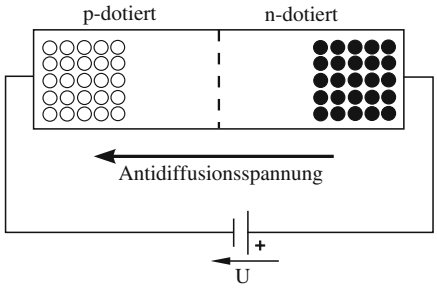


Abb. 32.9 pn-Übergang in Sperrrichtung

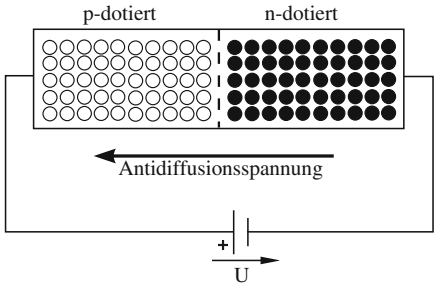


Abb. 32.10 pn-Übergang in Durchlassrichtung

Potentialdifferenz und damit ein elektrisches Feld, die eine weitere Diffusion unterbindet und somit die Breite der Sperrschicht begrenzt. Diese *Antidiffusionsspannung* ist von der Dotierung, vom Grundmaterial und der Temperaturspannung abhängig. Weil diese Spannung, die manchmal auch als Diffusionsspannung bezeichnet wird, nur unmittelbar an der Sperrschicht auftritt, kann sie von außen nicht gemessen werden.

Legt man nun nach Abb. 32.9 von außen an den pn-Übergang eine größere Spannung an, die dieselbe Polarität wie die Antidiffusionsspannung aufweist, also eine Spannungsquelle mit dem Minus-Pol an den p-Bereich und dem Plus-Pol an den n-Bereich, so werden Elektronen in den p-Bereich gepumpt, die dann ihrerseits mit den Löchern rekombinieren.

In den n-Bereich werden zusätzliche Löcher gepumpt (Elektronen abgesaugt). Die Folge ist eine Ausweitung der Sperrschicht, da diese Grenzschicht weiter an Ladungsträgern verarmt. Der Widerstand der Sperrschicht wird nahezu unendlich groß.

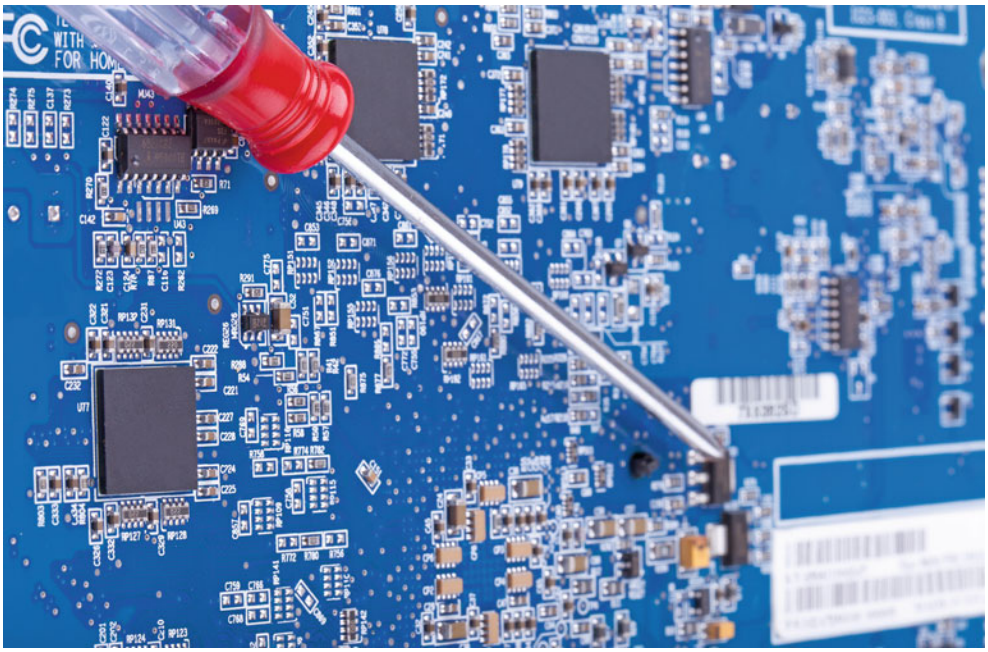
Polt man nun die Spannungsquelle nach Abb. 32.10 um (Plus-Pol an den p-Bereich und den Minus-Pol an den n-Bereich), werden Elektronen in den n-Bereich eingebracht, während Defektelektronen in den p-Bereich verbracht werden. Die Folge ist, dass rekombinierte Teilchen sofort durch neue ersetzt, die Sperrschicht abgebaut und bei genügend großer angelegter Spannung zum Verschwinden gebracht wird. Damit fließt ein größer werdender Strom durch den pn-Übergang. Überschreitet die Spannung den Wert der Antidiffusionsspannung, wird der Durchlassstrom schnell sehr groß.

Ein pn-Übergang ist also ein Bauelement, dessen Widerstand von der Richtung der angelegten Spannung abhängig ist. Man unterscheidet eine *Sperrrichtung* und eine *Durchlassrichtung*.

Dioden

33

Peter Döring



33.1 Kennlinien 398
33.2 Kenndaten und Grenzwerte 399
33.3 Kennzeichnung von Halbleiter-Bauelementen 401
33.4 Diodenarten 402
33.5 Anwendungsschaltungen 406

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_33

33.1 Kennlinien

Aus den *Diodenkennlinien* nach Abb. 33.1 lassen sich Daten für den Einsatz und das Verhalten der einzelnen Diodentypen ermitteln.

Der *Durchlassbereich* wird von den Größen Durchlassspannung U_F und Durchlassstrom I_F beschrieben (F steht für forward direction, engl.: Vorwärtsrichtung).

Legt man an die Kennlinie in Durchlassrichtung eine Tangente, wenn die Kennlinie in eine Gerade übergeht, so schneidet diese Tangente die Spannungsachse bei der Schleusenspannung $U_{(TO)}$.

Es gelten folgende Werte:

$$\begin{aligned} \text{Ge} \quad &\Rightarrow \quad U_{(TO)} = 0{,}2 \ldots 0{,}4\,\text{V} \\ \text{Si} \quad &\Rightarrow \quad U_{(TO)} = 0{,}5 \ldots 0{,}8\,\text{V} \end{aligned}$$

Der *Sperrbereich* wird von den Größen Sperrspannung U_R und Sperrstrom I_R beschrieben (R steht für reverse direction, engl.: Rückwärtsrichtung).

Es gelten folgende Werte:

$$\begin{aligned} \text{Ge} \quad &\Rightarrow \quad I_R = 10 \ldots 500\,\mu\text{A} \\ \text{Si} \quad &\Rightarrow \quad I_R = 5 \ldots 500\,\text{nA} \end{aligned}$$

Im Koordinatenursprung scheint in der Kennlinie ein Knick zu sein. Bei genauerem Hinsehen findet man als Ursache einen Wechsel der Einheitengröße bei Strom und Spannung. Die Spannung $U_{R(BR)}$ ist hier nach Abb. 33.1 die rückwärtige Durchbruchspannung.

Die *Durchbruchspannung* ist als die Spannung definiert, für die die maximale Feldstärke im pn-Übergang einen kritischen Wert erreicht und der Sperrstrom sehr plötzlich große Werte annimmt.

Der Strom am pn-Übergang setzt sich aus einem Löcherinjektionsstrom ins n-Gebiet, einem Elektroneninjektionsstrom ins p-Gebiet und einem Rekombinations-Generationsstrom in der Raumladungszone zusammen. Dieser Strom ist temperaturabhängig. Je höher die Temperatur wird, umso größer ist die Wahrscheinlichkeit, dass die Elektronen große Geschwindigkeiten annehmen.

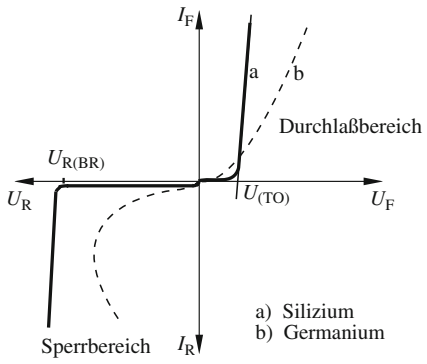


Abb. 33.1 Vierquadrantenkennlinie für Silizium- und Germanium-Dioden

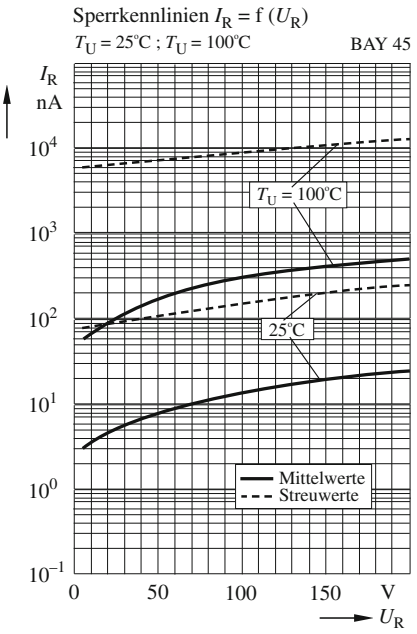


Abb. 33.2 Sperrkennlinien einer Diode als $f(T)$

Die *Temperaturspannung* U_T ist eine Funktion der Temperatur und lässt sich über den Energieerhaltungssatz aus der Wärmelehre ermitteln. Setzt man voraus, dass die Elektronen, bedingt durch die Temperatur T, eine bestimmte Energie besitzen, ohne dass eine elektrische Spannung vorhanden ist, so ist die Temperaturspannung als eine Spannung erklärbar.

$$\textit{Temperaturspannung} \quad U_T = \frac{k \cdot T}{|e|} \tag{33.1}$$

mit $k = 1{,}3804 \cdot 10^{-23}\,\text{W s/K}$ (Boltzmannkonstante) und $e = 1{,}602 \cdot 10^{-19}\,\text{A s}$ (Elementarladung) (für $T = 300\,\text{K}$ ist $U_T \approx 26\,\text{mV}$).

Tritt nun eine elektrische Feldstärke infolge einer angelegten Spannung auf, so wird die Bewegung der Elektronen zum positiven Potential gerichtet sein. Insgesamt kann der Strom mathematisch wie folgt als die Kennlinie einer Diode in Durchlassrichtung so beschrieben werden:

$$\textit{Diodenkennlinie} \quad I = I_S\{e^{U/U_T} - 1\} \tag{33.2}$$

Hierbei ist I_S eine Konstante (Sättigungsstrom), die von der Konstruktion der Diode und verschiedenen Materialparametern abhängt.

Sperr- und Durchlassverhalten einer Diode sind temperaturabhängig. Bei einem Anstieg der Sperrschichttemperatur ϑ_J um ca. 10 K bei einer Ge-Diode und ca. 7 K bei einer Si-Diode verdoppelt sich der Sperrstrom. Die Schleusenspannung verringert sich mit zunehmender Sperrschichttemperatur ϑ_J um ca. 2 mV/K.

Die Diagramme in Abb. 33.2 und 33.3 zeigen die Temperaturabhängigkeit der Sperr- beziehungsweise Durchlasskennlinien für 25 °C und 100 °C.

33.2 Kenndaten und Grenzwerte

Diodenhersteller geben in den Datenblättern eine Reihe von Kennwerten an. Diese Kennwerte sind getrennt nach Grenzdaten und Kenndaten zu bewerten.

Grenzwerte sind solche Werte, die auf keinen Fall überschritten werden dürfen, weil das Bauelement sonst zerstört werden könnte. Ein einzelner Grenzwert darf auch dann nicht überschritten werden, wenn andere Grenzwerte nicht voll ausgenutzt sind.

Dioden können auf verschiedene Arten elektrisch überlastet werden; durch einen zu großen Durchlassstrom I_F, eine zu hohe Sperrspannung U_R, eine zu große, in ihr umgesetzte Verlustleistung P_V oder durch eine zu hohe Sperrschichttemperatur $\vartheta_{J\max}$.

Die Datenblätter enthalten nach DIN 41782 Strom- und Spannungswerte mit folgenden Bedeutungen: Stoßspitzenspannung U_{RSM}, Periodische Spitzensperrspannung U_{RRM}, Dauergrenzstrom I_{FAV}, Durchlassstrom-Effektivwert I_{FRMS} und Stoßstrom-Grenzwert I_{FSM}.

Beim Betrieb einer Diode entsteht sowohl in Durchlassrichtung als auch in Sperrrichtung eine Verlustleistung P_V in der Diode, wobei bei niedrigen Frequenzen die Sperrverluste gegenüber den Durchlassverlusten vernachlässigbar klein sind.

$$\textit{Verlustleistung} \quad P_V = U_F \cdot I_F \tag{33.3}$$

Zur Berechnung dieser Verlustleistung muss U_F näher bestimmt werden. Es gilt:

$$U_F = U_{(\mathrm{TO})} + r_F \cdot I_F$$

Eingesetzt in (33.3) ergibt sich:

$$P_V = (U_{(\mathrm{TO})} + r_F \cdot I_F) \cdot I_F = U_{(\mathrm{TO})} \cdot I_F + r_F \cdot I_F^2 \tag{33.4}$$

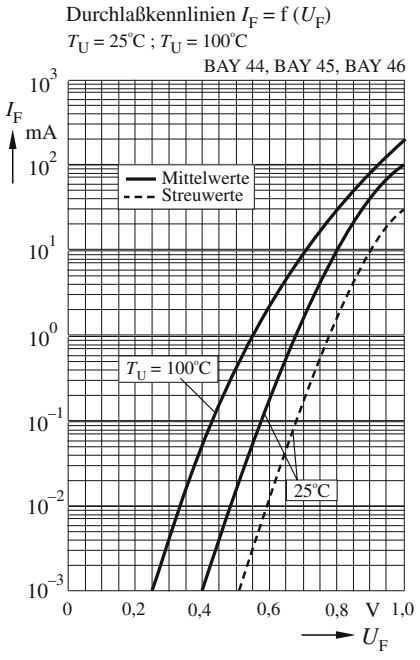


Abb. 33.3 Durchlasskennlinien einer Diode als $f(T)$

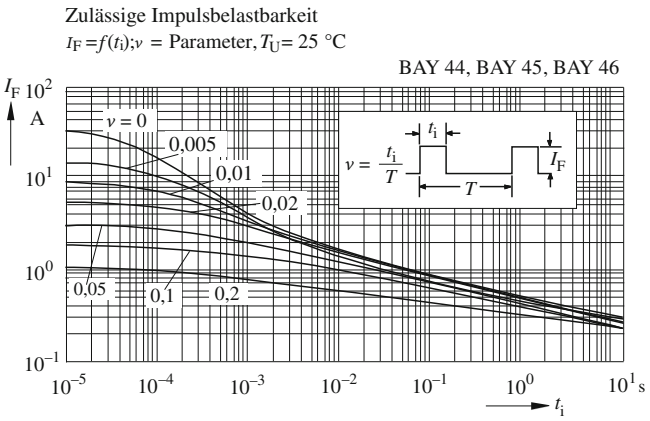


Abb. 33.4 Zulässige Impulsbelastbarkeit

Sowohl Spannung als auch Strom sind hier abhängig von der Zeit $f(t)$. Der 1. Teil der Summe in (33.4) beschreibt den DC-Anteil der Leistung, während der 2. Teil der Summe den AC-Anteil darstellt. Damit ergibt sich $P_V = U_{(\mathrm{TO})} \cdot I_{\mathrm{FAV}} + r_F \cdot I_{\mathrm{RMS}}^2$ und wegen $I_{\mathrm{RMS}} = F \cdot I_{\mathrm{FAV}}$ mit $F \equiv$ Formfaktor ergibt sich die

totale Verlustleistung

$$P_{\mathrm{tot}} \geq P_V = U_{(\mathrm{TO})} \cdot I_{\mathrm{FAV}} + r_F (F \cdot I_{\mathrm{FAV}})^2 \tag{33.5}$$

Diese *Verlustleistung* führt zur Erwärmung des Halbleiterkristalls. Die größte zulässige Verlustleistung wird als totale Verlustleistung P_{tot} vom Hersteller für bestimmte Temperaturbedingungen angegeben.

In diesem Zusammenhang muss auch die zulässige Impulsbelastbarkeit von Dioden gesehen werden. Wird die Diode mit einer rechteckförmigen Spannung betrieben, kann mit dem Diagramm nach Abb. 33.4 gearbeitet werden. Hier ist die Abhängigkeit des zulässigen Stroms I_F von der Impulsdauer t_i aufgetragen. Als Parameter wird hier $\nu = t_i/T$ angegeben. Je größer der Parameter ν wird, umso geringer wird der zulässige Strom in Durchlassrichtung.

Hauptsächlich durch die *Verlustleistung bei Durchlassbetrieb* entsteht in der Sperrschicht Wärme, durch die sich die Sperrschichttemperatur erhöht. Diese darf die maximale Sperrschichttemperatur $\vartheta_{J\max}$ nicht überschreiten, weil der Kristall seine Halbleitereigenschaften ansonsten stark verändert und die Diode zerstört wird.

Die höchste zulässige Sperrschichttemperatur hängt vom Material ab: $\vartheta_{J\max} \approx 70\,^{\circ}\mathrm{C}$ bis $90\,^{\circ}\mathrm{C}$ bei Germaniumdioden und $\vartheta_{J\max} \approx 150\,^{\circ}\mathrm{C}$ bis $200\,^{\circ}\mathrm{C}$ bei Siliziumdioden.

Besteht die Gefahr, dass die höchstzulässige Sperrschichttemperatur infolge der Verlustleistung und der Umgebungstemperatur überschritten werden kann, so muss die Diode „gekühlt“ werden.

Als *Kenndaten* werden Eigenschaften der Dioden angegeben, die das Verhalten in bestimmten Arbeitspunkten kennzeichnen. Die Kenndaten werden in dynamische und statische Daten unterteilt.

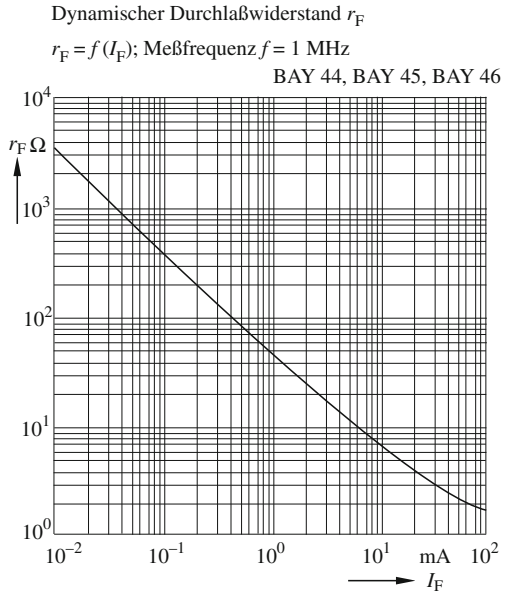


Abb. 33.5 Dynamischer Durchlasswiderstand r_F einer Diode

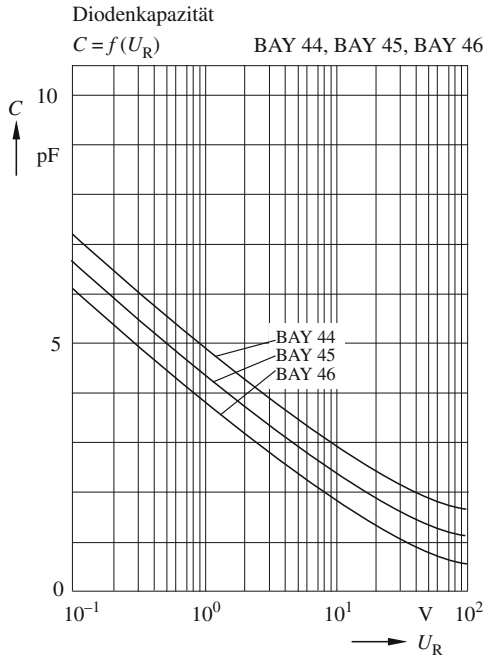


Abb. 33.6 Dioden-Kapazität als Funktion der Sperrspannung

Statische Kenndaten beschreiben das Gleichstromverhalten während *dynamische* Kenndaten über das Verhalten bei Wechselstrom- und Impulsbetrieb informieren.

Der statische Durchlasswiderstand R_F ist der Gleichstromwiderstand einer Diode. Er ist nicht konstant, sondern hängt vom gewählten Arbeitspunkt ab. Oberhalb der Schleusenspannung ist R_F sehr klein, unterhalb aber bereits relativ groß.

Der dynamische Durchlasswiderstand r_F ist der Wechselstromwiderstand der Diode. Er lässt sich aus der Stromänderung ΔI_F infolge einer Spannungsänderung ΔU_F für einen bestimmten Arbeitspunkt aus dem Kennlinienbild ermitteln. Der dynamische Durchlasswiderstand wird mit steigendem Durchlassstrom kleiner. Sein Wert wird vom Hersteller in den entsprechenden Datenblättern in doppelt-logarithmischen Diagrammen, zum Beispiel im Diagramm nach Abb. 33.5 dargestellt.

dynamischer Durchlasswiderstand

$$r_F = \frac{\Delta U_F}{\Delta I_F} \tag{33.6}$$

Der statische Sperrwiderstand R_R ist im gesamten zulässigen Sperrbereich sehr groß. Er ändert sich auch nur relativ wenig in Abhängigkeit von der Sperrspannung. Wird jedoch die höchstzulässige Spitzensperrspannung U_{RSM} überschritten, so wird der Sperrwiderstand plötzlich sehr klein. Eine normale Diode ist dann nicht mehr funktionsfähig.

Die *Sperrschicht*, die sich zwischen dem p-dotierten und dem n-dotierten Material ausbildet, wirkt als Isolator wie ein Dielektrikum, an das sich auf beiden Seiten ein gut leitendes Material anschließt. Jede Diode besitzt daher auch ohne angelegte Spannung eine *Sperrschichtkapazität*. Sie wird nach Abb. 33.6 mit zunehmender Sperrspannung kleiner, weil die Breite der Sperrschicht größer wird. Aus der Sperrschichtkapazität und der zwischen den Anschlüssen der Diode und dem Gehäuse auftretenden Kapazität ergibt sich die Diodenkapazität C_D.

Obwohl diese Diodenkapazität einen relativ kleinen Wert hat, kann sie beim Diodeneinsatz sehr störend wirken. So bildet eine Diode mit einem Vorwiderstand ein RC-Glied, das eine bestimmte Zeitkonstante hat. Bei Betrieb mit einer Wechselspannung sehr hoher Frequenz wirkt das RC-Glied als Tiefpass, bei Impulsbetrieb als Integrierglied, durch das eine Impulsverformung auftritt.

Beim Einsatz in der Digitaltechnik müssen also Dioden mit sehr kleiner Diodenkapazität ausgewählt werden.

Die Sperrschicht ist abgebaut, wenn eine Diode in Durchlassrichtung betrieben wird. Nun befinden sich zahlreiche Ladungsträger im Bereich der Grenzschicht zwischen dem p- und n-Material. Bei schlagartigem Umpolen der äußeren Spannung (Übergang vom Durchlass- in den Sperrbetrieb) kann eine *Sperrwirkung* des pn-Überganges erst dann entstehen, wenn alle Ladungsträger aus der Grenzschicht abgeflossen sind (Träger-Staueffekt, TSE).

Die Zeit, die das Ausräumen der Ladungsträger aus der Grenzschicht erfordert, wird *Sperrverzögerungszeit* t_{rr} (**r**everse **r**ecovery **t**ime) oder Rückwärtserholzeit t_{rr} genannt. Abb. 33.7 zeigt den Zusammenhang zwischen dem maximal auftretenden Sperrstrom I_{RM} und der Dauer des fließenden Sperrstroms, bis dieser auf ca. 10 % von I_{RM} abgesunken ist, bei der die Sperrverzögerungszeit endet.

Die Sperrverzögerungszeit soll aus mehreren Gründen möglichst klein sein. So kann unmittelbar nach dem Umschalten vom Durchlass- in den Sperrbereich bereits eine hohe Sperrspannung U_R anliegen, während noch für eine gewisse Zeit ein

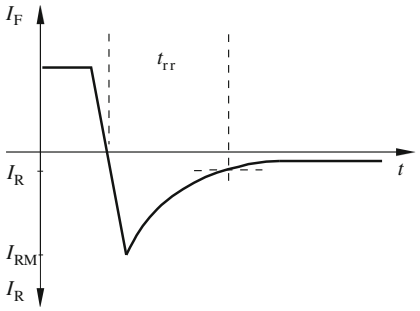


Abb. 33.7 Sperrverzögerungszeit t_{rr}

relativ großer Ausräumstrom I_R fließt. Kurzzeitig ergibt sich dadurch eine hohe Verlustleistung, durch die die Diode zerstört werden kann. Eine TSE-Beschaltung (RC-Reihenglied parallel zur Diode) mindert diesen Effekt.

Weiterhin bestimmt die Sperrverzögerungszeit aber auch die höchste Frequenz, die mit der betreffenden Diode noch gleichgerichtet werden kann. Eine Gleichrichtung ist nur möglich, wenn die Dauer der zu sperrenden Halbwelle größer als die Sperrverzögerungszeit ist.

Bei Betrieb einer Diode (aber auch aller anderen Halbleiterbauelemente) wird die in der (oder den) Sperrschicht(en) auftretende Verlustleistung P_V in Wärme umgewandelt. Diese Wärme muss über das den Kristall umschließende Gehäuse aus Metall, Kunststoff oder Glas an die Umgebungsluft abgeführt werden.

Aufgrund der thermischen Gesetze geschieht dieses in dem Maße, je größer die Temperaturdifferenz zwischen Sperrschicht und Umgebungsluft ist und je besser die Ableitung der erzeugten Wärme erfolgen kann.

Wie gut die Wärme von der Sperrschicht (J = junction) durch das Material an die Luft (U = Umgebung, auch A für engl. = ambient) abgeführt werden kann, wird durch den *Wärmewiderstand* $R_{\mathrm{th\,JU}}$ beschrieben, wobei der Index „th“ den thermischen Widerstand beschreibt.

Den mathematischen Zusammenhang beziehungsweise die Definition des Wärmewiderstandes kann man leicht formulieren.

$$\textit{Wärmewiderstand} \quad R_{\mathrm{th\,JU}} = \frac{\vartheta_J - \vartheta_U}{P_V} \tag{33.7}$$

Bei Dioden mit Verlustleistungen $P_V < 1\,\mathrm{W}$ genügt meist die Wärmeableitung über die Oberfläche des Gehäuses. Der Wärmewiderstand umfasst den Weg von der Sperrschicht über das Gehäuse an die Umgebungsluft. Bei Verlustleistungen $P_V > 1\,\mathrm{W}$ reicht die Wärmeableitung über das Gehäuse allein nicht mehr aus, weil die Oberfläche des Gehäuses zu klein ist. Um die abstrahlende Oberfläche zu vergrößern, wird die Diode auf einen zusätzlichen Kühlkörper gesetzt. Hierfür eignen sich Metalle, die eine gute Wärmeleitfähigkeit bei gleichzeitig geringem Gewicht haben.

Werden Dioden mit einem zusätzlichen Kühlkörper betrieben, so setzt sich der Wärmewiderstand $R_{\mathrm{th\,JU}}$ aus einer Reihenschaltung der drei Teilwiderstände $\rightarrow R_{\mathrm{th\,JG}}$ (Wärmewiderstand Sperrschicht – Gehäuse), $R_{\mathrm{th\,GK}}$ (Wärmewiderstand Gehäuse – Kühlkörper) und $R_{\mathrm{th}\,K}$ (Wärmewiderstand des Kühlkörpers) zusammen.

$$\rightarrow R_{\mathrm{th\,JU}} = R_{\mathrm{th\,JG}} + R_{\mathrm{th\,GK}} + R_{\mathrm{th}\,K}$$

Der gesamte Wärmewiderstand $R_{\mathrm{th\,JU}}$ zwischen Sperrschicht und Umgebungsluft ist nun wegen der großen Oberfläche kleiner als der Wärmewiderstand $R_{\mathrm{th\,JU}}$, der für das Halbleiterbauelement ohne Kühlkörper gilt. Je kleiner der Wärmewiderstand ist, umso größere Wärmemengen können von der Sperrschicht abgeführt werden. Die Wärmespeicherung kann durch entsprechende Ersatzschaltbilder (zum Beispiel „Kondensator“) dargestellt und berechnet werden.

Zulässige Verlustleistung

$$P_V = \frac{\vartheta_J - \vartheta_U}{R_{\mathrm{th\,JG}} + R_{\mathrm{th\,GK}} + R_{\mathrm{th}\,K}} \tag{33.8}$$

Der Wärmewiderstand $R_{\mathrm{th}\,K}$ eines Kühlkörpers ist aber nicht nur von der Größe der Oberfläche, sondern auch vom Material abhängig. Diese Materialabhängigkeit wird meistens in einem Diagramm angegeben. Heute werden in zunehmendem Umfang Rippenkühlkörper verwendet. Sie werden hauptsächlich aus einer *Aluminiumlegierung* gefertigt.

Zwischen dem Gehäuse des Bauelementes und dem Kühlkörper soll ein sehr enger Kontakt bestehen, damit der Wärmewiderstand $R_{\mathrm{th\,GK}}$ möglichst klein wird. Dies lässt sich mit einer *Wärmeleitpaste*, zum Beispiel Siliconfett erreichen.

33.3 Kennzeichnung von Halbleiter-Bauelementen

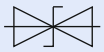
Die Typenbezeichnung setzt sich nach „Pro Electron“ aus *zwei* (für Typen, die vorwiegend in Rundfunk-, Fernseh- und Magnettongeräten verwendet werden) oder *drei* (Typen für kommerzielle Zwecke, beziehungsweise Industrietypen) Buchstaben und einer Zahlenkombination zusammen. Bei amerikanischen Herstellern ist die Bezeichnung 1N… für Dioden und 2N… für Transistoren üblich, wobei die Zahl vor dem „N“ die Anzahl der internen pn-Übergänge kennzeichnet. Hier ist das Ausgangsmaterial der Bauelemente nicht erkennbar. Die nachfolgende drei- oder vierstellige Zahl ist eine Registriernummer ohne besondere technische Aussage.

Das folgende Schema zeigt die Kennzeichnung von Halbleiter-Bauelementen der Firma Siemens und anderer Firmen. Darin bedeuten:

1. Buchstabe (Ausgangsmaterial):

A Germanium; B Silizium; C Material mit einem Energieabstand von mehr als 1,3 eV, zum Beispiel Gallium-Arsenid; D Material mit einem Energieabstand von weniger als 0,6 eV, zum Beispiel Indium-Antimonid; R Halbleiter-Material für Photoleiter und Hallgeneratoren

Tab. 33.1 Diodenarten und Anwendungsbereiche

Art	Schaltzeichen	Anwendungsbereich
Silizium-Flächendiode		Universaldiode zur Gleichrichtung von Spannungen und Strömen in der Elektrotechnik, Elektronik und Messtechnik
Silizium-Gleichrichterzelle		Gleichrichten von Strömen und Spannungen in der Leistungselektronik
Germanium-Spitzendiode		Gleichrichten von kleinen Wechselspannungen und -strömen bis zu hoher Frequenz. Anwendung als schnelle Schaltdiode in der Digitalelektronik
Schottky-Diode		Gleichrichten von kleinen Wechselspannungen und -strömen bis zu extrem hoher Frequenz. Anwendung als sehr schnelle Schaltdiode. Bestandteil monolithischer integrierter Schaltungen in der Digitalelektronik
Backwarddiode		Gleichrichten extrem kleiner Wechselspannungen bei hoher Frequenz
Tunneldiode		Schwingungserzeugung im Mikrowellenbereich. Schnelle Schaltdiode in der Digitalelektronik
Kapazitätsdiode		Abstimmung von Schwingkreisen, automatische Nachstimmschaltungen, Modulationsschaltungen, Mischer und Frequenzvervielfacher
Z-Diode		Begrenzen von Spannungen, Spannungs-Referenzelement in der Messtechnik, Schaltungen zur Spannungsneubildung
Magnetdiode		Magnetfeldabhängige Steuerungen
Suppressor-Diode		Begrenzen von Überspannungsspitzen im gesamten Bereich der Elektronik

2. Buchstabe (Verwendungszweck):

A Allgemeine Kleinsignalgleichrichtung, Schaltdiode; B Kapazitätsdiode; C NF-Transistor; D NF-Leistungstransistor; E Tunneldiode; F HF-Transistor; G Oszillatordiode für HF-Anwendung, auch Multichips; H Diode, die auf ein Magnetfeld anspricht, Hall-Feldsonde; L HF-Leistungstransistor; N Optokoppler; P Fotodiode, Fotoelement (strahlungsempfindliches Bauelement); Q Leuchtdiode (strahlungserzeugendes Bauelement); R Elektrisch ausgelöste Steuer- oder Schaltbauteile mit Durchbruchcharakteristik; S Schalttransistor; T Elektrisch oder mittels Licht ausgelöste Steuer- oder Schaltbauteile mit Durchbruchcharakteristik; U Leistungs-Schalttransistor; X Verfielfacherdiode; Y Leistungsdiode; Z Z-Diode, Referenzdiode

Der 3. Buchstabe (nur X, Y und Z) kennzeichnet kommerzielle Typen.

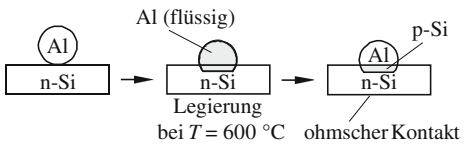


Abb. 33.8 Prinzip des Legierungsverfahrens

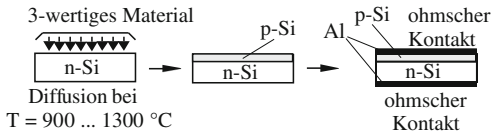


Abb. 33.9 Prinzip des Diffusionsverfahrens

33.4 Diodenarten

Dioden unterscheidet man zweckmäßigerweise nach ihrer Anwendung voneinander. Die bisher beschriebenen Grenz- und Kenndaten werden hierbei je nach Ausprägung der Dioden und der beabsichtigten Anwendung als Vor- oder Nachteile angesehen und zielgerichtet verwendet. Die folgenden Diodenarten stellen nur eine willkürliche Auswahl dar (Tab. 33.1).

Alle mit den Vorgängen im Halbleitermaterial zusammenhängenden elektrischen Eigenschaften sind durch geeignete Wahl der Konzentration der Fremdatome (Donatoren beziehungsweise Akzeptoren) beiderseits des pn-Überganges, durch Wahl der Dicke von p- und n-Bereich, durch Wahl der Kristallfläche und durch das Herstellungsverfahren in weiten Grenzen beeinflussbar. In der Art, wie die gewünschten Konzentrationen von Donatoren und Akzeptoren in das Grundmaterial Silizium eingebracht werden, unterscheiden sich die möglichen grundsätzlichen Herstellungsverfahren.

Beim *Legierungsverfahren* nach Abb. 33.8 wird eine kleine Aluminium- oder Indiumkugel (-draht) auf eine n-dotierte Siliziumscheibe gebracht. Beides zusammen wird soweit erhitzt, bis sich ein kleiner geschmolzener Al-Si-Bereich bildet, der nach dem Abkühlen eine p-leitende Zone bildet. Das Restaluminium bildet den Kontakt der p-Zone. Die genaue Lage, die Fläche und die Eindringtiefe des entstandenen pn-Überganges sind nur schwer exakt zu steuern. Das Legierungsverfahren wird deshalb nur noch in Sonderfällen eingesetzt.

Das *Festkörper-Diffusionsverfahren* erlaubt eine wesentlich bessere Kontrolle der Eindringtiefe des pn-Überganges. Bei diesem Verfahren werden nach Abb. 33.9 einkristalline *n-dotierte Si-Scheiben* bei hohen Temperaturen einer Dotierstoff enthaltenen Atmosphäre ausgesetzt. Die Fremdatome diffundieren in

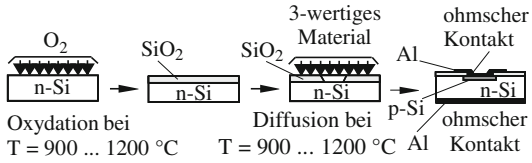


Abb. 33.10 Planar-Diffusionsverfahren

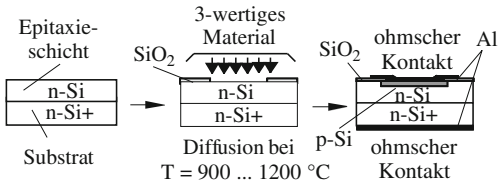


Abb. 33.11 Epitaxie-Planar-Diffusionsverfahren

das n-Silizium, wobei sich durch geeignete Wahl von Temperatur und Zeit die Eindringtiefe sehr genau einstellen lässt. Die Fläche des für eine einzelne Diode benutzten pn-Überganges wird hierdurch jedoch nicht bestimmt. Diese muss durch Zerkleinern in Einzelkristalle gewünschter Größe erfolgen. Der dann seitlich freiliegende pn-Übergang muss zusätzlich geätzt und abgedeckt werden.

Durch das Planarverfahren ist eine wesentlich bessere Kontrolle der Geometrie von pn-Übergängen möglich. Bei diesem Verfahren wird nach Abb. 33.10 auf der einkristallinen n-dotierten Si-Scheibe bei hohen Temperaturen durch Sauerstoffzufuhr eine SiO_2-Schicht aufgebracht, durch die quasi keine Fremdatome eindringen können. Durch einen *photolithographischen Prozess* wird in diese SiO_2-Schicht ein Fenster eingefügt, das einer Dotierstoff enthaltenen Atmosphäre ausgesetzt wird. Die Fremdatome diffundieren an dieser Stelle in das n-Silizium. So lässt sich durch geeignete Wahl von Temperatur und Zeit die Eindringtiefe und die Fläche des für eine einzelne Diode benutzten pn-Überganges sehr genau einstellen.

Hinsichtlich der inneren Verluste bringt das *Epitaxie-Verfahren* nach Abb. 33.11 eine weitere Verbesserung gegenüber dem einfachen Planar-Diffusionsverfahren mit sich. Hierbei wird auf der niederohmigen einkristallinen n-dotierten Si-Scheibe epitaktisch eine ebenfalls, allerdings hochohmige, einkristalline n-dotierte Si-Scheibe durch chemische Reaktion bei der Entstehung abgeschieden. Die weiteren Schritte entsprechen dem oben beschriebenen Planar-Diffusionsverfahren.

Diese grundsätzlichen Herstellungsverfahren können natürlich auch für die Fertigung von Transistoren und anderen HL-Bauelementen angewendet werden. Zur Herstellung von Feldeffekttransistoren und integrierten Schaltkreisen wird neuerdings die Ionenimplantation angewendet. Bei diesem Verfahren werden die Dotierungsatome in einem elektrischen Feld ionisiert und mit hoher Geschwindigkeit in das Grundmaterial „geschossen“. Da sich hierbei die Eindringtiefe sehr genau steuern lässt, sind die pn-Übergänge sehr scharf abgegrenzt, so dass unerwünschte Kapazitäten klein gehalten werden können. Auch die Größe des einzelnen Transistorsystems kann verringert werden, was die Integrationsdichte von ICs erhöht.

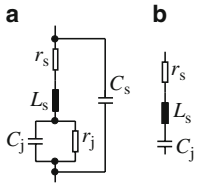


Abb. 33.12 Ersatzschaltungen von Kapazitätsdioden

33.4.1 Kapazitätsdioden

Bei den Kapazitätsdioden wird die Sperrschichtkapazität und die Abhängigkeit ihrer Größe von der anliegenden Sperrspannung als gewünschte spezielle Eigenschaft benutzt. Die in Sperrrichtung betriebene Kapazitätsdiode kann mit einem Plattenkondensator verglichen werden. Die Breite der Sperrschicht, das heißt hier der Plattenabstand, kann durch die Größe der Sperrspannung variiert werden. So wird mit zunehmender Sperrspannung die Breite der Sperrschicht größer und damit die Sperrschichtkapazität C_j kleiner. Um eine starke Spannungsabhängigkeit der Sperrschichtkapazität zu erreichen, bildet man hyperabrupte pn-Übergänge. Die Diffusionstechnik macht es möglich, Siliziumdioden mit eng tolerierter und recht verlustfreier Sperrschichtkapazität herzustellen.

Bei den Kenndaten ist die *Nennkapazität* von wesentlicher Bedeutung. Sie wird in den Datenblättern stets für bestimmte Sperrspannungen angegeben und wird nicht nur durch die Sperrschichtkapazität C_j gebildet. Die Sperrschicht bildet keinen idealen kapazitiven Blindwiderstand, sondern ist im Zusammenhang mit den anderen Anteilen der Diode in den *Ersatzschaltungen* nach Abb. 33.12 zu betrachten. Neben der Kapazität C_j ist auch die Güte Q einer Kapazitätsdiode beim praktischen Einsatz von Bedeutung, da Kapazitätsdioden für den Einsatz in HF-Schwingkreisen vorgesehen sind. Unter Vernachlässigung der Serieninduktivität L_S lässt sich die Güte nach Abb. 33.12a definieren.

$$\text{Güte} \quad Q = \frac{1}{\omega C_j \cdot r_s + \frac{1}{\omega C_j \cdot r_j}} \tag{33.9}$$

Gleichung (33.9) lässt erkennen, dass die Güte frequenzabhängig ist. Die Serieninduktivität (bedingt durch die Anschlussdrähte) ist bei hohen Frequenzen nicht vernachlässigbar. Die meisten Anwendungen ermöglichen die Verwendung eines Ersatzschaltbildes nach Abb. 33.12b. Hierbei kann der Sperrschichtwiderstand r_j und die Anschlusskapazität C_s vernachlässigt werden. In den Datenblättern der Hersteller wird die Abhängigkeit der Güte von der Frequenz in Diagrammen dargestellt, desgleichen die Temperatur- und Sperrspannungsabhängigkeit der ausnutzbaren Kapazität.

Wegen ihrer relativ kleinen Kapazitäten werden Kapazitätsdioden überwiegend als Schwingkreiskapazitäten in der Hochfrequenztechnik eingesetzt. Sie haben in Radio- und Fernsehgeräten den früher üblichen mechanischen Drehkondensator weitgehend verdrängt.

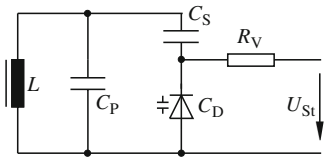


Abb. 33.13 Prinzipschaltung einer Nachstimmschaltung

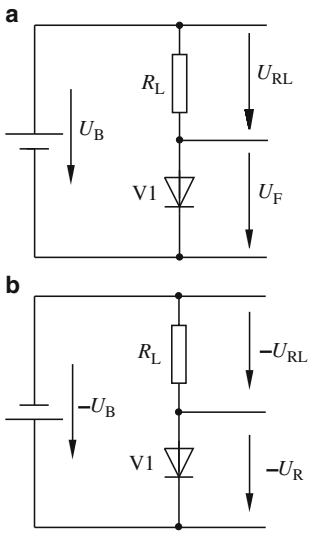


Abb. 33.14 Diode als elektronischer Schalter

Die Kapazität der Kapazitätsdiode ist in den Nachstimmschaltungen nach Abb. 33.13 meistens nur eine Teilkapazität. Durch Ansteuern mit einer veränderbaren Gleichspannung U_{st} wird die Sperrschichtkapazität und damit auch die Resonanzfrequenz des Schwingkreises verändert (siehe Teil XIII).

33.4.2 Schalterdioden

In einem Stromkreis muss ein Schalter in der Stellung „Geschlossen“ einen möglichst kleinen Widerstand (idealer Wert $R = 0\,\Omega$) und in der Stellung „Offen“ einen möglichst großen Widerstand (idealer Wert $R = \infty$) haben. Mit mechanischen Schaltern lassen sich die idealen Werte weitgehend erreichen, bei elektronischen Schaltern jedoch nur näherungsweise.

Die Diode stellt einen elektronischen Schalter dar, wenn ihr kleiner Durchlasswiderstand R_F und ihr großer Sperrwiderstand R_R als Schalterzustand benutzt werden. Der Schaltzustand der Diode als Schalter hängt also von der Polarität der angelegten Betriebsspannung U_B in der Schaltung nach Abb. 33.14a, b ab. Wird der Lastwiderstand R_L als Widerstandsgerade in das Kennlinienfeld einer Diode nach Abb. 33.15 eingetragen und eine Rechteckspannung U_B mit positiven und negativen Amplituden verwendet, so können die jeweils auftretenden Spannungen und Ströme in den beiden sich einstellenden Arbeitspunkten AP abgelesen und die zugehörigen Leistungen ermittelt werden.

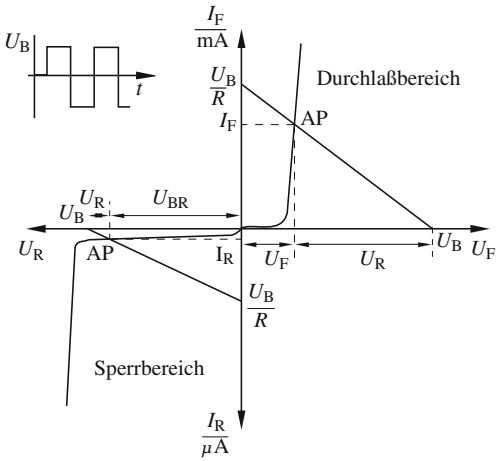


Abb. 33.15 Kennlinienfeld einer Diode mit Arbeitsgerade und Arbeitspunkten

Die Schnittpunkte der Durchlass- und Sperrkennlinien mit den Widerstandsgeraden ergeben die sich einstellenden Arbeitspunkte bei leitender Diode (Schalter geschlossen) und gesperrter Diode (Schalter offen).

Diodenschalter

$$U_B = I_F \cdot R_L + U_F \quad \text{(Schalter geschlossen)}$$
$$U_B = I_R \cdot R_L + U_R \quad \text{(Schalter offen)}$$

Beim Arbeitspunkt in Durchlassrichtung fällt an der Diode die Durchlassspannung U_{FA} ab, während am Widerstand die Spannung $(U_B - U_{FA})$ abfällt. Durch beide Bauelemente fließt der Strom I_{FA} (Durchlassstrom im Arbeitspunkt). Da auch in Sperrrichtung ein geringer Strom (Sättigungsstrom, Reststrom) fließt, fällt auch am Widerstand in Sperrrichtung eine minimale Spannung ab.

Wichtige Grenzdaten: maximale Sperrspannung U_{RRM}; maximaler Durchlassstrom I_{FAVM}; maximale Sperrschichttemperatur $\vartheta_{J\,max}$; maximale Verlustleistung P_{tot}; minimale Sperrverzögerungszeit $t_{rr\,min}$

$$\textit{Schaltleistung} \quad P_S = I_F^2 \cdot R_L \leq I_{FAVM}^2 \cdot R_L \qquad (33.10)$$

Die Schaltleistung P_S, die mit einer Diode geschaltet werden kann, ist wesentlich größer als die zulässige Verlustleistung der Diode. Sie hängt praktisch nur von der Betriebsspannung ab.

33.4.3 Schottky-Dioden

Unter bestimmten Bedingungen stellt sich eine Gleichrichterwirkung auch zwischen einer n-dotierten Siliziumschicht und einer aufliegenden Metallelektrode ein. Dieser Effekt wurde nach seinem Entdecker „Schottky-Effekt“ benannt. Weil die Elektronen im n-Silizium einen höheren Energiezustand haben als die Elektronen im Metall, wandern Elektronen aus dem n-Silizium in das Metall. Dadurch entsteht in der Grenzschicht eine Raumladungszone.

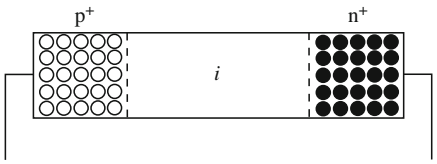


Abb. 33.16 Schematische Darstellung einer pin-Diode

In Durchlassrichtung erreichen die frei beweglichen Elektronen des Siliziums eine so hohe Energie, dass sie aus der n-Zone in das Metall hineinwandern. Die frei beweglichen Elektronen des Metalls können dagegen bei Raumtemperatur das Metall nicht verlassen. Daher kann auch nach Umpolen der angelegten Spannung kein Elektronenfluss in dem Silizium-Kristall zustandekommen. Der Übergang vom Sperr- in den Durchlasszustand und umgekehrt erfolgt sehr schnell, da es praktisch keine Sperrschichtkapazität gibt, die von Ladungsträgern geräumt werden muss. Die Schaltzeiten von Schottky-Dioden sind daher sehr klein.

Darum werden Schottky-Dioden als integrierte Bauelemente in integrierten Schaltkreisen der Digitalelektronik (zum Beispiel Schottky-TTL in 74LSxx-Bausteinen) verwendet.

Schottky-Dioden werden auch als „Hot-Carrier-Dioden“ (engl. = heiße Ladungsträger-Diode) bezeichnet aufgrund des höheren Energiezustandes der Elektronen im Silizium im Vergleich zum Metall.

Wichtige Kennwerte: Einschaltverzögerungszeit $t_{tr} \approx 50\,\text{ps}$; Ausschaltverzögerungszeit $t_{rr} \approx 100\,\text{ps}$; Diodenkapazität $C_D \approx 10\,\text{pF}$; Sperrstrom $I_R \approx 5\,\mu\text{A}$.

33.4.4 Gleichrichter-Dioden

Gleichrichterdioden lassen sich in Klein- und Leistungsgleichrichterdioden unterteilen. Als Kleingleichrichterdioden werden in der Regel Dioden mit höchstzulässigen Durchlassströmen von einigen mA bis zu etwa 5 A bezeichnet. Der Bereich der Leistungsgleichrichterdioden erstreckt sich von etwa 5 bis 2000 A.

Die Durchbruchspannung $U_{R(\text{BR})}$ eines normalen pn-Überganges liegt für sinnvolle Störstellenkonzentration von ca. $10^{16}\,\text{cm}^{-3}$ in der Größenordnung von einigen 10 V. Das reicht für Netzgleichrichter (mögliche Sperrspannungen bei einem 230-V-Netz liegen bei 650 V) nicht aus.

Deshalb sind für solche Gleichrichterdioden andere Konstruktionen notwendig und zwar die pin-Struktur nach Abb. 33.16. Hierbei befindet sich zwischen zwei hochdotierten p- und n-Gebieten noch eine eigenleitende Schicht (i-Gebiet) oder eine schwach n- beziehungsweise p-leitende Schicht. Diese Zwischenschicht nimmt nun die gesamte Sperrspannung auf.

Bei einer pin-Diode ist die Feldstärke über die gesamte i-Schicht konstant. Damit ist die Sperrspannung U_R proportional zur Dicke d_i der i-Schicht. Reine i-Schichten sind in der Praxis nicht

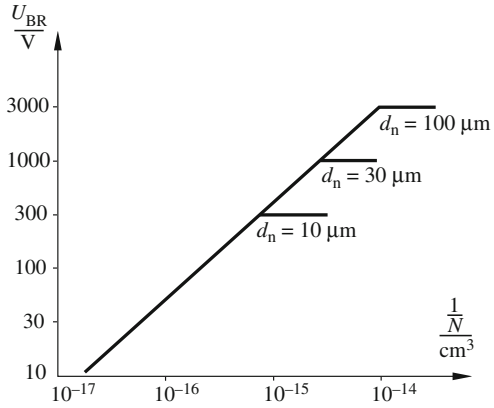


Abb. 33.17 Durchbruchspannung in Abhängigkeit von der Dotierung und der Dicke des Mittelgebietes

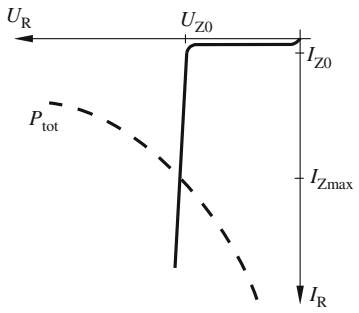


Abb. 33.18 Charakteristische Kennlinie einer Z-Diode mit Arbeitsbereich

erreichbar, denn hier dürften absolut keine Störstellen enthalten sein. Leistungsgleichrichter werden deshalb als p^+nn^+-Diode oder p^+pn^+-Diode realisiert.

Die Durchbruchspannung $U_{R(\text{BR})}$ von p^+nn^+-Dioden in Abhängigkeit von der Dotierung und der Dicke d_n des n-dotierten Mittelgebietes zeigt das Diagramm nach Abb. 33.17. Heute werden Leistungsgleichrichterdioden mit Durchbruchspannung in der Größe von 4800 V hergestellt.

33.4.5 Z-Dioden

Wird bei einer Si-Diode die Sperrspannung über den Wert U_{RRM} hinaus erhöht, steigt der Strom nach dem Überschreiten der Durchbruchspannung $U_{R(BR)}$ plötzlich sehr stark an. Dioden, die in Sperrrichtung einen sehr scharf einsetzenden, steilen Stromanstieg nach Abb. 33.18 besitzen, werden Z-Dioden genannt. Wegen ihrer charakteristischen Kennlinie im Sperrbereich (Bereich der Durchbruchspannung) kann die Z-Diode zur Stabilisierung oder Begrenzung von Gleichspannungen benutzt werden. Bei den Z-Dioden überlagern sich im Bereich der Durchbruchspannungen zwei unterschiedliche physikalische Effekte.

So tritt bei Z-Dioden mit einer Z-Spannung von $U_Z < 5\,\text{V}$ eine innere Feldemission auf, die als Zener-Effekt bezeichnet wird.

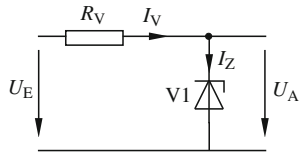


Abb. 33.19 Z-Diode mit Vorwiderstand

Ein Durchbruch aufgrund des Zener-Effektes erfolgt, weil beim Überschreiten einer bestimmten Feldstärke beim Silizium Elektronen aus dem Gitterverband herausgelöst werden. Die dabei freiwerdenden Elektronen und Löcher erhöhen die Anzahl der freien Ladungsträger und damit die Leitfähigkeit der Diode. Dadurch steigt der Sperrstrom I_Z stark an.

Bei Z-Dioden mit etwa $U_Z > 6\,\text{V}$ erfolgt dagegen der Durchbruch aufgrund des Lawineneffektes (engl.: avalanche breakdown) und führt zu einem scharfen Abknicken der Kennlinie. Beim Lawinen-Effekt ist die Geschwindigkeit der vorhandenen beweglichen Ladungsträger infolge der hohen Feldstärke so groß, dass sie beim Aufprall auf Atome Elektronen und Löcher herauslösen, die als frei bewegliche Ladungsträger nun ihrerseits weitere Ladungsträger aus dem Atomverband herausschlagen. Weil der Sperrstrom I_Z dadurch sehr schnell anwächst, wird dieser Vorgang als Lawinendurchbruch bezeichnet.

Zwischen etwa 5 und 6 V liegt ein Gebiet, in dem sich Zener-Effekt und Lawinen-Effekt überlagern.

Z-Dioden mit einer Durchbruchspannung $U_Z < 5\,\text{V}$ haben für die Durchbruchspannung einen negativen Temperaturkoeffizienten, weil eine Temperaturerhöhung eine verstärkte Feldemission bewirkt. Bei Z-Dioden mit $U_Z > 6\,\text{V}$ wird mit zunehmender Temperatur die freie Weglänge der Ladungsträger kleiner. Um den gleichen Durchbruchstrom aufgrund des Lawineneffekts zu erzielen, sind bei größeren Temperaturen größere Sperrspannungen notwendig. Diese Z-Dioden haben einen positiven Temperaturkoeffizienten α. Bei positiven Temperaturkoeffizienten verschiebt sich die Kennlinie zu höheren Z-Spannungen.

Der höchstzulässige Sperrstrom $I_{Z\,\text{max}}$ ergibt sich nach Abb. 33.18 als Schnittpunkt der Leistungshyperbel P_{tot} mit der Sperrkennlinie der Z-Diode. Z-Dioden werden in einem Arbeitsbereich betrieben, der zwischen dem Stromwert $I_{Z\,\text{max}}$ und einem Wert I_Z liegt, der unmittelbar hinter dem scharfen Kennlinienbereich ist. Lieferbar sind Z-Dioden mit Z-Spannungen U_{ZO} zwischen 2,4 V und einigen 100 V. Da ihr Arbeitspunkt immer im Bereich der Durchbruchspannung liegt, muss zur Strombegrenzung unbedingt ein Vorwiderstand R_V nach Abb. 33.19 verwendet werden.

Wichtig ist der differentielle Widerstand r_Z der Diode, der die Steilheit der Arbeitskennlinie beschreibt. Je kleiner dieser Widerstand ist, umso geringer ist die Spannungsänderung bei einer bestimmten Stromänderung. Der differentielle Widerstand r_Z liefert also eine Aussage über den Stabilisierungsgrad der Z-Diode. Er ist nicht völlig konstant, sondern hängt von der Temperatur, dem Z-Strom und der Z-Spannung nach Abb. 33.20 ab.

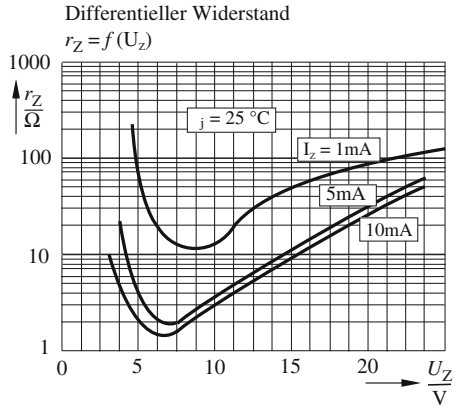


Abb. 33.20 Differentieller Widerstand r_Z

Die für eine Z-Diode in den Datenblättern angegebene Z-Spannung gilt oft für einen Strom $I_Z = 5\,\text{mA}$. Die Abstufung zwischen den Spannungen entspricht häufig der E-24-Reihe. Um die Toleranzen der angegebenen Z-Spannungen zu kennzeichnen, wird an die Typenbezeichnung des Grundtyps der Kennbuchstabe der Güteklasse angehängt.

Wichtige Grenzdaten: max. Sperrschichttemperatur $\vartheta_{J\,\text{max}}$; maximale Verlustleistung P_{tot}; maximaler Durchlassstrom I_{FAVM}; Wärmewiderstand $R_{\text{th JU}}$.

Wichtige Kenndaten: Nennspannung U_{ZO} bei $I_Z = 5\,\text{mA}$; differentieller Widerstand r_Z bei $I_Z = 5\,\text{mA}, 1\,\text{mA}$.

33.5 Anwendungsschaltungen

Dioden sind in vielen Funktionen und Leistungskategorien einsetzbar. Ihre Anwendung in der Nachrichtentechnik unterliegt anderen Gesichtspunkten als in der Leistungselektronik oder Digitaltechnik.

33.5.1 Begrenzerschaltungen

Ein Einsatzgebiet der Diode als Schalter sind Begrenzerschaltungen. Das sind Schaltungen, die Spannungsamplituden oberhalb oder unterhalb eines bestimmten positiven oder negativen Pegels durchlassen oder unterdrücken.

Im Bereich der Digitaltechnik müssen sowohl Störspannungsspitzen unterdrückt wie auch das Rauschen bei Signalen herausgefiltert werden. Diodenschalter wie Differenzier- und Integrierglieder lassen sich auch zur Impulsformung einsetzen.

Begrenzerschaltungen werden auch zur Unterdrückung von Störspannungsspitzen in der Signaltechnik eingesetzt, aber auch zur Trennung der einzelnen Amplitudenhöhen eines Fernseh-BAS-Signals benutzt.

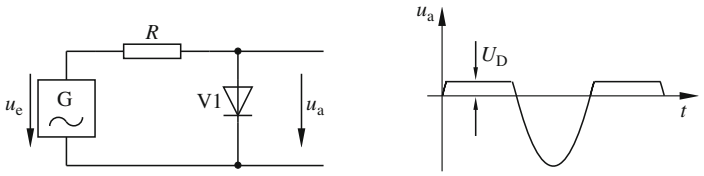


Abb. 33.21 Spannungsbegrenzung auf die Diodenschwellspannung U_D und zugehöriges Liniendiagramm

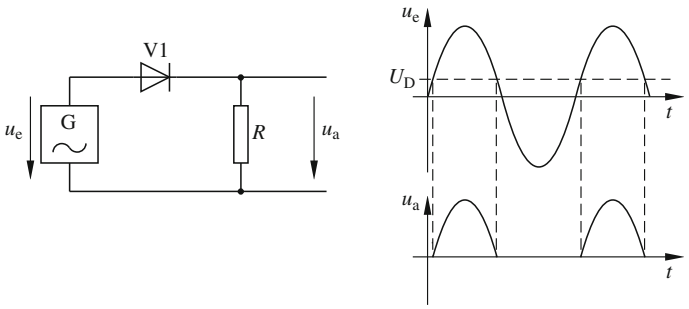


Abb. 33.22 Schaltung zur Restspannungsunterdrückung zugehöriges Liniendiagramm

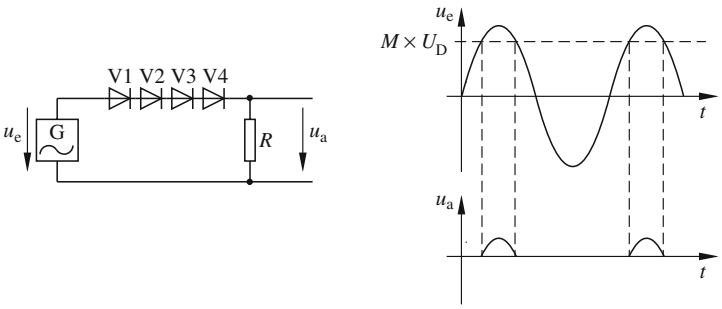


Abb. 33.23 Begrenzerschaltung mit zugehörigem Liniendiagramm

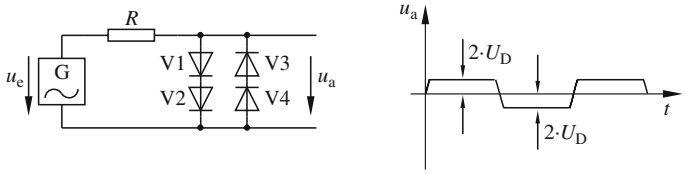


Abb. 33.24 Schaltung zur Begrenzung positiver und negativer Eingangssignale

Die sinusförmige Eingangsspannung u_e ist so klein, dass die Diodenschwellspannung U_D nicht vernachlässigbar klein ist. Abb. 33.21 zeigt eine Begrenzerschaltung, bei der die Ausgangsspannung u_a in Durchlassrichtung auf die Schwellspannung beschränkt ist. In Sperrrichtung liegt allerdings die Spannung u_e in voller Größe, also mit dem Scheitelwert, an. Die Schaltung dient zum Begrenzen positiver Spannungswerte auf den Wert der Diodenschwellspannung U_D, bei Reihenschaltung mehrerer Dioden auf die entsprechende höhere Spannungssumme.

In der Schaltung nach Abb. 33.22 sind Diode V1 und Widerstand R gegenüber Abb. 33.21 getauscht. Die Ausgangsspannung u_a ist gleich Null, solange die Diode noch gesperrt ist, so dass kein Strom fließt. Überschreitet die Eingangsspannung u_e den Wert der Diodenschwellspannung U_D, so fließt ein sinusförmiger Strom durch den Widerstand R und lässt eine proportionale Spannung abfallen. Bei kleinen Amplituden von u_e setzt der Stromfluss und damit die Ausgangsspannung u_a deutlich später als die Eingangsspannung ein. Da bei negativer Eingangsspannung u_e die Diode sperrt, ist die Ausgangsspannung u_a in der Zeit gleich Null.

Statt einer Diode sind in der Schaltung nach Abb. 33.23 mehrere in Reihe geschaltete Dioden V1 bis V4 eingesetzt, so dass außer der negativen Amplitude auch ein erheblicher Anteil (hier $4 \cdot 0{,}7\,\text{V}$) der positiven Spannungswerte unterdrückt wird.

Abb. 33.24 zeigt eine Schaltung, die sowohl positive als auch negative Eingangssignale u_e am Ausgang begrenzt. Die Ausgangs-

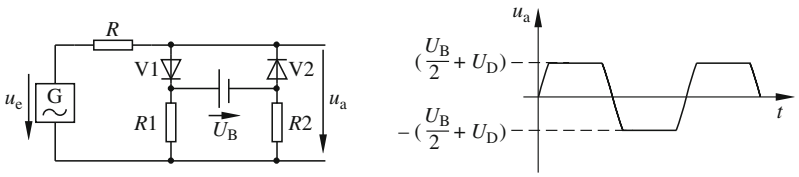


Abb. 33.25 Schaltung mit einstellbarem Begrenzungspegel der positiven und negativen Ausgangsspannung mit zugehörigem Liniendiagramm

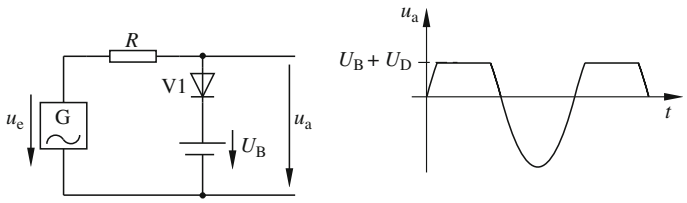


Abb. 33.26 Schaltung mit einstellbarem Begrenzungspegel der positiven Ausgangsspannung und zugehöriges Liniendiagramm

spannung u_a wird hier durch die antiparallele Reihenschaltung von jeweils 2 Dioden (V1, V2 und V3, V4) auf die doppelte Diodenschwellspannung U_D begrenzt.

Mit Hilfe der Schaltung nach Abb. 33.25 wird der Begrenzungspegel der Ausgangsspannung einstellbar gemacht. Die Spannung U_B wird durch die Widerstände R1 in zwei gleiche Teile aufgeteilt. Diese Teilspannungen spannen die Dioden V1 und V2 in Sperrrichtung vor, so dass immer dann eine der beiden Dioden durchlässig wird, wenn die Eingangsspannung u_e in positiver oder negativer Richtung den Wert $U_B/2 + U_D$ überschreitet. Ist die Spannung U_B einstellbar, kann der Begrenzungspegel variiert werden.

Abb. 33.26 zeigt eine Begrenzerschaltung, bei der die Diode V1 durch die Gleichspannungsquelle U_B in Sperrrichtung vorgespannt ist. Solange sie gesperrt ist, hat sie den größten Widerstand im Generator- und Batteriestromkreis. Die Diode wird leitend, sobald ihre Anode um den Betrag der Schwellspannung positiver ist als die Kathode. Dies ist der Fall, wenn die Eingangsspannung u_e mindestens einen positiven Wert von $U_B + U_D$ hat. Steigt u_e weiter an, bleibt die Ausgangsspannung u_a auf diesen Wert begrenzt.

33.5.2 Gleichrichter

Haupteinsatzgebiet der Klein- und Leistungsgleichrichterdioden ist die Gleichrichtung niederfrequenter Wechselspannungen, insbesondere von Netzspannungen.

Im Allgemeinen benötigen elektronische Schaltungen als Energieversorgung eine oder mehrere Gleichspannungsquellen. Bei höherem Energiebedarf ist die Verwendung von Batterien unwirtschaftlich. Darum erfolgt eine Speisung aus dem Wechselspannungsnetz (AC). Hierbei ist neben einer Abwärtstransformation auch eine Gleichrichtung der Wechselspannung zu Gleichspannung (DC) erforderlich. Halbleiterdioden sind hierfür wegen ihrer kleinen Durchlassspannung, ihrer hohen Belastbarkeit und ihres großen Verhältnisses zwischen Sperrwiderstand und Durchlasswiderstand sehr gut geeignet.

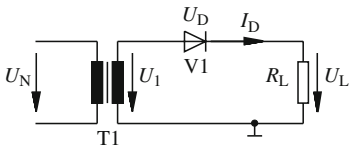


Abb. 33.27 Einweggleichrichterschaltung (M1) mit ohmscher Last

33.5.2.1 Einweggleichrichter (M1)

Bei der in Abb. 33.27 dargestellten Schaltung ist die Diode nur während der positiven Halbwellen der AC leitend. Die Schwellspannung U_D der Diode V1 sei vernachlässigbar klein gegenüber der Spitzenspannung U_{S1} der vom Transformator gelieferten Spannung U_1 ($= U_{RMS}$).

Daher fließt auch nur während dieser Zeitdauer ein Strom I_D durch den Lastwiderstand R_L. An diesem Widerstand kann aber nur dann eine Spannung auftreten, wenn ein Strom fließt. Der Spannungsverlauf am Lastwiderstand entspricht also einer pulsierenden DC, die nur aus positiven sinusförmigen Halbwellen besteht.

Während der negativen Halbwelle liegt die Diode in Sperrrichtung an dem Spitzenwert U_{S1} der AC, d. h., die Diode muss für $U_{S1} = U_{RRM} = (\sqrt{2} \cdot U_1)$ geeignet sein.

Die am Lastwiderstand R_L auftretende Spannung U_L ist nach Abb. 33.28 eine pulsierende Gleichspannung (DC). Wird sie mit einem Gleichspannungs-Messinstrument (Drehspulinstrument) gemessen, bildet dieses den arithmetischen Mittelwert U_{AVL} beziehungsweise den arithmetischen Mittelwert I_{AVL} aus der pulsierenden DC. Wird sie dagegen mit einem Dreheiseninstrument gemessen, bildet dieses den quadratischer Mittelwert (Effektivwert) U_{RMSL} aus der pulsierenden DC.

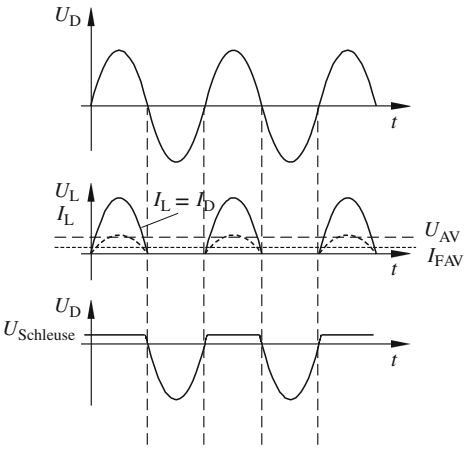


Abb. 33.28 Ausgangsspannung der Einweggleichrichterschaltung nach Abb. 33.27

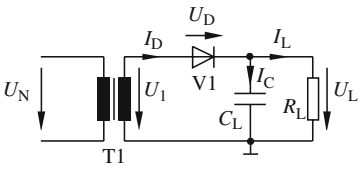


Abb. 33.29 Einweggleichrichterschaltung (M1) mit ohmscher Last und Ladekondensator

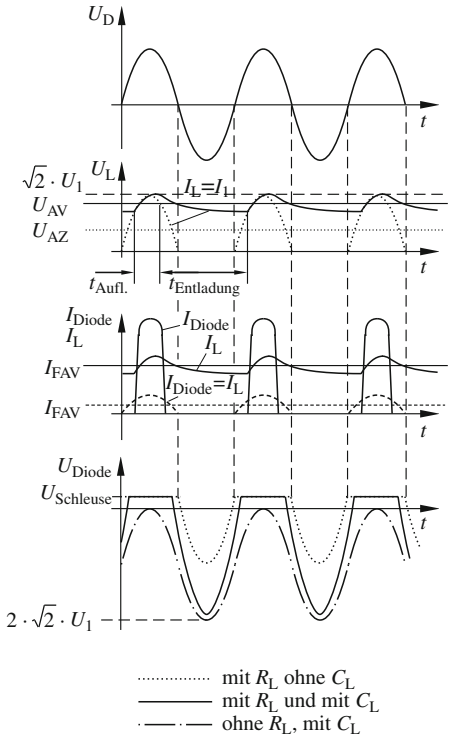


Abb. 33.30 Strom- und Spannungsverlauf für verschiedene Werte von R_L und C_L

Während der positiven Halbwelle der Trafospannung ist die Diode jeweils leitend, und es fließt durch die Diode ein periodischer Spitzendurchlassstrom

$$I_{FRM} = (\sqrt{2} \cdot I_D).$$

Auswahlkriterien:

U_{RRM} periodische Spitzensperrspannung
I_{FRM} periodischer Spitzendurchlassstrom

Die Ausgangsspannung U_L einer Einweggleichrichterschaltung pulsiert so stark, dass sie zur Energieversorgung von elektronischen Geräten nicht geeignet ist. Teilweise Abhilfe schafft die Schaltung nach Abb. 33.29.

Der Ladekondensator C_L wirkt in dieser Schaltung als Speicher, der jeweils in der positiven Halbwelle aufgeladen wird. Durch die Diode V1 fließt dann gleichzeitig der Ladestrom I_C für den Kondensator C_L und der Laststrom I_L. In den negativen Halbwellen, in denen die Diode sperrt, entlädt sich der Kondensator über den Lastwiderstand, so dass dieser auch in den Sperrzeiten der Diode von einem Strom durchflossen wird. Auf diese Art wird die pulsierende DC geglättet.

Im Abb. 33.30 sind die Strom- und Spannungsverläufe für verschiedene Fälle dargestellt.

Wie gut der Ladekondensator C_L diese Glättung bewirkt, hängt hauptsächlich von zwei Bedingungen ab:

1. Kapazität des Ladekondensators: Je kleiner seine Kapazität C_L ist, desto stärker entlädt er sich in der negativen Halbwelle. Damit wird die Welligkeit größer und der arithmetische Mittelwert U_{AVL} kleiner.
2. Größe des Lastwiderstandes R_L: Je hochohmiger der Lastwiderstand R_L ist, umso geringer ist der Strom, den er in den negativen Halbwellen aus dem Ladekondensator C_L zieht, das heißt, desto weniger wird der Ladekondensator entladen.

Tab. 33.2 Charakteristische Werte und Faktoren für einfache Gleichrichterschaltungen

Schaltungsart	M1	M2	B2
Leerlaufspannung U_{AVL}	$\sqrt{2} \cdot U_1$	$\sqrt{2} \cdot U_1$	$\sqrt{2} \cdot U_1$
Ausgangsgleichspannung U_{AV2}	$1{,}2 \cdot U_1$	$1{,}3 \cdot U_1$	$1{,}3 \cdot U_1$
Durchlassstrom I_{FAVM}	I_L	$0{,}5 \cdot I_L$	$0{,}5 \cdot I_L$
Max. Sperrspannung U_{RRM}	$2 \cdot (\sqrt{2} \cdot U_1)$	$2 \cdot (\sqrt{2} \cdot U_1)$	$\sqrt{2} \cdot U_1$
Faktor k (U_{Br})	$4{,}8 \cdot 10^{-3}$ s	$1{,}8 \cdot 10^{-3}$ s	$1{,}8 \cdot 10^{-3}$ s
Faktor k ($U_{Br\,ss}$)	$14 \cdot 10^{-3}$ s	$5 \cdot 10^{-3}$ s	$5 \cdot 10^{-3}$ s
Frequenz f_{Br}	50 Hz	100 Hz	100 Hz

Die Größe der mittleren Ausgangsgleichspannung U_{AVL}, aber auch die ihr überlagerte Brummspannung U_{Br}, ist sehr stark abhängig von der Kapazität des Ladekondensators und vom Laststrom. Die folgende Tab. 33.2 gilt unter der Voraussetzung, dass der Kondensator eine Kapazität von 1 mF je 1 mA Laststrom hat.

Unter der oben genannten Bedingung kann für die Ausgangsgleichspannung U_{AVL} mit hinreichender Genauigkeit nach Tab. 33.2 $U_{AVL} \approx 1{,}2 \cdot U_1$ angenommen werden.

Bei $R_L = \infty$ wird während der negativen Halbwellen kein Strom aus dem Kondensator gezogen, so dass für die Leerlauf-Ausgangsgleichspannung gilt

$$U_{AVL} = \sqrt{2} \cdot U_1 - U_D = U_{S1} - U_D$$

Für große Werte von U_1 kann die Spannung U_D vernachlässigt werden.

In einer Einweggleichrichterschaltung mit Ladekondensator ist die Diode nicht mehr während einer gesamten Halbwelle, sondern nach Abb. 33.30 nur noch während eines kurzen Zeitabschnittes leitend. Dieser Zeitabschnitt hängt von den Augenblickswerten der Spannungsdifferenz zwischen Kathode und Anode ab.

Die Nachladung des Ladekondensators beginnt jeweils während der positiven Halbwelle der Trafospannung zu dem Zeitpunkt, in dem die Spannung an der Anode um etwa 0,6 V größer wird als die Spannung an der Kathode. Die Nachladung wird beendet, sobald die Spannung an der Anode wieder unter die Spannung an der Kathode, also unter die Kondensatorspannung, abgesunken ist.

Da hier in sehr kurzer Zeit eine große Energie nachgeladen werden muss, ist der Ladestrom erheblich größer als der entnommene Gleichstrom. Darum geben die Hersteller in den Datenblättern oft einen Höchstwert für den Ladekondensator an.

Die maximal mögliche Sperrspannung im Leerlauf ist $U_{RRM} = 2 \cdot (\sqrt{2} \cdot U_1)$ nach Tab. 33.2, da der Ladekondensator dann auf $+U_{S1}$ aufgeladen ist, jedoch der negative Spitzenwert der Eingangsspannung U_1 den Wert $-U_{S1}$ hat. Abb. 33.30 zeigt diesen Umstand im unteren Teil auf.

Wegen der fortlaufenden Auf- und Entladung des Kondensators ist der mittleren Ausgangsgleichspannung U_{AV2} immer noch eine Wechselspannung (Restwelligkeit) überlagert, die nichtsinusförmig ist, aber eine Frequenz $f = 50\,\text{Hz}$ hat, wenn die Gleichrichterschaltung mit Netzfrequenz betrieben wird. Diese überlagerte Wechselspannung wird als *Brummspannung* U_{Br} bezeichnet, da sie zum Beispiel in einem Lautsprecher als tiefes Brummen gut hörbar ist.

Die am Ladekondensator abfallende Brummspannung lässt sich mit den in Tab. 33.2 angegebenen K-Faktoren hinreichend genau berechnen.

$$\textit{Brummspannung} \quad U_{Br} = K \cdot \frac{I_L}{C_L} \tag{33.11}$$

$$\textit{periodischer Spitzenstrom} \quad I_{FRM} \leq \frac{U_{AVL}}{\sqrt{R_i \cdot R_L}} \tag{33.12}$$

Der Wert der Brummspannung hängt also im Wesentlichen von der Kapazität des Ladekondensators, der Größe des Laststroms und einer schaltungsspezifischen Konstanten ab.

$$\textit{Einschaltspitzenstrom} \quad I_{FRME} \leq \frac{U_{S1}}{R_i} \tag{33.13}$$

Der periodische Spitzenstrom I_{FRM} lässt sich nach (33.12) berechnen, der Einschaltspitzenstrom I_{FRME} nach (33.13).

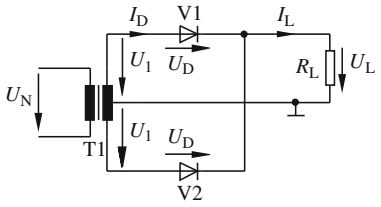


Abb. 33.31 Mittelpunktgleichrichterschaltung (M2) mit ohmscher Last

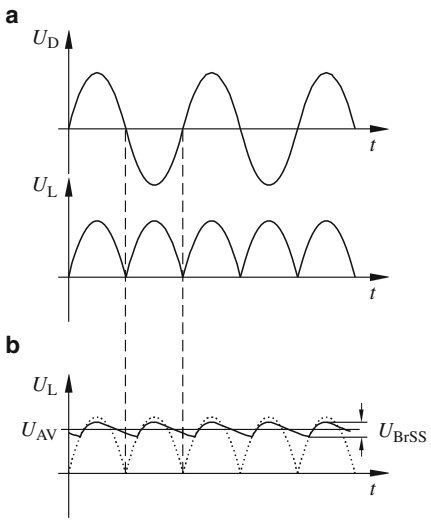


Abb. 33.32 Ausgangsspannung einer M2-Schaltung mit/ohne Ladekondensator

33.5.2.2 Mittelpunktschaltung (M2)

Um die Höhe der Brummspannung grundsätzlich zu verkleinern, werden nach Abb. 33.31 zwei Einweggleichrichter zu einer Mittelpunktgleichrichterschaltung zusammengefasst. Für einen solchen Zweiweggleichrichter wird ein Transformator mit einer Mittelanzapfung benötigt, die auf Masse gelegt wird und als Bezugspunkt dient.

Eine Analyse der Schaltung ergibt, dass die Diode D1 für die Dauer der positiven Halbwelle der Spannung U_1 leitend ist, während die Diode D2 gesperrt ist. In der negativen Halbwelle verhalten sich die Dioden umgekehrt. Jede Diode wird aber nur noch mit einem Diodenstrom I_D belastet, der dem halben Laststrom I_L entspricht.

Dies führt dazu, dass sowohl in der positiven als auch in der negativen Halbwelle ein Strom durch den Lastwiderstand R_L fließt. Da das in derselben Richtung geschieht, muss von DC gesprochen werden.

Die am R_L abfallende Spannung U_2 besteht nach Abb. 33.32 nur aus positiven sinusförmigen Halbwellen mit einer Frequenz $f = 100\,\text{Hz}$, das heißt, die negativen sinusförmigen Halbwellen werden einfach „nach oben„ geklappt.

Genau wie bei einem Einweggleichrichter kann auch hier nach der Schaltung in Abb. 33.33 ein Ladekondensator parallel zum

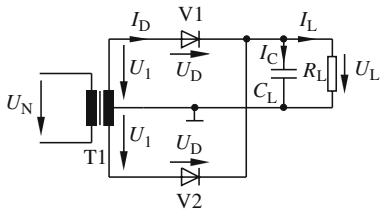


Abb. 33.33 Mittelpunktgleichrichterschaltung mit ohmscher Last und Ladekondensator

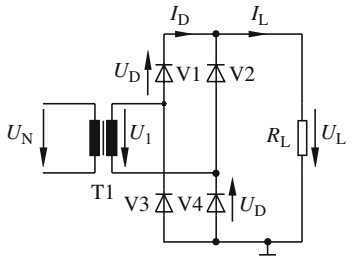


Abb. 33.34 Brückengleichrichterschaltung (B2) mit ohmscher Last

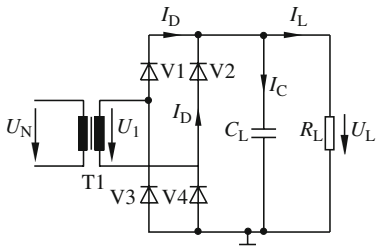


Abb. 33.35 Brückengleichrichterschaltung (B2) mit ohmscher Last und Ladekondensator

Lastwiderstand geschaltet werden, der die pulsierende DC entsprechend Abb. 33.32 weiter glättet.

Die Mittelpunktschaltung wird nur noch in Ausnahmefällen eingesetzt. Ein Transformator mit Mittelanzapfung ist teurer und hat wegen der zweiten Sekundärwicklung ein größeres Gewicht.

33.5.2.3 Brückengleichrichterschaltung (B2)

Die am häufigsten verwendete Gleichrichterschaltung ist die Brückenschaltung. Sie kommt mit einem normalen Transformator aus, benötigt aber vier Dioden.

In der Schaltung nach Abb. 33.34 kann man bei angenommener Polarität der Spannung U_1 den Strom I_L in seinem Flussverlauf verfolgen und so erkennen, dass der Laststrom I_L, der sich aus der Summe der beiden Diodenströme I_D bildet, stets in der gleichen Richtung durch den Lastwiderstand R_L fließt. Der Spannungsabfall am Lastwiderstand ist also eine pulsierende DC, die nur aus positiven sinusförmigen Halbwellen besteht und eine Frequenz $f = 100\,\text{Hz}$ hat. Die Ausgangsgrößen dieser Schaltung sind identisch mit denen der Mittelpunktschaltung.

Die pulsierende DC wird nach Abb. 33.35 durch die Parallelschaltung eines Ladekondensators C_L zum Lastwiderstand R_L

Tab. 33.3 Technische Daten der Baureihe B40C... Blockgleichrichter

Typ	U_1	I_{FRM}	Strombegrenzungswiderstand R_i	$C_{L\,max}$
B40C800	40 V	0,8 A	1 Ω	2500 µF
B40C1000	40 V	1,0 A	1 Ω	2500 µF
B40C1500/1000	40 V	1,5 A	1 Ω	2500 µF
B40C3200/2200	40 V	3,2 A/2,2 A	0,6 Ω	5000 µF
B40C5000/3300	40 V	5,0 A/3,3 A	0,5 Ω	10.000 µF

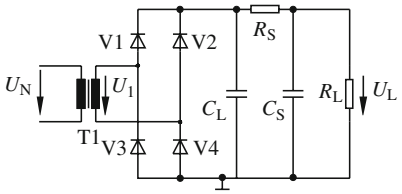


Abb. 33.36 B2-Schaltung mit RC-Siebglied

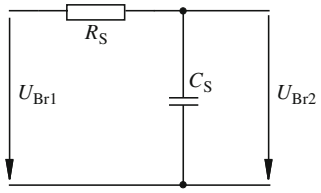


Abb. 33.37 Schaltung eines RC-Siebgliedes

geglättet. Bezüglich der Auflade- und Entladevorgänge ergeben sich ähnliche Verhältnisse wie bei der Mittelpunktschaltung. Die Auf- und Entladung des Kondensators erfolgt bei der Brückenschaltung in jeder Halbwelle. Auch bei der Brückengleichrichterschaltung mit Ladekondensator hat die Ausgangsspannung noch eine Restwelligkeit. Entsprechende Werte können der Tab. 33.2 entnommen werden.

Um den Aufbau von Gleichrichterschaltungen zu vereinfachen, bieten die Hersteller komplette Gleichrichtersätze in einem Kunststoff- oder Metallgehäuse an. Lieferbar sind verschiedene Baureihen. Der Einsatzbereich kann bei diesen Baureihen von Brückengleichrichtern direkt aus der Typenbezeichnung abgelesen werden.

So bedeutet B = Brückenschaltung. Der Buchstabe C bedeutet, dass der Gleichrichter für eine kapazitive Last, das heißt, für den Betrieb mit einem Ladekondensator vorgesehen ist. Bei Doppelwerten des Laststroms gilt der erste Wert für eine Chassismontage des Gleichrichters, der zweite Wert für eine freistehende Montage (Tab. 33.3).

Um die Brummspannung am Ausgang der oben beschriebenen Schaltungen zu senken, schaltet man an den Ausgang der Schaltung nach Abb. 33.35 einen Filter.

Der in die Schaltung nach Abb. 33.36 integrierte RC-Tiefpass reduziert die Brummspannung um den Glättungsfaktor g.

Der Kondensator in der RC-Tiefpass-Filterschaltung nach Abb. 33.37 stellt für höhere Frequenzen einen Kurzschluss beziehungsweise einen sehr geringen kapazitiven Blindwiderstand

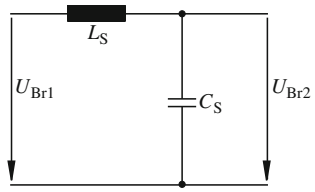


Abb. 33.38 Schaltung eines LC-Siebgliedes

dar. Die Schaltung kann als Spannungsteiler mit U_{Br2} als Ausgangsspannung betrachtet werden.

Glättungsfaktor

$$g = \frac{U_{Br1}}{U_{Br2}} = \sqrt{\omega_{Br}^2 \cdot R_S^2 \cdot C_S^2 + 1} \qquad (33.14)$$

Der RC-Filter nach Abb. 33.37 kann auch durch einen LC-Filter nach Abb. 33.38 ersetzt werden mit dem Vorteil einer effektiveren Filterung und geringen ohmschen Verlusten.

$$\textit{Glättungsfaktor} \quad g = \sqrt{\omega_{Br}^4 \cdot L_S^2 \cdot C_S^2 + 1} \qquad (33.15)$$

Ein Filter mit der Schaltung nach Abb. 33.19 hat zusätzlich den Vorteil, dass auch eine Stabilisierung der Ausgangsspannung vorgenommen wird, wenn vorausgesetzt werden kann, dass die Brummspannung kleine Amplituden aufweist.

$$\textit{Glättungsfaktor} \quad g = 1 + \frac{R_V}{r_Z} \qquad (33.16)$$

Wird mit einem Zwei-Kanal-Oszilloskop die Wechselspannung über dem Ladekondensator C_L und dem Siebkondensator C_S (Abb. 33.36) gemessen, also die Spannung U_L am Ausgang der Schaltung, so zeigen nach Abb. 33.39 beide Spannungen deutliches Einschwingverhalten.

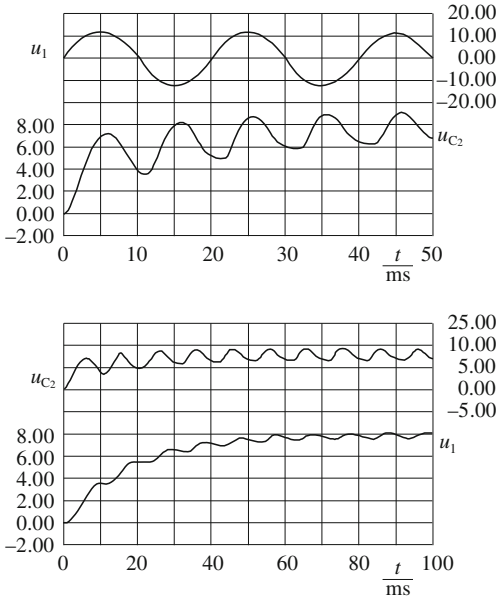


Abb. 33.39 Einschwingverhalten einer B2-Gleichrichterschaltung mit Ladekondensator und RC-Siebglied

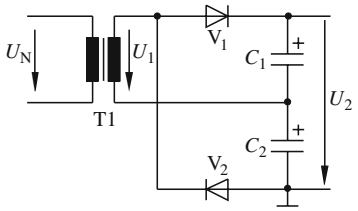


Abb. 33.40 Spannungsverdopplerschaltung (Delon-Schaltung)

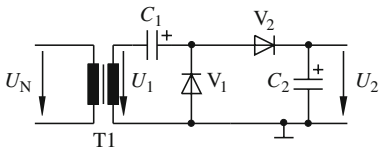


Abb. 33.41 Spannungsverdopplerschaltung (Villard-Schaltung)

Schaltungsdaten: Spannung $U_{S1} = 12\,\text{V}$, Ladekondensator $C_L = 10\,\mu\text{F}$, Siebwiderstand $R_S = 160\,\Omega$, Siebkondensator $C_S = 50\,\mu\text{F}$, Lastwiderstand $R_L = 1\,k\Omega$.

Nach ungefähr 100 ms befindet sich die Schaltung im „stationären Zustand„, die Schaltung ist eingeschwungen, und es zeigt sich am Ausgang eine Gleichspannung $U_{AVL} = 8\,\text{V}$, die von einer geringen Brummspannung überlagert ist.

33.5.3 Spannungsvervielfacher

Zum Betrieb von Oszilloskopen und Bildröhren oder Geiger-Müller-Zählrohren werden Gleichspannungen benötigt, die erheblich größer als die Netzspannungen sind und damit bereits im Bereich der Hochspannung liegen.

Da diese Spannungsquellen nur sehr kleine Lastströme liefern müssen, ist es meist günstiger, keine speziellen Hochspannungstransformatoren einzusetzen, sondern die hohen Gleichspannungen mit Hilfe von Spannungsverdoppler- oder Spannungsvervielfacherschaltungen zu erzeugen.

Die Delon-Schaltung besteht aus zwei Einwegschaltungen, die nach Abb. 33.40 zuammengeschaltet sind. Je nach Polarität der Spannung U_1 ist entweder die Diode V1 oder die Diode V2 in Durchlassrichtung geschaltet. Infolgedessen werden die Kondensatoren C_1 und C_2 mit der angegebenen Polarität geladen. Die Ladespannungen beider Kondensatoren liegen in Reihe, so dass sie sich zur Gesamtspannung $U_2 = U_{AV2}$ (im unbelasteten Zustand) addieren.

$$\textit{Gesamtspannung} \quad U_{AV2} = U_{C1} + U_{C2} = 2 \cdot U_{S1} \qquad (33.17)$$

Diese Schaltung ist also eine Spannungsverdopplerschaltung im Leerlaufbetrieb.

Abb. 33.41 stellt eine Villard-Schaltung dar. Bei entsprechender Polarität der Spannung U_1 liegt die Diode V1 in Durchlassrichtung und lädt den Kondensator C_1 in der eingezeichneten

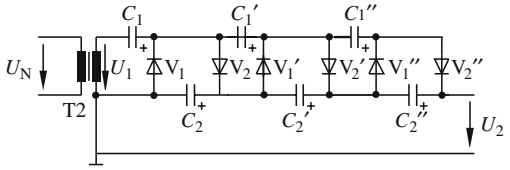


Abb. 33.42 Spannungsvervielfacherschaltung nach Villard

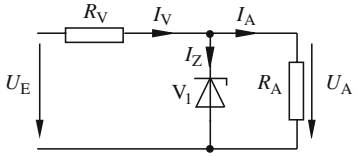


Abb. 33.43 Grundschaltung einer Spannungsstabilisierung mit Z-Diode

Polarität auf den Wert $U_{C1} = U_{S1}$. In der nächsten Halbwelle liegt die Diode V1 in Sperrrichtung. An V1 liegt dann die Summe aus $U_{S1} + U_{C1} = 2 \cdot U_{S1}$ als Sperrspannung an. Diese Spannung ist jetzt die Speisespannung für die zweite Gleichrichterstufe. Über V2 wird der Kondensator C_2 dann auf die Spannung $2 \cdot U_{S1}$ aufgeladen.

Die Dioden müssen eine periodische Spitzensperrspannung $U_{RRM} = 2 \cdot U_{S1}$ in beiden Schaltungen ermöglichen, während der Strom $I_{FM} = I_L$ beträgt. Beiden Schaltungen können nur kleine Lastströme entnommen werden. Die Belastung der Schaltungen erfolgt in der Praxis so, dass die Ausgangsspannung minimal sinkt.

maximale Ausgangsgleichspannung

$$U_{AV2} \sim 0{,}8 \ldots 0{,}9 \cdot (2 \cdot U_{S1}) \qquad (33.18)$$

Die Villard-Schaltung lässt sich zur Vervielfacherschaltung erweitern. Werden nach Abb. 33.42 zwei oder mehr Villard-Schaltungen hintereinander geschaltet (hier drei Villard-Schaltungen), so entsteht eine Hochspannungskaskadenschaltung. Für die Ausgangsspannung U_2 einer unbelasteten Kaskadenschaltung gilt (33.17), wobei n die Anzahl der aneinandergereihten Villard-Schaltungen bedeutet.

maximale Ausgangsgleichspannung

$$U_{AV2} \approx n \cdot 2 \cdot U_{S1} \qquad (33.19)$$

Die Dioden müssen eine periodische Spitzensperrspannung $U_{RRM} = 2 \cdot U_{S1}$ ermöglichen, während der Strom $I_{FM} = n \cdot I_L$ beträgt. Es sind nur geringe Lastströme möglich. Kaskadenschaltungen mit vielen Stufen zeigen deutlich ein Hochlaufen der Ausgangsspannung, weil die einzelnen Kondensatoren nacheinander aufgeladen und die Ladungen von einem Kondensator auf den nächsten verschoben werden.

33.5.4 Diode als Konstantspannungsquelle (Z-Diode)

Abb. 33.43 zeigt die Grundschaltung für die Spannungsstabilisierung mit Hilfe einer Z-Diode. Als stabilisierte Ausgangsspannung U_A tritt die an der Z-Diode liegende Spannung U_Z auf. Im zulässigen Arbeitsbereich bleibt sie sowohl bei einer Änderung der Eingangsspannung U_E als auch bei einer Änderung des Laststroms I_A nahezu konstant ($I_V = I_Z + I_A = \text{const}$).

Dieses kann aber nur eintreten, wenn der Diodenstrom I_Z jeweils um den gleichen Betrag kleiner wird wie der Laststrom I_A größer. Der Arbeitspunkt der Z-Diode wandert daher nach Abb. 33.44a auf der Kennlinie in Richtung auf A2.

Infolge der steilen Kennlinie der Z-Diode ändert sich die Ausgangsspannung U_A nur um einen kleinen Betrag, das heißt, U_Z bleibt nahezu konstant, wenn sich der Laststrom I_A ändert. Wird dabei jedoch $I_Z < I_{Z\,min}$, so wandert der Arbeitspunkt in den waagerechten Bereich der Kennlinie, so dass sich U_A sehr stark ändert, das heißt, die stabilisierende Wirkung der Z-Diode ist beendet ($I_{Z\,min} \approx 0{,}1 \cdot I_{Z\,max}$, $I_{A\,max} \approx 0{,}9 \cdot I_{Z\,max}$).

Ändert sich die Eingangsspannung U_E um den Betrag $\pm\Delta U_E$ bei konstantem Laststrom I_A nach Abb. 33.44b, so verschiebt sich der Arbeitspunkt A0 aufgrund der Parallelverschiebung der Arbeitsgeraden nach A1 oder A2, je nachdem, ob es sich um eine Spannungserhöhung oder -minderung handelt. Aufgrund der steilen Kennlinie ändert sich U_Z deutlich weniger als U_E, so dass die Ausgangsspannung $U_A = U_Z$ nahezu gleich bleibt ($U_E \approx 2 \ldots 4 \cdot U_A$, $U_{E\,min} \approx 1{,}2 \ldots 2 \cdot U_A$).

Beim praktischen Einsatz der Stabilisierungsschaltung nach Abb. 33.43 überlagern sich beide Fälle, so dass eine exakte Trennung kaum noch möglich ist. Der gewählte Vorwiderstand R_V muss beiden Erwartungen gerecht werden. Er lässt sich näherungsweise mit der Gleichung II.18 berechnen. Zu verwenden ist dann ein Widerstandswert nach der E-Reihe, der zwischen den beiden Werten, aber näher an $R_{V\,max}$ liegt.

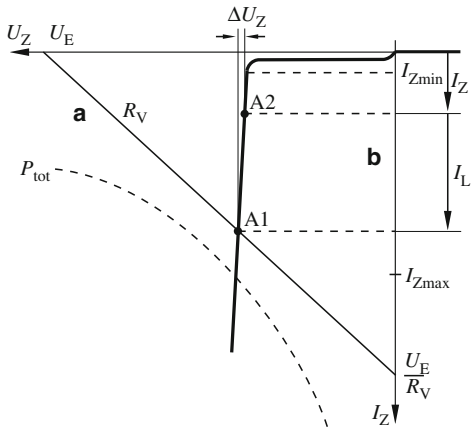


Abb. 33.44 Ermittlung der Arbeitspunktverschiebung. **a** konstante Eingangsspannung und variable Last, **b** konstante Last und schwankende Eingangsspannung

Teil V

Vorwiderstand

$$R_{V\,\min} \approx \frac{U_{E\,\max} - U_Z}{I_{Z\,\max} + I_{A\,\min}} \qquad R_{V\,\max} \approx \frac{U_{E\,\min} - U_Z}{I_{Z\,\min} + I_{A\,\max}} \tag{33.20}$$

Die Stabilisierung wird bei einer vorgegebenen Z-Diode umso besser, je größer die Eingangsspannung gewählt wird. Die Verlustleistung am Vorwiderstand R_V nimmt allerdings schneller zu als der relative Stabilisierungsfaktor S.

Es hat sich als zweckmäßig erwiesen, das Verhältnis der Relativwerte der Schwankungen von Eingangs- und Ausgangsspannung als Stabilisierungsfaktor S einzuführen.

Stabilisierungsfaktor

$$S = \frac{\frac{\Delta U_E}{U_E}}{\frac{\Delta U_A}{U_A}} = \left(1 + \frac{R_V}{r_Z}\right) \cdot \frac{U_A}{U_E} \tag{33.21}$$

Der differentielle Widerstand r_Z einer Z-Diode wird für bestimmte Stromwerte in den Datenblättern der Hersteller genannt. Auch die Z-Spannung U_{ZO} wird für konkrete Werte des Z-Stroms I_Z angegeben. Damit ergibt sich die Möglichkeit, eine Ersatzschaltung für die Z-Diode in der Schaltung nach Abb. 33.43 zu finden. Die Schaltung nach Abb. 33.45 zeigt eine solche Ersatzschaltung, in der die Z-Diode als Spannungsquelle mit der Quellenspannung U_{ZO} und dem Innenwiderstand $R_i = r_Z$ verwendet wird. Mit ihrer Hilfe lässt sich der Ausgangswiderstand einer solchen Stabilisierungsschaltung bestimmen.

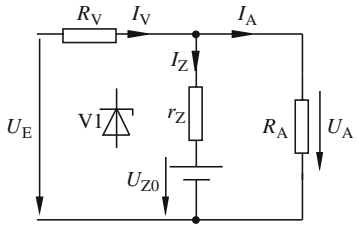
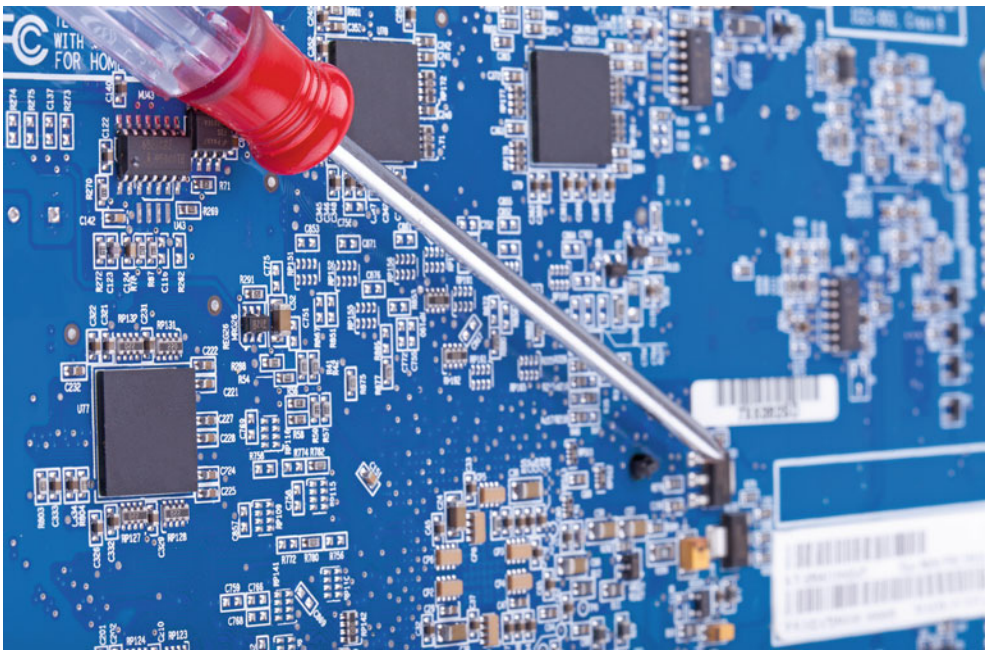


Abb. 33.45 Ersatzschaltung einer Z-Diode mit Vorwiderstand und ohmscher Last

Mehrschichtdioden und -trioden

34

Peter Döring



34.1 Vierschichtdioden 416
34.2 Thyristoren 417
34.3 Diac 420
34.4 Triac 421
34.5 Schutz der Dioden und Trioden 421
34.6 Zündmethoden 422

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_34

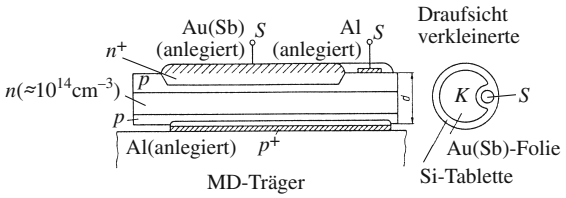


Abb. 34.1 Realer Aufbau des Thyristors

Silizium-Gleichrichterdioden mit „pin"-Übergang werden wegen ihrer hohen Durchlassströme und verhältnismäßig hohen Sperrspannungen überwiegend in industriellen Anlagen zur Gleichrichtung der Netzwechselspannung eingesetzt. Für zahlreiche Anwendungen werden aber regelbare, steuerbare Gleichspannungen gefordert.

Bereits im Jahre 1958 wurde ein steuerbarer Silizium-Gleichrichter (**SCR** = **S**ilicon-**c**ontrolled-**r**ectifier) entwickelt, der ähnliche Eigenschaften zeigte wie die erwähnte Gleichrichterzelle. Dieses elektronische Bauelement erhielt die Bezeichnung „Thyristor".

Heute geht man bei der Herstellung von Thyristoren von einem hochohmigen n-dotierten Siliziumplättchen aus, in das von beiden Seiten Akzeptoren (B, Al, Ga) eindiffundiert werden. Anschließend wird als Kathode eine Sb-dotierte Au-Folie an der Oberseite einlegiert, wodurch dort ein stark dotiertes n-Gebiet (n^+) entsteht. Auch die Al-Kontakte für die Steuerelektrode und die Anode werden anlegiert. Die entstandene „Silizium-Tablette" nach Abb. 34.1 wird anodenseitig auf einen Molybdänträger aufgebracht. Die Tablette ist meist rund, da so das elektrische Feld an den Rändern besser zu handhaben ist.

Unter dem Sammelbegriff „Thyristoren" werden häufig die Mehrschicht-Halbleiter zusammengefasst, die einfach Mehrschichtdioden und -trioden genannt werden. Sie sind jeweils aus drei oder mehr pn-Übergängen aufgebaut und haben ein ausgeprägtes Schalt- beziehungsweise Kippverhalten. Diese Mehrschicht-Halbleiter besitzen zwei Schaltzustände, und zwar den Schaltzustand „Ein" = leitend und den Schaltzustand „Aus" = gesperrt.

Das Umschalten vom gesperrten in den leitenden Zustand wird als „Zünden" und das Umschalten vom leitenden in den gesperrten Zustand als „Löschen" bezeichnet. Sowohl der Zündvorgang als auch der Löschvorgang laufen in sehr kurzer Zeit ab.

Bei den Bauelementen mit *Einrichtungs-Betrieb* (Vierschichtdiode, Thyristor) werden die beiden Hauptanschlüsse wie bei der Diode mit Anode und Kathode bezeichnet. Ein *Stromfluss* ist nur möglich, wenn die Anode positiv gegenüber der Kathode ist. Bei den Bauelementen mit Zweirichtungs-Betrieb (Diac, Triac) werden die Hauptanschlüsse dagegen Anode 1 (A1) und Anode 2 (A2) genannt.

Thyristor und Triac haben einen Steueranschluss, der als Gate (G) bezeichnet wird. Die Zündung erfolgt mit einem Steuerstrom über das Gate und löst einen wesentlich größeren Hauptstrom durch das Bauelement aus.

Von großer Bedeutung ist ihr Einsatz als Leistungsschalter. Durch die gezielte und vermehrte Anwendung von Thyristoren hat sich inzwischen ein eigenes Fachgebiet der Elektronik, die Leistungselektronik, entwickelt. Triacs haben wegen ihres Zweirichtungs-Betriebes (aber relativ geringem Durchlassstrom) und ihrer recht einfachen Ansteuermöglichkeit zum Beispiel über Diacs eine weite Verbreitung in Phasenanschnittschaltungen zur Leistungssteuerung von Lampen (Helligkeit) und zur Drehzahlsteuerung von Kleinmotoren gefunden.

34.1 Vierschichtdioden

Vierschichtdioden haben nach Abb. 34.2 eine pnpn-Zonenfolge in einem Silizium-Kristall. An den beiden äußeren Zonen sind die Anschlüsse sperrschichtfrei angebracht.

Beim Anlegen einer kleinen Spannung U_D in der eingezeichneten Polarität (Anode positiver als die Kathode) liegen die Sperrschichten I und III in Durchlassrichtung, während die Sperrschicht II in Sperrrichtung gepolt ist. An dieser Schicht fällt dementsprechend die gesamte anliegende Spannung U_D ab. Die Vierschichtdiode befindet sich im Blockierbereich (hochohmiger Bereich), so dass lediglich ein sehr kleiner Sperrstrom I_D durch das Bauelement fließt.

Wird die Spannung U_D weiter erhöht, so steigt die Feldstärke in der Sperrschicht II an. Bei einem bestimmten Spannungswert, der Nullkippspannung $U_{(\mathrm{BO})\,O}$, wird die Feldstärke jedoch so groß, dass ein Durchbruch an Sperrschicht II auftritt und plötzlich ein stark ansteigender Strom I_T durch die Vierschichtdiode fließt. Dieser Vorgang wird als „Zünden" bezeichnet. Der Strom I_T muss durch einen Vorwiderstand R_V begrenzt werden.

Die Vierschichtdiode geht in den Durchlassbereich (niederohmiger Bereich) über, und die Spannung U_D an der Diode sinkt plötzlich auf einen kleinen Wert ab. Diese Durchlassspannung U_T beträgt etwa 0,5 ... 1,2 V. Der Durchlassbereich einer Vierschichtdiode, also ihr weiterer Kennlinienverlauf nach dem Zünden, ähnelt dem der normalen Silizium-Diode.

Die Vierschichtdiode kippt in den Blockierbereich zurück, wenn der Strom I_T durch die Diode den Haltestrom I_H unterschreitet. Dieser Zusammenhang wird als „Löschen" bezeichnet.

Im Sperrbereich der Vierschichtdiode, also im 3. Quadranten der Kennlinie nach Abb. 34.3, kann man erkennen, dass erst

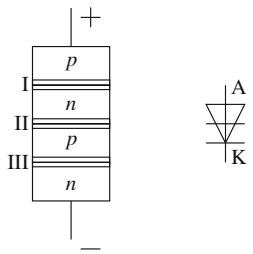


Abb. 34.2 Zonenfolge in einem Silizium-Kristall

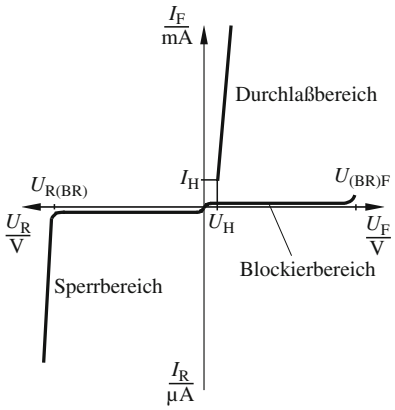


Abb. 34.3 Kennlinie einer Vierschichtdiode

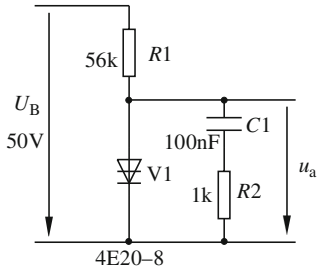


Abb. 34.4 Sägezahngenerator mit Vierschichtdiode

bei einer sehr hohen Durchbruchspannung $U_{(BR)\,R}$ der Durchbruchstrom zum Fließen gelangt. Bei dieser Polarität sind die Sperrschichten I und III in Sperrrichtung gepolt.

Eine Vierschichtdiode „zündet“ also nur, wenn sie in Vorwärtsrichtung betrieben wird und die anliegende Spannung U_D größer als die Kippspannung $U_{(BO)\,O}$ wird. Auch bei einer Verringerung der Spannung U_D bleibt die Diode leitend, sie „löscht“ erst wieder, wenn der Haltestrom I_H unterschritten wird.

Kennwerte: Schaltspannung $U_{(BO)\,O}$ ($\approx 20 \ldots 200$ V), Haltestrom I_H ($\approx 1 \ldots 45$ mA), Schaltzeit (100 bis 600 ns)

Die Schaltung in Abb. 34.4 zeigt einen Sägezahngenerator mit Vierschichtdiode. Der Kondensator C wird über die Widerstände R1 und R2 aufgeladen, bis dieser den Wert $U_{(BO)\,O}$ der Vierschichtdiode erreicht. Nach der erfolgten Zündung entlädt sich C sehr schnell über die Diode und den Widerstand R2, der den Entladestrom begrenzt. Sobald bei der Entladung der Haltestrom I_H unterschritten wird, kippt die Diode in den Sperrzustand zurück, und es beginnt ein neuer Ladevorgang.

Die sägezahnförmige Ausgangsspannung nach Abb. 34.5 kann linearer gemacht werden, indem man das Verhältnis von U_B zu $U_{(BO)\,O}$ größer macht. In dem Falle wird nur der nahezu lineare Beginn der nach der e-Funktion verlaufenden Ausgangsspannung benutzt. Der gleiche Effekt wird erzielt, wenn die Spannungsquelle durch eine Konstantstromquelle ersetzt wird.

$$\textit{Lastwiderstand} \quad \frac{U_B}{I_H} \geq R_1 \geq \frac{U_B}{I_{T\,\max}} \qquad (34.1)$$

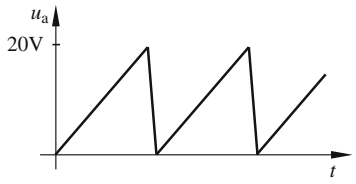


Abb. 34.5 Liniendiagramm der sägezahnförmigen Ausgangsspannung

Ist der Lastwiderstand der Vierschichtdiode zu hochohmig, kann nicht der erforderliche Haltestrom I_H fließen. Es kommt nicht zum dauerhaften Zünden der Diode. Eine rechnerische Einschätzung ermöglicht (34.1).

34.2 Thyristoren

Die Einrichtungs-Thyristortrioden werden allgemein als Thyristoren bezeichnet und sind generell aufgebaut wie Vierschichtdioden. Das einzig neue ist das als Steueranschluss zusätzlich herausgeführte Gate (G).

Der Anschluss ist nach Abb. 34.6 sowohl an der innenliegenden p-Schicht (kathodenseitig gesteuerter Thyristor) als auch an der innenliegenden n-Schicht (anodenseitig gesteuerter Thyristor) möglich. Auch zwei Gateanschlüsse sind möglich, jedoch wird die Thyristortetrode selten verwendet. Die wichtigste Bauart ist der kathodenseitig gesteuerte, rückwärtssperrende Thyristor.

Unabhängig von der Polarität der zwischen Anode und Kathode liegenden Spannung ist bei kleinen Spannungswerten mindestens ein pn-Übergang in Sperrrichtung. Der Thyristor ist daher sowohl im Vorwärtsbetrieb (Anode positiv gegenüber Kathode) als auch im Rückwärtsbetrieb (Anode negativ gegenüber Kathode) zunächst gesperrt. Wird im Vorwärtsbetrieb die Spannung weiter erhöht, kippt der Thyristor genau wie die Vierschichtdiode bei $U_{(BO)\,O}$ schlagartig in den leitenden Zustand. Dieser Vorgang ist zwar technisch zulässig, aber im Sinne einer Steuerung nicht erwünscht.

$$\textit{maximale Betriebsspannung} \quad U_{DRM} \leq \frac{2}{3} U_{(BO)\,O} \qquad (34.2)$$

Um ein Überkopfzünden zu vermeiden, darf die maximale Betriebsspannung des Thyristors die Nullkippspannung nicht erreichen. Für ausreichende Sicherheit kann gesorgt werden, indem ein Thyristortyp mit höherer Nullkippspannung verwendet wird.

In Rückwärtsrichtung verhält sich ein Thyristor wie eine gesperrte Diode. Solange die maximal zulässige Sperrspannung

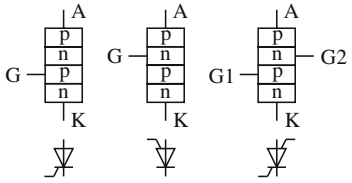


Abb. 34.6 Anschlüsse der Einrichtungs-Thyristortrioden

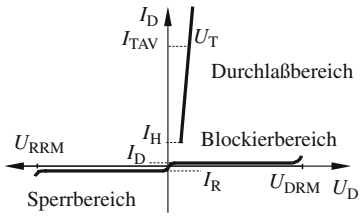


Abb. 34.7 Charakteristische Kennlinie eines Thyristors

$U_{(BR)R}$ nicht überschritten wird, fließt nur ein sehr kleiner Sperrstrom I_R. Wird dieser Spannungswert überschritten, so steigt der Sperrstrom lawinenartig an, und der Thyristor wird zerstört. Die charakteristische Kennlinie eines Thyristors nach Abb. 34.7 ähnelt also stark der Kennlinie einer Vierschichtdiode.

Thyristoren können durch Ansteuerung des Gates gezündet werden, wenn die Anode positiver ist als die Kathode. Bei den kathodenseitig gesteuerten Thyristoren ist eine positive Gatespannung U_G erforderlich, die einen Gatestrom I_G in den Thyristor hineinfließen lässt. Der Zündvorgang hängt wesentlich von der Größe des jeweiligen Gatestroms I_G ab. Durch ihn gelangen Ladungsträger in die Halbleiterzonen, und infolge der gegenseitigen Beeinflussung der pn-Übergänge kippt der Thyristor bei niedrigeren Werten von U_D in den leitenden Zustand. Den Zusammenhang zwischen Kippspannung $U_{(BO)O}$ und Zündstrom I_{GT} zeigt das Diagramm nach Abb. 34.8. Diese Art der Zündung bezeichnet man als Vertikalsteuerung. Sie wird selten angewendet, da sich der Zündzeitpunkt nur ungenau einstellen lässt.

Um den Zündzeitpunkt genauer einstellen zu können, geben die Hersteller Zünddiagramme für Thyristoren an, die den Zusammenhang zwischen der Gatespannung und dem Gatestrom beschreiben. Diese weisen nach Abb. 34.9 drei wichtige Zonen auf:

- Unterhalb der Zündspannung U_{GT} ist eine Zündung nicht möglich, weil die erforderliche Zündenergie nicht zur Verfügung steht.

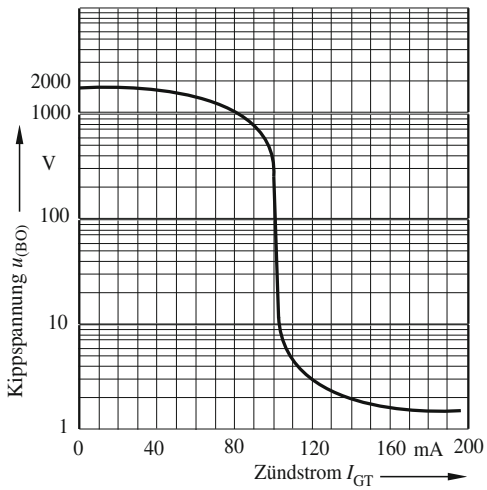


Abb. 34.8 Zündkennlinie eines Thyristors zur Vertikalsteuerung

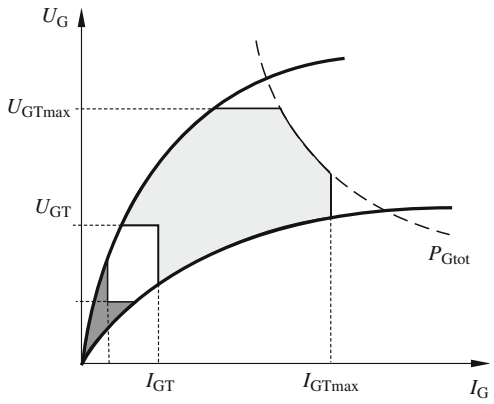


Abb. 34.9 Eingangskennlinie und Zünddiagramm eines Thyristors

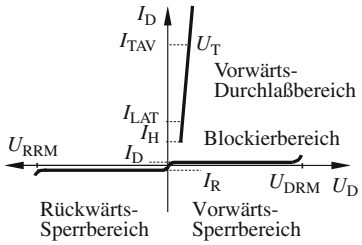


Abb. 34.10 Darstellung der wichtigsten Kenndaten eines Thyristors

- Oberhalb der Zündspannung U_{GT} liegt der Bereich, in dem zwar eine Zündung erfolgen kann, die aber von mehreren zusätzlichen Faktoren abhängt, wie zum Beispiel der anliegenden Durchlassspannung, der Sperrschichttemperatur, u. a.
- Sichere Zündung ist gewährleistet, wenn die Gatespannung größer als U_{GT} und der Gatestrom größer als I_{GT} sind. Begrenzt wird dieser Bereich durch die Grenzwerte U_{GTM}, I_{GTM} und die maximal zulässige Gate-Verlustleistung P_{Gtot}.

Das Zündverhalten hängt relativ stark von der Sperrschichttemperatur des Thyristors ab. Je höher die Temperatur, desto kleiner werden die zur sicheren Zündung erforderliche obere Zündspannung U_{GT} und der obere Zündstrom I_{GT}. Nach dem Zünden kippt der Thyristor sehr schnell vom Blockierbereich in den Durchlassbereich. Die Durchlasskennlinie ähnelt der von Silizium-Dioden, jedoch ist die Durchlassspannung etwas größer. In den Datenblättern werden die Strom- und Spannungswerte bei gezündeten Thyristoren (Durchlasszustand) als I_T und U_T anstelle von I_D und U_D bezeichnet. Die wichtigsten Kenndaten von Thyristoren werden im Kennlinienfeld nach Abb. 34.10 dargestellt.

Der hohe Durchlassstrom stellt genügend Ladungsträger zur Verfügung, um die pn-Übergänge im leitfähigen Zustand zu halten, wenn der Einraststrom I_{LAT} kurzzeitig überschritten wird. Um den Thyristor wieder in den Sperrzustand zu bringen, muss der Durchlassstrom kleiner als der Haltestrom I_H werden. Dieser Strom hängt vom jeweiligen Thyristortyp ab und liegt zwischen einigen mA und etwa 100 mA. Auch der Haltestrom ist stark temperaturabhängig.

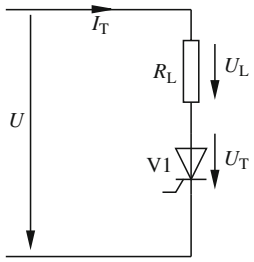


Abb. 34.11 Thyristor mit Lastwiderstand an Wechselspannung

Wird ein Thyristor nach Abb. 34.11 mit sinusförmiger Wechselspannung betrieben, so erfolgt ein Löschen bei jedem Nulldurchgang der Betriebsspannung, weil dadurch zwangsläufig der Haltestrom unterschritten wird. Der Thyristor muss also nach jedem positiven Nulldurchgang erneut gezündet werden.

Beim Betrieb mit Gleichspannung wird üblicherweise der Thyristor gelöscht, indem man einen kurzen negativen Spannungsimpuls an die Anode gibt. Dadurch wird der Thyristor kurzzeitig im Rückwärts-Sperrbereich betrieben und somit gelöscht.

Das Löschen über das Gate ist bei den meisten Thyristoren nicht möglich (Ausnahme: GTO-Thyristor).

Bei Wechselspannungsbetrieb ist der Thyristor in der negativen Halbwelle stets gesperrt, und eine Zündung ist nur in der positiven Halbwelle möglich. Der Zündvorgang kann jeweils bei einem bestimmten Phasenwinkel α der anliegenden Wechselspannung erfolgen. Abb. 34.12 zeigt hier die entsprechenden Zusammenhänge auf.

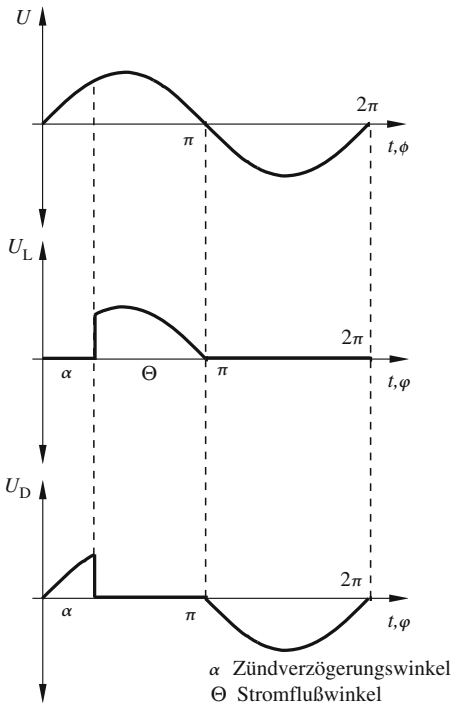


Abb. 34.12 Liniendiagramm einer sinusförmigen Spannung am Thyristor

Der Thyristor wird jeweils bei $\alpha = 90°$ gezündet und automatisch bei $\omega t = 180°$, also beim nächsten Nulldurchgang, gelöscht.

Ein Strom fließt also nur während $180° \geq \omega t \geq 90°$ durch den Thyristor. Diese Zeitspanne wird auch als Stromflusswinkel Θ bezeichnet. Der Stromflusswinkel Θ wird zur Berechnung des arithmetischen Mittelwertes des Stroms I_{TAV} benötigt, aber natürlich auch zur Berechnung des quadratischen Mittelwertes. Um diese Berechnung zu vermeiden, wird der Zusammenhang zwischen Verlustleistung, Stromflusswinkel Θ und Strom I_{TAV} in Diagrammen angegeben.

Die hier gezeigte Leistungssteuerung wird als *Phasenanschnittsteuerung* bezeichnet und ermöglicht eine stufenlose Leistungssteuerung.

Grenzdaten: periodische Spitzensperrspannung U_{DRM} und U_{RRM}; maximale Stoßspitzenspannung U_{DSM} und U_{RSM}; Dauergrenzstrom I_{TAV}; maximaler periodischer Spitzenstrom I_{TRM}; maximaler Durchlassstrom I_{TRMS}.

Kenndaten: Nullkippspannung $U_{(BO)O}$; Sperrströme I_D, I_R; Durchlassspannung U_T, Haltestrom I_H; Einraststrom I_{LAT}; Zündstrom I_{GT}; Zündspannung U_{GT}.

Beim Betrieb des Thyristors sind dynamische Kennwerte zu beachten. Der zum Zünden erforderliche Gatestrom muss mindestens so lange fließen, bis die Zündung erfolgt ist. Diese Zeit wird als Zündzeit t_{gt} bezeichnet. Zwischen dem Löschzeitpunkt im Nulldurchgang des Durchlassstroms und der Wiederkehr einer Spannung in Vorwärtsrichtung muss eine Mindestwartezeit liegen, die Freiwerdezeit oder Erholzeit t_q genannt wird.

Aber auch ein zu schneller Strom- oder Spannungsanstieg im Zünd- beziehungsweise Löschzeitpunkt kann zu unkalkulierbaren Zuständen einschließlich der Zerstörung des Thyristors führen (Strom- und Spannungssteilheit).

Durch entsprechende Schutzbeschaltungen des Thyristors (siehe Abschn. 34.5) kann dem entgegengewirkt werden.

Thyristoren werden in großem Umfang als Leistungsschalter und zur Leistungssteuerung in Gleich-, Wechsel- und Drehstromkreisen eingesetzt.

Mit einem Thyristor als Steuerelement nach Abb. 34.11 ist nur Halbwellenbetrieb möglich, das heißt, nur die positiven Halbwellen (max. Stromflusswinkel $\Theta = 180°$) gelangen an die Last und erbringen maximal eine Halbierung der wirksamen Leistungsaufnahme. Eine Verringerung des Stromflusswinkels bedeutet eine weitere Verringerung der Leistung.

Eine Grundschaltung für den Vollwellenbetrieb nach Abb. 34.13 ist die Antiparallelschaltung zweier Thyristoren. Hier ist zu beachten, dass die Gatespannungen U_{G1} und U_{G2} jeweils potentialfrei zugeführt werden müssen. Dies erfolgt in modernen Steuerungen mit Hilfe von Optokopplern, aber auch mit Übertragern. Die Synchronisation muss so erfolgen, dass jeweils mit dem gleichen Phasenanschnittwinkel α der V1 in der positiven und der V2 in der negativen Halbwelle der Betriebsspannung gezündet wird. Im Liniendiagramm nach Abb. 34.14 wird ein Zündwinkel $\alpha = 90°$ angenommen. Bei Vollwellenbetrieb ist

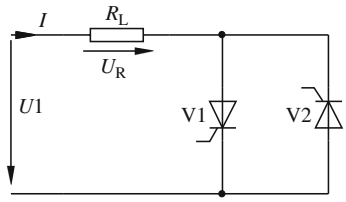


Abb. 34.13 Antiparallelschaltung zweier Thyristoren für Vollwellenbetrieb

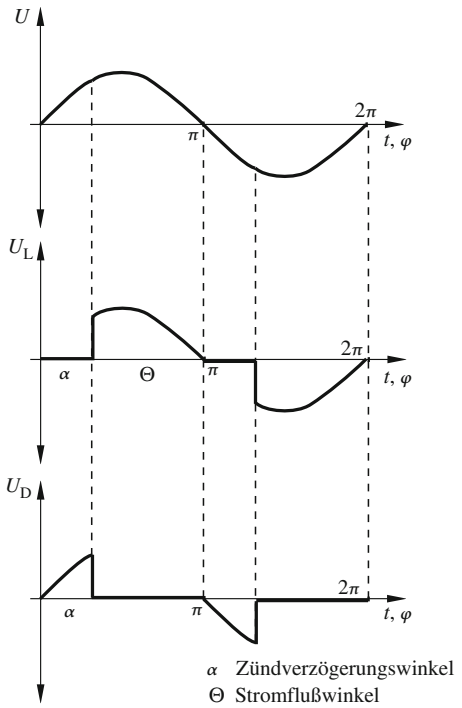


Abb. 34.14 Liniendiagramm einer sinusförmigen Spannung bei Vollwellenbetrieb

eine Leistungssteuerung zwischen $P = 0\,\mathrm{W}$ und $P = P_{\max}$ möglich. Der Zusammenhang zwischen der gesteuerten Leistung P und dem Phasenanschnittwinkel a ist nichtlinear aufgrund der Nichtlinearität der sinusförmigen Versorgungsspannung.

Bei kleineren Leistungen werden wegen des Schaltungsaufwandes statt dieser Antiparallelschaltung von Thyristoren Triacs und Diacs verwendet.

34.3 Diac

Schaltet man zwei Vierschichtdioden antiparallel zueinander intern in einen Silizium-Kristall, so erhält man eine Zweirichtungs-Thyristordiode, die sowohl in Vorwärts- wie auch in Rückwärtsrichtung gezündet werden kann. Die übliche Bezeichnung ist „Diac“ (= **di**ode **a**lternating **c**urrent switch = Diodenwechselstromschalter), aber auch die Bezeichnung Triggerdiode ist üblich. Prinzipieller Aufbau und Schaltzeichen sind

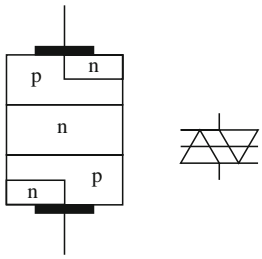


Abb. 34.15 Aufbau und Schaltung eines Diacs

in Abb. 34.15 zu sehen. Während dieser Diac insgesamt 5 Schichten hat, gibt es auch solche mit 3 Schichten (pnp).

Aufgrund der Antiparallelschaltung ergibt sich für den Rückwärtsbereich (3. Quadrant) ein spiegelbildlicher Verlauf der Kennlinie im Vorwärtsbereich (1. Quadrant). Nach Überschreiten der Durchbruchspannungen $U_{(\mathrm{BR})R}$ und $U_{(\mathrm{BR})F}$ kommt ein Strom in der jeweiligen Richtung schlagartig zum Fließen, der durch einen Vorwiderstand begrenzt werden muss. Die charakteristische Kennlinie eines Diacs mit Blockier- und Durchlassbereich zeigt Abb. 34.16. Sowohl bei Betrieb im Vorwärts- als auch im Rückwärtsbereich kann der gezündete und damit niederohmige Diac nur durch Unterschreiten seines Haltestroms wieder in den Blockierzustand gebracht werden. Bei Diacs mit drei Schichten erfolgt der Spannungseinbruch nach dem Zünden nur um einen bestimmten Spannungsbetrag ΔU.

Diacs lassen sich als elektronische Wechselstromschalter verwenden. Sie werden überwiegend für relativ kleine Ströme gefertigt und hauptsächlich zur Ansteuerung von Triacs eingesetzt.

Kennwerte: Durchbruchspannungen $U_{(\mathrm{BR})F}$, $U_{(\mathrm{BR})R}$; Durchbruchströme $I_{(\mathrm{BR})F}$, $I_{(\mathrm{BR})R}$; Halteströme I_{HF}, I_{HF}; maximaler Spitzenstrom $I_{\max}$.

Wegen seines Kippverhaltens in Vorwärts- und Rückwärtsrichtung werden Diacs hauptsächlich in Zündschaltungen für Triacs eingesetzt. Die Zündschaltung nach Abb. 34.17 zeigt das Prinzip eines Dimmers, wobei der Widerstand R_L für den Gate-Widerstand eines Triacs steht.

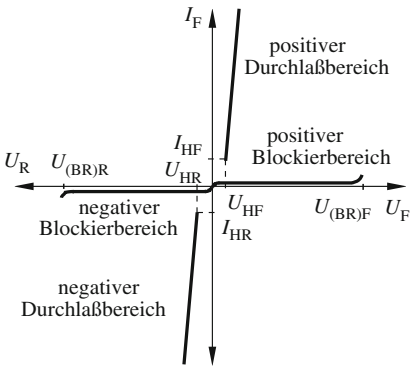


Abb. 34.16 Kennlinie einer Zweirichtungs-Thyristordiode

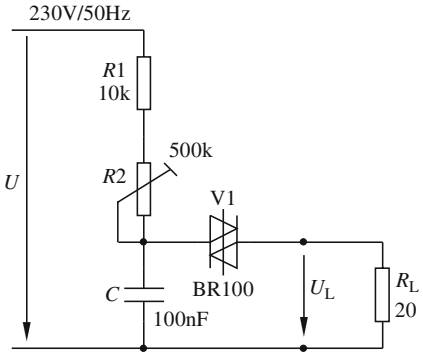


Abb. 34.17 Prinzip der Zündschaltung eines Dimmers

Der Kondensator C wird in dieser Schaltung über R1 und R2 durch die anliegende AC jeweils auf- und umgeladen. Die RC-Schaltung muss so ausgelegt werden, dass die Spannung U_C in jeder Halbwelle $U_{(BR)F}$ beziehungsweise $U_{(BR)R}$ erreicht, da der Diac sonst nicht zündet. Wenn der Diac zündet, entlädt sich der Kondensator C schlagartig über den Diac und den Widerstand R_L. Dadurch entsteht an R_L ein kurzer positiver beziehungsweise negativer Stromimpuls. Durch Verstellen des Ladewiderstandes R2 kann der Zündzeitpunkt des Diacs während der Halbwellen verändert werden. Die Spannungshöhe und die Impulsdauer hängen im Wesentlichen von der Kapazität des Kondensators, dem Durchlasswiderstand des Diacs und der Spannungsdifferenz ΔU des Diacs ab.

Der Diac bleibt gezündet, bis sein Haltestrom unterschritten wird, was beim nächsten Nulldurchgang der Wechselspannung eintritt.

34.4 Triac

Ein Triac (= **tri**ode **a**lternating **c**urrent switch) kann von seinem technologischen Aufbau her, aber auch aufgrund seiner Funktion nach Abb. 34.18 als eine Antiparallelschaltung zweier Thyristoren aufgefasst werden.

Der Triac hat zwei Vorwärtsrichtungen und kann am Gate mit einem Zündstrom beziehungsweise einer Zündspannung beliebiger Polarität gezündet werden.

Die Kennlinie eines Triacs nach Abb. 34.19 zeigt einen spiegelbildlichen Verlauf im I. und III. Quadranten des Kennlinienfeldes auf der Basis der Kennlinie eines Thyristors. Ein gezündeter Triac kippt genau wie ein Thyristor nur dann wieder in den Sperrzustand zurück, wenn sein Haltestrom unterschritten wird.

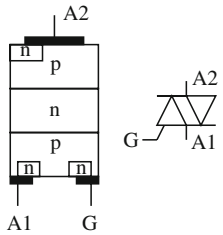


Abb. 34.18 Aufbau und Schaltzeichen eines Triacs

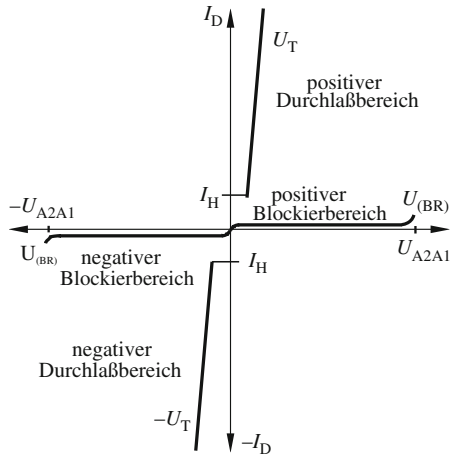


Abb. 34.19 Kennlinie eines Triacs

Triacs werden meistens mit einem Impuls gezündet, dessen Polarität mit der Polarität der an Anschlüssen A1 und A2 anliegenden Spannung (I. und III. Quadrant) übereinstimmt. Eine entsprechende Zündung kann auch in den anderen Quadranten erfolgen, erfordert aber größere Steuerleistungen. Wegen der vier möglichen Steuerungsarten werden für die Gateströme vier zugehörige Werte angegeben.

Mit Triacs lassen sich in recht einfacher Weise Phasenanschnittsteuerungen für Vollwellenbetrieb aufbauen. Wegen der niedrigen zulässigen kritischen Spannungssteilheit ist der Einsatz bei Lasten mit stark induktivem Anteil nicht gut möglich.

34.5 Schutz der Dioden und Trioden

Dioden, Triacs und Thyristoren müssen, wie alle Bauelemente in der Elektronik, gegen Überlastung geschützt werden. Den Überstromschutz übernehmen superflinke Sicherungen und Sicherungsautomaten mit magnetischer und thermischer Schnellauslösung.

Zum Schutz der Bauelemente gegen hohe Spannungsanstiegsgeschwindigkeit (du/dt) werden RC-Glieder eingesetzt und den Ventilen nach Abb. 34.20 parallel geschaltet. Die Hersteller ge-

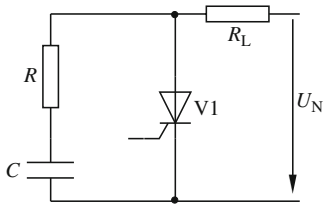


Abb. 34.20 Schutzbeschaltung gegen hohe Spannungsanstiegsgeschwindigkeit (du/dt)

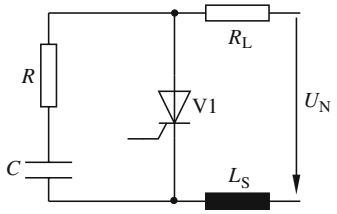


Abb. 34.21 Schutzbeschaltung gegen hohe Stromanstiegsgeschwindigkeit (di/dt)

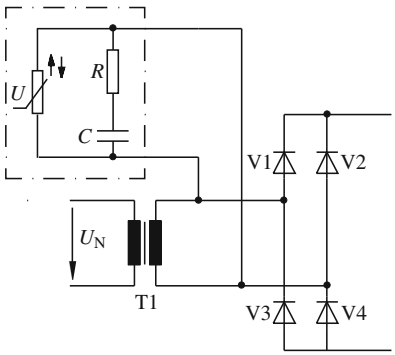


Abb. 34.22 Fertige Baugruppen zum Schutz gegen Überspannungen

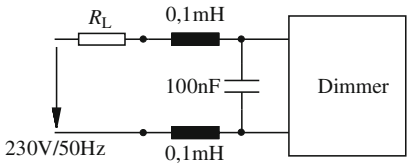


Abb. 34.23 Entstörmaßnahme für Dimmerschaltungen

ben hier Dimensionierungshilfen in Tabellen- oder Diagrammform. Die RC-Beschaltung dient gleichzeitig dem Schutz vor den Folgen des Trägerstaueffektes.

Zum Schutz der Bauelemente gegen hohe Stromanstiegsgeschwindigkeit (di/dt) werden kleine Schutzinduktivitäten L_S nach Abb. 34.21 in Reihe zu den Ventilen geschaltet. Hier reichen unter Umständen bereits die Induktivität des Netztransformators oder die Anschlussdrähte der Bauelemente aus.

Um einen gewissen Schutz gegen Überspannungen sicherzustellen, werden von den Herstellern fertig zugeschnittene Baugruppen geliefert, die zusätzlich einen Varistor parallel zum RC-Glied enthalten. Auch andere Schutzorgane finden hier Verwendung. Solche Baugruppen können natürlich auch in Gleichrichterschaltungen wie hier nach Abb. 34.22 integriert werden.

Da Triacs und Thyristoren in wenigen Mikrosekunden zünden, muss auch der Strom in sehr kurzer Zeit auf den Wert ansteigen, den die Last aufgrund der zum Zündzeitpunkt anliegenden Spannung „zieht“.

Derartig kurze Stromanstiege verursachen jedoch Hochfrequenzstörungen, die sich beim Rundfunk- und Fernsehempfang störend bemerkbar machen. Daher sind Entstörmaßnahmen für Dimmer zum Beispiel nach Abb. 34.23 vorzunehmen (Entstörfilter).

34.6 Zündmethoden

Bei Thyristoren und Triacs muss zwischen der Vertikal- und der Horizontalsteuerung unterschieden werden. In der Praxis wird der Steuerstrom größer als der obere Zündstrom gewählt. Begrenzt wird der Steuerstrom durch die zulässige Steuerverlustleistung P_G, die in Datenblättern für definierte Impulsdauern nach Abb. 34.24 angegeben wird. Die Einschaltzeit des Thyristors muss beendet sein, bevor der Impuls seinen Scheitelwert erreicht hat.

Die Vertikalsteuerung (Abb. 34.8) wird nur selten angewendet, weil eine genaue Bestimmung der Zündwinkel nicht erreicht werden kann und nur Steuerwinkel bis $\alpha = 90°$ möglich sind. Die Horizontalsteuerung weist diese Mängel nicht auf. Die Ansteuerung erfolgt häufig durch Zündimpulse und ermöglicht das Zünden auch bei kleiner Spannung und niedriger Temperatur. Die Steuerung kann mit Einzelbauelementen oder mit Ansteuer-IC's verwirklicht werden.

Bei Betrieb der Schaltung nach Abb. 34.25 an Wechselspannung zündet der Triac bei geschlossenem Schalter S mit dem Anstieg der sinusförmigen Betriebsspannung. Der Zündzeitpunkt kann durch die Wahl des Gatevorwiderstandes R_G und damit der Einstellung des Gatestroms für Steuerwinkel bis $\alpha = 90°$ bestimmt werden. Er löscht wieder, wenn beim nächsten Nulldurchgang der Wechselspannung sein Haltestrom zwangsläufig unterschritten wird. Solange Schalter S geschlossen ist, fließt ein Wechselstrom durch den Lastwiderstand.

Die Schaltung ermöglicht das Einschalten von Wechselstromlasten. Ersetzt man die Spannungsquelle und den Schalter S

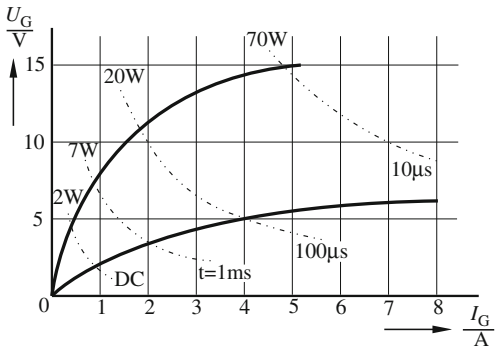


Abb. 34.24 Gate-Eingangskennlinie

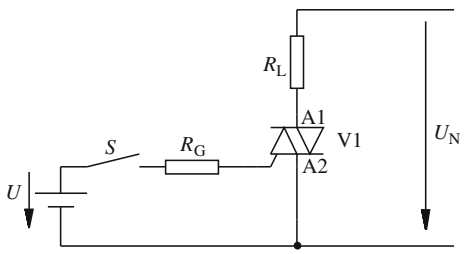


Abb. 34.25 Zünden eines Triacs durch DC am Wechselstromnetz

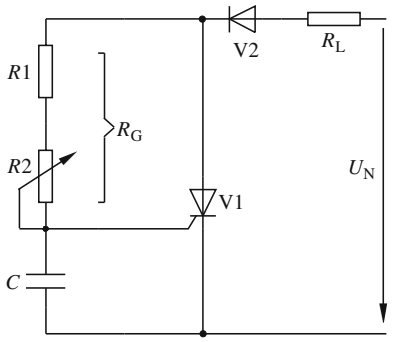


Abb. 34.26 Grundprinzip einer Phasenanschnittsteuerung

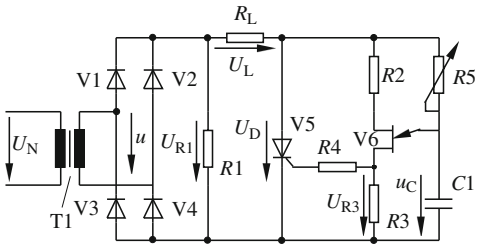


Abb. 34.27 Schaltung zum Vollwellenbetrieb mit einem Thyristor

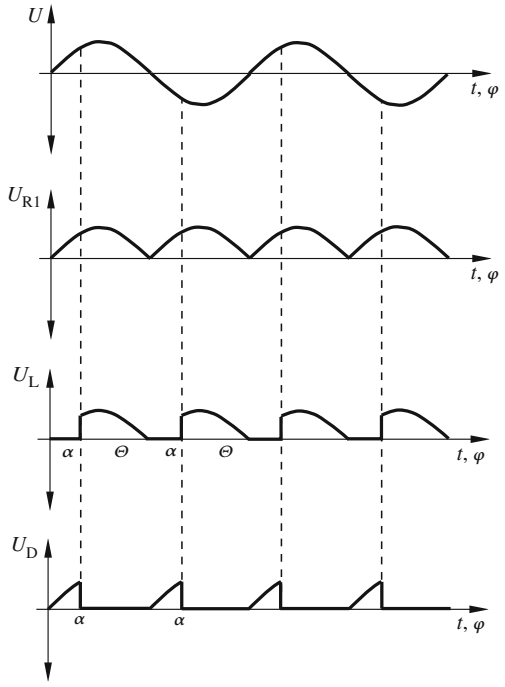


Abb. 34.28 Liniendiagramm einer Vollwellenschaltung mit Thyristor

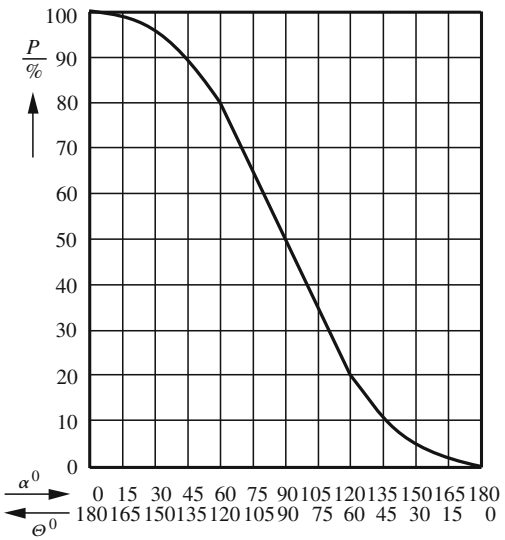


Abb. 34.29 Zusammenhang zwischen Anschnittwinkel, Stromflusswinkel und Leistung

durch ein Digitalgatter, kann eine Last kontaktlos und fast leistungslos eingeschaltet werden. Wird die Stromzufuhr ins Gate unterbrochen, so wird der Triac mit dem nächstfolgenden Nulldurchgang des Laststroms löschen (Vertikalsteuerung).

Die Schaltung nach Abb. 34.26 zeigt ein Grundprinzip einer Phasenanschnittsteuerung mit Kondensatorladung. In der positiven Halbwelle der Betriebsspannung wird der Kondensator C über den veränderbaren Vorwiderstand R_G aufgeladen. Je nach Widerstandswert von R2 steigt die Spannung am Kondensator schnell oder langsam an. Dadurch wird die zum Zünden des Thyristors erforderliche Gatespannung erst mit einer einstellbaren Verzögerung gegenüber dem Beginn der positiven Halbwelle der Betriebsspannung erreicht. Auf diese Weise kann der Zündzeitpunkt „beliebig“ eingestellt werden. Sobald der Thyristor gezündet hat, wird der Kondensator über die Gate-Kathodenstrecke des Thyristors teilweise wieder entladen. Der Widerstand R1 begrenzt den maximalen Gatestrom.

Der Lastwiderstand ist vor das RC-Glied gezogen worden und bietet den Vorteil, dass die Ladespannung des Kondensators nach dem Zünden auf die geringe Durchlassspannung des Thyristors begrenzt wird. Mit dieser Schaltung lassen sich Stromflusswinkel $20° < \Theta < 160°$ erreichen.

Bei der Horizontalsteuerung erfolgt die Ansteuerung zweckmäßigerweise durch Zündimpulse und ermöglicht präzises Zünden zu jedem gewünschten Zeitpunkt innerhalb der Halbwellen. Dies lässt sich durch eine Synchronisierung des Impulsgenerators mit der Netzspannung erreichen.

Die Schaltung nach Abb. 34.27 lässt den Vollwellenbetrieb mit einem Thyristor zu. Der B2-Gleichrichter lässt eine Spannung am Widerstand R1 abfallen. Mit Hilfe des RC-Gliedes (R5 und C1) kann nun der Zeitpunkt eingestellt werden, an dem der Unijunction-Transistor (UJT) durchsteuert und Spannungsimpulse an R3 hervorgerufen werden, die den Thyristor am Gate ansteuern. Der Widerstand R4 dient zur Begrenzung des Gatestroms.

Aufgrund der Spannung U_{R1} ist der Thyristor immer im Vorwärtsbetrieb und kann darum in jeder Halbwelle gezündet werden. Die Leistung ist theoretisch zwischen 0 W und P_{max} steuerbar, zum Beispiel kann die Helligkeit einer Lampe von

0 bis 100 % gesteuert werden. Die Grenzbereiche ($\alpha < 15°$ und $\alpha > 165°$) lassen sich mit dieser Schaltung nicht realisieren.

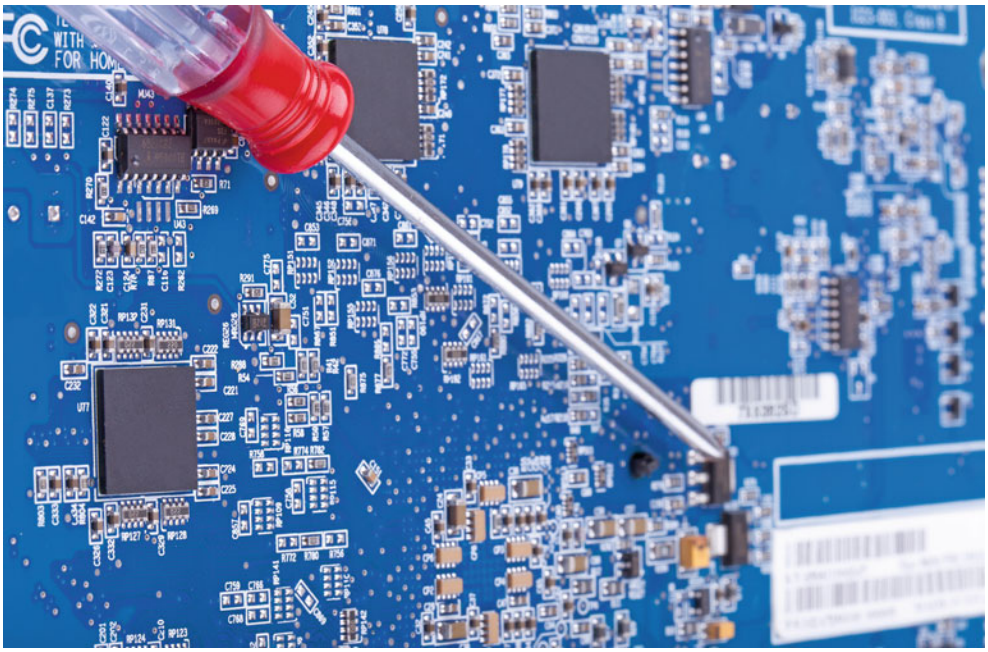
Im Liniendiagramm nach Abb. 34.28 ist ein Phasenanschnittwinkel $\alpha = 90°$ angenommen, so dass hier I_L und U_L jeweils mit halbierten positiven Halbwellen wirksam sind, also die Hälfte der maximalen Leistung an die Last (Lampe) gelangt.

Der Zusammenhang zwischen P_{max} und dem jeweiligen Winkel ist im Abb. 34.29 dargestellt. Hier ist erkennbar, dass der Zusammenhang nichtlinear ist.

Transistoren

35

Peter Döring



35.1 Bipolare Transistoren 426
35.2 Feldeffekttransistoren (FET) 432

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_35

Der Prototyp des Transistors wurde im Jahre 1948 gefunden, als die US-Amerikaner John Bardeen und Walter H. Brattain bei der Untersuchung von Diodenübergängen eine Metallprüfspitze auf die n-Schicht aufsetzten und so unbeabsichtigt eine pnp-Schichtfolge entstand. Zu ihrer Überraschung stellten sie bei weiteren Messungen fest, dass bei Widerstandsänderungen in der einen Grenzschicht auch der Widerstand der anderen Grenzschicht beeinflusst wird. Damit war das Grundprinzip des Transistors entdeckt. Das Kunstwort „Transistor“ entstand aus der Bezeichnung „transfer resistor“, was soviel wie „Widerstandsänderungen von einer Grenzschicht zur anderen übertragen“ bedeutet.

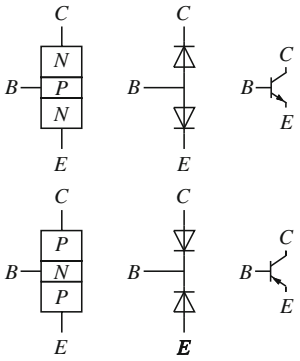


Abb. 35.1 Schichtenfolge, Ersatzschaltbild und Schaltzeichen von npn- und pnp-Transistoren

35.1 Bipolare Transistoren

Ergänzt man die pn-Schichtenfolge einer Diode um eine weitere n- oder p-Schicht, so erhält man die Schichtenfolge eines Transistors. Entsprechend der Schichtenfolge unterscheidet man npn- oder pnp-Transistoren.

Am Ladungstransport sind aufgrund der verschiedenen Dotierungsarten sowohl Löcher als auch Elektronen beteiligt. Darum spricht man hier von „bipolaren“ Transistoren im Gegensatz zu den „unipolaren“ Transistoren, die im Abschn. 35.2 beschrieben werden.

Die mittlere Schicht bezeichnet man als Basis. Sie dient zum Steuern des Transistors. Die beiden äußeren Schichten heißen Emitter (sendet Ladungsträger aus) und Kollektor (sammelt Ladungsträger ein). Der Pfeil am Emitter gibt die technische Stromrichtung des Emitterstroms an.

Als Ersatzschaltbild nach Abb. 35.1 kann man die Reihenschaltung zweier Dioden verwenden, die eine gemeinsame Kathode beziehungsweise Anode haben. Die eine Diode ist die Basis-Emitter-Diode, die andere die Basis-Kollektor-Diode.

Anhand dieses Ersatzschaltbildes nach Abb. 35.1 soll die grundsätzliche Polung eines als Verstärker betriebenen Transistors nach folgenden Bedingungen beschrieben werden:

1. Die Basis-Emitter-Diode ist grundsätzlich in Durchlassrichtung zu betreiben.
2. Die Basis-Kollektor-Diode ist grundsätzlich in Sperrrichtung geschaltet.

Um den Transistor als Verstärker betreiben zu können, müssen für die Basis zwei Voraussetzungen erfüllt sein:

1. Sie muss sehr gering dotiert sein; Dotierungsgrad 1/100 des Emitters!
2. Sie muss sehr dünn sein gegenüber der mittleren freien Weglänge der Majoritätsträger (1 bis 100 µm).

Sind diese beiden Voraussetzungen erfüllt, kann mit einem kleinen Eingangsstrom (Basisstrom) ein großer Ausgangsstrom (Kollektorstrom) gesteuert werden.

35.1.1 Transistoreffekt

Durch die in Durchlassrichtung betriebene Basis-Emitter-Diode eines npn-Transistors fließt ein großer Durchlassstrom (Emitterstrom), der so gerichtet ist, dass die durch Dotierung vorhandenen Elektronen zur Basis wandern. In der Basis findet diese große Menge an Elektronen wegen des geringen Dotierungsgrades der Basis nur wenig Löcher als Rekombinierungspartner vor. Es kommt deshalb nur zu einer geringen Anzahl von Rekombinationen. Der größte Teil der vom Emitter kommenden Elektronen überschwemmt die Basis als freie Ladungsträger. Da die Basis sehr dünn ist, werden die Elektronen aufgrund ihrer Eigengeschwindigkeit zum Kollektor hin abgedrängt. Das Überschreiten der Kollektor-Basis-Sperrschicht bildet für die Elektronen kein Problem, da sie in der Basis Minoritätsträger sind. Die Sperrschicht eines pn-Überganges stellt jedoch nur für die Majoritätsträger ein Hindernis dar.

Wegen der geringen Anzahl von Rekombinationen in der Basis gelangen also fast alle vom Emitter ausgesandten Elektronen zum Kollektor. Es fließt ein Kollektorstrom, der etwas kleiner ist als der Emitterstrom, denn die durch Rekombination in der Basis gebliebenen Elektronen würden die Basis negativer machen, wenn sie nicht nach außen abgeführt würden. Dieser Elektronenstrom, der aus der Basis herausfließen muss, damit Basispotential und Emitterstrom konstant bleiben, stellt den Basisstrom dar. Er bestimmt den Arbeitspunkt des Transistors.

Wird der Basisstrom größer, werden mehr Elektronen aus der Basis entnommen, als zur Erhaltung eines stabilen Zustandes notwendig wäre. Die Basis wird positiver und die Basis-Emitter-Spannung und der Emitterstrom werden größer. Werden weniger Elektronen entnommen, so wird der Emitterstrom kleiner. Durch diesen kleinen Basisstrom, der zum Emitterstrom im gleichen Verhältnis steht wie der Dotierungsgrad der Basis zum Dotierungsgrad des Emitters, kann der große Emitter- beziehungsweise Kollektorstrom gesteuert werden.

Beim npn-Transistor sind vorwiegend Elektronen am Ladungstransport beteiligt, während beim pnp-Transistor vornehmlich Löcher am Ladungstransport beteiligt sind.

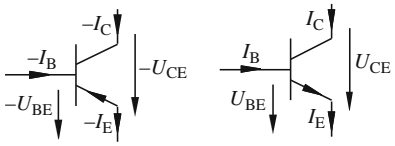


Abb. 35.2 Spannungs- und Strompfeile an Transistoren

Deren Bewegungsrichtung ist wegen der umgekehrten Polung des pnp-Transistors jedoch die gleiche wie die der Elektronen beim npn-Transistor, nämlich vom Kollektor zum Emitter. Somit erübrigt sich die Beschreibung der physikalischen Wirkungsweise des pnp-Transistors. Letztlich ist ja auch der Löcherstrom ein Elektronenstrom, der auf der indirekten Elektronenbewegung beruht (darum sind Elektronen beweglicher als Löcher).

Wegen der unterschiedlichen Beweglichkeit von Elektronen und Löchern sind npn-Transistoren für hohe Frequenzen besser geeignet als pnp-Transistoren. Viele elegante und einfache Schaltungen lassen sich mit „Komplementärpärchen“ (je ein npn- und pnp-Transistor mit identischen Daten) realisieren.

Die *Spannungen* am Transistor werden nach Abb. 35.2 durch zwei Indices gekennzeichnet, nämlich durch die Indices der Elektroden, zwischen denen die Spannung gemessen wird. Die Reihenfolge der Indices beziehungsweise die Richtung der Spannungspfeile gibt die Messrichtung an: Kollektor-Basisspannung U_{CB}, Basis-Emitterspannung U_{BE} und Kollektor-Emitterspannung U_{CE}.

Die *Ströme* am Transistor werden durch einen Index gekennzeichnet, und zwar durch den Index der Elektrode, durch die sie fließen.

Die Festlegung der Richtungspfeile kann man nach verschiedenen Gesichtspunkten vornehmen: Entweder nach der tatsächlichen Bewegungsrichtung der Ladungsträger im Kristall und in den Zuleitungen (zur physikalischen Deutung) oder nach der allgemein üblichen technischen Stromrichtung, vom Pluspol zum Minuspol. Transistoren sind in zahlreichen Gehäuseformen (Bauformen) lieferbar, die nur zum Teil genormt sind.

35.1.2 Transistorkennlinien

Um das Betriebsverhalten eines Transistors im Normal- und Grenzfall beschreiben zu können, ist die Kenntnis der Ein- und Ausgangsgrößen erforderlich.

Eingangsgrößen: Basis-Emitter-Spannung U_{BE} Basisstrom I_B
Ausgangsgrößen: Kollektor-Emitter-Spannung U_{CE} Kollektorstrom I_C

Die Eingangskennlinie $I_B = f(U_{BE})$ eines Transistors ist in Abb. 35.3 dargestellt. Mit dem Basisstrom verändert sich zwangsläufig auch der Kollektorstrom. Darum müssen Kollektorspannung und Sperrschichttemperatur konstant gehalten werden.

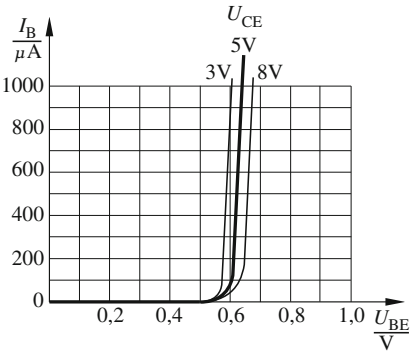


Abb. 35.3 Eingangskennlinie eines Transistors

Da die Basis-Emitter-Strecke eines Transistors eine Diode in Durchlassrichtung ist, hat die Eingangskennlinie die Form einer Diodenkennlinie in Durchlassrichtung. Der Basisstrom steigt stark an, sobald die Basis-Emitter-Spannung einen bestimmten Wert (die Diffusionsspannung) überschritten hat. Für andere Werte von U_{CE} ergeben sich andere Eingangskennlinien, wie Abb. 35.3 zeigt.

$$U_{BE} \approx 0{,}2 \text{ bis } 0{,}3\,\text{V bei Ge-Transistoren}$$
$$U_{BE} \approx 0{,}6 \text{ bis } 0{,}7\,\text{V bei Si-Transistoren}$$

Hieraus lässt sich der Ersatzwiderstand r_{BE} zwischen Basis- und Emitteranschluss des Transistors berechnen, durch den die Steuerspannungsquelle belastet wird.

$$\textit{DC-Eingangswiderstand} \quad R_{BE} = \frac{U_{BEA}}{I_{BA}} \tag{35.1}$$

Dieser Gleichstromwiderstand R_{BE} darf nicht mit dem differentiellen Eingangswiderstand r_{BE} (Kleinsignal-Eingangswiderstand) des Transistors verwechselt werden. Er stellt die Steigung der Eingangskennlinie im jeweiligen Kennlinienpunkt, sprich Arbeitspunkt, dar.

Für eine Diodenkennlinie gilt (33.2), die dementsprechend auch für eine Transistor-Eingangskennlinie gilt. Berechnet man mathematisch die 1. Ableitung der Kurve, so gilt:

$$\frac{dI_B}{dU_{BE}} = \frac{1}{r_{BE}} = \frac{I_{BS} \cdot e^{\frac{U_{BE}}{U_T}}}{U_T} \approx \frac{I_{BA}}{U_T} \quad \text{für } U_{BE} \gg U_T$$
$$\textit{differentieller Widerstand} \quad r_{BE} \approx \frac{U_T}{I_{BA}} \tag{35.2}$$

Der differentielle Widerstand r_{BE} ist bedeutend kleiner als der Basis-Emitter-Ersatzwiderstand R_{BE}.

Bei der Aufnahme der Stromsteuerkennlinie $I_C = f(I_B)$ nach Abb. 35.4 wird der Basisstrom I_B verändert, während U_{CE} jeweils konstant gehalten wird. Bereits bei der Untersuchung der Funktionsweise des Transistors ist erkennbar, dass der Kollektorstrom mit dem Basisstrom verknüpft ist.

Die Stromverstärkung B wird durch den Dotierungsgrad der einzelnen Schichten des Transistors eingestellt, ist also in erster Linie hiervon abhängig und weniger von anderen Einflussgrößen.

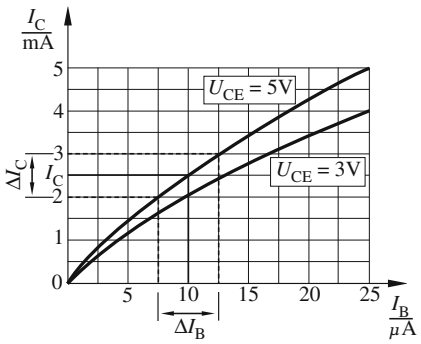


Abb. 35.4 Stromsteuerkennlinie

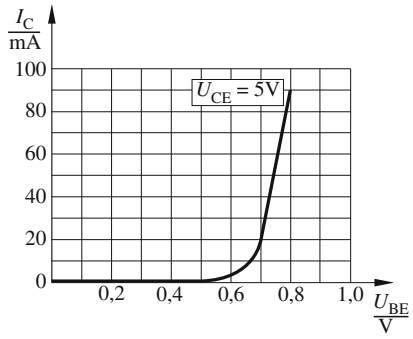


Abb. 35.5 Spannungssteuerkennlinie

Tatsächlich ist zwischen I_C und I_B über einen weiten Bereich eine Proportionalität vorhanden, der Graph ergibt annähernd eine Gerade. Für verschiedene Werte von U_{CE} ergeben sich unterschiedliche Stromverstärkungskennlinien, wie Abb. 35.4 zeigt.

$$\textit{Gleichstromverstärkung} \qquad B = \frac{I_C}{I_B} \tag{35.3}$$

$$\textit{Kleinsignalverstärkung} \qquad \beta = \frac{\Delta I_C}{\Delta I_B} \tag{35.4}$$

Die Spannungssteuerkennlinie $I_C = f(U_{BE})$ ergibt annähernd den exponentiellen Verlauf einer Diodenkennlinie (siehe Eingangskennlinie). Die *Steigung S* der Kennlinie nach Abb. 35.5 wird in Anlehnung an die Röhrentechnik als *Steilheit S* bezeichnet. Beim „idealen Transistor", bei dem der Kollektorstrom exponentiell mit der Basis-Emitter-Spannung steigt, ist das Verhältnis von Steilheit zu Kollektorstrom konstant.

Die Gleichung ergibt sich aus der 1. Ableitung der Diodenkennlinie unter Berücksichtigung der Temperaturspannung U_T.

$$\textit{Steilheit} \quad S = \frac{dI_C}{dU_{BE}} \approx \frac{I_C}{U_T} \quad [S] = \frac{\text{mA}}{\text{V}} \tag{35.5}$$

Eine Sperrschichttemperatur von 45 °C ergibt $U_T = 24{,}4\,\text{mV}$. Die theoretisch mögliche Steilheit eines Transistors für Überschlagsrechnungen beträgt bei dieser Temperatur

$$\textit{Steilheit} \quad S = 36 \cdot \frac{1}{V} \cdot I_C \tag{35.6}$$

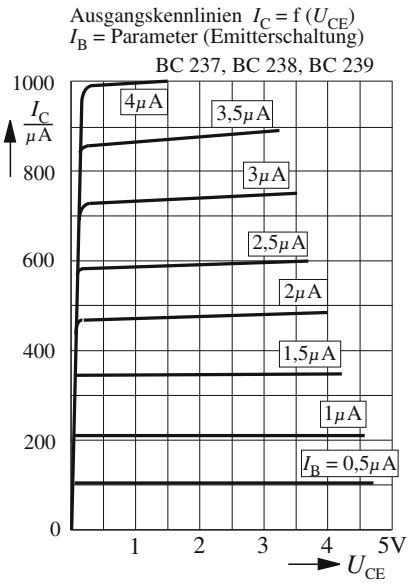


Abb. 35.6 Ausgangskennlinienfeld mit I_B als Parameter

Für Kleinsignal-Transistoren stimmt diese Gleichung recht gut, bei Leistungs-Transistoren weichen die errechneten Werte von den tatsächlichen Größen erheblich ab.

Die Ausgangskennlinien beschreiben den Zusammenhang zwischen den beiden Ausgangsgrößen, dem Kollektorstrom I_C und der Kollektor-Emitter-Spannung U_{CE}. Bei konstanter Sperrschichttemperatur muss entweder der Basisstrom oder die Basis-Emitter-Spannung konstant gehalten werden, wenn die Kennlinien eindeutig sein sollen.

Man unterscheidet deshalb zwei Ausgangskennlinienfelder:

1. das Ausgangskennlinienfeld, bei dem I_B konstant gehalten wird (I_B als Parameter, Abb. 35.6)
2. das Ausgangskennlinienfeld, bei dem U_{BE} konstant gehalten wird (U_{BE} als Parameter, Abb. 35.7)

Der Verlauf der beiden Ausgangskennlinienfelder ist grundsätzlich gleich. Bei beiden Kennlinienfeldern steigt der Kollektorstrom schon bei kleinen Werten der Kollektor-Emitter-Spannung stark an, um dann ab Erreichen eines bestimmten Wertes von U_{CE} in die Sättigung überzugehen.

Dieser Wert von U_{CE}, der vom jeweiligen Basisstrom abhängig ist, heißt Sättigungsspannung $U_{CE\,sat}$. Wird U_{CE} größer als $U_{CE\,sat}$, steigt der Kollektorstrom nur noch sehr gering an.

Die Kennlinien beider Kennlinienfelder zeigen nach Überschreiten der Sättigungsspannung einen sehr flachen Verlauf. Wenn der Arbeitspunkt eingangsseitig festgehalten wird (U_{BE}, I_B), ändert sich der Kollektorstrom kaum. Die Stromverstärkung wird durch die Dotierung eingestellt, kann also durch U_{CE} kaum beeinflusst werden.

Um zu untersuchen, welche Rückwirkung die Ausgangsspannung U_{CE} des Transistors auf die Eingangsspannung U_{BE} ausübt, wird U_{BE} in Abhängigkeit von U_{CE} dargestellt. Damit

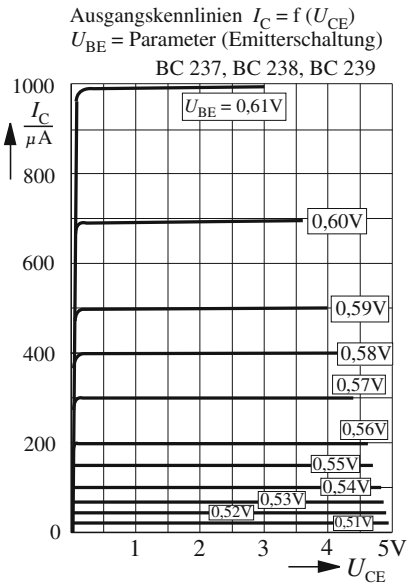


Abb. 35.7 Ausgangskennlinienfeld mit U_{BE} als Parameter

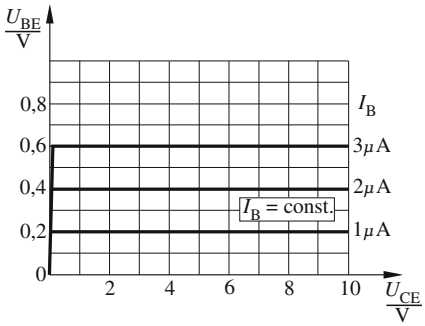


Abb. 35.8 Spannungsrückwirkungskennlinie

die Kennlinien eindeutig sind, wird der Basisstrom I_B konstant gehalten. Abb. 35.8 zeigt für verschiedene Werte von I_B das Kennlinienfeld der Spannungsrückwirkung D_u. Es ist erkennbar, dass die Spannungsrückwirkung sehr gering ist. Das Verhältnis von $\Delta U_{BE}/\Delta U_{CE}$ liegt bei etwa 10^{-4} bis 10^{-6}.

Eine Darstellungsweise, die die Zusammenhänge zwischen den einzelnen Kennlinien besonders deutlich macht, ist das Vierquadranten-Kennlinienfeld. Hier werden alle vier Quadranten eines Koordinatenkreuzes zur Darstellung der verschiedenen Kennlinien ausgenutzt, wobei die gegenseitige Abhängigkeit in Abb. 35.9 deutlich wird.

Im I. Quadranten wird das Ausgangskennlinienfeld mit I_B als Parameter dargestellt, im II. Quadranten die Stromsteuerkennlinie, im III. Quadranten die Eingangskennlinie und im IV. Quadranten die Rückwirkungskennlinien.

Je flacher die Ausgangskennlinien verlaufen, desto weniger hängt die Stromsteuerkennlinie vom Parameter U_{CE} ab.

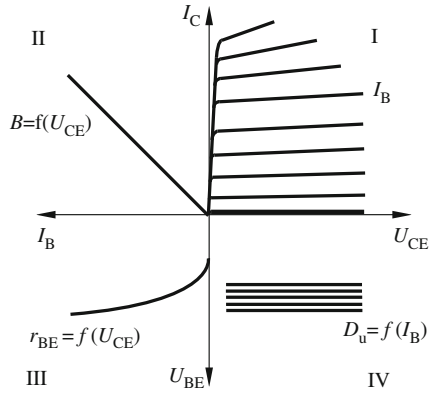


Abb. 35.9 Vier-Quadranten-Kennlinienfeld eines Transistors

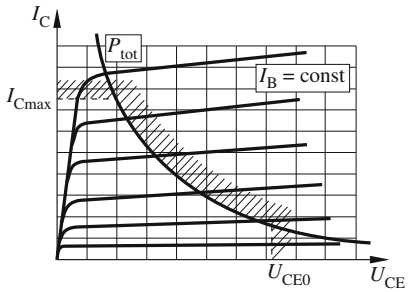


Abb. 35.10 Zulässiger Arbeitsbereich eines Transistors

35.1.3 Kenn- und Grenzwerte des Transistors

Um den Arbeitsbereich eines Transistors nach Abb. 35.10 einzugrenzen, sind Grenzwerte und Kenndaten zu ermitteln und zuzuordnen.

Grenzwerte: zu hohe Kollektor-Emitter-Spannung $U_{CE} > U_{CE\,max}$, zu großer Kollektorstrom $I_C > I_{C\,max}$, zu große Verlustleistung $P_V > P_{tot}$, zu hohe Sperrschichttemperatur $\vartheta_J > \vartheta_{J\,max}$, zu große Basis-Emitter-Spannung $U_{BE} > U_{BE\,max}$, zu großer Basisstrom $I_B > I_{B\,max}$ können zur Zerstörung des Transistors führen.

Das Diagramm nach Abb. 35.11 zeigt, wie hoch die Verlustleistung bei einer bestimmten Umgebungstemperatur sein darf. Wird jedoch die im Transistor erzeugte Wärme durch einen Kühlkörper $R_{th\,JG}$ oder durch eine Montage auf ein Chassis besser abgeführt, so ist die Gesamtverlustleistung auch noch bei höheren Umgebungstemperaturen zulässig.

Gesamtverlustleistung

$$P_V = U_{CE} \cdot I_C + U_{BE} \cdot I_B \leq P_{tot} \tag{35.7}$$

Wegen

$$U_{CE} \cdot I_C \gg U_{BE} \cdot I_B$$

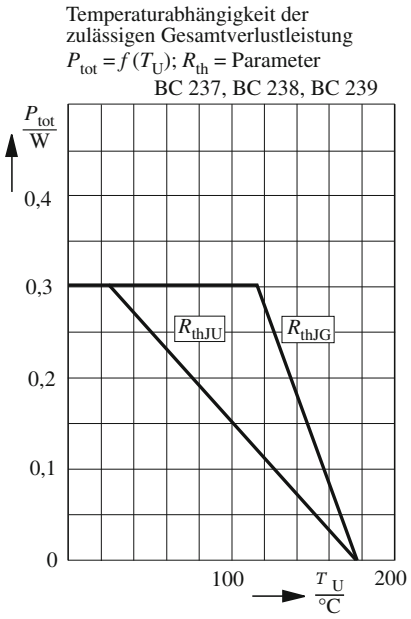


Abb. 35.11 Verlustleistung in Abhängigkeit von der Umgebungstemperatur

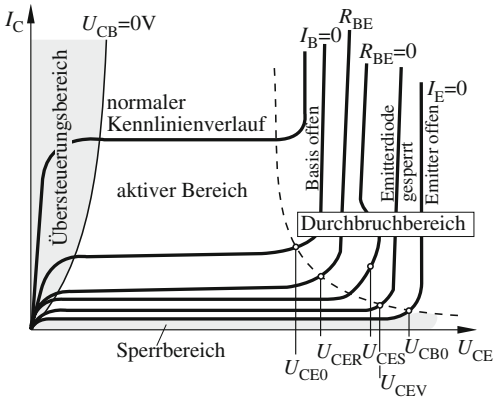


Abb. 35.12 Durchbruchspannungen und Restströme eines Transistors

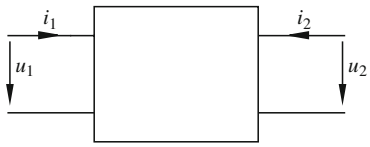


Abb. 35.13 Transistor als Vierpol

gilt

$$P_V \approx U_{CE} \cdot I_C \leq P_{tot} \tag{35.8}$$

Durch die Verlustleistung bei Dauerbetrieb entsteht in der Sperrschicht Wärme; die Sperrschichttemperatur erhöht sich. Sie darf bestimmte Werte nicht überschreiten, weil der Kristall ansonsten seine Halbleitereigenschaften stark verändert und der Transistor zerstört wird. Die zulässige Sperrschichttemperatur T_J hängt vom Halbleitermaterial ab.

Sperrschichttemperatur $\vartheta_J \approx 90\,°C$ bei Germanium-Transistoren und $\vartheta_J \approx 150$ bis $200\,°C$ bei Silizium-Transistoren.

Oberhalb einer bestimmten, vom Basisstrom abhängigen Kollektor-Emitter-Spannung wächst der Kollektorstrom plötzlich sehr stark mit U_{CE} an. Dieses Verhalten wird als „2. Durchbruch“ bezeichnet. Hier darf ein Transistor nicht mehr betrieben werden.

Abb. 35.12 zeigt die wichtigsten Durchbruchspannungen und die zugehörigen Restströme.

Es gilt folgende Beziehung:

$$U_{CEO} < U_{CER} < U_{CES} < U_{CEV}$$

Statische Kenndaten: Gleichstromverstärkung B, Sättigungsspannung $U_{CE\,sat}$, Restströme I_{CEO}.

Die *Sättigungsspannung* ist vor allem dann von Interesse, wenn ein Transistor im Schalterbetrieb arbeitet. Sie ist die Spannung, die zwischen Kollektor und Emitter abfällt, wenn der Transistor voll durchgesteuert ist. Sie wird auch Restspannung genannt.

Gerade bei kleinen Kollektorstromwerten macht sich in den pn-Übergängen der Reststrom störend bemerkbar. Wenn der Basisstrom noch Null ist, fließt zum Beispiel schon ein geringer Kollektorstrom. Dies ist der Kollektor-Emitter-„Reststrom“ I_{CEO}. Das ist der auf der Eigenleitung beruhende Sperrstrom der Kollektor-Emitter-Strecke; er ist stark temperaturabhängig.

Die oben beschriebenen Kennlinien liefern die Daten für die Gleichstromsteuerung und den Großsignalbetrieb. Infolge der Krümmung der meisten Kennlinien muss man zur Beschreibung des Kleinsignalverhaltens des Transistors die Steigung der Kennlinien im Arbeitspunkt verwenden (Tangente an die Kennlinie im AP).

Die grafische Ermittlung dieser Kennwerte ist meist ungenau. Sie werden darum in den Datenblättern der Hersteller als Zahlenwert genannt. Die Datenblätter geben die *dynamischen Kenngrößen* in Form der h-Parameter oder y-Parameter an, die nur für einen bestimmten Arbeitspunkt, eine bestimmte Temperatur und eine bestimmte Frequenz gelten. Sie stellen die Wechselstrom-(Signal-)kennwerte dar.

Während diese Transistor-Kenndaten im NF-Bereich meist als h-Parameter angegeben sind, werden sie im HF-Bereich und bei den Feldeffekttransistoren als y-Parameter dargestellt. Diese Parameter sind Verhältnisgrößen und beschreiben wegen der Vierpoltheorie das Gesamtverhalten eines Transistors als Kleinsignalverstärker, wenn man ihn als Vierpol nach Abb. 35.13 betrachtet.

$$u_1 = h_{11} \cdot i_1 + h_{12} \cdot u_2$$
$$i_2 = h_{21} \cdot i_1 + h_{22} \cdot u_2$$

Die Begriffe kurzgeschlossener und offener Ein- gang oder Ausgang beziehen sich immer auf den Wechselspannungsanteil eines Signals. Ein solcher Kurzschluss lässt sich zum Beispiel leicht mit Hilfe eines ausreichend großen Kondensators herstellen.

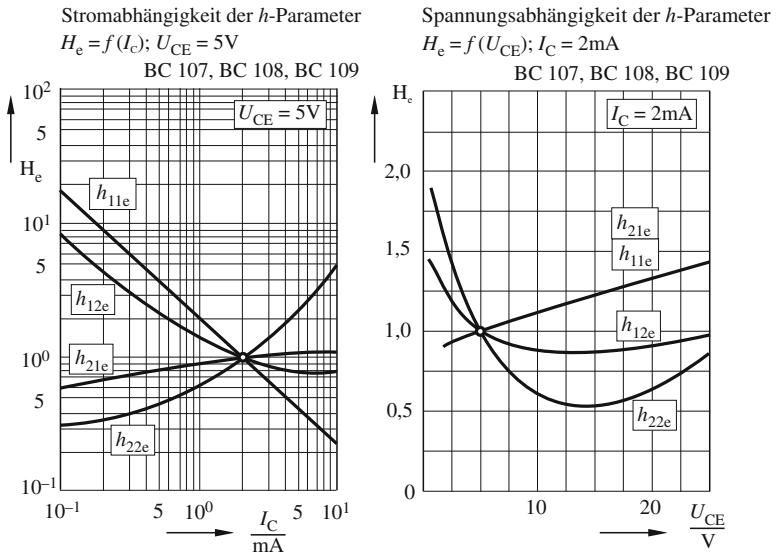


Abb. 35.14 Diagramme zur Umrechnung der h-Parameter für andere Arbeitspunkte

Tab. 35.1 Bedeutung der h-Parameter

Ermittlung aus Kennlinien/Messgrößen	
$h_{11} = r_{BE} = \frac{\Delta U_{BE}}{\Delta I_B} = \frac{u_{BE}}{i_B}$	Kurzschluss-Eingangswiderstand (Einheit Ω)
Für U_{CE} = const und $u_{CE} = 0$	
$h_{12} = D_u = \frac{\Delta U_{BE}}{\Delta U_{CE}} = \frac{u_{BE}}{u_{CE}}$	Leerlauf-Spannungsrückwirkung (dimensionslos)
Für I_B = const und $i_B = 0$	
$h_{21} = \beta = \frac{\Delta I_C}{\Delta I_B} = \frac{i_C}{i_B}$	Kurzschluss-Stromverstärkung (dimensionslos)
Für U_{CE} = const und $u_{CE} = 0$	
$h22 = \frac{1}{r_{CE}} = \frac{\Delta I_C}{\Delta U_{CE}} = \frac{i_C}{u_{CE}}$	Leerlauf-Ausgangsleitwert (Einheit eines Leitwertes)
Für I_B = const und $i_B = 0$	

Auf Transistoren angewendet:

$$u_{BE} = h_{11} \cdot i_B + h_{12} \cdot u_{CE}$$
$$i_C = h_{21} \cdot iB + h_{22} \cdot u_{CE} \qquad (35.9)$$

Die in den Datenblättern angegebenen Parameter gelten immer nur für einen ganz genau definierten Arbeitspunkt. Soll der Transistor in einem anderen Arbeitspunkt betrieben werden, müssen die h-Parameter für den neuen Arbeitspunkt umgerechnet werden.

Aus den Diagrammen nach Abb. 35.14 lässt sich jeweils der Faktor H_e für den neugewählten Kollektorstrom oder für die neugewählte Kollektorspannung ablesen. Daraus lassen sich die h-Parameter für den neuen Arbeitspunkt berechnen.

umgerechnete h-Parameter

$$h_{neu} = h_{alt} \cdot H_{ei}$$
$$h_{neu} = h_{alt} \cdot H_{eu}$$
$$h_{neu} = h_{alt} \cdot H_{ei} \cdot H_{eu} \qquad (35.10)$$

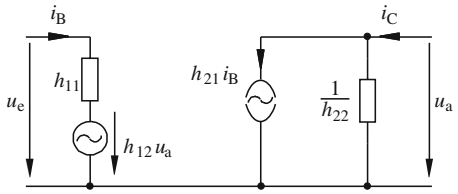


Abb. 35.15 Wechselstrom-Ersatzschaltbild eines Transistors

Mit den h-Parametern lässt sich das für alle Grundschaltungen des Transistors gültige Wechselstrom-Ersatzschaltbild nach Abb. 35.15 beschreiben.

In den Datenblättern werden häufig die für die Emitterschaltung gültigen h-Parameter (h-Parameter mit dem Index „e“ gekennzeichnet) angegeben, die sich mit Hilfe der Gleichungen in Tab. 35.2 leicht in die h-Parameter für die anderen Grundschaltungen umrechnen lassen.

Wichtig für das Wechselstromverhalten sind die Sperrschichtkapazitäten des Transistors; ihr Wert hängt von der anliegenden Sperrspannung ab. Die Kollektor-Basis-Kapazität C_{CB} liegt zwischen Emitter- und Basis-Anschluss im Gegenkopplungskreis und bildet mit dem Ausgangswiderstand des Transistors einen Hochpass für die Signalfrequenzen.

Zusätzlich macht sich bei höheren Frequenzen auch die Laufzeit der Ladungsträger durch den Transistor bemerkbar.

Aus diesen Gründen wird die Stromverstärkung β des Transistors mit zunehmender Frequenz kleiner. Bei der Grenzfrequenz f_g ist die Verstärkung um 3 dB gesunken.

Die Frequenzabhängigkeit der Stromverstärkung eines Transistors ist in Abb. 35.16 im Prinzip dargestellt.

Tab. 35.2 Umrechnungsfaktoren

Basisschaltung		Kollektorschaltung	
$h_{11b} = \frac{h_{11e}}{\sum h_e}$	$h_{12b} = -\frac{h_{12e}-\Delta h_e}{\sum h_e}$	$h_{11c} = h_{11e}$	$h_{12c} = -h_{12e}+1$
$h_{21b} = -\frac{h_{21e}+\Delta h_e}{\sum h_e}$	$h_{22b} = \frac{h_{22e}}{\sum h_e}$	$h_{21c} = -h_{21e}-1$	$h_{22c} = h_{22e}$
$\Delta h_e = h_{11e} \cdot h_{22e} - h_{12e} \cdot h_{21e}$		$\sum h_e = 1 - h_{12e} + h_{21e} + \Delta h_e$	

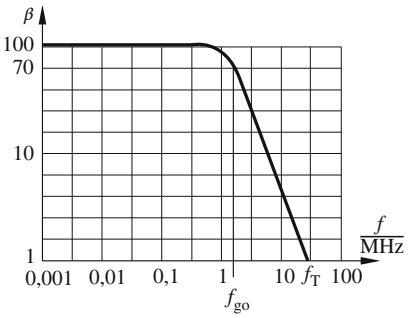


Abb. 35.16 Stromverstärkung in Abhängigkeit von der Frequenz

Die Frequenz, bei der $\beta = 1$ ist, wird als Transitfrequenz f_T bezeichnet. Die Transitfrequenz f_T hängt sowohl vom Kollektorstrom als auch von der Spannung ab.

Bei Kleinsignalverstärkern mit geringen Eingangsleistungen ist das *Rauschen* von Transistoren im NF- und HF-Bereich von Bedeutung.

Das Stromrauschen ist dabei auf Unregelmäßigkeiten im Ladungsträgerfluss zurückzuführen. Solche Effekte treten in Halbleitern durch Generation und Rekombination von Ladungsträgern auf, aber auch an Metall-Halbleiterübergängen.

Thermisches Rauschen dagegen beruht auf der regellosen Bewegung thermisch angeregter Ladungsträger.

Mit der Zunahme der Temperatur steigt auch das thermische Rauschen an.

Ein Maß für das in einem Transistor erzeugte Rauschen ist die *Rauschzahl F*, die in den Datenblättern meistens im logarithmischen Wert in dB angegeben ist. Hiermit kann nun bei einer gegebenen Verstärkerschaltung die Mindestnutzspannung berechnet werden, die am Eingang anliegen muss, damit am Ausgang des Verstärkers das Nutzsignal um einen bestimmten Faktor größer als das Rauschen ist.

Die so erzeugte Rauschspannung deckt unter Umständen ein weites Frequenzband ab und führt bei sehr geringen Signalspannungen in der Nachrichtentechnik zu erheblichen Problemen.

35.2 Feldeffekttransistoren (FET)

Neben den bipolaren Transistoren gibt es die „unipolaren" Transistoren, bei denen am Ladungsträgertransport nur eine Ladungsträgerart, also entweder Löcher oder Elektronen, beteiligt

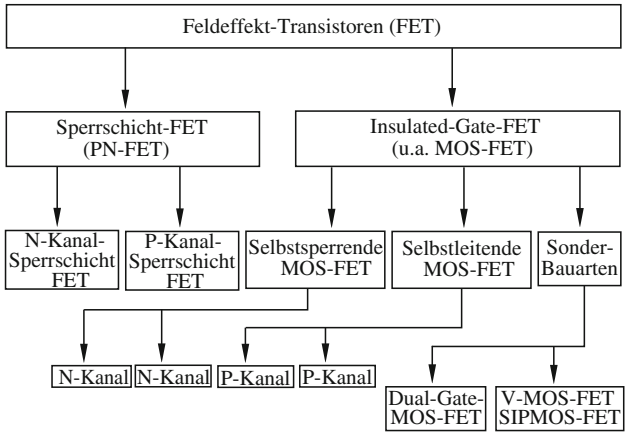


Abb. 35.17 Übersicht zu den Transistoren

sind. Die Steuerung des Stromflusses erfolgt bei ihnen durch ein elektrisches Feld. Deshalb wird bei den unipolaren Transistoren auch von „Feld-Effekt-Transistoren" (FET) gesprochen.

Grundsätzlich muss zwischen zwei FET-Gruppen unterschieden werden, und zwar zwischen den Sperrschicht-FET (auch J-FET genannt) und den IG-FET (IG = isoliertes Gate). Bei den IG-FET unterscheidet man zwischen den Typen je nach technischer Ausführung der Isolierschicht und dem verwendeten dotierten Halbleitermaterial. Eine Übersicht bietet Abb. 35.17.

In der heutigen Technologie verwendet man als Isolierschicht fast ausschließlich Siliziumdioxyd. Daher kommen MOS-FET (**m**etal-**o**xide-**s**emiconductor) am häufigsten vor.

Da die Steuerung des Stroms im FET über ein elektrisches Feld erfolgt, fließt in den Steuereingang des FET (Gate G) quasi kein Strom. Folglich hat der FET einen sehr hohen Eingangswiderstand und wird nahezu leistungslos angesteuert.

Der Einsatz als Wechselspannungsverstärker beschränkt sich dabei fast nur auf die Vorverstärkerstufe im HF- und NF-Bereich. Das Hauptanwendungsgebiet der FET liegt jedoch bei den integrierten Schaltungen der Analogtechnik und insbesondere der Digitaltechnik, also Verwendung als Schalter.

Die Weiterentwicklung der FET hat zu Typen wie VMOSFET, SIPMOSFET und IGBT geführt, die die Einsatzgebiete ausgedehnt haben.

35.2.1 Aufbau und Wirkungsweise des Sperrschicht-FET

Sperrschicht-FET werden als P-Kanal- oder N-Kanal-Typ hergestellt. Ihre Schaltzeichen sind in Abb. 35.18 dargestellt. Im Folgenden wird ein *N-Kanal-Typ* beschrieben.

Bei einem *P-Kanal-Typ* sind lediglich die Kanal- und Gatedotierung sowie die Polaritäten der Spannungen zu vertauschen.

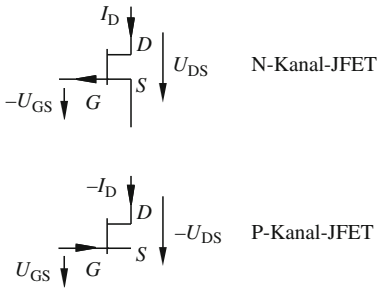


Abb. 35.18 Schaltzeichen eines Sperrschicht-FET vom N-Kanal-Typ oder P-Kanal-Typ

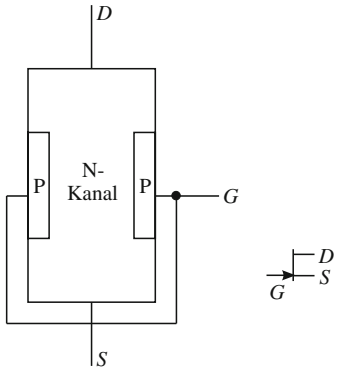


Abb. 35.19 Prinzipieller Aufbau eines N-Kanal-Sperrschicht-FET

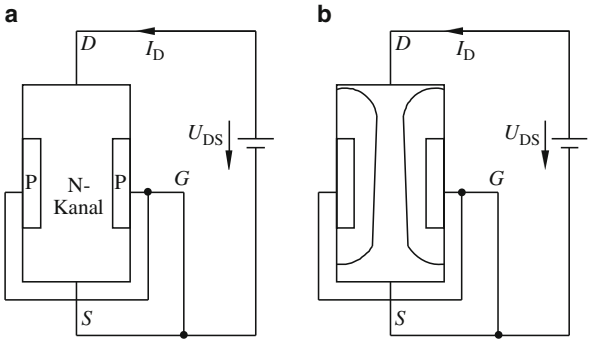


Abb. 35.20 Abhängigkeit der Kanalbreite von der Drain-Source-Spannung U_{DS}

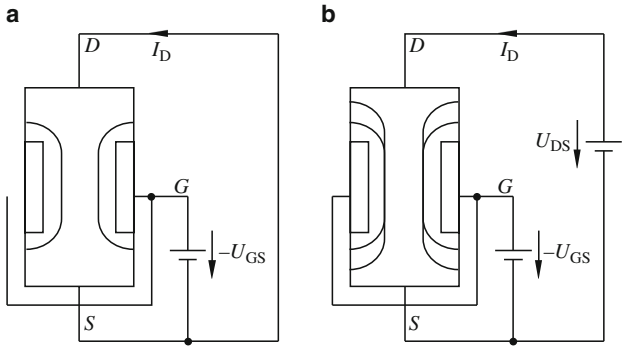


Abb. 35.21 Abhängigkeit der Kanalbreite von U_{DS} und der Steuerspannung U_{GS}

Abb. 35.19 zeigt den prinzipiellen Aufbau eines N-Kanal-Sperrschicht-FET mit den drei Anschlüssen Gate G, Drain D und Source S. Er besteht im Wesentlichen aus einem länglichen Stück n-dotierten Halbleitermaterial, dessen Anschlüsse an den Enden Drain (engl.: Abfluss, Senke) und Source (engl.: Quelle) genannt werden. In die Seiten wird p-dotiertes Halbleitermaterial eindiffundiert, so dass pn-Übergänge entstehen. Diese sind miteinander verbunden und bilden den Gate-Anschluss. Durch eine entsprechende Steuerspannung U_{GS} zwischen dem Gate- und Source-Anschluss wird der pn-Übergang in Sperrrichtung betrieben, so dass in den Steueranschluss lediglich ein geringer Sperrstrom fließt.

Der pn-Übergang darf nicht in Durchlassrichtung betrieben werden! Verbindet man nun das Gate mit dem Source und legt zwischen Drain und Source eine Spannung U_{DS} nach Abb. 35.20a, so bildet sich an beiden pn-Übergängen eine Sperrschicht. Dadurch reduziert sich die für die Leitfähigkeit wirksame Kanalbreite, so dass der Kanalwiderstand erhöht wird.

Wird die Spannung U_{DS} fortwährend erhöht, so wird der Kanal nach Abb. 35.20b schmaler, bis die beiden Sperrschichten aneinander stoßen und den Kanal abschnüren. Am Drain-Anschluss ist die Abschnürung deutlicher.

Dieser Wert der Spannung U_{DS} wird maximale Sättigungsspannung U_{DSS} genannt. Den zugehörigen Wert des Stroms I_D nennt man maximaler Sättigungsstrom I_{DS}. Für Werte $U_{DS} > U_{DSS}$ bleibt der Stromfluss erhalten.

Durch eine negative Steuerspannung U_{GS} nach Abb. 35.21a wird die Breite der Sperrschicht an den pn-Übergängen beeinflusst. Im Betrieb überlagern sich die Wirkungen der beiden Spannungen und zeigen die in Abb. 35.21b beschriebene Eingrenzung des Kanals.

Die Steuerspannung U_{GS} darf bei einem N-Kanal-PN-FET nicht positiv werden, da dieses zur Zerstörung des Bauelementes führen kann.

35.2.2 Aufbau und Wirkungsweise des MOSFET

Beim Sperrschicht-FET wird die Veränderung des wirksamen Kanalquerschnittes zur Einstellung des Widerstandes der Drain-Source-Strecke benutzt. Beim FET mit isoliertem Gate hingegen wird die Variation des Kanalwiderstandes mit Hilfe der Ladungsträgerdichte betrieben, wobei die Anzahl der geeigneten Ladungsträger zwischen Drain und Source durch ein elektrisches Feld gesteuert wird. Hierbei unterscheidet man generell zwei Versionen.

Der IG-FET vom Normally-On-Typ (auch Verarmungstyp, Depletion-Typ genannt) nach Abb. 35.22 ist genau wie der Sperrschicht-FET ein selbstleitender Typ. Bei einer Steuerspan-

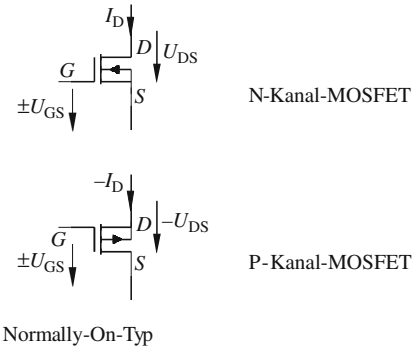


Abb. 35.22 Schaltzeichen selbstleitender IG-FET

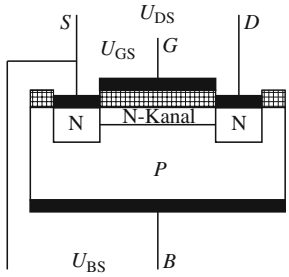


Abb. 35.23 Prinzipieller Aufbau eines selbstleitenden N-Kanal-MOS-FET

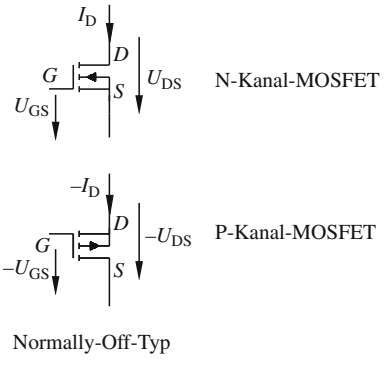


Abb. 35.24 Schaltzeichen selbstsperrender IG-FET

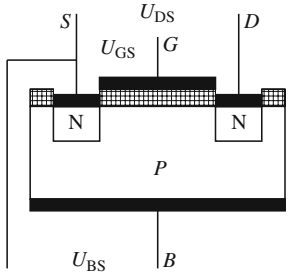


Abb. 35.25 Prinzipieller Aufbau eines selbstsperrenden N-Kanal-MOS-FET

nung $U_{GS} = 0\,V$ existiert bereits ein leitender Kanal zwischen Drain und Source, dessen Ladungsträgerdichte je nach Kanalart (N-Kanal oder P-Kanal) und der Polarität der Steuerspannung weiter erhöht oder verringert werden kann.

Der prinzipielle Aufbau eines selbstleitenden N-Kanal-MOS-FET ist in Abb. 35.23 dargestellt. Der Gate-Anschluss ist vom Halbleitermaterial durch eine Isolierschicht getrennt. Bei einer Steuerspannung $U_{GS} = 0\,V$ fließt durch den N-Kanal bereits ein Drainstrom. Durch eine negative Steuerspannung U_{GS} nimmt die Elektronendichte im Kanal ab, da die Elektronen durch das elektrische Feld abgedrängt werden. Damit nimmt aber der Widerstand zwischen Drain und Source zu. Bei $U_{GS} = U_P$ existiert kein N-Kanal mehr zwischen Drain und Source, das heißt, der Drainstrom I_D wird zu Null. Legt man eine positive Steuerspannung an das Gate, so wird die Ladungsträgerdichte im Kanal größer; die Leitfähigkeit des Kanals wird vergrößert.

Der IG-FET kann sowohl über die Gate-Source-Spannung als auch über eine Spannung zwischen Bulk B und Source S gesteuert werden. Der Einfluss dieser Spannung ist aber geringer, und diese Steuerungsmöglichkeit wird seltener genutzt. Meist sind Bulk und Source miteinander intern verbunden.

Die Funktion eines selbstleitenden P-Kanal-MOS-FET entspricht der des selbstleitenden N-Kanal-MOS-FET.

Der IG-FET vom Normally-Off-Typ (auch Anreicherungs-, Enhancement- oder selbstsperrender Typ genannt) nach Abb. 35.24 besitzt bei einer Steuerspannung $U_{GS} = 0$ V keinen leitfähigen Kanal zwischen Drain und Source. Er bildet sich erst, wenn die angelegte Steuerspannung einen Schwellenwert $U_{GS(TO)}$ überschreitet.

Der prinzipielle Aufbau eines selbstsperrenden N-Kanal-MOS-FET ist in Abb. 35.25 dargestellt. Bei einer Steuerspannung $U_{GS} = 0\,V$ fließt kein Drainstrom, da ein Kanal nicht durch entsprechende Dotierung eingebaut wurde.

Legt man eine positive Steuerspannung an das Gate, so werden Elektronen aus dem p-dotierten Substrat in Richtung Gate gezogen und führen dort zur Ausbildung eines N-Kanals. Der zum Fließen kommende Drainstrom wird umso größer, je größer die positive Steuerspannung wird.

Die Funktion eines selbstsperrenden P-Kanal-MOS-FET entspricht der des selbstsperrenden N-Kanal-MOS-FET.

35.2.3 Kennlinien von FET

Bei einem J-FET fließt in das Gate nur der Sperrstrom des pn-Übergangs, so dass eine Eingangskennlinie für den PN-FET nicht existiert. Interessant sind die Steuer- und Ausgangskennlinie eines J-FET nach Abb. 35.26.

Diesen Kennlinien können alle relevanten Gleichspannungs- und -stromwerte entnommen werden.

Abb. 35.27 zeigt die Steuerkennlinie $I_D = f(U_{GS})$ eines J-FET. Die Steuerkennlinie gilt für $U_{DS} = \text{const}$. Bei $U_{GS} = 0\,V$ hat

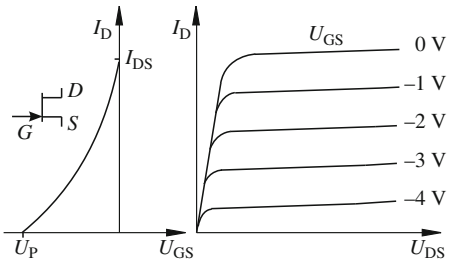


Abb. 35.26 Steuer- und Ausgangskennlinie eines J-FET

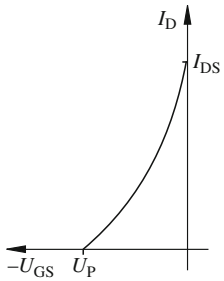


Abb. 35.27 Steuerkennlinie $I_D = f(U_{GS})$ eines J-FET

der Drainstrom I_D eines J-FET seinen höchsten Wert I_{DS} und verringert sich mit negativer werdender Steuerspannung. Die Spannung U_{GS}, bei der der Drainstrom auf Null reduziert ist, wird Abschnürspannung U_P (engl. pinch-off-voltage) genannt und liegt je nach Typ bei Werten zwischen 2 und 10 V.

Während bei den bipolaren Transistoren der Kollektorstrom mit steigender Temperatur ebenfalls ansteigt, wird der Drainstrom beim J-FET mit steigender Temperatur kleiner.

Dies hat seine Ursache darin, dass die frei beweglichen und durch Dotierung entstandenen Ladungsträger durch die thermische Eigenbewegung der Atome in ihrer Beweglichkeit gemindert werden. Mit steigender Temperatur wird aber auch die Abschnürspannung größer. Beide Effekte zusammen haben zur Folge, dass es einen Drainstromwert gibt, der temperaturunabhängig ist, das heißt, der Temperaturkoeffizient wird Null. Abb. 35.28 zeigt den Temperatureinfluss auf die Übertragungskennlinie.

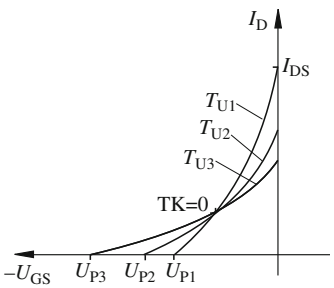


Abb. 35.28 Temperatureinfluss auf die Übertragungskennlinie

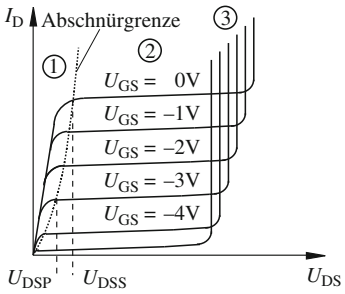


Abb. 35.29 Ausgangskennlinienfeld mit U_{GS} als Parameter

Die Gate-Source-Spannung U_{GSK} für den temperaturunabhängigen Kompensationspunkt lässt sich in Näherung berechnen.

$$\textit{Gate-Source-Spannung} \quad U_{GSK} \approx U_P + 0{,}7\,\text{V}$$

Soweit möglich sollte dieser Wert mit der Steuerspannung eingestellt werden, um eine thermische Arbeitspunktstabilisierung unnötig zu machen. Von Nachteil ist hier, dass die Kennlinie keine Gerade ist und somit das Ausgangssignal verzerrt sein kann.

Das Ausgangskennlinienfeld (I_D-U_{DS}-Kennlinie mit U_{GS} als Parameter) nach Abb. 35.29 lässt sich mit Unterstützung durch Abb. 35.26 qualitativ erklären.

Steigert man die Spannung U_{DS} von Null an, so steigt der Strom I_D bei U_{GS} = const zunächst linear mit der Spannung, der Kanalwiderstand bleibt in etwa konstant (ohmscher Bereich). Mit wachsendem Strom beginnt sich der Kanal merklich zu verengen, der Kanalwiderstand nimmt zu, und der Anstieg von I_D mit U_{DS} wird immer geringer, bis beide Sperrschichten sich nahezu berühren. Man sagt, der Kanal ist abgeschnürt.

Die Drain-Source-Spannung U_{DSP}, bei der diese Abschnürung (pinch off) eintritt, ist in Abb. 35.29 eingezeichnet. U_{DSP} wird auch Kniespannung genannt.

$$\textit{Abschnürspannung} \quad U_{DSP} \equiv U_{DSS} + U_{GS} \qquad (35.11)$$

Mit wachsender Steuerspannung U_{GS} wird die Abschnürgrenze früher, das heißt bei kleinerem Strom I_D und kleineren Werten von U_{DSP} erreicht. Die $I_D - U_{DS}$-Kennlinien verschieben sich daher mit zunehmender Gate-Source-Sperrspannung zu kleineren Drainstromwerten.

Verbindet man alle Abschnürpunkte nach Abb. 35.29 miteinander, so erhält man die *Abschnürgrenze*, die den ohmschen Bereich (Bereich 1) vom aktiven Bereich (Bereich 2) – auch Sättigungsbereich genannt – trennt. Oberhalb von U_{DSP} steigt der Drainstrom nur noch wenig an.

Wird die Spannung U_{DS} zu weit erhöht, so kommt es zu einem Durchbruch (Bereich 3) zwischen Drain und Source, wobei der Drainstrom so stark ansteigt, dass die Gefahr der Zerstörung des FET eintritt.

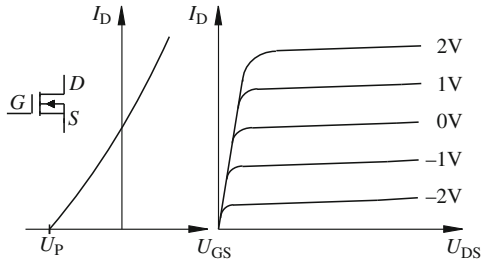


Abb. 35.30 Steuer- und Ausgangskennlinie eines selbstleitenden N-Kanal-MOS-FET

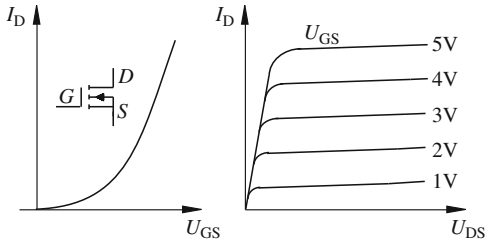


Abb. 35.31 Steuer- und Ausgangskennlinie eines selbstsperrenden N-Kanal-MOS-FET

Da der J-FET eine quadratische Abhängigkeit des Drainstroms von der Steuerspannung hat, sind seine nichtlinearen Verzerrungen bei einer Spannungsaussteuerung wesentlich geringer als beim bipolaren Transistor.

Für das Betreiben des J-FET im aktiven Bereich, sprich Verstärkerbereich, lässt sich somit die Steilheit S errechnen. Praktische Werte für die maximale Steilheit liegen zwischen 1 mA/V und 50 mA/V. Die Steilheit S ist ein direktes Maß für die Spannungsverstärkung einer J-FET-Verstärkerstufe. Je größer S, umso größer ist auch die zu erwartende Spannungsverstärkung V_u.

Im ohmschen Bereich kann der J-FET als spannungsgesteuerter Widerstand eingesetzt werden.

Bei den IG-FET fließt kein Strom in das Gate, da es durch eine Isolationsschicht vom übrigen Halbleiter getrennt ist, so dass auch hier keine Eingangskennlinie existiert.

Die Steuer- und Ausgangskennlinie eines selbstleitenden N-Kanal-MOS-FET, also vom Normally-On-Typ, zeigt Abb. 35.30.

Diesen Kennlinien können alle relevanten Gleichspannungs- und Stromwerte entnommen werden. Da auch der PN-FET ein selbstleitender FET ist, ergeben sich hier generell zahlreiche Übereinstimmungen im Verhalten.

Für die Ausgangskennlinie ($I_D - U_{DS}$-Kennlinie) gelten im Prinzip die Aussagen zum J-FET. Hier wird die Spannung U_{GS} als Parameter verwendet, während die Drain-Source-Spannung verändert wird.

Deutlich erkennbar ist die Abhängigkeit des Drainstroms I_D vom Wert der negativen *und* positiven Steuerspannung U_{GS}. Der Normally-On-Typ kann also sowohl als Verarmungs-FET als auch wie ein Anreichherungs-FET betrieben werden.

Der Drainstrom kann durch eine positive Gatespannung U_{GS} nicht weiter beliebig erhöht werden, weil ab einem bestimmten Spannungswert eine Vergrößerung der Leitfähigkeit des Kanals nicht mehr möglich ist.

Die Steuerkennlinie (I_D-U_{DS}-Kennlinie) des selbstleitenden N-Kanal-MOS-FET zeigt, dass bereits bei $U_{GS} = 0\,\text{V}$ ein Drainstrom fließt. Selbstleitende MOS-FET haben wie die J-FET einen geringen positiven Temperaturkoeffizienten. Der Drainstrom wird also mit steigender Temperatur kleiner. Die Gate-Source-Spannung U_{GSK} für den temperaturunabhängigen Kompensationspunkt lässt sich in Näherung berechnen.

$$\textit{Gate-Source-Spannung} \quad U_{GSK} \approx U_P + 2{,}5\,\text{V} \qquad (35.12)$$

Die Steuer- und Ausgangskennlinie eines selbstsperrenden N-Kanal-MOS-FET, also vom Normally-Off-Typ, zeigt Abb. 35.31. Diesen Kennlinien können alle relevanten Gleichspannungs- und Stromwerte entnommen werden.

Die Schwellenspannung $U_{GS(TO)}$ kann als die Spannung an der Gateelektrode verstanden werden, die die Störstellenladung und die Raumladung kompensiert.

Sie ist nur geringfügig von der Temperatur abhängig. Darum lässt sich die Steuerspannung U_{GSK} für einen temperaturkompensierten Arbeitspunkt gut näherungsweise berechnen.

$$\textit{Steuerspannung} \quad U_{GSK} \approx U_{GS(TO)} + 2{,}5\,\text{V} \qquad (35.13)$$

Je größer die Steuerspannung U_{GS} ist, umso größer wird der Drainstrom. Ein maximaler Sättigungsstrom lässt sich daher nicht angeben.

35.2.4 Kennwerte von FET

Wie bei den bipolaren Transistoren und Dioden unterscheidet man auch hier Kenn- und Grenzwerte. Für die Kenngrößen werden in den Datenblättern je nach Bedarf Mittelwerte und/oder obere und/oder untere Streuwerte angegeben. Die Kennliniendarstellungen sind, falls nicht anders vermerkt, Mittelwert-Darstellungen und gelten nur für die jeweils angegebene Temperatur. J-FET können auf verschiedene Weise elektrisch überlastet werden:

Drain-Source-Spannung	$U_{DS} > U_{DS\,max}$
Drain-Gate-Spannung	$U_{DG} > U_{DG\,max}$
Gate-Source-Spannung	$U_{GS} > U_{GS\,max}$
Drain-Strom	$I_D > I_{D\,max} = I_{DSS}$
Gate-Strom	$I_G > I_{G\,max}$
Verlustleistung	$P_V > P_{tot}$
Sperrschichttemperatur	$\vartheta_J > \vartheta_{J\,max}$

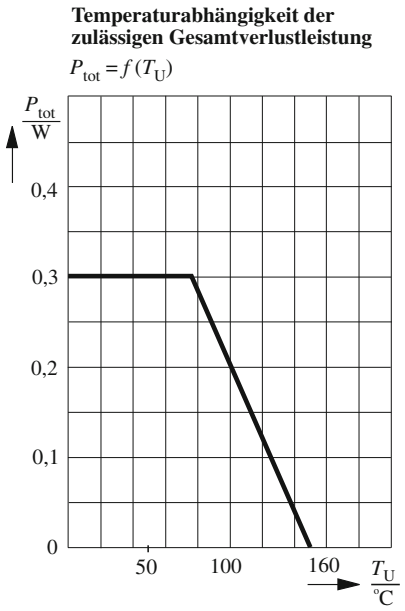


Abb. 35.32 Temperaturabhängigkeit der zulässigen Verlustleistung

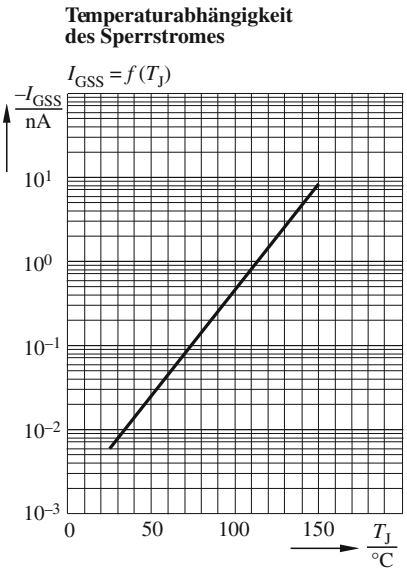


Abb. 35.33 Temperaturabhängigkeit des Sperrstroms

Liegt zwischen dem Kanal- und dem Gate-Anschluss eine zu hohe Spannung, kann der in Sperrrichtung betriebene pn-Übergang zerstört werden. Die Spannungswerte $U_{DS\,max}$, $U_{DG\,max}$ und $U_{GS\,max}$ dürfen nicht überschritten werden und liegen bei den heute verwendeten N-Kanal-J-FET etwa zwischen 20 und 30 V.

Der durch das Gate fließende Sperrstrom ist normalerweise vernachlässigbar klein. Bei kurzzeitigem Betrieb des pn-Überganges in Durchlassrichtung darf der dann fließende Strom I_G den Wert $I_{G\,max}$ nicht überschreiten. I_G liegt meistens in der Größenordnung von 10 mA.

Beim Betrieb eines FET liegt zwischen Drain und Source die Spannung U_{DS}, und es fließt der Strom I_D. Daraus ergibt sich eine Verlustleistung, die im FET zur Erwärmung des Kristalls führt. Diese Verlustleistung P_V muss kleiner als die größte zulässige Verlustleistung P_{tot} sein.

totale Verlustleistung

$$PV = U_{DS} \cdot I_D \leq P_{tot} \tag{35.14}$$

Hauptsächlich durch die Verlustleistung im Dauerbetrieb entsteht in der Sperrzone und im Kanal Wärme, durch die sich die Temperatur im Bauelement erhöht. Die in den Datenblättern angegebene Sperrschichttemperatur ϑ_J bezieht sich auf die höchste im Kanal auftretende Kanaltemperatur $\vartheta_{J\,max}$. Bei den üblichen FET beträgt $\vartheta_{J\,max}$ etwa 150 bis 200 °C.

In den Datenblättern einiger Hersteller wird ein Diagramm für die Temperaturabhängigkeit der zulässigen Verlustleistung angegeben. In Abb. 35.32 kann direkt abgelesen werden, wie groß P_{tot} bei einer bestimmten Umgebungstemperatur ist.

Mit den statischen Kenndaten wird das Gleichstromverhalten eines FET beschrieben. Die Transistoren werden nach dem Drain-Source-Kurzschlussstrom I_{DSS} ausgesucht und eventuell in Gruppen A, B und C eingeteilt. Der Strom I_{DSS} ist der Drainstrom, der bei einem selbstleitenden FET mit Drain-Source-Kurzschluss fließt.

Der pn-Übergang wird zwischen Gate und Kanal in Sperrrichtung betrieben. Trotzdem fließt hier ein Gate-Reststrom $-I_{GSS}$. Wie Abb. 35.33 zeigt, hängt die Größe dieses Sperrstroms von der Temperatur ab.

Um einen FET zu sperren, ist die Gate-Source-Abschnürspannung $-U_P$ erforderlich. Diese Spannung ist arbeitspunktabhängig und unterliegt einer großen Streuung.

Den statischen Eingangswiderstand R_{GS} des gesperrten pn-Übergangs kann man in Abhängigkeit von der Spannung U_{GS} aus dem entsprechenden Diagramm in den Datenblättern herauslesen. Er liegt in der Größenordnung von 10^9 bis $10^{11}\,\Omega$ und kann in vielen Anwendungsfällen vernachlässigt werden.

Für den N-Kanal-J-FET ergeben sich etwa die folgenden praktischen Richtwerte:

Drain-Source-Spannung	$U_{DS\,max} \approx 25$ bis 30 V
Gate-Source-Spannung	$U_{GS\,max} \approx -8$ V
Drain-Strom	$I_{D\,max} \approx 25$ mA
Gate-Strom	$I_{G\,max} \approx 10$ mA
Verlustleistung	$P_{tot} \approx 200$ mW
Sperrschichttemperatur	$\vartheta_{J\,max} \approx 125$ °C

Aufgrund des vorhandenen pn-Übergangs zwischen Bulk und den anderen Anschlüssen müssen die Grenzwerte für diese

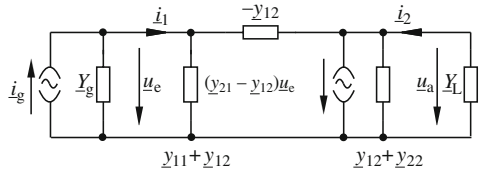


Abb. 35.34 AC-Ersatzschaltbild eines FET

Spannungen beachtet werden.

Drain-Bulk-Spannung	$U_{DB} > U_{DB\,max}$
Gate-Bulk-Spannung	$U_{GB} > U_{GB\,max}$

Infolge des isolierten Gate ergibt sich für den MOS-FET ein sehr hoher Eingangswiderstand r_{GS} in der Größenordnung von etwa $10^{15}\,\Omega$. Statische Aufladungen können dementsprechend nicht entladen werden, was zu sehr hohen Feldstärken führt, die bei Überschreitung einer bestimmten Feldstärke zu einem Durchschlag und damit zur Zerstörung des MOS-FET führen.

Für den N-Kanal-MOS-FET ergeben sich etwa die folgenden praktischen Richtwerte:

Drain-Source-Spannung	$U_{DS\,max} \approx 22$ bis 35 V
Drain-Strom	$I_{D\,max} \approx 20$ bis 50 mA
Verlustleistung	$P_{tot} \approx 150$ bis 200 mW
Sperrschichttemperatur	$\vartheta_{J\,max} \approx$ bis 200 °C
Drain-Bulk-Spannung	$U_{DB\,max} \approx 30$ bis 35 V

Bei der Kleinsignalaussteuerung eines FET können die Kennlinien in einem Bereich um den Arbeitspunkt durch Geraden angenähert werden. Der FET verhält sich dann wie ein linearer, aktiver Vierpol.

Für diesen Vierpol kann das AC-Ersatzschaltbild bei niedrigen, aber auch bei hohen Frequenzen bestimmt werden. Um das Wechselstromverhalten eines FET zu beschreiben, werden seine dynamischen Kennwerte nach Abb. 35.34 angegeben.

Üblich ist es, an Stelle der Hybrid-Koeffizienten (*h*-Parameter) die *y*-Leitwertparameter anzugeben. Sie sind bei FET sinnvoller, da die Stromverstärkung für FET's nicht angegeben werden kann.

Die *y*-Parameter haben komplexe Werte, das heißt, die Ein- und Ausgangsspannungen u_1 und u_2 beziehungsweise die Ein- und Ausgangsströme i_1 und i_2 sind ebenfalls komplexe Größen.

Vierpolgleichungen

$$\begin{aligned} \underline{i}_1 &= \underline{y}_{11} \cdot \underline{u}_e + \underline{y}_{12} \cdot \underline{u}_a \\ \underline{i}_2 &= \underline{y}_{21} \cdot \underline{u}_e + \underline{y}_{22} \cdot \underline{u}_a \end{aligned} \qquad (35.15)$$

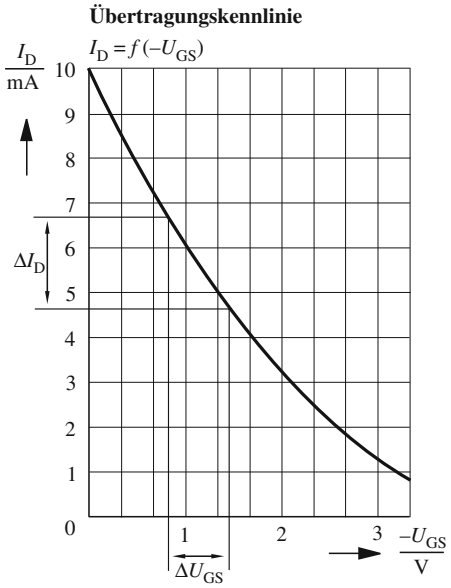


Abb. 35.35 Übertragungskennlinie

Aus der Steuerkennlinie nach Abb. 35.35 kann grafisch die Vorwärtssteilheit y_{21} ermittelt werden. Hierbei wird die Steilheit der Kennlinie im Arbeitspunkt bei $U_{DS} = \text{const}$ ermittelt.

$$\textit{Vorwärtssteilheit} \quad y_{21} = \frac{\Delta I_D}{\Delta U_{GS}} \qquad (35.16)$$

Da eine Annäherung durch eine Gerade im AP häufig zu ungenau ist, steht meistens ein entsprechendes Diagramm nach Abb. 35.36 zur Verfügung, aus dem die Steilheit für jeden Drainstrom und die A-, B- oder C-Typen ablesbar ist.

Wichtig ist in diesem Zusammenhang der differentielle Ausgangswiderstand r_{ds}. Er ist der Kehrwert des Ausgangsleitwertes y_{22}.

$$\textit{Ausgangsleitwert} \quad y_{22} = \frac{\Delta I_D}{\Delta U_{DS}} \qquad (35.17)$$

Der Ausgangsleitwert beschreibt die Steilheit der Ausgangskennlinie im Arbeitspunkt für $U_{GS} = \text{const}$. Seine Bestimmung wird im Diagramm nach Abb. 35.37 dargestellt.

Da eine Annäherung durch eine Gerade im AP häufig zu ungenau ist, steht i. d. R. ein entsprechendes Diagramm nach Abb. 35.38 zur Verfügung, aus dem der dynamische Drain-Source-Widerstand für jede Gate-Source-Spannung und die A-, B- oder C-Typen ablesbar ist.

Erheblichen Einfluss haben im HF-Bereich die dort wirksam werdenden Kapazitäten, die beim J-FET durch den pn-Übergang und bei den IG-FET durch die Isolierschichten entstehen. Die Kapazitäten sind es auch, die die *y*-Parameter komplex werden lassen. Wir haben Ein- und Ausgangskapazität sowie die Rückwirkungskapazität nach Abb. 35.39 zu berücksichtigen.

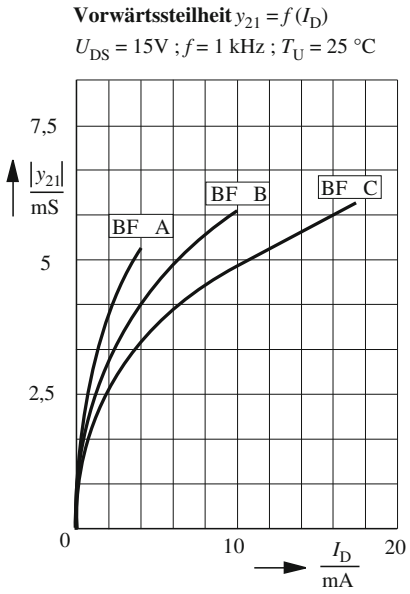


Abb. 35.36 Vorwärtssteilheit

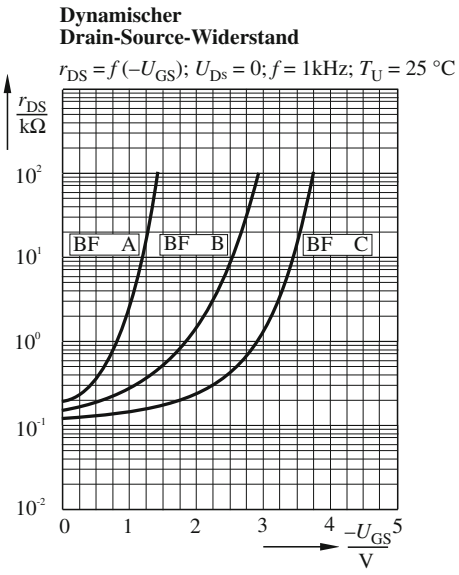


Abb. 35.38 Dynamischer Drain-Source-Widerstand

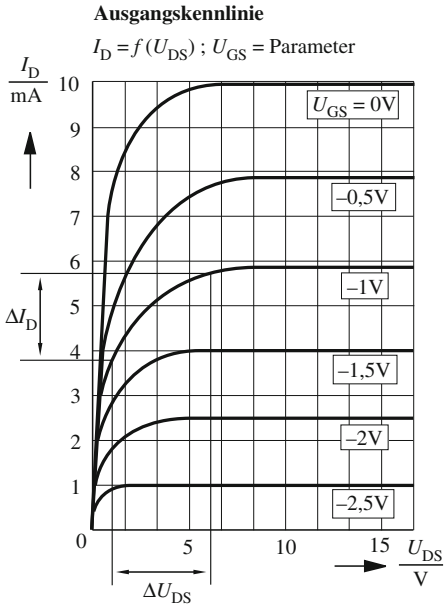


Abb. 35.37 Ausgangskennlinien

In den Datenblättern der Hersteller werden *arbeitspunktbezogen die Werte für die Kurzschluss-Eingangskapazität* C_{iss}, die *Kurzschluss-Ausgangskapazität* C_{oss} und die *Rückwirkungskapazität* C_{rss} angegeben.

Analog zum bipolaren Transistor lassen sich auch für den FET die *Betriebsgrößen des Vierpols*, nämlich die *Wechselspannungsverstärkung sowie die Ein- und Ausgangsleitwerte* anhand des AC-Ersatzschaltbildes nach Abb. 35.34 wie folgt berechnen.

Die betriebliche Spannungsverstärkung des Vierpols ist das Verhältnis von Ausgangsspannung zur Eingangsspannung.

Spannungsverstärkung

$$\underline{V}_u = \frac{\underline{u}_a}{\underline{u}_e} = -\frac{\underline{y}_{21}}{\underline{y}_{22} + \underline{Y}_L} \tag{35.18}$$

Der Eingangsleitwert des Vierpols ist das Verhältnis von Eingangsstrom zur Eingangsspannung.

Eingangsleitwert

$$\underline{y}_i = \underline{y}_{11} + \frac{\underline{y}_{12} \cdot \underline{y}_{21}}{\underline{y}_{22} + \underline{Y}_L} = \underline{y}_{11} + \underline{y}_{12} \cdot \underline{V}_u \tag{35.19}$$

Der Ausgangsleitwert des Vierpols ist das Verhältnis von Ausgangsstrom zur Ausgangsspannung. Um ihn bestimmen zu können, wird an den Ausgang eine Spannung u_a angelegt, während die Stromquelle eine Unterbrechung darstellt.

Ausgangsleitwert

$$\underline{y}_0 = \frac{\underline{i}_2}{\underline{u}_a} = \underline{y}_{22} - \frac{\underline{y}_{12} \cdot \underline{y}_{21}}{\underline{y}_{11} + \underline{Y}_g} \tag{35.20}$$

Aufgrund der spezifischen Kennwerte von J-FET lässt sich das *Wechselstromersatzschaltbild stark vereinfachen* und für die NF- und HF-Bereiche getrennt darstellen in Abb. 35.40.

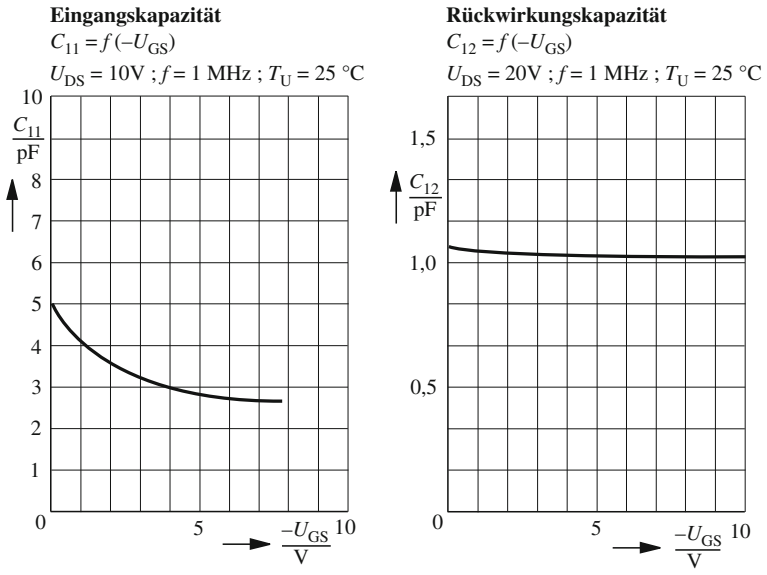


Abb. 35.39 Kennlinien der Eingangs- und Rückwirkungskapazität

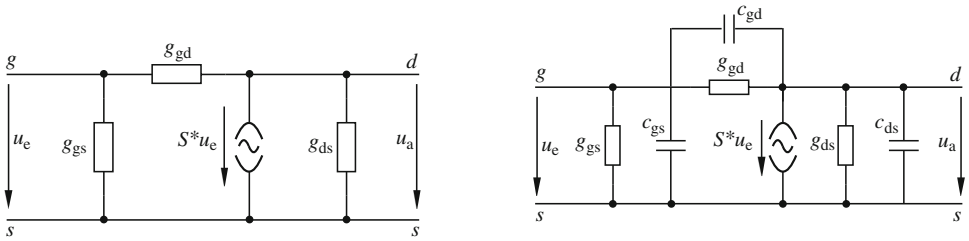


Abb. 35.40 AC-Ersatzschaltbilder für den NF- und HF-Bereich von J-FET

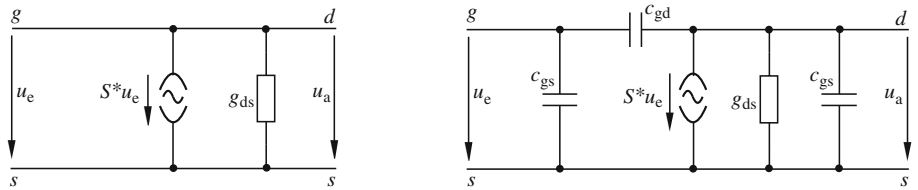


Abb. 35.41 AC-Ersatzschaltbilder für den NF- und HF-Bereich von MOS-FET

Die Daten von FET mit isoliertem Gate unterscheiden sich nur unwesentlich von J-FET. Die Steuerkennlinien unterscheiden sich in der Lage zum Nullpunkt, der Kennlinienverlauf ist ähnlich. Auch für MOS-FET lässt sich das Wechselstromersatzschaltbild stark vereinfachen und für die NF- und HF-Bereiche getrennt darstellen in Abb. 35.41.

Aufgrund der Isolierschicht zwischen Kanal und Gate sind beim MOS-FET die Leitwerte g_{gs} und g_{gd} zwischen Gate und Source beziehungsweise Gate und Drain so gering, dass sie vernachlässigt werden können. Damit ergibt sich dieses AC-Ersatzschaltbild.

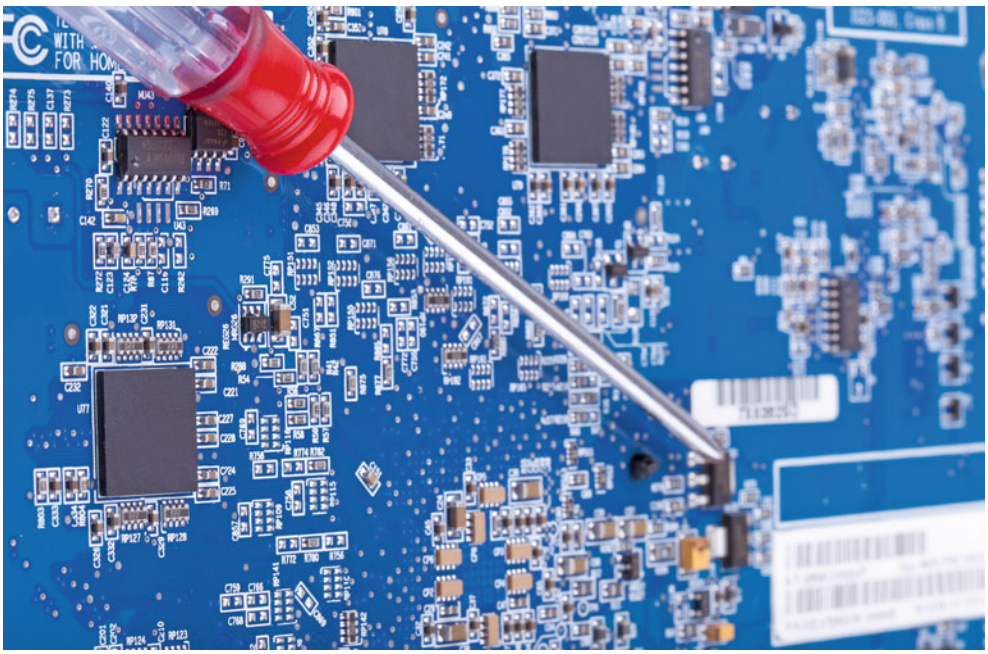
Bei höheren Frequenzen wird die *Gate-Kanal-Kapazität durch die Kondensatoren* c_{gs} und c_{gd} *berücksichtigt*. Der Kanal wird in diesem Ersatzschaltbild durch den Ausgangsleitwert g_{ds}, die Kapazität c_{ds} und die Stromquelle $S \cdot u_e$ elektrisch dargestellt.

Vergleicht man diese AC-Ersatzschaltbilder mit dem aktiven Vierpol nach Abb. 35.34, so lassen sich die *y-Parameter, also die Leitwertparameter, ermitteln*. Für die Sourceschaltung von FETs werden diese *y*-Parameter mit einem zusätzlichen Index *s* gekennzeichnet, wie wir es schon von den Bipolaren Transistoren her kennen.

Besondere Halbleiter-Bauelemente

36

Peter Döring



36.1 Unijunction-Transistor (Doppelbasisdiode) 442
36.2 Darlington-Transistor 443
36.3 VMOS-Transistoren 444
36.4 SIPMOS-Transistoren 444
36.5 IGBT 446

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_36

36.1 Unijunction-Transistor (Doppelbasisdiode)

Unijunction-Transistoren (UJTs) werden aus einem homogenen, n-dotierten Si-Kristall als Planartyp oder Legierungstyp nach Abb. 36.1 hergestellt. An zwei gegenüberliegenden Seiten sind sperrschichtfreie Anschlüsse angebracht, die als Basis 1 (B1) und Basis 2 (B2) bezeichnet werden. Unsymmetrisch zu diesen Basisanschlüssen ist als Emitter (E) eine p-Zone angeordnet. Dadurch entsteht ein pn-Übergang, der die Funktion einer Diode hat.

Aus der physikalischen Wirkungsweise lässt sich das Ersatzschaltbild des UJT nach Abb. 36.2 herleiten, wobei sich zwei interne Vorgänge überlagern.

Legt man zwischen B2 und B1 eine positive Spannung U_{BB} bei offenem Emitteranschluss, so fließt ein relativ kleiner Elektronenstrom von B1 nach B2, der von der Größe der Spannung U_{BB} und der des statischen Interbasiswiderstandes R_{BB} abhängt. Wird dagegen eine Spannung U_{EB1} zwischen E und B1 gelegt ohne eine Spannung U_{BB}, so zeigt diese Strecke das Verhalten einer Diode.

In einer Schaltung liegen aber beide Spannungen am UJT. Für die Betriebsspannung wirkt der UJT wie ein Spannungsteiler, der die Spannung U_{BB} im Verhältnis der Teilwiderstände aufteilt. Der mit „A“ gekennzeichnete Punkt in der Schaltung liegt also auf einem Ruhepotential U_A.

Solange die Eingangsspannung U_{EB1} kleiner ist als U_A, wird die Diode in Sperrrichtung betrieben und es fließt ein sehr kleiner Sperrstrom. Wird die Eingangsspannung U_{EB1} aber um die Schwellenspannung der Ersatzdiode größer als U_A, so wird die Diode in Durchlassrichtung betrieben und es fließt ein Emitterstrom I_E.

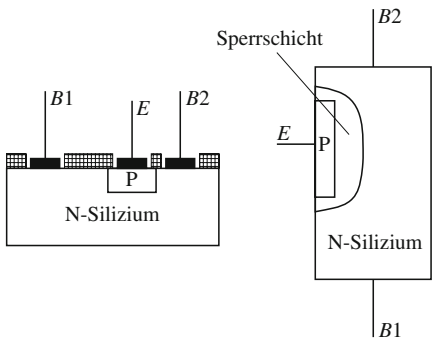


Abb. 36.1 Schichtfolge beim UJT als Planartyp oder Legierungstyp

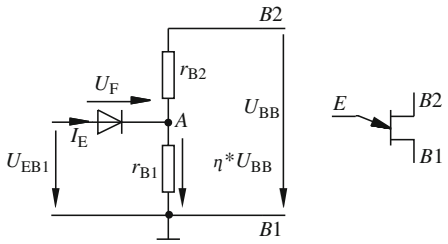


Abb. 36.2 Ersatzschaltbild und Schaltzeichen eines UJT

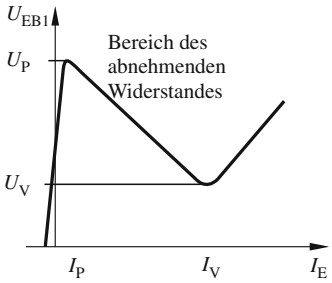


Abb. 36.3 Kennlinie eines UJT

Höckerspannung

$$U_{EB1} \geq 0{,}7\,\mathrm{V} + U_A = U_{EP} \tag{36.1}$$

Wird $U_{EB1} > U_{EP}$, so fließt ein Emitterstrom und es werden Löcher in das n-dotierte Silizium injiziert. Damit vergrößert sich die Leitfähigkeit der Strecke zwischen E und B1. Daher fließt jetzt bei einer geringeren Spannung U_{EB1} ein größerer Strom I_E als vorher bei einer größeren Spannung U_{EB1}.

Mit dem Talpunkt („Valley“) beginnt der Sättigungsbereich, weil eine weitere Verkleinerung von r_{B1} durch Ladungsträgerinjektion nicht mehr möglich ist.

Die UJT-Kennlinie nach Abb. 36.3 geht daher in die Kennlinie einer normalen Si-Dioden-Kennlinie über, und der Strom I_E steigt mit steigender Spannung U_{EB1} wieder an.

Sowohl das Umschalten vom hochohmigen Zustand in den niederohmigen Zustand als auch das Umschalten vom niederohmigen in den hochohmigen Zustand erfolgt sehr schnell, wenn die Spannung U_{EB1} die Höckerspannung U_{EP} überschreitet beziehungsweise unter den Wert der Talspannung U_{EV} absinkt, also $U_{EB1} < U_{EV}$ wird.

Die Hersteller geben eine charakteristische Kennlinie nach Abb. 36.3 an, in der die Definitionen der einzelnen Kennwerte angegeben sind.

Grenzwerte: Zu hohe Spannung U_{BB}; zu hohe Spannung $-U_{EB1}$; zu hoher Emitterstoßstrom I_E; zu hohe Gesamtverlustleistung P_V; zu hohe Sperrschichttemperatur ϑ_J.

Statische Kenndaten: Interbasiswiderstand r_{BB}; Temperaturbeiwert α.

Aus dem Kennlinienfeld nach Abb. 36.4 ist zu erkennen, dass sowohl Höckerspannung und Talspannung als auch der Talstrom von der Spannung U_{BB} abhängig sind.

Anwendungen: Wegen ihres Schaltverhaltens werden UJTs hauptsächlich in Sägezahn- oder Zündgeneratoren in der Schaltung nach Abb. 36.5 für Thyristoren beziehungsweise Triacs eingesetzt. Die Spannung u_C wirkt in dieser Schaltung als Steuerspannung des UJT.

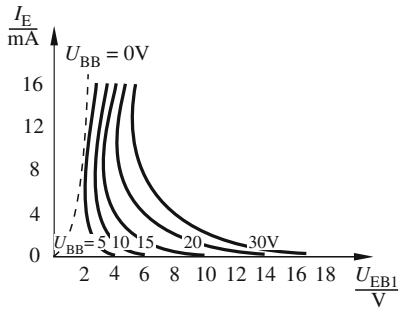


Abb. 36.4 Kennlinienfeld eines handelsüblichen UJT

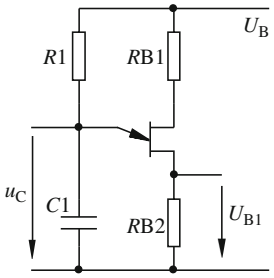


Abb. 36.5 Sägezahnspannungsgenerator beziehungsweise Zündgenerator mit UJT

Der Kondensator C wird über R1 geladen, und die Spannung u_C steigt entsprechend einer e-Funktion nach Abb. 36.6 an. Sobald $u_C > U_{EP}$ wird, zündet der UJT, und es fließt sowohl der Entladestrom des Kondensators als auch der Transistorstrom über den UJT ab. Aus diesem Grunde treten an R_{B2} Spannungsimpulse auf. Wird nun u_C kleiner als die Talspannung U_{EV}, so sperrt der UJT wieder und ein neuer Aufladevorgang des Kondensators beginnt.

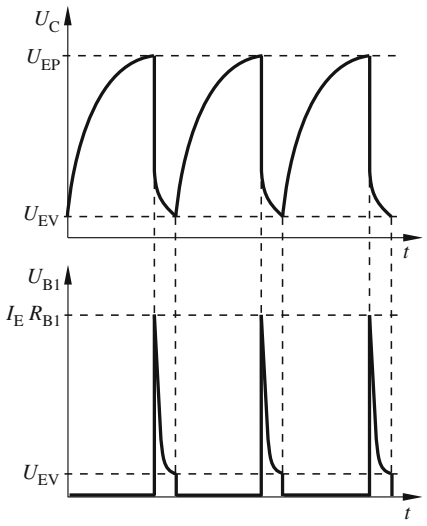


Abb. 36.6 Ein- und Ausgangsspannungen der Schaltung nach Abb. 36.5

36.2 Darlington-Transistor

Eine Möglichkeit, höhere Gleichstrom-Verstärkung zu erhalten, bietet der Darlington-Transistor, der in der Schaltung in Abb. 36.7 zu sehen ist. Der zweite Transistor arbeitet hier als Lastwiderstand für den ersten Transistor, der als Emitterfolger (Kollektorschaltung, siehe Kap. 34) geschaltet ist. Durch eine derartige Kopplung zweier Transistoren mit gleicher Schichtfolge ergeben sich für den Darlington-Transistor interessante Werte für den Eingangswiderstand und den Ausgangswiderstand der Schaltung.

Eingangswiderstand

$$R_{\text{ein}} = r_{\text{BE}} + (\beta + 1)(R_E \parallel R_A) \tag{36.2}$$

Ausgangswiderstand

$$R_{\text{aus}} = \frac{r_{\text{BE}} + R_g}{\beta + 1} \tag{36.3}$$

mit $h_{11} = r_{\text{BE}} = r_{\text{BE1}} + \beta_1 \cdot r_{\text{BE2}}$ und $\beta = \beta_1 \cdot \beta_2$.

Schleusenspannung $U_{\text{BE}} = U_{\text{BE1}} + U_{\text{BE2}}$

Interessant wird der Darlington-Transistor neben hohem Eingangswiderstand und kleinem Ausgangswiderstand nach Abb. 36.8 durch seine hohe Stromverstärkung, die i. d. R. bei $B > 2000$ liegt. Dieser Transistor gilt aus diesem Grunde als DC-Verstärker und wird insbesondere in Leistungsverstärkern (siehe Kap. 38) eingesetzt.

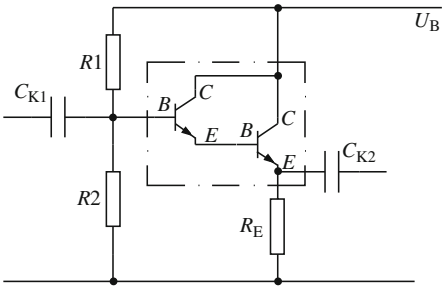


Abb. 36.7 Verstärker in Kollektorschaltung für höhere Gleichstrom- und Wechselstrom-Verstärkung

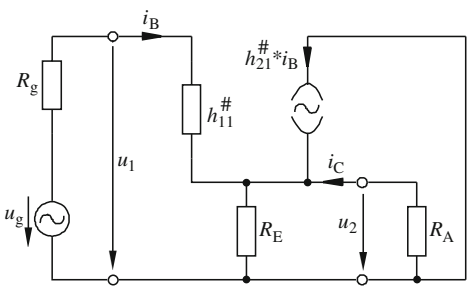


Abb. 36.8 AC-Ersatzschaltbild eines Darlington-Transistors

Werden zwei Transistoren mit unterschiedlicher Schichtfolge zusammengeschaltet, so entsteht ein Komplementär-Darlington-Transistor. Das Verhalten eines Darlington-Transistors wird im Wesentlichen durch die Eigenschaften des ersten Transistors bestimmt. Komplementär-Darlington-Transistoren haben den Vorteil, dass sie nur eine sehr kleine Schleusenspannung U_{BE} haben.

Darlington-Transistoren beziehungsweise Komplementär-Darlington-Transistoren können durch zusammengeschaltete Einzeltransistoren realisiert werden, in den meisten Fällen empfiehlt sich jedoch die Benutzung eines fertig konfektionierten Typs, wie er von zahlreichen Herstellern (zum Beispiel Siemens) angeboten wird.

36.3 VMOS-Transistoren

Die Grenzwerte des Drainstroms I_{DM} und der Drain-Source-Spannung U_{DSS} begrenzen die mit einem Feldeffekt-Transistor zu steuernde Leistung. Die höchste mit einem FET zu steuernde Leistung ergibt sich also als $P_{max} = I_{DM} \cdot U_{DSS}$.

Höhere Grenzwerte des Drainstroms I_{DM} erfordern eine höhere Steilheit. Die lässt sich aber nur durch eine Verkürzung des Kanals erreichen. Eine Verkürzung des Kanals geht aber zu Lasten der Drain-Source-Spannung U_{DSS}. In der Konsequenz bedeutet dies aber, dass die maximal zu steuernde Leistung gleich bleibt. Offensichtlich lässt sich mit der bisher beschriebenen Technologie ein FET zur Steuerung höherer Leistungen nicht verwirklichen.

Abb. 36.9 beschreibt den grundsätzlichen Schichtaufbau eines VMOS-FET (**v**ertical-**m**etal-**o**xide-**s**emiconductor-**f**ield-**e**ffekt-**t**ransistor). In eine Vier-Lagen-Struktur ist ein V-förmiger Graben eingeätzt. Hier fließt der Drainstrom I_D senkrecht, also vertikal durch durch den Kristall. Bei positiver Ansteuerung des Gate bildet sich ein durchgehender Strompfad beiderseits des Grabens mit einem sehr niedrigen Bahnwiderstand aus, so dass der Drainstrom, durch den V-förmigen Einschnitt geteilt, zu beiden Source-Anschlüssen fließt. Somit ist der Kanalquerschnitt groß und die Kanallänge kurz.

Im Ergebnis bedeutet dies ein Bauteil mit hoher zulässiger Verlustleistung bei hoher Spannungsfestigkeit und großem Verstärkungsfaktor. Der selbstsperrende N-Kanal-VMOS-FET weist praktisch keine Schaltverzögerung auf.

Tab. 36.1 Unterschied zwischen Standard-MOS-FET und Leistungs-MOS-FET

	MOS-FET	VMOS-FET
Drainstrom I_D	mA	A
Einschaltwiderstand $R_{DS\,(On)}$	100 … 500 Ω	0,1 … 10 Ω
Gate-Source-Spannung U_{GS}	0 … 5 V	0 … 10 V

Die Source-Elektrode überbrückt die obere N- und P-Schicht, was eine parasitäre Diode zwischen Gate und Source ergibt. Bei Anwendungen mit induktiver Last funktioniert die Diode als Freilaufdiode, so dass eine zusätzliche Schutzbeschaltung der Drain-Source-Strecke i. d. R. unnötig ist.

Kennlinien und Grenzwerte von VMOS-FET sind identisch mit denen der MOS-FET. Der Unterschied besteht lediglich in den Werten (Tab. 36.1).

Die Parallelschaltung von VMOS-FET ist fast problemlos möglich, da der positive Temperaturkoeffizient des Kanals für thermische Stabilität sorgt.

Aufgrund des hohen Eingangswiderstandes, der hohen Spannungsfestigkeit und der hohen Stromverstärkung ist der VMOS-FET ein sehr gut geeignetes Bauteil für Pegelumsetzer und Treiberschaltungen (zum Beispiel TTL-, TTL-LS- und CMOS-Familie).

Für höhere Betriebsspannungen ist die VMOS-Technik nicht gut geeignet, da die Isolierschicht in der Spitze der V-Rinne nicht die erforderliche gleichmäßige Dicke besitzt, so dass es dort zu Durchschlägen kommt.

36.4 SIPMOS-Transistoren

Eine wesentlich andere Konstruktion kommt bei den SIPMOS-Leistungstransistoren (**Si**emens-**P**ower-MOS; eingetragenes Warenzeichen der Siemens AG) nach Abb. 36.10 zur Anwendung.

Diese sind vertikal in Planartechnik aufgebaut und haben eine doppelt implantierte Kanalstruktur. Bei einem N-Kanal-FET dient das niederohmige n^+-Substrat mit der darunter liegenden Metallschicht als Träger und Drainanschluss. Über dem niederohmigen n^+-Substrat liegt eine n^--Epitaxieschicht, die je nach

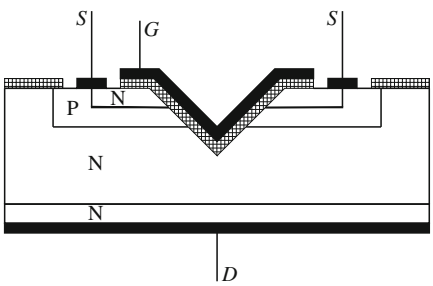


Abb. 36.9 Grundsätzlicher Schichtaufbau eines VMOS-FET

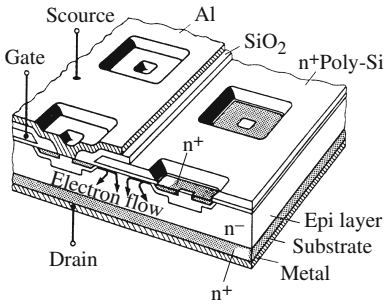


Abb. 36.10 Aufbau eines SIPMOS-FET (Siemens-Bild)

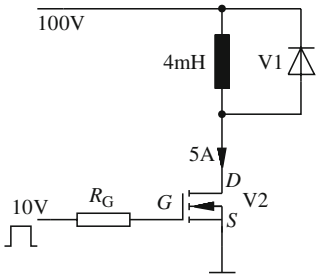


Abb. 36.11 Ansteuerschaltung für SIPMOS-FET

zu realisierender Sperrspannung verschieden dick und entsprechend dotiert ist. Darüber liegt das Gate aus n^+-polykristallinem Silizium mit hoher Dotierung, eingebettet in isolierendes Siliziumdioxid.

Das Gate bildet eine Gitterstruktur. In die Gitteröffnungen sind die Sourcezellen hineindotiert. Diese werden durch eine abdeckende Metallisierung (hier mit Aluminium) parallel geschaltet. Also kann man sagen, dass ein SIPMOS-FET aus der Parallelschaltung einiger tausend einzelner MOS-Systeme in einem Chip besteht. Die Gateschicht wird in einer Ecke des Chips unbedeckt gelassen und kann dort kontaktiert werden.

SIPMOS-FET haben sehr hohe Eingangswiderstände, aber Durchlasswiderstände $R_{D\,(on)}$ von nur wenigen Ohm. Ihre Schaltzeit liegt in der Größenordnung von einigen hundert Nanosekunden und ist mit einer Ansteuerschaltung nach Abb. 36.11 variabel machbar, um zu große Stromsteilheiten beim Einschalten nach Abb. 36.12 zu vermeiden beziehungsweise zu variieren. SIPMOS-FET benötigen nur beim Einschalten Energie und zwar hauptsächlich zum Aufladen der Gate-Drain-Kapazität. Wie der VMOS-FET hat auch dieser FET einen positiven Temperaturbeiwert und kann sich somit gegen thermische Überlastung selber schützen (Tab. 36.2).

SIPMOS-FET sind selbstsperrende FET und können auch aus diesem Grunde als schneller Leistungsschalter in der Leistungselektronik Verwendung finden. Reicht der Drainstrom eines SIPMOS-FET nicht aus, so können nahezu beliebig viele nach Abb. 36.13 parallel zueinander geschaltet werden und gleichzeitig angesteuert werden. Zum Ausgleich von Unsymmetrien werden Gate-Widerstände ($R = 4{,}7$ bis $200\,\Omega$) empfohlen.

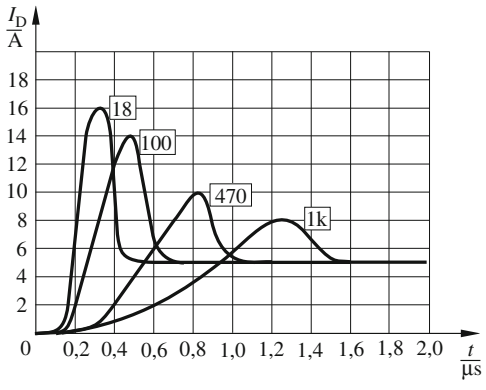


Abb. 36.12 Schaltverhalten des Drainstroms für verschiedene Gatewiderstände

Tab. 36.2 Typenübersicht und Kenndaten (Fa. Siemens)

SIPMOS-FET				
Typ	U_{DS} (V)	$R_{DS(on)}$ (Ω)	$I_{D(on)}$ (A)	Gehäuse
BUZ 10	50	0,1	12	TO-220
BUZ 20	100	0,2	8	TO-220
BUZ 31	200	0,2	12,5	TO-220
BUZ 45	500	0,8	7,5	TO-3
BUZ 84	800	2,0	4,7	TO-3
BUZ 54	1000	2,0	4,7	TO-3

IGBT			
Typ	U_{CE} (V)	I_C (A)	Gehäuse
BUP 304	1000	25	TO-218

IGBT-Module				
Typ	U_{CE} (V)	I_C (A)	$U_{CE(sat)}$ (V)	Bemerkungen
BSM50GB100D	1000	2×50	5	Halbbrücke
BSM15GD100D	1000	6×15	5	Vollbrücke

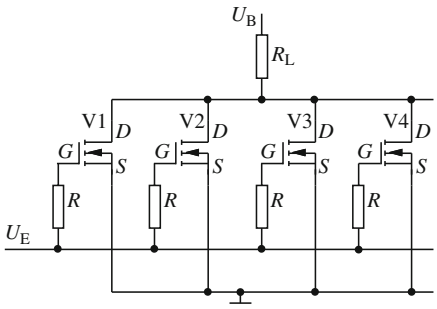


Abb. 36.13 Parallel geschaltete SIPMOS-FET für höhere Stromleistungen

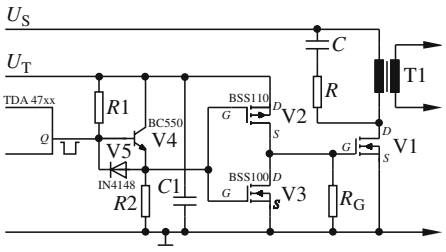


Abb. 36.14 Treiberschaltung für schnelles Ein- und Ausschalten von SIPMOS-FET (Fa. Siemens)

Gleichzeitig wird eine Drainstromreduzierung um den Faktor 0,8 empfohlen.

$$\rightarrow I_{D\,\mathrm{ges}} = 0{,}8 \cdot I_D \cdot n \quad (n = \text{Anzahl der FET})$$

Schnelles Ein- und Ausschalten, aber auch sicherer Schutz gegen unbeabsichtigtes Einschalten müssen gewährleistet werden. Die komplementären SIPMOS-Kleinsignaltransistoren BSS100 und BSS110 sind gut geeignet, um auch SIPMOS-FET mit großer Eingangskapazität oder parallel geschaltete SIPMOS-FET schnell zu steuern. Die Treiberschaltung nach Abb. 36.14

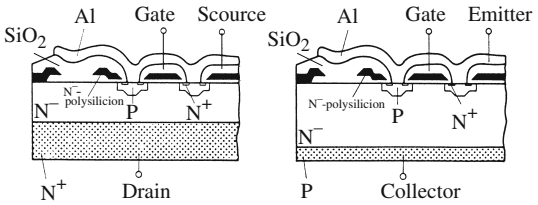


Abb. 36.15 Aufbau eines SIPMOS-FET und eines IGBT

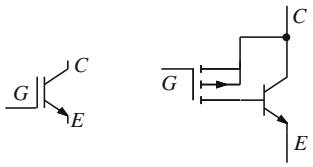


Abb. 36.16 Schaltzeichen und Ersatzschaltbild eines IGBT

(Fa. Siemens) bietet auch im spannungslosen Zustand einen sicheren Schutz gegen unbeabsichtigtes Einschalten.

Die Versorgungsspannung für die Ansteuerschaltung kann hier $U_T = 10$ bis 15 V betragen. Der Widerstandswert von R_G liegt ungefähr bei 10 kΩ, während $R1 = 2{,}2\,\text{k}\Omega$ und $R2 = 1\,\text{k}\Omega$ betragen. Der Kondensator C1 hat einen Wert von 0,47 µF. Die Werte für R und C bemessen sich nach der Größe des Transformators T1 beziehungsweise seiner Leistungsbeanspruchung.

36.5 IGBT

Ein IGBT (**i**nsulated-**g**ate-**b**ipolar-**t**ransistor) beruht grundsätzlich auf der Technologie der SIPMOS-Transistoren.

Abb. 36.15 macht hier die Unterschiede deutlich. Während der SIPMOS-FET auf einem n^+n^--Substrat am Drain aufgebaut ist,

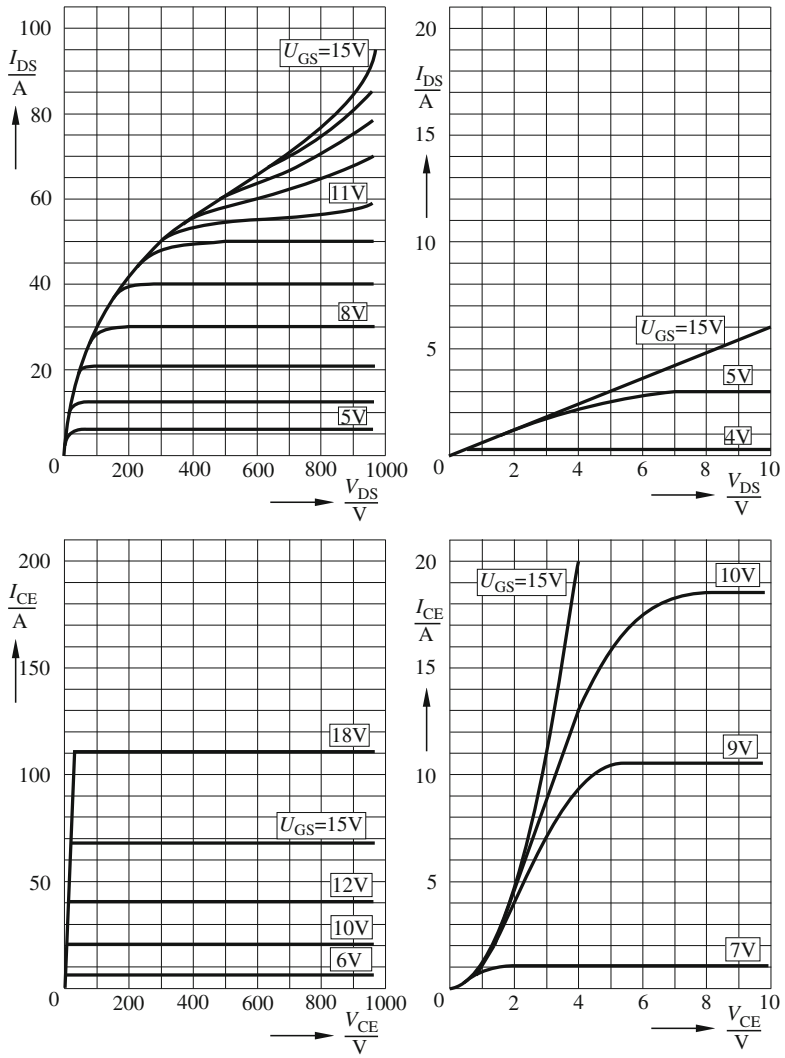


Abb. 36.17 Ausgangskennlinienfeld eines SIPMOS-FET (obere Reihe) und eines IGBT (untere Reihe)

hat der IGBT ein homogenes Substrat mit einem speziell ausgebildeten pn-Übergang an dieser Seite. Deshalb wird dieser Anschluss auch Kollektor genannt. In eingeschaltetem Zustand bewirkt der pn-Übergang die Reduzierung des Einschaltwiderstandes $R_{D(on)}$ durch Ladungsträgerinjektion.

Der IGBT ist eine Kombination aus MOS-FET und Bipolarem Transistor. Entsprechend verhält sich der IGBT am Eingang wie ein selbstsperrender MOS-FET und kann nahezu leistungslos gesteuert werden. Ausgangsseitig ist der IGBT einem bipolaren Leistungstransistor ähnlich. Deshalb kann er relativ hohe Spannungen (zur Zeit bis 1700 V) und Ströme (zur Zeit ungefähr 600 A) pro Chip schalten mit zunehmender Tendenz. Aus diesem Zusammenhang resultiert das Schaltzeichen und Ersatzschaltbild eines IGBT nach Abb. 36.16.

Die Ausgangskennlinien nach Abb. 36.17 zeigen einen SIPMOS BUZ 54 und einen IGBT BUP 304 bei gleicher Chipgröße. Das Schaltverhalten der beiden Bauelemente unterscheidet sich hauptsächlich durch den Spannungsabfall im eingeschalteten Zustand. So hat der BUZ 54 bereits bei 5 A einen Spannungsabfall von 8 V, während der BUP 304 bei 10 A auf weniger als 3 V Spannungsabfall kommt. Erkennbar ist außerdem, dass der Einschaltwiderstand $R_{D(on)}$ weniger spannungsabhängig ist als der der SIPMOS-FET.

Im Schaltbetrieb ist die Gesamtverlustleistung quasi identisch mit der temperaturabhängigen Durchlassverlustleistung. Der Einschaltwiderstand $R_{D(on)}$ wird in den Datenblättern für 25 °C angegeben.

IGBT haben im Gegensatz zu SIPMOS-FET aufgrund der unterschiedlichen Herstellungstechnologie keine integrierte Inversdiode. Deshalb muss bei Verwendung des IGBT als Schalter mit induktiven Lasten eine separate Freilaufdiode geschaltet werden.

In „Modulen“ sind IGBT, Inversdioden und andere erforderliche Schutzbausteine zusammengefasst. Sie sind rationelle Bausteine im Leistungsteil von selbstgeführten Stromrichtern.

Grundsätzlich kann man einen IGBT dadurch abschalten, indem man die Gate-Emitter-Spannung auf „Null“ setzt. Um das Abschaltverhalten zu verbessern und auch Notabschaltungen leichter zu ermöglichen, wird der IGBT nach Abb. 36.18 auch

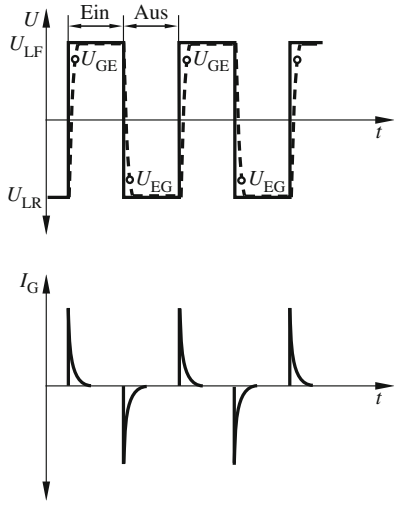


Abb. 36.18 Liniendiagramm von Steuerstrom und -spannung

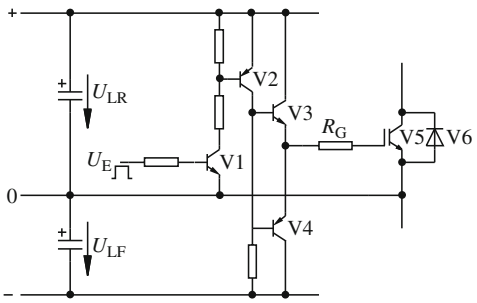


Abb. 36.19 Prinzipschaltung einer Treiberstufe mit komplementären Transistoren

mit negativen Spannungen an der Gate-Emitter-Strecke angesteuert.

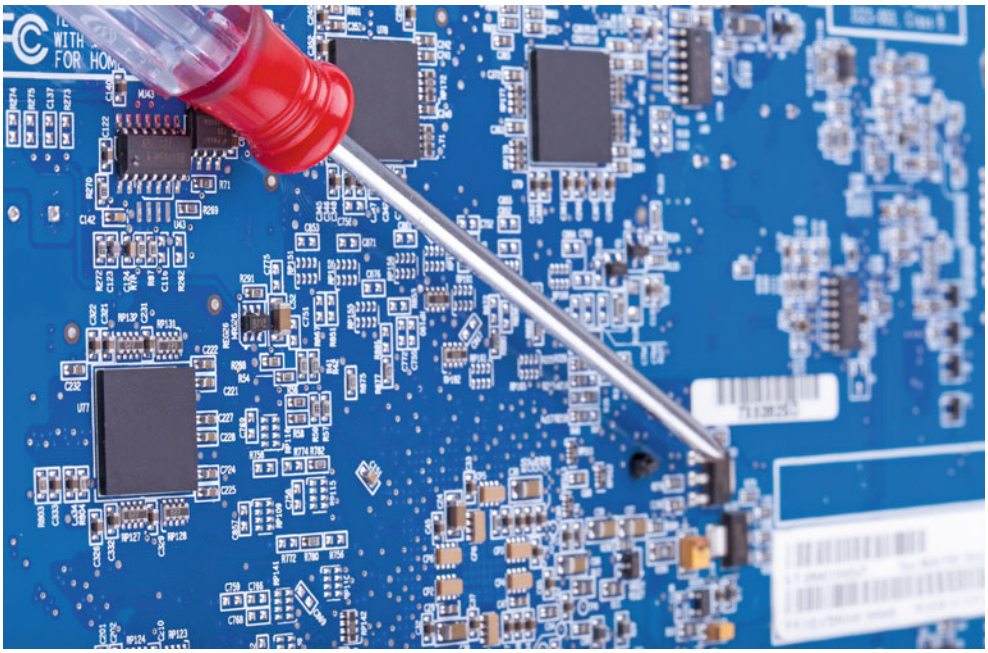
Der Gatevorwiderstand R_G begrenzt (wie beim SIPMOS) die entstehenden Steuerstromimpulse.

Um die Steuerspannungen zu erzeugen, verwendet man Treiberstufen mit komplementären Transistoren, wobei die Schaltung nach Abb. 36.19 nur eine mögliche Prinzipschaltung darstellt.

Analoge Verstärker

37

Peter Döring



37.1 Bipolarer Transistor als Verstärker 450
37.2 Feldeffekt-Transistor als Verstärker 461
37.3 Mehrstufige Verstärker 470

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_37

Der *Transistor* ist ein aktives Bauelement im Gegensatz zu passiven Bauelementen wie zum Beispiel Dioden oder Widerstände. Einer Transistorschaltung kann am Ausgang mehr Signalleistung entnommen werden als ihr am Eingang zur Verfügung steht.

In den meisten Anwendungsfällen werden Transistoren verwendet, um kleine Eingangsspannungen und -ströme zu großen Ausgangsspannungen und -strömen zu verstärken. Der Signalverlauf, also die Form der Signale, bleibt erhalten, während sich der Amplitudenwert der Signale vergrößert. Da Eingangs- und Ausgangssignale einander ähnlich sind, arbeitet der Transistor als Analogverstärker.

Die zur Umformung (Verstärkung) eines anliegenden Signals erforderliche Hilfsenergie liefert die Betriebsspannung. Unter Signal versteht man eine bestimmte Eingangsspannung, einen bestimmten Eingangsstrom oder eine bestimmte Eingangsleistung mit den entsprechenden Ausgangsgrößen.

Zu verstärkende Signale liefern Mikrophone, Tonabnehmersysteme von Plattenspielern, CD-Playern, Tonköpfen in Recordern, Antennen, Vorverstärker. Die verstärkten Signale (Ausgangssignale) können über entsprechende Wandler (Lautsprecher, Bildröhren, Motoren und andere Ausgabeeinheiten) hörbar beziehungsweise sichtbar gemacht werden.

37.1 Bipolarer Transistor als Verstärker

Der bipolare Transistor muss als gleichstromsteuerndes Bauelement verstanden werden.

Wechselspannungen, also Signale mit positiven und negativen Amplituden, können nur mit Hilfe von Schaltungskniffen verstärkt werden. Bildet man zumBeispiel eine Mischspannung aus einer Gleich- und einer Wechselspannung, indem man einem Gleichstrom die Signalspannung entsprechend der Schaltung nach Abb. 37.1 überlagert, so verstärkt der Transistor diese pulsierende Gleichspannung.

Abb. 37.2a stellt die sinusförmigen Ein- und Ausgangsspannungen an den Punkten 1 und 4 dar. Die Phasenverschiebung zwischen den beiden Kurven ergibt sich aus der generellen Phasenverschiebung bei der Emitterschaltung und der Phasenverschiebung aufgrund der beiden RC-Glieder.

Abb. 37.2b zeigt die Mischspannungen an der Basis (Punkt 2) und am Kollektor (Punkt 3) des Transistors.

Der Gleichanteil muss so groß bemessen sein, dass die Mischspannung keine negativen Anteile hat. Ohne diese Maßnahme würden nur die positiven Amplituden verstärkt werden. Durch einen Kondensator kann der Gleichstromanteil am Ausgang der Verstärkerschaltung unterdrückt werden.

Der besondere Vorteil des Vierquadranten-Kennlinienfeldes eines bipolaren Transistors liegt darin, dass bestimmte Daten von einem Quadranten direkt in einen anderen Quadranten, also

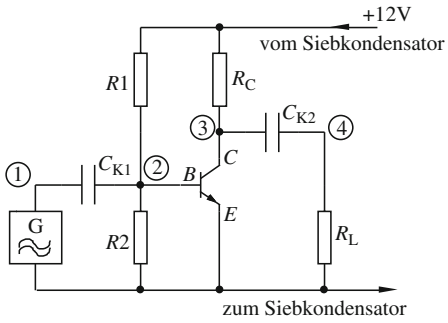


Abb. 37.1 Transistor als AC-Verstärker

von einem Kennlinienfeld in ein anderes, übertragen werden können. Daher lässt sich die Verstärkerwirkung eines bipolaren Transistors im Vierquadranten-Kennlinienfeld besonders gut darstellen und erläutern.

In Abb. 37.3 sind die Zusammenhänge für eine prinzipiell zu erstellende Verstärkerschaltung dargestellt und graphisch erläutert.

Zunächst muss die *Arbeitsgerade* des Kollektorwiderstandes R_C (5 kΩ gewählt) in das Ausgangskennlinienfeld eingezeichnet werden. Sie verläuft von links oben nach rechts unten und endet

auf der X-Achse bei

$$U_{CE} = U_B \quad (10\,\text{V gewählt}) \tag{37.1}$$

auf der Y-Achse bei

$$I_C = \frac{U_B}{R_C} \quad (2\,\text{mA errechnet}) \tag{37.2}$$

Der Arbeitspunkt AP1 sollte liegen bei

$$U_{CEA} = \frac{U_B}{2} \quad (5\,\text{V errechnet}) \tag{37.3}$$

Diese Wahlen ergeben im I. Quadranten einen $I_C = 1\,\text{mA}$, im II. Quadranten einen $I_B \approx 5\,\mu\text{A}$ und weitergeführt in den III. Quadranten ein $U_{BE} = 0{,}5\,\text{V}$.

Der Arbeitspunkt AP liegt nun fest bei $I_{CA} = 1\,\text{mA}$ und $U_{CEA} = 5\,\text{V}$.

Legt man nun an die Basis des Transistors eine Eingangswechselspannung (AC) $U_{ess} = 0{,}1\,\text{V}$, so wird die Gleichspannung von ihr überlagert.

Für die AC wandert nun der AP im III. Quadranten an der Kennlinie entlang zwischen $I_B = 2\,\mu\text{A}$ und $8\,\mu\text{A}$. Somit ändert sich auch der I_C und wandert nun zwischen $I_C = 0{,}5$ und $1{,}25\,\text{mA}$.

Die Änderung des Kollektorstroms ruft am Arbeitswiderstand R_C eine Spannungsänderung hervor, so dass sich auch U_{CE} zwischen $U_{CE} = 7{,}25$ und $1{,}25\,\text{V}$ um $\Delta U_{CE} = 6\,\text{V}$ ändert.

Bei einer Änderung der Eingangsspannung U_e um einen bestimmten Wert ergibt sich bei diesem Transistor eine Änderung

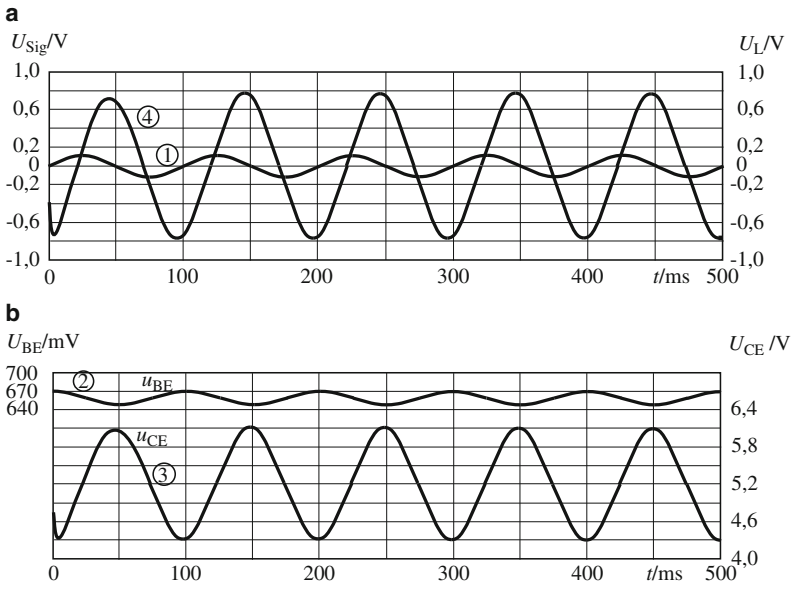


Abb. 37.2 Sinusförmige Ein- und Ausgangsspannungen sowie Mischspannungen an Basis und Kollektor

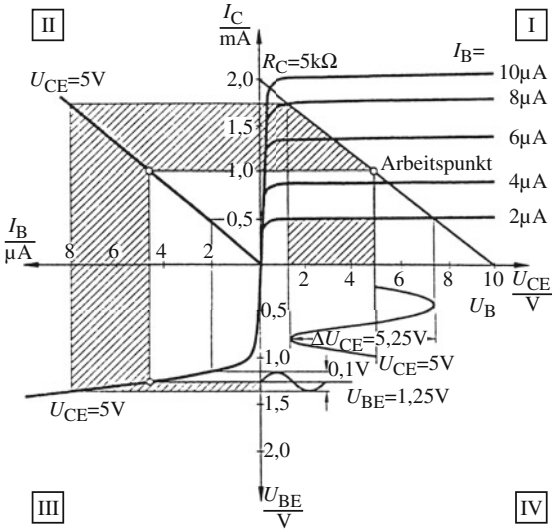


Abb. 37.3 Verstärkerwirkung eines bipolaren Transistors im Vier-Quadranten-Kennlinienfeld

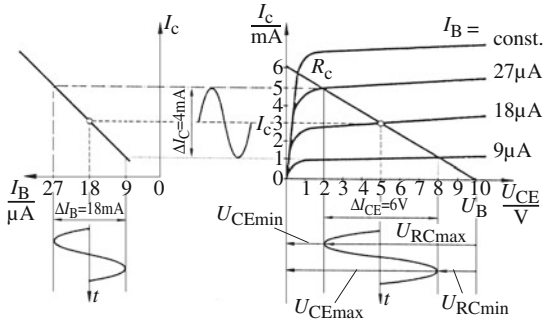


Abb. 37.4 Stromsteuerung

der Ausgangsspannung U_a um einen entsprechend vergrößerten Betrag.

Es ist erkennbar, dass die Ausgangsspannung gegenüber der Eingangsspannung verzerrt ist. Es handelt sich demnach um eine Spannungssteuerung des Transistors.

Im Falle einer *Stromsteuerung* des Transistors (der Basisstrom entstammt einer „Stromquelle“) kann man von einer Betrachtung nach Abb. 37.4 ausgehen, bei der die Ausgangsspannung nahezu unverzerrt am Kollektorwiderstand abfällt.

37.1.1 Grundschaltungen

Ein Verstärker hat stets zwei Anschlüsse für den Eingang und zwei Anschlüsse für den Ausgang. Transistoren haben aber nur drei Elektroden (Beinchen). Wird ein Transistor als Verstärker eingesetzt, muss eine dieser Elektroden sowohl zu den Eingangsanschlüssen als auch zu den Ausgangsanschlüssen gehören.

Damit ergeben sich drei Möglichkeiten, Transistoren als Verstärker zu betreiben. Diese drei Schaltungen werden als Transistor-Grundschaltungen bezeichnet. Die jeweilige Bezeichnung der Transistor-Grundschaltung ist von derjenigen Elektrode abgeleitet, die gemeinsamer Bezugspunkt für das Eingangs- und Ausgangssignal ist.

Jede der drei Grundschaltungen hat besondere Eigenschaften. Sie werden durch eine Reihe von Kenngrößen näher beschrieben.

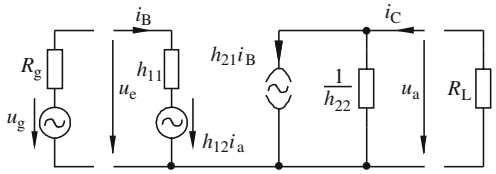


Abb. 37.5 Vierpol-Ersatzschaltbild mit Last und Signalquelle

Wichtige Kenngrößen sind: Wechselstromeingangswiderstand r_e, Wechselstromausgangswiderstand r_a, Spannungsverstärkung V_u, Stromverstärkung V_i, Leistungsverstärkung V_p, Phasendrehung des Signals φ und Grenzfrequenz der Schaltung f_g.

Das Vierpol-Ersatzschaltbild nach Abb. 37.5 in Anlehnung an Abb. 35.15 gilt ebenso wie die folgenden Betriebsgrößen für alle Grundschaltungen.

Für einen beliebigen Abschlusswiderstand R_L kann mit Hilfe der Vierpolgleichungen (Tab. 35.1) die Stromverstärkung V_i des Transistors bestimmt werden.

$$\textit{Stromverstärkung} \quad V_i = \frac{i_C}{i_B} = \frac{h_{21}}{1 + h_{22} \cdot R_L} \tag{37.4}$$

Die betriebliche Stromverstärkung ist also stets kleiner als die Kurzschlussstromverstärkung β des Transistors als Bauelement.

Auch für die Berechnung der betrieblichen Spannungsverstärkung V_u werden die Vierpolgleichungen benutzt. Durch geschicktes Einsetzen ergibt sich dann

Spannungsverstärkung

$$V_u = \frac{u_a}{u_e} = \frac{h_{21} \cdot R_L}{\Delta h \cdot R_L + h_{11}} \tag{37.5}$$

mit $\Delta h = h_{11} \cdot h_{22} - h_{12} \cdot h_{21}$.

Strom- und Spannungsverstärkung sind ebenso wie der Eingangswiderstand r_e lastabhängige Größen.

$$\textit{Eingangswiderstand} \quad r_e = \frac{\Delta h \cdot R_L + h_{11}}{1 + h_{22} \cdot R_L} \tag{37.6}$$

Der Ausgangswiderstand r_a ist vor allem vom Innenwiderstand R_g der Signalquelle abhängig. Er lässt sich nach dem Ohm'schen Gesetz vom Ausgang her berechnen.

Ausgangswiderstand

$$r_a = \frac{u_a}{i_C} = \frac{h_{11} + R_g}{\Delta h + h_{22} \cdot R_g} \tag{37.7}$$

Unter Leistungsverstärkung versteht man das Verhältnis von Ausgangs- zur Eingangsleistung.

Leistungsverstärkung

$$V_P = \frac{h_{21}^2 \cdot R_L}{(1 + h_{22} \cdot R_L)(h_{11} + \Delta h \cdot R_L)} \tag{37.8}$$

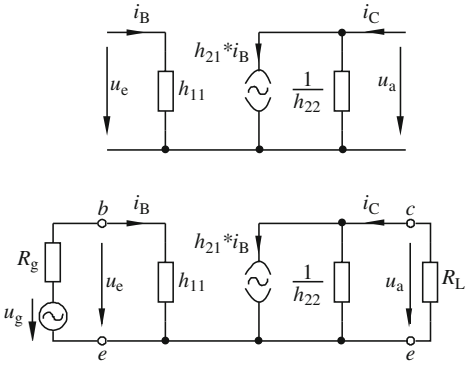


Abb. 37.6 Ersatzschaltungen bei Vernachlässigung der Spannungsrückwirkung

Mit Hilfe von (37.4) bis (37.8) lassen sich die betrieblichen Übertragungseigenschaften eines Transistorverstärkers berechnen. Dabei muss berücksichtigt werden, dass die h-Parameter nur für den Kleinsignalbetrieb im NF-Bereich gelten.

Die *Spannungsrückwirkung* des Transistors ist in den meisten Fällen so gering, dass man sie vernachlässigen kann und die Ersatzschaltungen nach Abb. 37.6 gültig sind. Aufgrund dieser Tatsache lassen sich Vereinfachungen vornehmen.

So beschreiben folgende sechs Werte das Verhalten des Transistors selber.

Stromverstärkung

$$B = \frac{I_C}{I_B} \quad \text{und} \quad \beta = \frac{\Delta I_C}{\Delta I_B} = h_{21}$$

Eingangswiderstand

$$R_{BE} = \frac{U_{BE}}{I_B} \quad \text{und} \quad r_{BE} = \frac{\Delta U_{BE}}{\Delta I_B} = h_{11}$$

Ausgangswiderstand

$$R_{CE} = \frac{U_{CE}}{I_C} \quad \text{und} \quad r_{CE} = \frac{\Delta U_{CE}}{\Delta I_C} = \frac{1}{h_{22}}$$

Um die Betriebsgrößen der Grundschaltungen bestimmen zu können, müssen die Grundbedingungen für den Betrieb der Schaltungen definiert werden. Der Arbeitspunkt wird generell durch eine Gleichspannungsquelle am Eingang eingestellt. Er wird jeweils so eingestellt, dass die Schaltung im linearen Teil der Eingangskennlinie arbeitet und Kleinsignalbetrieb vorliegt, also der Arbeitspunkt nicht in den Sättigungs- beziehungsweise Sperrbereich gelangen kann.

Zu der **Emitter-Grundschaltung** nach Abb. 37.7 zählt auch der Kollektorwiderstand R_C. Dieser ist unbedingt erforderlich, damit eine Wechselspannung, also ein Ausgangssignal am Kollektor entstehen kann. Ohne den Arbeitswiderstand R_C liegt der Kollektor des Transistors über die Gleichspannungsquelle wechselstrommäßig auf Masse, was einen wechselstrommäßigen Kurzschluss bedeuten würde.

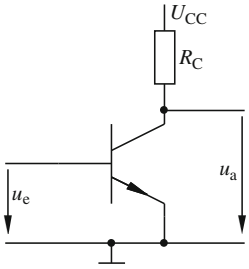


Abb. 37.7 Emitter-Grundschaltung

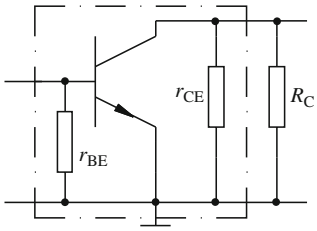


Abb. 37.8 Ein Wechselstrom-Ersatzschaltbild der Emitter-Grundschaltung

Die Eigenschaften der Emitter-Grundschaltung hängen nicht nur von den Kennwerten des Transistors selbst, sondern auch von der Größe des Kollektorwiderstandes ab. Die Kennwerte dieser gesamten Grundschaltung bezeichnet man als Betriebsgrößen. Ein mögliches Wechselstrom-Ersatzschaltbild, basierend auf Abb. 37.6 (Emitter-Grundschaltung) zeigt Abb. 37.8.

Wechselstromeingangswiderstand

$$r_{eE} = r_{BE} = h_{11e} \tag{37.9}$$

Wechselstromausgangswiderstand

$$r_{aE} = r_{CE} \parallel R_C \tag{37.10}$$

Spannungsverstärkung

$$|V_{uE}| = \frac{-\beta}{r_{BE}} \cdot r_{CE} \parallel R_C = \frac{\beta \cdot R_C \cdot r_{CE}}{r_{BE}(R_C + r_{CE})} \tag{37.11}$$

Stromverstärkung

$$V_{iE} = \beta \cdot \frac{r_{CE}}{R_C + r_{CE}} \tag{37.12}$$

Leistungsverstärkung

$$V_{pE} = \beta^2 \cdot \frac{r_{CE}^2 \cdot R_C}{(R_C + r_{CE})^2 \cdot r_{BE}} \tag{37.13}$$

maximale Leistungsverstärkung $V_{p\max}$ bei Leistungsanpassung, also wenn $R_C = r_{CE}$ · Phasendrehung des Signals 180°.

Die *Emitterschaltung* liefert sowohl eine Strom- und als auch eine Spannungsverstärkung, die beide > 1 sind. Somit liefert diese Schaltung die größtmögliche Leistungsverstärkung und wird deshalb häufig angewendet.

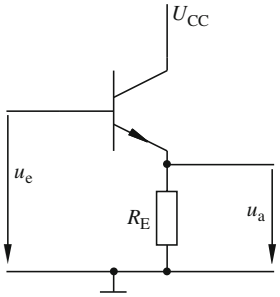


Abb. 37.9 Kollektor-Grundschaltung

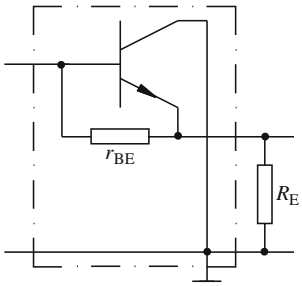


Abb. 37.10 Wechselstrom-Ersatzschaltbild der Kollektor-Grundschaltung

In der **Kollektor-Grundschaltung** nach Abb. 37.9 liegt der Kollektor wechselstrommäßig auf Masse.

Wird der Basisstrom erhöht, so erhöhen sich auch der Kollektorstrom und der Spannungsabfall am Emitterwiderstand R_E. Die Phasenverschiebung zwischen Ausgangs- und Eingangsspannung beträgt also 0°.

Die Eigenschaften der Kollektor-Grundschaltung hängen nicht nur von den Kennwerten des Transistors selbst, sondern auch von der Größe des Emitterwiderstandes ab. Ein Wechselstrom-Ersatzschaltbild der Kollektor-Grundschaltung zeigt Abb. 37.10.

Da in den Datenblättern der Hersteller die Kenndaten des Transistors nur für die Emitter-Grundschaltung angegeben werden, ist es sinnvoll, die Betriebsgrößen der Kollektor- und Basis-Grundschaltung auf die Kennwerte der Emitter-Grundschaltung zurückzuführen.

Wechselstromeingangswiderstand

$$r_{eC} = r_{eE} + \beta \cdot R_E \tag{37.14}$$

Wechselstromausgangswiderstand

$$r_{aC} = R_E \parallel \frac{r_{BE} + R_g}{\beta} \tag{37.15}$$

Spannungsverstärkung

$$V_{uC} \approx 1 \tag{37.16}$$

Stromverstärkung

$$V_{iC} = (\beta + 1) \cdot \frac{r_{CE}}{(R_E + r_{CE})} \tag{37.17}$$

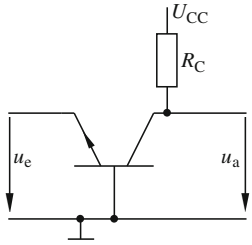


Abb. 37.11 Basis-Grundschaltung

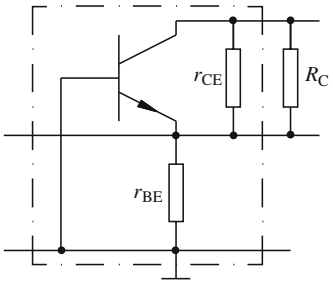


Abb. 37.12 Wechselstrom-Ersatzschaltbild der Basis-Grundschaltung

Leistungsverstärkung

$$V_{pC} \approx V_{iE} \tag{37.18}$$

Die Kollektorschaltung hat bei hohem Eingangswiderstand einen niedrigen Ausgangswiderstand. Sie findet also häufig Anwendung als Impedanzwandler, zum Beispiel zur Anpassung hochohmiger Generatoren an niederohmige Verbraucher.

Auch in Endstufen von Leistungsverstärkern findet sie Anwendung, da ihre Stromverstärkung etwa so groß ist wie die der Emitterschaltung.

In der **Basis-Grundschaltung** nach Abb. 37.11 liegt die Basis *wechselstrommäßig* auf Masse. Wird die Eingangsspannung (Spannung zwischen Emitter und Basis) größer, der Emitter also negativer, so wird der Transistor niederohmiger. Es fließt daher ein erhöhter Kollektorstrom, der am Kollektorwiderstand R_C einen erhöhten Spannungsabfall bewirkt. Die Phasenverschiebung zwischen Ausgangs- und Eingangsspannung beträgt dann 0°.

Die Eigenschaften der Basis-Grundschaltung hängen nicht nur von den Kennwerten des Transistors selbst, sondern auch von der Größe des Kollektorwiderstandes ab. Ein Wechselstrom-Ersatzschaltbild der Basis-Grundschaltung zeigt Abb. 37.12.

Wechselstromeingangswiderstand

$$r_{eB} = \frac{r_{eE}}{\beta} \tag{37.19}$$

Wechselstromausgangswiderstand

$$r_{aB} = R_C \parallel r_{CE} \cdot \left(1 + \beta \cdot \frac{R_g}{r_{BE}}\right) \tag{37.20}$$

und für $R_g \Rightarrow 0\,\Omega$:

$$r_{aB} \approx R_C \parallel r_{CE} \tag{37.21}$$

Spannungsverstärkung

$$V_{uB} = V_{uE} - 1 \tag{37.22}$$

Stromverstärkung

$$V_{iB} \approx 1 \tag{37.23}$$

Leistungsverstärkung

$$V_{pB} \approx V_{uB} \tag{37.24}$$

Die Stromverstärkung ist quasi gleich eins und der Ausgangsstrom praktisch identisch mit dem Eingangsstrom. Damit erfolgt eine schnellere Steuerung des Ausgangsstroms durch den Eingangsstrom als bei den anderen Grundschaltungen.

Die obere Grenzfrequenz der Basisschaltung hängt von der Stromverstärkung des Transistors ab.

$$\textit{obere Grenzfrequenz} \quad f_a \approx \beta \cdot f_g \tag{37.25}$$

Wegen ihrer hohen Grenzfrequenz f_a wird die Basisschaltung vorzugsweise in der Hochfrequenztechnik eingesetzt. Sie hat bei einem niedrigen Eingangswiderstand einen verhältnismäßig hohen Ausgangswiderstand.

37.1.2 Arbeitspunktstabilisierung

Zum Betrieb eines Transistors ist nach Abb. 37.1 und 37.3 neben der Spannungsversorgung U_B auch ein Widerstand R_C in der Kollektorzuleitung erforderlich.

Der *Kollektorwiderstand* R_C hat zwei verschiedene Aufgaben zu erfüllen. Er muss den Kollektorstrom I_C begrenzen und die im Transistor auftretende Stromverstärkung β in eine Spannungsverstärkung V_u umwandeln.

Der Transistor muss mit genau festgelegten Werten für I_C und U_{CE} betrieben werden. Das Bestimmen und Einstellen dieser Werte bezeichnet man als „in den Arbeitspunkt bringen“ oder „den Arbeitspunkt einstellen“.

Der *Arbeitspunkt AP* (siehe Abschn. 37.1) wird durch U_{CEA} und I_{CA} festgelegt und durch I_{BA} eingestellt und gehalten. Bei einer vorgegebenen Betriebsspannung U_B lässt sich der erforderliche R_C errechnen.

$$R_C = \frac{U_B - U_{CEA}}{I_{CA}} \tag{37.26}$$

$$I_{BA} = \frac{I_{CA}}{B} \tag{37.27}$$

Der Arbeitspunkt eines Transistors im Kleinsignalbetrieb als NF-Verstärker sollte möglichst in der Mitte der Widerstandsgeraden liegen (37.3). Dadurch wird verhindert, dass ein Signal

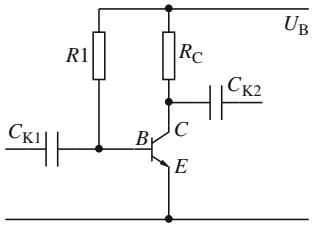


Abb. 37.13 Einstellung des Arbeitspunktes mit Basisvorwiderstand

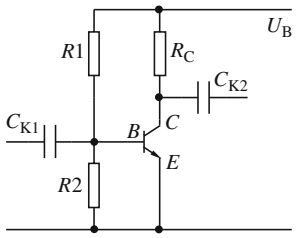


Abb. 37.14 Einstellung des Arbeitspunktes mit Basisspannungsteiler

infolge der Kennlinienkrümmungen unsymmetrisch verstärkt wird.

Bei Ansteuerung des Transistors mit einem sinusförmigen Signal an der Basis ändert sich der Kollektorstrom um einen bestimmten Wert nach oben und unten. Man kann sagen, dass sich der AP an der Widerstandsgeraden entlang um den AP herum hin und her bewegt.

Der Arbeitspunkt des Transistors kann durch Einstellung des Basisstroms und durch Einstellung der Basis-Emitter-Spannung festgelegt werden.

Die Basis-Emitter-Strecke muss in Durchlassrichtung betrieben werden, damit ein Kollektorstrom fließen kann. Das wird durch eine Basisvorspannung U_{BE} erreicht, die doppelt so groß sein muss wie die Schleusenspannung der Basis-Emitter-Diode.

Die einfachste Schaltung zur Erzeugung dieser Basisvorspannung und zur Einstellung des Basisstroms mit Hilfe des Basisvorwiderstandes R1 ist in Abb. 37.13 dargestellt.

Der Spannungsabfall an R1 muss so groß sein, dass von der Spannung U_B gerade noch die Basisvorspannung übrig bleibt, die die Basis-Emitter-Diode in Durchlassrichtung betreibt.

$$\textit{Basisvorwiderstand} \quad R_1 = \frac{U_B - U_{\mathrm{BEA}}}{I_{\mathrm{BA}}} \tag{37.28}$$

mit I_{BA} nach (37.27).

Durch diese Schaltungsmaßnahme (Konstantstrom) wird der Arbeitspunkt des Transistors eingestellt.

Die Basisvorspannung bleibt sowohl bei Schwankungen von U_B als auch bei Temperaturschwankungen nahezu konstant, weil der Spannungsabfall bei einem Ansteigen von I_{BA} an R1 höher wird. Dadurch wird U_{BE} wieder kleiner, verbunden mit einem Rückgang von I_{BA} auf seinen ursprünglichen Wert.

Von Vorteil ist auch der stets sehr hochohmige Widerstand R1, weil dadurch die Spannungs- und die Signalquelle nur gering belastet werden.

Nachteilig ist dagegen, dass aufgrund der Exemplarstreuungen der Transistoren jede einzelne Transistorstufe abgeglichen werden muss, was nach einem möglichen Auswechseln eines Transistors wiederholt werden muss.

Die Kondensatoren C_{K1} und C_{K2} verhindern, dass eine Gleichspannung an den Signalgenerator oder eine davor liegende Verstärkerstufe gelangt.

Eine weitere Möglichkeit, einen Transistor in den AP zu bringen, ist die Erzeugung der Basisvorspannung durch einen Basisspannungsteiler nach Abb. 37.14.

Der Strom durch R1 teilt sich an der Basis auf in den Basisstrom I_{BA} und den Querstrom I_q, so dass durch R1 die Summe aus I_q und I_{BA} fließt, durch R2 jedoch nur der I_q.

Der *Querstrom* I_q sollte verhältnismäßig groß gegenüber dem Basisstrom gewählt werden. Dies erreicht man, indem der Spannungsteiler entsprechend niederohmig gemacht wird, so dass er als unbelastet gelten kann und somit eine von Belastungsschwankungen weitgehend unabhängige U_{BE} liefern kann.

$$\textit{Richtwert} \quad I_q = (2 \ldots 10) \cdot I_B$$

Die Berechnung der Teilerwiderstände ergibt sich aus:

$$R_1 = \frac{U_B - U_{\mathrm{BEA}}}{I_q + I_{\mathrm{BA}}} \tag{37.29}$$

$$R_2 = \frac{U_{\mathrm{BEA}}}{I_q} \tag{37.30}$$

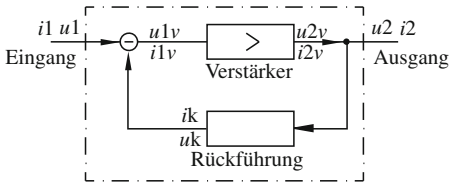
Da die Eingangskennlinie sehr steil verläuft, ergibt schon eine kleine Änderung von U_{BEA} eine große Änderung des Basisstroms. Deshalb erfolgt die genaue Einstellung des Arbeitspunktes bei dieser Schaltung durch ein Potentiometer anstelle von R_2.

Die Lage des Arbeitspunktes ist nicht temperaturstabil. Eine einfache Möglichkeit zur thermischen Arbeitspunkt-Stabilisierung lässt sich realisieren, indem man parallel zum Widerstand R_2 einen NTC-Widerstand schaltet, der unmittelbar an dem Transistorgehäuse oder an dem Kühlkörper des Transistors installiert wird. Mit steigender Temperatur wird der Widerstandswert des NTC aufgrund des negativen Temperaturkoeffizienten kleiner.

Da der Ersatzwiderstand einer Parallelschaltung immer kleiner ist als der kleinste Teilwiderstand, wird auch der Spannungsabfall an dieser Parallelschaltung kleiner, so dass I_B heruntergeregelt wird und damit auch I_C nicht weiter ansteigt.

Mit dem Widerstand R_2 lässt sich die *Wirksamkeit* des NTC einstellen. Je niederohmiger R_2 ist, umso geringer ist der Einfluss des NTC auf den Widerstandswert der Parallelschaltung.

Diese Art der Arbeitspunkt-Stabilisierung wird betrieben, wenn es sich um eine Verstärkerstufe mit hoher Ausgangsleistung handelt, man also auf den Emitterwiderstand R_E verzichten möchte (wegen zu großer Leistungsverluste in ihm) und die



$i1v = i1 - ik$
$u1v = u1 - uk$

Abb. 37.15 Blockschaltbild des Signalflusses bei einem gegengekoppelten Verstärker

infolge der Verlustleistung des Transistors abgegebene Wärmemenge groß genug ist, um eine ausreichend große Wirkung auf den NTC-Widerstand auszuüben.

In den meisten Emitterschaltungen wird eine Arbeitspunktstabilisierung durch Gegenkopplung angestrebt.

Man spricht von einer Rückkopplung, wenn ein Teil der Ausgangsenergie eines Verstärkers auf seinen Eingang zurückgeführt wird. Liegt das rückgekoppelte Signal mit dem Eingangssignal in Phase, wirkt es also gleichsinnig mit dem Eingangssignal, spricht man von Mitkopplung (Schalter- und Trigger-Anwendungen). Wirkt das rückgekoppelte Signal jedoch dem Eingangssignal entgegen, so spricht man von Gegenkopplung.

Abb. 37.15 soll den Signalfluss bei einem gegengekoppelten Verstärker verdeutlichen. Der strichpunktierte Kasten stellt den gegengekoppelten Verstärker dar, während der mit „Verstärker“ gekennzeichnete Kasten den rückkopplungsfreien Verstärker darstellt. Das mit „Rückführung“ gekennzeichnete Übertragungsglied koppelt nun eine Ausgangsgröße – entweder dem Ausgangsstrom oder der Ausgangsspannung proportional – aus und führt diese Größe mit umgekehrtem Vorzeichen dem Verstärkereingang zu.

Vorteile der Gegenkopplung sind: Die Verstärkerkennwerte werden gegenüber Temperatureinflüssen und Exemplarstreuung stabilisiert. Die Ein- und Ausgangswiderstände und die Verstärkung lassen sich in ihrer Größe beeinflussen. Die nichtlinearen Verzerrungen (Klirrfaktor) werden verringert. Der Übertragungsbereich (Bandbreite) eines Verstärkers kann erhöht werden (Verringerung der linearen Verzerrungen).

Eine Gegenkopplung hat immer eine Verringerung der Verstärkung zur Folge. Gegenkopplung kann durch das Einfügen eines Emitterwiderstandes R_E in die Schaltung nach Abb. 37.14 erreicht werden.

Der Emitterstrom I_E lässt am Widerstand R_E in der Schaltung nach Abb. 37.16 eine Spannung U_{RE} abfallen. Ändert sich der Strom I_E im Transistor aufgrund einer Temperaturänderung, so regelt sich die Transistorschaltung selber aus.

Beschreibung der Regelung

Aufgrund einer Temperaturerhöhung erhöht sich der Strom I_C. Daraufhin wird die Spannung U_{RE} beispielsweise von 1 auf 1,1 V ansteigen. Da U_{R2} durch den Basisspannungsteiler konstant gehalten wird, muss die Basisvorspannung U_{BE} um 0,1 auf

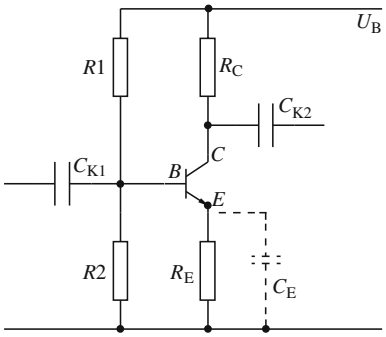


Abb. 37.16 Transistorverstärker mit Emitterwiderstand und Emitterkondensator

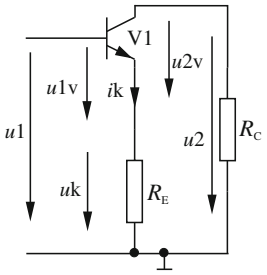


Abb. 37.17 Strom-Spannungs-Gegenkopplung (Regelkreis)

0,6 V sinken. Das wiederum führt zu einem Absinken des Basisstroms I_B.

Durch die Gleichstromverstärkung B sinkt der Kollektorstrom I_C um einen entsprechenden Betrag ab. Das führt zu einer Senkung der Spannung an R_E; der ursprüngliche Zustand ist wieder erreicht.

Da sich U_{BE} bei jeder kleinsten Emitterstromänderung sofort mit verändert, wird die thermische Arbeitspunktstabilisierung mit Hilfe eines Emitterwiderstandes in fast allen Verstärkerschaltungen angewandt.

Diesen Effekt bezeichnet man als *Strom-Spannungs-Gegenkopplung*, da eine Stromänderung am Ausgang ($I_E = i_k$) in eine Spannungsänderung am Eingang ($U_{\mathrm{BE}} = u_1 - u_k$) umgewandelt wird (Abb. 37.17).

Die Güte der Stabilisierung des eingestellten Arbeitspunktes hängt von der Spannung am Widerstand $R_E$und damit vom Widerstandswert ab. Die Änderung des Kollektorstroms infolge einer Erwärmung des Transistors kann als Störgröße im Gegenkopplungskreis aufgefasst werden.

$$\textit{Richtwert} \quad U_{\mathrm{RE}} \approx (0{,}1 \ldots 0{,}2) \cdot U_B$$

Der Widerstand R2 muss so berechnet werden, dass an ihm die Summe der Spannungen aus U_{RE} und U_{BE} abfällt.

$$R_1 = \frac{U_B - (U_{\mathrm{BEA}} + U_{\mathrm{RE}})}{I_q + I_{\mathrm{BA}}} \tag{37.31}$$

$$R_2 = \frac{U_{\mathrm{BEA}} + U_{\mathrm{RE}}}{I_q} \tag{37.32}$$

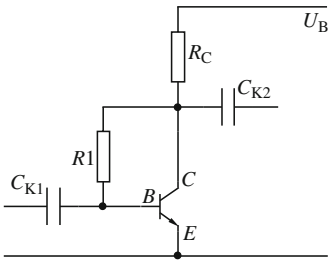


Abb. 37.18 Thermische Arbeitspunktstabilisierung mit Spannungs-Strom-Gegenkopplung

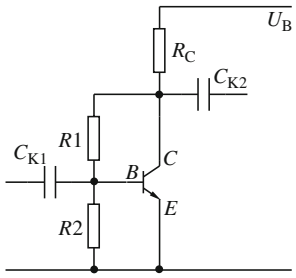


Abb. 37.19 Arbeitspunkteinstellung durch Basisspannungsteiler bei Spannungs-Strom-Gegenkopplung

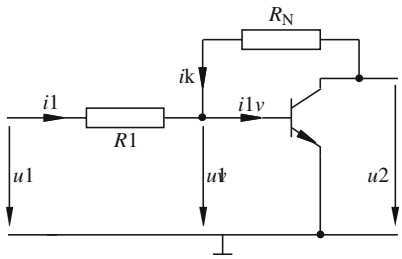


Abb. 37.20 Spannungs-Strom-Gegenkopplung (Regelkreis)

In Abb. 37.16 ist der Emitterwiderstand nur dann gleichstrommäßig wirksam, wenn er mit einem Kondensator C_E überbrückt ist, der auch bei der unteren Grenzfrequenz der Verstärkerstufe noch sehr niederohmig gegen R_E ist und deshalb jede Wechselspannung (= Signal), die am Emitter auftritt, nach Masse kurzschließt.

Befindet sich in der Schaltung nach Abb. 37.16 der Kondensator C_E, so ist wechselstrommäßig nur der Kollektorwiderstand R_C wirksam. Diese Schaltung verhält sich wechselstrommäßig wie eine Emitterschaltung ohne R_E, bei der die Batteriespannung um den Betrag U_{REA} kleiner ist als im vorliegenden Fall.

Eine weitere Möglichkeit zur thermischen Arbeitspunktstabilisierung zeigt die Schaltung nach Abb. 37.18. Im Gegensatz zur Stromgegenkopplung, bei der eine dem Ausgangsstrom proportionale Spannung zum Eingang zurückgekoppelt wird, koppelt man hier einen Teil der gegenphasigen Ausgangsspannung zum Eingang zurück (Spannungs-Strom-Gegenkopplung).

Die Stabilisierung wird umso besser, je größer das Verhältnis R_C/R_1 wird. Die Gegenkopplung geschieht durch den Widerstand R1, der den Basisstrom I_{BA} zur Arbeitspunkteinstellung liefert. Dieser ist jedoch nicht an die konstante Betriebsspannung angeschlossen, sondern an das veränderliche Kollektorpotential. Da der Widerstand R1 gleich- und wechselspannungsmäßig wirksam ist, spricht man von einer Gleich- und Wechselspannungsgegenkopplung.

Ändert sich in dieser Schaltung die Temperatur, so ändert sich auch der eingestellte Kollektorstrom, damit auch die Spannung U_{CE} und als Folge davon die Basisvorspannung U_{BE}.

$$R_C = \frac{U_B - U_{\mathrm{CEA}}}{I_{\mathrm{CA}} + I_{\mathrm{BA}}} \tag{37.33}$$

$$R_1 = \frac{U_{\mathrm{CEA}} - U_{\mathrm{BEA}}}{I_{\mathrm{BA}}} \tag{37.34}$$

Zur Berechnung von R_C kann I_{BA} wegen $B > 100$ vernachlässigt werden. In der Schaltung nach Abb. 37.19 wird die Arbeitspunkteinstellung wieder durch einen Basisspannungsteiler unter Ausnutzung der Gegenkopplung vorgenommen.

Die Wechselspannungsgegenkopplung lässt sich hier nur schwer unwirksam machen, so dass diese Schaltung nicht häufig Verwendung findet.

Kollektorwiderstand

$$R_C = \frac{U_B - U_{\mathrm{CEA}}}{I_{\mathrm{CA}} + I_{\mathrm{BA}} + I_q} \tag{37.35}$$

Basisspannungsteilerwiderstände

$$R_1 = \frac{U_{\mathrm{CEA}} - U_{\mathrm{BEA}}}{I_{\mathrm{BA}} + I_q} \tag{37.36}$$

$$R_2 = \frac{U_{\mathrm{BEA}}}{I_q} \tag{37.37}$$

Die Wirkung *der Spannungs-Strom-Gegenkopplung* lässt sich nach Abb. 37.20 mit Unterstützung von Abb. 37.15 nachvollziehen. Der Widerstand R_N ist der Gegenkopplungswiderstand R1 nach Abb. 37.19.

37.1.3 Emitterschaltungen

Die Emitterschaltung nach Abb. 37.16 ist die am häufigsten eingesetzte Grundschaltung. Der Emitter ist bei dieser Schaltung wechselstrommäßig der gemeinsame Anschluss für den Eingang und Ausgang des Transistors.

Diese Schaltung wird mit dem Abschlusswiderstand R_A belastet und von einer Signalquelle u_S mit dem Innenwiderstand R_S gespeist.

Abb. 37.21a zeigt das AC-Ersatzschaltbild der Emitterschaltung mit dem integrierten Emitterkondensator C_E, wobei das AC-Ersatzschaltbild des Transistors nach Abb. 37.6 verwendet

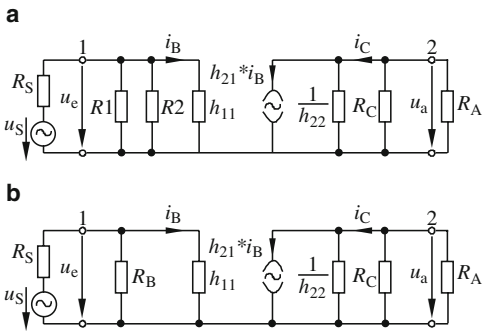


Abb. 37.21 AC-Ersatzschaltbild der Emitterschaltung mit Emitterkondensator

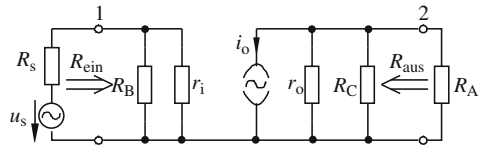


Abb. 37.22 Vollständige AC-Ersatzschaltung

wird. Der Emitterwiderstand R_E erscheint nicht in diesem Ersatzschaltbild, da er durch C_E kurzgeschlossen wird.

Die Spannungsrückwirkung wird hier vernachlässigt. Der Innenwiderstand der Gleichspannungsquelle kann für Wechselspannungen zu Null genommen werden, so dass der Kollektorwiderstand und der Widerstand R1 gegen Masse liegen.

Fasst man die Parallelschaltung der Widerstände R1 und R2 zum Ersatzwiderstand R_B zusammen, so erhält man Abb. 37.21b. Somit ergibt sich für diese Transistorstufe die Ersatzschaltung nach Abb. 37.22.

Der Wechselstrom-Eingangswiderstand R_{ein} wird bei der Emitterschaltung hauptsächlich durch den dynamischen Eingangswiderstand r_{BE} des Transistors bestimmt. Der Ausgangswiderstand R_{aus} ergibt sich als Parallelschaltung des Kollektorwiderstandes R_C mit dem Ausgangswiderstand r_0 (Abb. 37.21) des Transistors nach Abb. 37.22.

Eingangswiderstand

$$R_{ein} = R_1 \parallel R2 \parallel r_i = R_B \parallel r_{BE} \tag{37.38}$$

Ausgangswiderstand

$$R_{aus} = R_C \parallel r_0 = R_C \parallel r_{CE} \tag{37.39}$$

Der Wechselstrom-Ausgangswiderstand bei der Emitterschaltung wird also wegen $r_{aE} \gg R_C$ durch den Kollektorwiderstand R_C bestimmt.

Nunmehr lässt sich die AC-Ersatzschaltung nach Abb. 37.22 vollständig umrechnen nach Abb. 37.23 mit den gegebenen Werten.

Mit

$$R_L = R_C \parallel R_A; \quad R_g = R_1 \parallel R_2 \parallel R_S$$

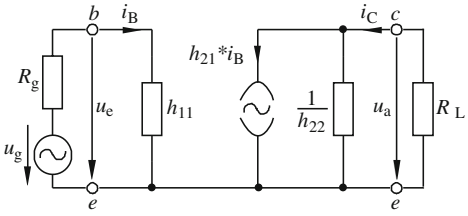


Abb. 37.23 AC-Ersatzschaltung mit umgerechneter Signalquelle

und

$$u_g = u_S \cdot \frac{R_1 \parallel R_2}{R_S + R_1 \parallel R_2}$$

gilt für die Spannungsverstärkung

$$V_u = \frac{\beta}{r_{BE}} \cdot r_{CE} \parallel R_L \tag{37.40}$$

und die Stromverstärkung

$$V_i = \beta \cdot \frac{r_{CE}}{R_L + r_{CE}} \tag{37.41}$$

Die Leistungsverstärkung der Stufe, also von Klemme 1 nach Klemme 2 gemäß Abb. 37.22, ist geringer als die Multiplikation der errechneten Spannungs- und Stromverstärkungswerte.

Leistungsverstärkung

$$V_{Stufe} = V_u^2 \cdot \frac{R_B \cdot r_{eE}}{R_A \cdot (R_B + r_{eE})} \tag{37.42}$$

Hier sind jetzt auch die Wechselstromverluste in den Widerständen R_C, R1 und R2, die zur Arbeitspunkteinstellung erforderlich sind, berücksichtigt.

Bei der Emitterschaltung wird das Ausgangssignal gegenüber dem Eingangssignal um 180° phasenverschoben verstärkt wiedergegeben.

Die Koppelkondensatoren C_{K1} und C_{K2} stellen im Idealfall für die Signalspannung einen Kurzschluss dar. Sie bestimmen die untere Übertragungsfrequenz des Verstärkers ebenso wie der Emitterkondensator, denn je größer ihre Wechselstromwiderstände mit abnehmender Frequenz werden, umso geringer wird die Spannungsverstärkung der Schaltung. Darum muss für die Koppelkondensatoren C_{K1} und C_{K2} eine untere Grenzfrequenz f_u angegeben werden, für die diese zu berechnen sind. Als Grenzfrequenz ist die Frequenz festgelegt, bei der die Spannungsverstärkung V_u auf 0,707 der Verstärkung bei mittleren Frequenzen abgesunken ist. Das entspricht einer Dämpfung von 3 dB.

Koppelkondensatoren

$$C_{K1} = C_{K2} = \frac{1}{2\pi f_u \cdot (R_{aus} + R_{ein})} \tag{37.43}$$

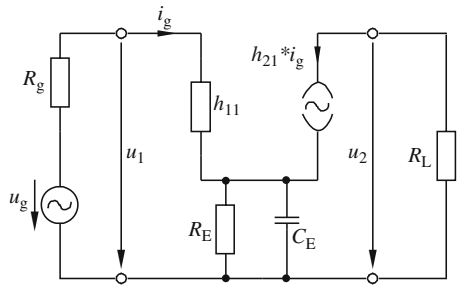


Abb. 37.24 AC-Ersatzschaltung für Emitterschaltung mit Emitterkondensator

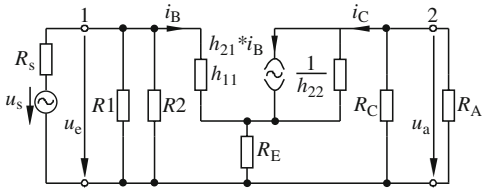


Abb. 37.25 AC-Ersatzschaltung für Emitterschaltung ohne Emitterkondensator

In (37.43) beschreibt R_{aus} den Ausgangswiderstand der signalgebenden Schaltung und R_{ein} den Eingangswiderstand der aufnehmenden Schaltung beziehungsweise des Lastwiderstandes. Die Kapazität des Emitterkondensators ergibt sich für eine vorgegebene untere Grenzfrequenz f_u mit folgender Überlegung nach Abb. 37.24.

Faktisch liegt der Kondensator C_E parallel zum Widerstand R_E und muss somit bei der Berechnung der Spannungsverstärkung berücksichtigt werden. Bei der Grenzfrequenz ist die Verstärkung um 3 dB abgesunken. Bei Frequenzen unterhalb der Grenzfrequenz ist der Emitterkondensator nicht mehr wirksam.

Emitterkondensator

$$C_E = \frac{R_g + r_{BE} + \beta \cdot R_E}{2\pi f_u \cdot R_E \cdot (R_g + r_{BE})} \tag{37.44}$$

Wird der Emitterwiderstand R_E nicht durch den Kondensator C_E nach Abb. 37.16 überbrückt, so ist er nach Abb. 37.25 auch wechselstrommäßig wirksam. In diesem Fall liegt eine Gleich- und Wechselstromgegenkopplung vor. Ein Teil der AC-Ausgangsspannung u_a, der Signalspannungsabfall an R_E, wird auf den Eingang zurückgekoppelt.

Diese AC-Gegenkopplung hat Auswirkungen auf die Spannungsverstärkung.

Vereinfacht gilt:

mit

$$R_L = R_C \parallel R_A: \quad u_a \approx -(i_B \cdot h_{21e}) \cdot R_L$$

und

$$u_e = i_B \cdot [h_{11e} + (h_{21e} + 1) \cdot R_E]$$

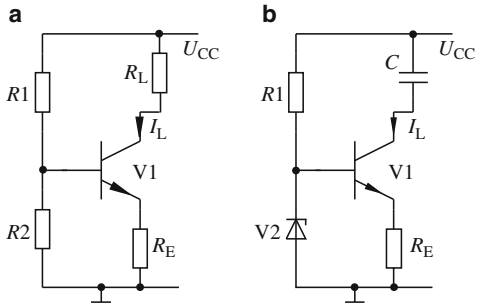


Abb. 37.26 Konstantstromquellen

Spannungsverstärkung

$$V'_u = -\frac{h_{21e} \cdot R_L}{[h_{11e} + (h_{21e} + 1) \cdot R_E]} \tag{37.45}$$

Durch die AC-Gegenkopplung wird die AC-Verstärkung der Transistorstufe herabgesetzt. Daraus folgt, dass die Spannungsverstärkung V'_u einer Transistorstufe mit R_E ohne C_E kleiner ist als bei einer Stufe mit $C_E \parallel R_E$.

$$\textit{Näherung} \quad V'_u \approx -\frac{R_L}{R_E} \tag{37.46}$$

wenn $R_E \gg \frac{r_{BE}}{\beta}$.

Die Wechselstromgegenkopplung macht zwar die Spannungsverstärkung kleiner, ergibt jedoch zwei erhebliche Vorteile:

Die *Spannungsverstärkung* lässt sich auf einen gewünschten Wert mit Hilfe der beiden Widerstände R_C und R_E einstellen! Die Spannungsverstärkung wird praktisch *unabhängig* von den Daten des Transistors und hängt nur von den beiden Widerständen ab (wichtig beim Bau von Seriengeräten und bei Reparaturen)!

Es ist erkennbar, dass der AC-Eingangswiderstand der Schaltung nach Abb. 37.26 andere Werte hat als in der Schaltung nach Abb. 37.22. Der Eingangswiderstand des Transistors als solcher ist durch die AC-Wirkung des R_E größer geworden.

Eingangswiderstand

$$R_{ein} = R_B \parallel [h_{11e} + (h_{21e} + 1) \cdot R_E] \tag{37.47}$$

Der Eingangswiderstand der gegengekoppelten Schaltung wird aber weiterhin im Wesentlichen durch die Größe der Spannungsteiler-Widerstände bestimmt.

Der Ausgangswiderstand der Schaltung ohne Beachtung von R_C und R_A wird durch den R_E erhöht. Es gilt vereinfacht:

Ausgangswiderstand

$$R_{aus} \approx r_{CE} \cdot \left[1 + \frac{h_{21e} \cdot R_E}{(h_{11e} + R_E + R_S)}\right] \tag{37.48}$$

Dieser „widerstandsbehaftete“ Transistor lässt sich nutzen. Die Transistorschaltung mit R_E hat das Verhalten einer Stromquelle, da der Widerstand R_E die Eigenschaft hat, den Signalstrom

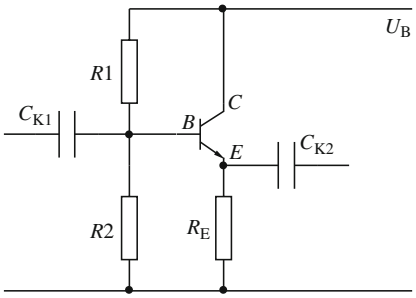


Abb. 37.27 Kollektorschaltung mit Basisspannungsteiler

zu stabilisieren beziehungsweise bei konstanter Eingangsspannung konstant zu halten. Konstantstromquellen nach Abb. 37.26 haben einen sehr großen, theoretisch unendlich hohen Innenwiderstand.

Der Kondensator als Lastwiderstand in der Schaltung nach Abb. 37.26b wird linear und nicht nach einer Exponentialfunktion geladen. Die Z-Diode V2 stabilisiert die Versorgungsspannung V_{CC}.

Der differentielle Innenwiderstand dieser Konstantstromquelle lässt sich leicht berechnen.

Innenwiderstand

$$r_i = \frac{dU_L}{dI_L} = r_{CE} \cdot \left[1 + \frac{h_{21e} \cdot R_E}{(h_{11e} + R_E)}\right] \tag{37.49}$$

Wegen $I_L \approx I_E$ lässt sich der Strom dieser Konstantstromquelle berechnen.

Laststrom

$$I_L \approx \frac{U_{RE}}{R_E} = \frac{U_{R2} - U_{BEA}}{R_E} \tag{37.50}$$

37.1.4 Kollektorschaltungen

Eine Kollektorschaltung mit Basisspannungsteiler ist in Abb. 37.27 dargestellt. Bei der gezeigten Schaltung wird die Basisvorspannung durch den Vorwiderstand R2 erzeugt und der AP durch den Emitterwiderstand R_E stabilisiert. Der Emitterwiderstand darf nicht durch einen Kondensator gebrückt werden, weil an R_E das Ausgangssignal abgegriffen wird. Die Kondensatoren C_{k1} und C_{k2} trennen den Generator und den Lastwiderstand gleichstrommäßig von der Schaltung.

Diese Schaltung wird nun (vgl. Abschn. 37.1.3) mit dem Lastwiderstand R_L belastet und von einer Signalquelle u_S mit dem Innenwiderstand R_S gespeist. Die Parallelschaltung der Widerstände R1 und R2 fasst man zum Ersatzwiderstand R_B zusammen.

Nun lässt sich die AC-Ersatzschaltung nach Abb. 37.28 mit den gegebenen Werten vollständig berechnen.

$$R_g = R_1 \parallel R_2 \parallel R_S, \quad u_g = u_S \cdot \frac{R_1 \parallel R_2}{R_S + R_1 \parallel R_2}$$

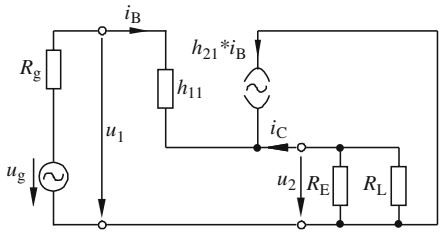


Abb. 37.28 AC-Ersatzschaltung der Kollektorschaltung

Die Kennwerte der vollständigen Schaltung lassen sich aus dem AC-Ersatzschaltbild nach Abb. 37.28 herleiten und erklären.

Unter Berücksichtigung des Lastwiderstandes und $r_{CE} \gg R_E \parallel R_L$ gilt dann:

AC-Eingangswiderstand

$$R_{ein} \approx \left[r_{BE} + \beta \cdot \frac{R_E \cdot R_L}{(R_E + R_L)}\right] \tag{37.51}$$

Der Wechselstrom-Eingangswiderstand der Kollektorschaltung ist folglich größer als der einer Emitterschaltung, wird jedoch durch die Parallelschaltung des Basisspannungsteilers R_B deutlich harabgesetzt.

Der Wechselstrom-Ausgangswiderstand der Kollektorschaltung ist laut Ersatzschaltbild in seinem Wert kleiner als bei der Emitterschaltung.

AC-Ausgangswiderstand

$$R_{aus} = R_E \parallel \frac{r_{BE} + R_g}{\beta} \tag{37.52}$$

Für die Spannungsverstärkung gilt grundsätzlich:

Spannungsverstärkung

$$V_u = \frac{u_a}{u_e} = \frac{\Delta U_{RE}}{\Delta U_{RE} + \Delta U_{BE}} \approx 1 \tag{37.53}$$

Da die Summe der Spannungen unter dem Bruchstrich größer ist als der Wert oberhalb, gilt $V_u < 1$!

Für die Stromverstärkung gilt grundsätzlich:

$$V_i = \frac{i_a}{i_e} = \frac{\Delta I_C + \Delta I_B}{\Delta I_B} = \beta + 1 \tag{37.54}$$

Näherungsweise kann mit $V_i \approx \beta$ gerechnet werden.

Da die Leistungsverstärkung V_p das Produkt aus V_u und V_i ist, gilt bei $V_u \approx 1$ und $V_i \approx \beta$ eine Leistungsverstärkung

$$V_p \approx V_i \approx \beta \tag{37.55}$$

Da ein größerer Basisstrom bei der Kollektorschaltung auch einen höheren Kollektorstrom und damit Emitterstrom bewirkt, wird auch bei steigendem Basisstrom der Spannungsabfall an R_E größer. Bei der Kollektorschaltung entsteht somit *keine* Phasenverschiebung zwischen dem Ein- und Ausgangssignal.

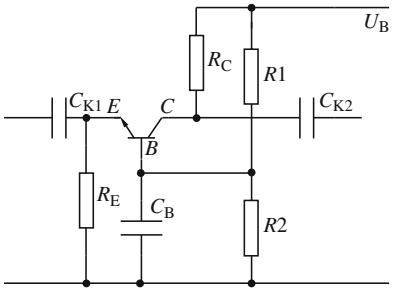


Abb. 37.29 Verstärkerstufe in Basisschaltung mit Basisspannungsteiler

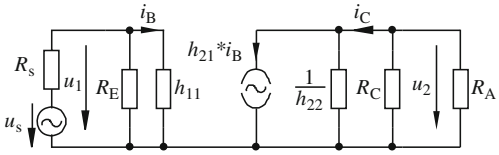


Abb. 37.30 AC-Ersatzschaltbild der Basisschaltung

Die *Grenzfrequenz* unterscheidet sich hier kaum von der der Emitterschaltung. Aufgrund des hohen Eingangswiderstandes und des niedrigen Ausgangswiderstandes wird die Kollektorschaltung als *Impedanzwandler* verwendet; sie ermöglicht die Anpassung hochohmiger Signalquellen an kleine Verbraucherwiderstände. Durch die geringe Belastung des Signalgenerators ist sie gut verwendbar als Eingangsstufe für mehrstufige Verstärkerschaltungen.

37.1.5 Basisschaltung

Bei der Basisschaltung ist die Basis gemeinsamer Bezugspunkt (Elektrode) für das Ein- und Ausgangssignal.

Abb. 37.29 zeigt eine vollständige Verstärkerstufe, die mit einem Basisspannungsteiler (R1 und R2) zur Basisvorspannungserzeugung versehen ist. Beachtenswert ist hier der Anschlusspunkt für die Versorgungsspannung U_B.

Der Arbeitspunkt wird durch den Widerstand R_E stabilisiert. Der Kondensator C_B sorgt dafür, dass die Basis des Transistors wechselspannungsmäßig auf Masse gelegt wird. Für Wechselspannungen schließt der Kondensator C_B den Widerstand R2 des Spannungsteilers kurz, während der Widerstand R1 durch die Batterie (Netzteil) kurzgeschlossen wird. Somit ist der Basisspannungsteiler im AC-Ersatzschaltbild nicht vorhanden.

Die Kennwerte lassen sich aus dem AC-Ersatzschaltbild nach Abb. 37.30 herleiten und erklären.

AC-Eingangswiderstand

$$R_{\text{ein}} = \frac{u_1}{i_1} = \frac{R_E \cdot r_{\text{BE}}}{r_{\text{BE}} + (\beta + 1) \cdot R_E} \tag{37.56}$$

Der Wechselstrom-Eingangswiderstand der Basisschaltung ist wesentlich niedriger als der der beiden anderen Grundschaltungen.

Generell gilt für den Wechselstrom-Ausgangswiderstand der Basisschaltung

AC-Ausgangswiderstand

$$R_{\text{aus}} \approx \frac{r_{\text{BE}} + (\beta + 1) \cdot R_S}{r_{\text{CE}} \cdot R_S} \parallel R_C \approx R_C \tag{37.57}$$

Die Spannungsverstärkung einer Basisschaltung ergibt sich aus zur Emitterschaltung ähnlichen Zusammenhängen. Da bei der Basisschaltung $r_{\text{CE}} \gg R_C$ ist, gilt

$$\textit{Spannungsverstärkung} \quad V_u \approx \frac{\beta \cdot R_C}{r_{\text{BE}}} \tag{37.58}$$

Ist ein Lastwiderstand vorhanden, so liegt dieser parallel zu R_C und muss entsprechend berücksichtigt werden.

$$\textit{Stromverstärkung} \quad V_i = \frac{\beta}{\beta + 1} \approx 1 \tag{37.59}$$

Da die Leistungsverstärkung das Produkt aus Spannungs- und Stromverstärkung ist, gilt näherungsweise:

$$\textit{Leistungsverstärkung} \quad V_p \approx V_u \approx \frac{\beta \cdot R_C}{r_{\text{BE}}} \tag{37.60}$$

Aus der Art der Ansteuerung ist erkennbar, dass es keine Phasendrehung zwischen Eingangs- und Ausgangssignal geben kann.

Die höchste übertragbare Frequenz f_{go} liegt bei der Basisschaltung in der Nähe der Transitfrequenz, aber auf jeden Fall oberhalb der Grenzfrequenz der Emitterschaltung. Den Grund dafür findet man im Wechselstrom-Ersatzschaltbild. Da die Basis direkt an Masse liegt, kann über die interne Kapazität C_{CB} des Transistors das Ausgangssignal nicht auf den Eingang zurückgeführt werden.

Die Grenzfrequenz der Basisschaltung errechnet sich daher zu

$$f_{\text{go}} = \beta \cdot f_{\text{go Emitter}} \tag{37.61}$$

Aufgrund der hohen Grenzfrequenz wird diese Schaltung fast ausschließlich in Hoch- und Höchstfrequenzschaltungen eingesetzt (Eingangsstufe für Signalfrequenzen größer 100 MHz).

Ein weiterer Vorteil ist der niedrige Eingangswiderstand der Basisschaltung, der eine sehr gute und einfache Anpassung an die niedrige Antennenimpedanz von $Z = 75\,\Omega$ ermöglicht.

37.2 Feldeffekt-Transistor als Verstärker

Feldeffekt-Transistoren haben im Vergleich zu den bipolaren Transistoren für die Anwendung als AC-Verstärker zwei erhebliche Vorteile. Auf die geringeren nichtlinearen Signalverzerrungen beim FET ist bereits hingewiesen worden. Die große

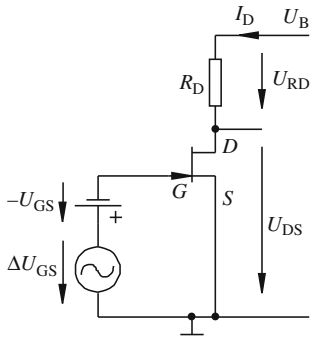


Abb. 37.31 Prinzipschaltung eines Verstärkers mit J-FET

Eingangsimpedanz der FET ist ein weiterer Vorteil, denn sie ermöglicht eine große Spannungsverstärkung.

Der Verstärkungsvorgang bei den FET ist im Prinzip identisch mit dem der bipolaren Transistoren (Abb. 37.31).

Durch eine negative Gate-Source-Spannung $-U_{GS}$ wird bei diesem N-Kanal-J-FET der Arbeitspunkt eingestellt, der damit sowohl für die Übertragungskennlinie wie auch für das Ausgangskennlinienfeld bestimmt ist. Dort liegt der Arbeitspunkt im Schnittpunkt der Arbeitsgeraden mit der I_D-U_{DS}-Kennlinie mit U_{GS} als Parameter. Die Arbeitsgerade gehorcht den gleichen Bedingungen wie bei den bipolaren Transistoren.

Das zu verstärkende Signal ΔU_{GS} wird der Gate-Source-Spannung überlagert und lässt nunmehr in der Summe den Arbeitspunkt auf der Arbeitsgeraden wandern. Die positive Amplitude des Signals lässt die Gate-Source-Spannung U_{GS} positiver werden und somit den Drainstrom I_D ansteigen. Dadurch erhöht sich der Spannungsabfall am Drainwiderstand R_D. Infolge dieser Spannungserhöhung muss die Drain-Source-Spannung U_{DS} kleiner werden. Ein- und Ausgangsspannung haben hier also entgegengesetzte Polarität, was von der Emitterschaltung bei den bipolaren Transistoren her bekannt ist.

Eine nahezu verzerrungsfreie Verstärkung des Signals ΔU_{GS} ist nur zu erwarten, wenn der Arbeitspunkt im Ausgangskennlinienfeld auch bei maximaler Amplitude des Signals nicht über die Abschnürgrenze hinaus in den ohmschen Bereich wandert. Abb. 37.32 veranschaulicht den beschriebenen Vorgang. Die Analogie zu den bipolaren Transistoren ist unübersehbar.

37.2.1 Arbeitspunkteinstellung und -stabilisierung

Bei der Verwendung des FET als Verstärker wird nur der Abschnürbereich des Ausgangskennlinienfeldes benutzt. Der Arbeitsbereich, also der Bereich, in dem der Arbeitspunkt liegen darf, ist durch folgende Kurven beziehungsweise Werte begrenzt: die Verlusthyperbel, die Abschnürgrenze und die Kennlinien für die maximale und minimale Gate-Source-Spannung.

Bei Kleinsignal-Aussteuerung legt man den Arbeitspunkt in den oberen Bereich der Übertragungskennlinie bei einem großen

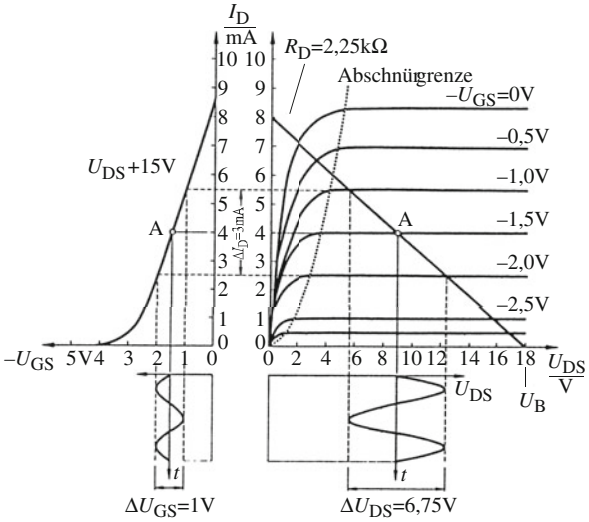


Abb. 37.32 Verstärkungsvorgang beim J-FET

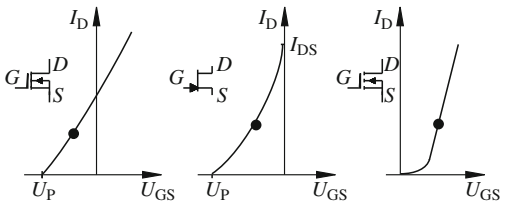


Abb. 37.33 Steuerkennlinien der drei N-Kanal-Typen

Ruhestrom I_{DA}, da hier die Kennlinie nahezu linear verläuft. Muss ein temperaturunabhängiger Arbeitspunkt gewählt werden, kann der Kompensationspunkt genommen werden.

Für diesen Arbeitspunkt wird in Abschn. 35.2 die Gate-Source-Spannung U_{GSK} ermittelt. Größere Nichtlinearität ist hier allerdings die Folge.

Bei der Aussteuerung muss gewährleistet werden, dass die Drainspannung die im Datenblatt angegebene Durchbruchspannung $U_{(BR)DS}$ nicht übersteigt.

In Abb. 37.32 ist der Arbeitspunkt so gewählt, dass $U_{DS} = U_B/2$ ist. Wie bei einem Verstärker mit bipolarem Transistor ist das ein guter Arbeitspunkt.

Dieser Arbeitspunkt muss nun mit der Gate-Source-Spannung U_{GS} eingestellt werden. Abb. 37.33 zeigt die Steuerkennlinien der drei möglichen *N-Kanal-Typen* mit günstigen Arbeitspunkten für den Betrieb in AC-Verstärkern.

Beim J-FET und beim MOS-FET vom Normally-On-Typ ist die Steuerspannung U_{GS} negativ, beim MOS-FET vom Normally-Off-Typ dagegen positiv zu machen.

Eine Arbeitspunktstabilisierung ist nur bei der Source Schaltung, die der Emitterschaltung entspricht, vorzusehen. Die beiden anderen Grundschaltungen sind infolge der internen Gegenkopplung thermisch stabil.

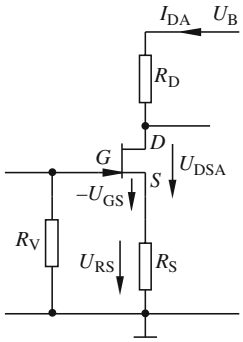


Abb. 37.34 Arbeitspunktstabilisierung mit automatischer Gate-Source-Spannung

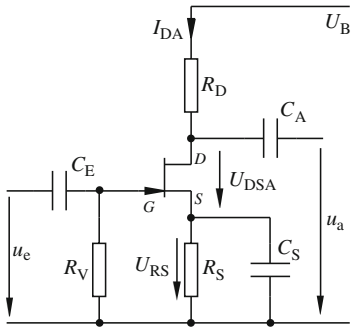


Abb. 37.35 Vollständige Verstärkerstufe mit J-FET

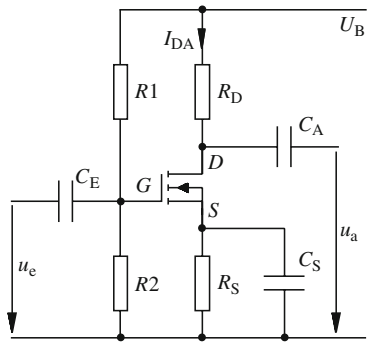


Abb. 37.36 Verstärkerschaltung mit selbstsperrenden MOS-FET

J-FET sind selbstleitende Typen, bei denen auch ohne Gate-Source-Spannung ein großer Drainstrom fließt. Dies ist auch bei den MOS-FET vom Normally-On-Typ der Fall.

Dieser Umstand ermöglicht die Arbeitspunkteinstellung und -stabilisierung mit Hilfe der automatischen Gate-Source-Spannung. Die automatische Erzeugung dieser Spannung mit Hilfe eines Widerstandes R_S in der Sourceleitung zeigt die Schaltung nach Abb. 37.34. Der durch die Widerstände R_D und R_S fließende Drainstrom I_{DA} lässt am Sourcewiderstand R_S eine Spannung U_{RS} abfallen. Die Source-Elektrode ist daher um den Wert dieser Spannung positiver als das über den Widerstand R_V auf Masse gelegte Gate, wobei daran erinnert sei, dass die Gate-Source-Strecke quasi einen unendlich hohen Widerstand darstellt und somit kein Gatestrom fließt. Damit wird deutlich, dass mit Hilfe des Sourcewiderstandes R_S die Gate-Source-Spannung eingestellt werden kann.

Gate-Source-Spannung

$$U_{RS} = -U_{GS} = I_{DA} \cdot R_S \tag{37.62}$$

Ein steigender Drainstrom führt zu einer negativer werdenden Gate-Source-Spannung, die ihrerseits den Drainstrom begrenzt beziehungsweise verkleinert. Ein sinkender Drainstrom führt zu einem entgegengesetzten Effekt. Es handelt sich also bei diesem Sourcewiderstand R_S um eine *Gleichstrom-Gegenkopplung*, wie sie beim bipolaren Transistor zur thermischen Arbeitspunktstabilisierung bereits behandelt worden ist.

Wird der FET mit einer Wechselspannung angesteuert, fällt an R_S eine Wechselspannung ab, die eine gegenkoppelnde Wirkung hat. Wie in der Emitterschaltung (bipolare Transistoren) wird auch hier die AC-Verstärkung gemindert. Will man eine AC-Gegenkopplung vermeiden, kann durch einen genügend großen Kondensator C_S der Widerstand R_S wechselspannungsmäßig kurzgeschlossen werden. Die Kondensatoren C_E und C_A nach Abb. 37.35 dienen zur Ein- und Auskopplung des Signals.

Je größer der Widerstand R_S gewählt wird, umso geringer wird eine mögliche Arbeitspunktverschiebung aufgrund von Exemplarstreuungen des FET sein. Nachteilig ist dann jedoch die sich ergebende Lage des Arbeitspunktes im unteren Teil der Übertragungskennlinie.

Bei den selbstsperrenden MOS-FET ist die automatische Vorspannungserzeugung nicht möglich. Aus diesem Grunde werden die Normally-Off-Typen immer und die anderen FET bei bestimmten Gelegenheiten mit Hilfe eines Gate-Spannungsteilers nach Abb. 37.36 „in den Arbeitspunkt gebracht“ und mit dem Sourcewiderstand R_S gegen Arbeitspunktschwankungen stabilisiert.

Für direkt gekoppelte Verstärkerschaltungen können nur FET verwendet werden, deren Drain-Source- und Gate-Source-Spannung gleiche Polarität haben. Je nach Verhältnis der Widerstände R_1, R_2 und R_S lassen sich mit dem Spannungsteiler entsprechende Vorspannungen am Gate einstellen.

Jetzt lassen sich auch Arbeitspunkte im oberen Teil der Übertragungskennlinie einstellen. Unter Vernachlässigung des Gatestroms I_G lässt sich der Drainstrom I_D berechnen.

Drainstrom

$$I_D = \frac{\frac{R_2}{R_1+R_2} \cdot U_B - U_{GS}}{R_S} = \frac{U_0 - U_{GS}}{R_S} \tag{37.63}$$

Je größer der Spannungsabfall an den Widerständen R_2 und R_S gewählt wird, umso besser lassen sich Änderungen des Arbeitspunktes durch Exemplarstreuungen des FET oder Temperaturschwankungen reduzieren. Der Spannungsteiler sollte aber nicht zu hoch gewählt werden, da sonst die exponentielle Temperaturabhängigkeit des Gatestroms wirksam wird. Wird der Spannungsteiler zu niederohmig, wird der Wechselstrom-Eingangswiderstand der Verstärkerschaltung wesentlich reduziert.

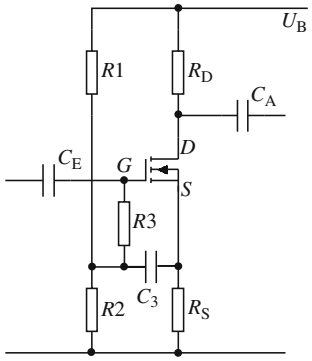


Abb. 37.37 Bootstrap-Schaltung mit selbstsperrenden FET

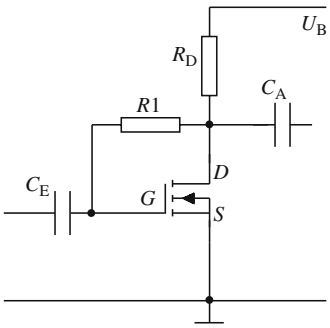


Abb. 37.38 Verstärkerstufe mit selbstsperrenden FET und Spannungsgegenkopplung

Bei vorgegebenen Werten für den Bereich des gewünschten Arbeitspunktes (Wandern des AP auf der Arbeitsgeraden) lässt sich der Sourcewiderstand durch Vorgabe des maximalen und minimalen Ruhestroms (I_{DA^+}, I_{DA^-}) berechnen.

$$\textit{Sourcewiderstand} \quad R_S = \frac{U_{GSA^+} - U_{GSA^-}}{I_{DA^+} - I_{DA^-}} \tag{37.64}$$

Der Spannungsteiler kann als unbelastet angenommen werden.

$$\textit{Teilerverhältnis} \quad \frac{R_1}{R_2} = \frac{U_B}{I_{DA} \cdot R_S + U_{GSA}} - 1 \tag{37.65}$$

Häufig wird die Spannung U_{RS} in der Größenordnung um 1 V gewählt, während der Querstrom I_q durch den Teiler im Mikroampere-Bereich liegt.

Da sich der Drainstrom bei Temperaturerhöhung nur geringfügig verringert, können FET ohne zusätzliche Sicherheitsmaßnahmen parallelgeschaltet werden.

Weitere Schaltungen zur Arbeitspunkteinstellung und -stabilisierung haben zusätzlich zum Ziel, einen hohen Eingangswiderstand des Verstärkers zu realisieren. Die Bootstrap-Schaltung nach Abb. 37.37 lässt sich genauso wie die Spannungs-Gegenkopplung nach Abb. 37.38 mit selbstsperrenden FET anwenden. Die Spannungsverstärkung wird durch die Gegenkopplung herabgesetzt, was bereits von den bipolaren Transistoren her bekannt ist.

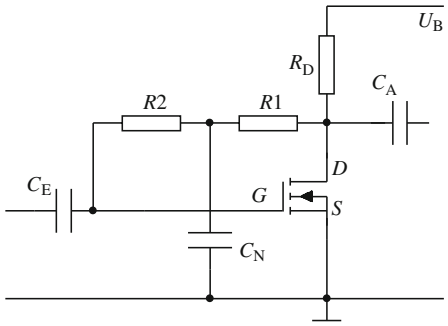


Abb. 37.39 Spannungsgegenkopplung ohne Signalgegenkopplung

Die Spannungs-Gegenkopplung kann für Signalspannungen mit höherer Frequenz aufgehoben werden. Hierzu wird der Gegenkopplungswiderstand in der Schaltung nach Abb. 37.38 in zwei Widerstände mit gleichem Widerstandswert aufgeteilt und ein Kondensator vom Mittelpunkt des Teilers gegen Masse geschaltet, wie in Abb. 37.39 zu sehen ist. Für die Signalspannung bestimmt der Widerstand R_2 den Eingangswiderstand der Schaltung.Der Arbeitspunkt bei Drainschaltungen, die den Kollektorschaltungen entsprechen, lässt sich auf die gleiche Weise wie zuvor beschrieben einstellen.

37.2.2 Grundschaltungen von FET

Das Betriebsverhalten von FET wird durch die vier typischen Eigenschaften dieser Transistoren bestimmt:

- hoher Eingangswiderstand im NF-Bereich
- verhältnismäßig geringe Vorwärts-Steilheit
- hoher Ausgangswiderstand (da spannungsgesteuerte Stromquelle)
- vernachlässigbar kleine Spannungsrückwirkung im NF-Bereich.

Entsprechend der Emitter-, Kollektor- und Baisschaltung des bipolaren Transistors lässt sich der FET als Verstärker in der Source-, Drain- und Gateschaltung betreiben. Jede dieser Grundschaltungen weist hinsichtlich der Betriebsgrößen typische Eigenschaften auf.

Die wichtigsten Kenngrößen dieser Grundschaltungen sind: Wechselstrom-Eingangswiderstand R_{ein}, Wechselstrom-Ausgangswiderstand R_{aus}, Spannungsverstärkung V_u, Stromverstärkung V_i, Phasenverschiebung φ. In den folgenden Abschnitten werden die Betriebsgrößen (Kenngrößen) der drei Grundschaltungen mit N-Kanal-J-FET beschrieben beziehungsweise berechnet.

37.2.2.1 Sourceschaltung

Die Source-Schaltungen nach Abb. 37.35 und 37.41 sind die am meisten verwendeten Verstärkerschaltungen des FET.

Mit Hilfe der Widerstände R_V, R_S und R_D oder R_1, R_2, R_S und R_D wird der Arbeitspunkt des FET eingestellt und stabilisiert.

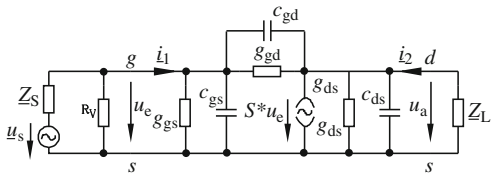


Abb. 37.40 AC-Ersatzschaltbild zur Source- Schaltung nach Abb. 37.35

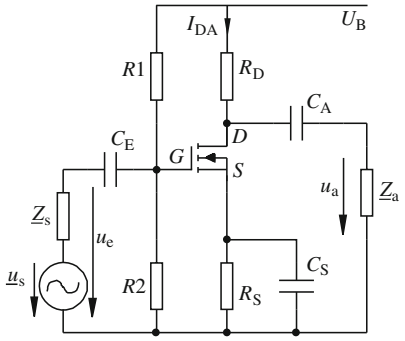


Abb. 37.41 Source-Schaltung mit Gatespannungsteiler

Um eine AC-Gegenkopplung zu verhindern, ist der Sourcewiderstand in beiden Schaltungen durch einen genügend großen Kondensator C_S kurzgeschlossen. Die Kondensatoren C_E und C_A dienen in beiden Schaltungen zur gleichstrommäßigen Entkopplung des FET-Verstärkers von der Signalquelle und vom Lastwiderstand.

Aufgrund der komplexen y-Parameter der Vierpol-Ersatzschaltung soll hier für die vorläufige Betrachtung ein komplexer Innenwiderstand $\underline{Z}_S$ der Signalquelle und ein komplexer Lastwiderstand $\underline{Z}_A$ als Abschlusswiderstand angenommen werden.

Zur Source-Schaltung nach Abb. 37.35 gehört das AC-Ersatzschaltbild nach Abb. 37.40. Der Widerstand $\underline{Z}_L$ steht hier als Ersatzwiderstand für die Parallelschaltung von $\underline{Z}_A$ und R_D.

In der Schaltung nach Abb. 37.41 wird der Arbeitspunkt durch den Gatespannungsteiler eingestellt.

Zur Source-Schaltung nach Abb. 37.41 gehört das AC-Ersatzschaltbild nach Abb. 37.42a. Am Eingang des J-FET liegt die Parallelschaltung von R_1 und R_2, am Ausgang liegt wieder die Parallelschaltung von $\underline{Z}_A$ und R_D.

Als Vereinfachung findet man R_B für $R_1 \parallel R_2$ und $\underline{Z}_L$ als Ersatzwiderstand für $\underline{Z}_A \parallel R_D$ in Abb. 37.42b.

Nimmt man (37.45) in Verbindung mit dem Ersatzschaltbild, so erhält man die betriebliche Spannungsverstärkung der Source-Schaltung.

Spannungsverstärkung

$$V_u = \frac{S + g_{gs}}{S + g_{ds} + g_{gs} + G_L} \tag{37.66}$$

Bei niedrigen Frequenzen können die kapazitiven Elemente des Ersatzschaltbildes ebenso vernachlässigt werden wie der sehr

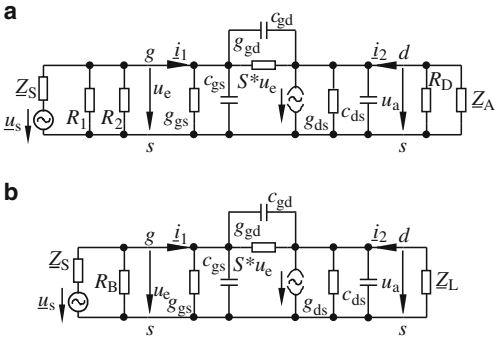


Abb. 37.42 AC-Ersatzschaltbilder zur Source-Schaltung nach Abb. 37.41

geringe Rückwirkungsleitwert g_{gd}, so dass sich die Gleichung mit einer rein ohmschen Last R_L vereinfachen lässt.

Spannungsverstärkung (NF)

$$V_u = -\frac{S}{g_{ds} + G_L} \tag{37.67}$$

Das negative Vorzeichen in der Gleichung weist auf die Phasenverschiebung φ zwischen Ein- und Ausgangssignal hin.

Der Kanalwiderstand $r_{ds} = 1/g_{ds}$ ist sehr hochohmig, so dass gilt: $r_{ds} \gg R_L$.

$$\textit{Spannungsverstärkung (NF)} \quad V_u \approx -S \cdot R_L \tag{37.68}$$

Die betriebliche Stromverstärkung der Source-Schaltung lässt sich mit (37.41) für niedrige Frequenzen bei rein ohmscher Last berechnen.

$$\textit{Stromverstärkung (NF)} \quad V_i = \frac{S \cdot G_L}{g_{gs}(g_{ds} + G_L)} \tag{37.69}$$

Mit $g_{ds} \ll G_L$ gilt dann:

$$\textit{Stromverstärkung (NF)} \quad V_i \approx \frac{S}{g_{gs}} \tag{37.70}$$

Da der Leitwert g_{gs} sehr klein ist, ergibt sich rein rechnerisch eine große Stromverstärkung bei tiefen Frequenzen. Bei hohen Frequenzen bewirken die kapazitiven Elemente des Ersatzschaltbildes eine Reduzierung der Stromverstärkung. Der AC-Eingangswiderstand R_{ein} ist derjenige Widerstand, mit dem die Verstärkerschaltung die Signalquelle belastet. Der komplexe Eingangsleitwert $\underline{y}_i$ der Source-Schaltung lässt sich nach (37.47) berechnen. Es ergibt sich folgende Vereinfachung:

Eingangsleitwert y_i

$$y_i = g_{gs} + g_{gd} + j\omega C_{iss} - (g_{gd} + j\omega C_{rss}) \cdot V_u \tag{37.71}$$

Es ist erkennbar, dass sich der kapazitive Anteil des Eingangsleitwertes bei größeren Werten für die Spannungsverstärkung beträchtlich erhöht, wodurch die obere Grenzfrequenz erheblich reduziert wird. Bei Annahme eines rein ohmschen Generatorleitwertes G_g ergibt sich folgender Zusammenhang:

obere Grenzfrequenz

$$f_o = \frac{g_{gs} + G_g + g_{gd}(1 - V_u)}{2\pi[C_{gs} + C_{gd}(1 - V_u)]} \tag{37.72}$$

Die kapazitive Rückwirkung über den Kondensator C_{gd} bewirkt also die niedrige Grenzfrequenz der Source-Schaltung. Der komplexe Generatorleitwert $\underline{Y}_g$ lässt sich als Parallelschaltung $\underline{Z}_S \parallel R_B$ berechnen.

Bei niedrigen Frequenzen ist der Eingangsleitwert rein reell und sehr klein

Eingangsleitwert

$$y_i = g_{gs} + g_{gd}(1 + \frac{S - g_{gd}}{g_{ds} + g_{gd} + G_L}) \tag{37.73}$$

und kann in vielen Fällen näherungsweise gelten als

$$\textit{Eingangsleitwert} \quad y_i \approx g_{gs} \tag{37.74}$$

Der AC-Ausgangswiderstand R_{aus} ist derjenige Widerstand, der als Innenwiderstand der Verstärkerstufe für einen Lastwiderstand wirksam ist. Der komplexe Ausgangsleitwert $\underline{y}_a$ der Source-Schaltung lässt sich nach (37.39) berechnen.

Ausgangsleitwert

$$y_a = g_{ds} + j\omega C_{oss} + \frac{j\omega C_{rss}(S - j\omega C_{rss})}{g_{gs} + j\omega C_{iss} + Y_g} \tag{37.75}$$

Bei niedrigen Frequenzen ergibt sich näherungsweise:

$$\textit{Ausgangsleitwert} \quad y_a \approx g_{ds} \tag{37.76}$$

Der Ausgangsleitwert lässt sich aus den Kennlinien zum Beispiel nach Abb. 37.33 ermitteln.Die wesentlichen Merkmale der Source-Schaltung im NF-Betrieb sind der geringe Eingangsleitwert, die große Spannungsverstärkung bei hochohmigem Lastwiderstand, die große Stromverstärkung und mittlere Werte für den Ausgangsleitwert.

AC-Eingangswiderstand

$$R_{ein} \approx R_V \parallel r_{gs} \quad \text{oder} \quad R_{ein} \approx R_B \parallel r_{gs}$$

da $r_{gs} \gg R_V$ oder $r_{gs} \gg R_B$ gilt folgende Näherung:

$$R_{ein} \approx R_V \quad \text{oder} \quad R_{ein} \approx R_B$$

AC-Ausgangswiderstand

$$R_{aus} \approx R_D \parallel r_{ds}$$

da $r_{ds} \gg R_D$ gilt folgende Näherung:

$$R_{aus} \approx R_D$$

Obwohl eine AC-Gegenkopplung die Spannungsverstärkung mindert, wird sie häufig angewendet, da sie die Source-Schaltung positiv beeinflusst. Die AC-Gegenkopplung bewirkt eine Verbesserung des Frequenzganges und der Stabilität. Außerdem ergeben sich geringere nichtlineare Verzerrungen des Signals.

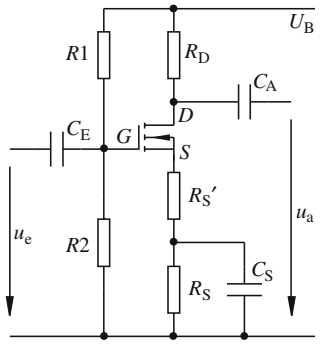


Abb. 37.43 Verstärkerschaltung mit einstellbarer Wechselstrom-Gegenkopplung

Um nicht die volle Gleichstrom-Gegenkopplung für Wechselsignale wirksam werden zu lassen, benutzt man häufig eine Schaltung nach Abb. 37.43. Hier wird eine Strom-Spannungs-Gegenkopplung durch einen Sourcewiderstand R'_S bewirkt. Die kapazitiven Elemente des Ersatzschaltbildes können bei niedrigen Frequenzen vernachlässigt werden. Bei ohmscher Last ergibt sich dann folgende Gleichung:

Spannungsverstärkung (NF)

$$V'_u \approx \frac{-S}{G_L + g_{ds} + R'_S G_L(S + g_{ds})} \tag{37.77}$$

Bei großer Steilheit S und einem großen Sourcewiderstand R'_S wird die Verstärkung nur durch das Verhältnis der Widerstände R_L und R'_S bestimmt.

Spannungsverstärkung (NF)

$$V'_u \approx -\frac{R_L}{R'_S} \tag{37.78}$$

Der Ausgangswiderstand des Transistors wird durch die „stabilisierende Wirkung“ des Gegenkopplungswiderstandes erhöht. Angenähert ergibt sich für den Ausgangswiderstand des Transistors folgender Wert:

Ausgangswiderstand

$$\frac{1}{y'_a} = r'_a \approx r_{ds}(1 + S \cdot R'_S)$$

Der Wert in der Klammer ist die Kreisverstärkung in der Gegenkopplungsschleife.

Wie in Verstärkerschaltungen mit bipolaren Transistoren wird die untere Grenzfrequenz durch die Widerstände und Kapazitäten (Hochpässe) in den Ein- und Ausgangskreisen bestimmt. FETs sind gleichstromsteuernde Bauelemente und haben keinen Einfluss auf die untere Grenzfrequenz f_u.

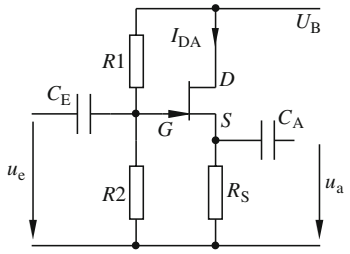


Abb. 37.44 J-FET als Verstärkerbauelement in Drainschaltung

Die Kapazitätswerte für C_E und C_A lassen sich berechnen nach folgenden Gleichungen.

Koppelkondensatoren

$$C_E = \frac{1}{2\pi \cdot f_u(R_B + Z_S)}$$

$$C_A = \frac{1}{2\pi \cdot f_u(R_D + Z_A)}$$

Jeder Hochpass in der Verstärkerschaltung bewirkt bei der unteren Grenzfrequenz eine Absenkung der Verstärkung um 3 dB. Soll in jeder Verstärkerstufe insgesamt nur eine Absenkung von 3 dB zugelassen sein, müssen die Kondensatoren 1,7mal so groß wie berechnet gewählt werden.

Der Sourcekondensator, der die Gegenkopplung für die Signale aufheben soll, kann nach folgender Gleichung berechnet werden:

$$C_S = \frac{S}{2\pi \cdot f_u}$$

37.2.2.2 Drainschaltung

Die Schaltung nach Abb. 37.44 zeigt einen J-FET als Verstärker in einer Drain-Schaltung. Hier darf der Sourcewiderstand nicht durch einen Kondensator gebrückt werden, weil in dieser Schaltung das Ausgangssignal am Sourcewiderstand abgegriffen wird. Der Koppelkondensator C_E verhindert eine Gleichstromkopplung mit dem Generator.

Da das Sourcepotential dem Gatepotential folgt ($u_a = u_e - u_{gs}$), wird die Drainschaltung auch Sourcefolger genannt.

Nimmt man nun (37.45) in Verbindung mit dem Grundersatzschaltbild, so ergibt sich die betriebliche Spannungsverstärkung der Drain-Schaltung.

Spannungsverstärkung

$$V_u = \frac{S + g_{gs} + j\omega(C_{iss} - C_{rss})}{S + g_{gs} + g_{ds} + j\omega(C_{iss} + C_{oss} - 2C_{rss}) + Y_L} \quad (37.79)$$

Die Spannungsverstärkung der Drain-Schaltung ist immer kleiner als Eins. Da die Kapazitäten im NF-Bereich vernachlässigt werden können, vereinfacht sich die Gleichung VI.79 bei rein ohmscher Last erheblich.

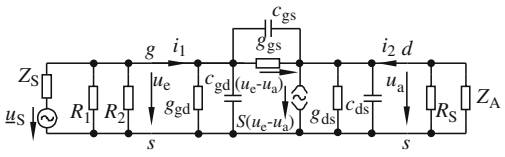


Abb. 37.45 AC-Ersatzschaltbild zur Drain-Schaltung

Spannungsverstärkung (NF)

$$V_u = \frac{S + g_{gs}}{S + g_{ds} + g_{gs} + G_L} \quad (37.80)$$

Das AC-Ersatzschaltbild zur Drain-Schaltung nach Abb. 37.44 zeigt Abb. 37.45. Der Leitwert G_L ist hier der Ersatzleitwert für $R_A \parallel R_S$. Sind S und G_L sehr viel größer als g_{gs} und g_{ds}, kann (37.80) weiter vereinfacht werden. Analog zur Source-Schaltung lässt sich auch die Stromverstärkung der Drain-Schaltung ermitteln.

Wie bei der Source-Schaltung lässt sich auch der komplexe Eingangsleitwert der Drain-Schaltung ermitteln. Für den NF-Bereich und bei rein ohmscher Last vereinfacht sich die Gleichung.

Eingangsleitwert (NF)

$$y_i = g_i = g_{gs} + g_{gd} - \frac{g_{gs}(S + g_{gs})}{S + g_{gs} + g_{ds} + G_L} \quad (37.81)$$

Sind S und G_L sehr viel größer als g_{gs} und g_{ds}, ergibt sich mit $g_{gd} \approx 0$ eine weitere zulässige Vereinfachung.

Eingangsleitwert (NF)

$$y_i = g_i = \frac{g_{gs} \cdot G_L}{S + G_L} \quad (37.82)$$

Wiederum analog zur Source-Schaltung lässt sich auch der komplexe Ausgangsleitwert der Drain-Schaltung ermitteln. Für den NF-Bereich ergibt sich mit $g_{gd} \approx 0$ und bei rein ohmschem Generatorleitwert G diese Gleichung.

Ausgangsleitwert (NF)

$$y_a = g_a = S + g_{ds} + \frac{g_{gs}(G_g - S)}{g_{gs} + G_g} \quad (37.83)$$

Der AC-Eingangswiderstand R_{ein} ist derjenige Widerstand, mit dem die Verstärkerschaltung die Signalquelle belastet.

AC-Eingangswiderstand

$$R_{ein} \approx R_V \parallel (1 + S \cdot R_S) r_{gs}$$

oder

$$R_{ein} \approx R_B \parallel (1 + S \cdot R_S) r_{gs}$$

also: $R_{ein} \approx R_V$ oder $R_{ein} \approx R_B$.

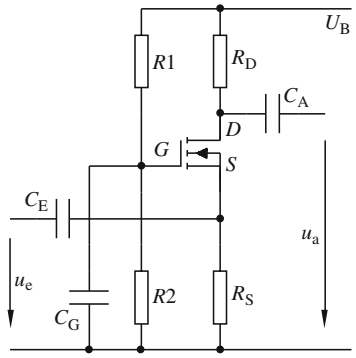


Abb. 37.46 Gate-Schaltung

Der AC-Ausgangswiderstand R_{aus} ist derjenige Widerstand, der als Innenwiderstand der Verstärkerstufe für einen Lastwiderstand wirksam ist.

AC-Ausgangswiderstand

$$R_{aus} \approx R_S \parallel 1/S \quad \text{also: } R_{aus} \approx 1/S$$

Die Drain-Schaltung ist vergleichbar mit der Kollektorschaltung von bipolaren Transistoren. Während der Eingangswiderstand sehr groß ist, wird der Ausgangswiderstand klein. Im Gegensatz zur Kollektorschaltung ist der Ausgangswiderstand unabhängig von Widerständen im Eingang.

Der Vorteil der Drain-Schaltung gegenüber der Source-Schaltung ist der geringere Ausgangswiderstand. Deshalb wird die Drain-Schaltung als Impedanzwandler und bei Verwendung eines Leistungs-FET auch als Leistungsverstärker benutzt.

37.2.2.3 Gateschaltung

Bei der Gate-Schaltung nach Abb. 37.46 liegt das Eingangssignal zwischen Source und Gate, während das Ausgangssignal zwischen Drain und Gate abgegriffen wird.

Das Eingangssignal soll wieder von einer Signalquelle U_S mit dem komplexen Innenwiderstand $\underline{Z}_S$ geliefert werden. Belastet wird die Verstärkerstufe mit dem komplexen Abschlusswiderstand $\underline{Z}_A$.

Das AC-Ersatzschaltbild nach Abb. 37.47 zeigt die Besonderheiten dieser Schaltung und macht die Berechnung der Betriebsgrößen im Vergleich mit der Vierpol-Ersatzschaltung möglich.

Nimmt man nun (37.45) in Verbindung mit dem Ersatzschaltbild, so ergibt sich die betriebliche Spannungsverstärkung der Gate-Schaltung.

Spannungsverstärkung

$$V_u = \frac{S + g_{ds} + j\omega C_{ds}}{g_{gd} + g_{ds} + j\omega(C_{gd} + C_{ds}) + Y_L} \tag{37.84}$$

Im NF-Bereich können wiederum die kapazitiven Leitwerte vernachlässigt werden, so dass sich bei rein ohmscher Last und $R_L = R_D \parallel R_A$ folgende Gleichung ergibt.

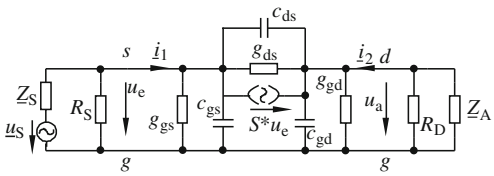


Abb. 37.47 AC-Ersatzschaltbild zur Gate-Schaltung nach Abb. 37.46

Spannungsverstärkung (NF)

$$V_u = \frac{S + g_{ds}}{g_{gd} + g_{ds} + G_L} \tag{37.85}$$

Sehr stark vereinfacht ergibt sich dann:

$$V_u(\text{NF}) \approx S \cdot R_L$$

Bei niedrigen Frequenzen liegt die Stromverstärkung bei ungefähr $V_i \approx -1$.

Der AC-Eingangswiderstand R_{ein} ist derjenige Widerstand, mit dem die Verstärkerschaltung die Signalquelle belastet. In der Schaltung nach Abb. 37.46 ist erkennbar, dass der Kondensator C_G die Widerstände R_1 und R_2 des Gatespannungsteilers für Signale kurzschließt, so dass diese Widerstände im AC-Ersatzschaltbild nach Abb. 37.47 nicht mehr erscheinen.

Eingangsleitwert

$$y_i = S + g_{gs} + g_{ds} + j\omega(C_{gs} + C_{ds}) - \frac{(g_{ds} + j\omega C_{ds})(S + g_{ds} + j\omega C_{ds})}{g_{gd} + g_{ds} + j\omega(C_{gd} + C_{ds}) + Y_L} \tag{37.86}$$

Für den NF-Bereich sind die Kapazitäten wieder vernachlässigbar.

Eingangsleitwert (NF)

$$y_i = g_i = S + g_{gs} + \frac{g_{ds}(G_L - S + g_{gd})}{g_{gd} + g_{ds} + G_L} \tag{37.87}$$

Analog dazu erhalten wir nun nach (37.33) den Ausgangsleitwert des FET.

Ausgangsleitwert

$$y_a = g_{gd} + g_{ds} + j\omega(C_{gd} + C_{ds}) - \frac{(g_{ds} + j\omega C_{ds})(S + g_{ds} + j\omega C_{ds})}{S + g_{gs} + g_{ds} + j\omega(C_{gs} + C_{ds}) + Y_g} \tag{37.88}$$

Für den NF-Bereich sind die Kapazitäten wieder vernachlässigbar.

Ausgangsleitwert (NF)

$$y_a = g_a = g_{gd} + \frac{g_{ds}(g_{ds} + G_g)}{S + g_{gs} + g_{ds} + G_g} \tag{37.89}$$

Die wesentlichen Merkmale der Gate-Schaltung sind ein großer Eingangs- und ein kleiner Ausgangsleitwert. Mit der Gate-Schaltung ist die höchste Grenzfrequenz aller Grundschaltungen zu erreichen. Sie wird darum fast ausschließlich bei hohen Frequenzen verwendet.

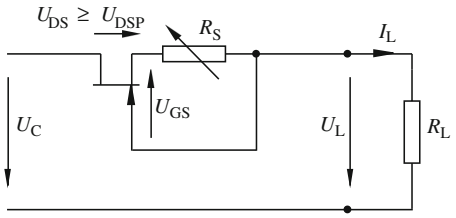


Abb. 37.48 Konstantstromquelle mit J-FET

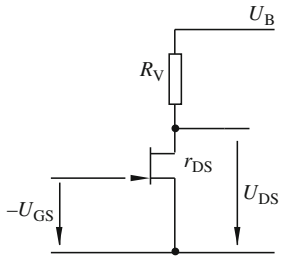


Abb. 37.49 J-FET als spannungsgesteuerter Widerstand

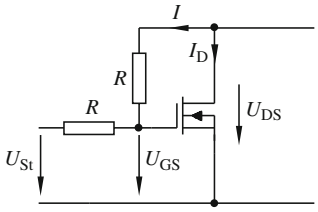


Abb. 37.50 Selbstleitender IG-FET als spannungsgesteuerter Widerstand

37.2.3 Weitere Anwendungen

Konstantstromquellen sind Schaltungen mit einem sehr hochohmigen Innenwiderstand, die in einem größeren Spannungsbereich nahezu gleichbleibende Ströme zur Verfügung stellen. Mit selbstleitenden Feldeffekt-Transistoren lassen sich solche Konstantstromquellen besonders einfach realisieren.

Bei der Spannung $U_{GS} = 0\,V$ fließt durch einen J-FET der maximale Drainstrom. In der Schaltung nach Abb. 37.48 lässt sich die Gate-Source-Spannung mit Hilfe des Sourcewiderstandes R_S variieren und damit wiederum der Ausgangsstrom I_L veränderbar gestalten. Der Arbeitspunkt des FET muss im Abschnürbereich des Ausgangskennlinienfeldes liegen, dass heißt, die Drain-Source-Spannung U_{DS} darf nicht kleiner sein als U_{DSP}.

Unter Vernachlässigung des Gatestroms ergibt sich eine einfache Gleichung.

$$U_B = U_{DS} - U_{GS} + R_L \cdot I_L$$

Die Gatespannung wird meistens in der Weise gewählt, dass der Arbeitspunkt in der Mitte der Steuerkennlinie liegt. Hier lässt sich auch ermitteln, welche Gate-Source-Spannung für einen bestimmten Laststrom I_L bei vorgegebenem Sourcewiderstand erforderlich ist.

$$\textit{Laststrom} \quad I_L = I_D = \frac{-U_{GS}}{R_S} \tag{37.90}$$

Bei vorgegebenem Ausgangsstrom darf der Lastwiderstand einen bestimmten maximalen Wert nicht übersteigen.

maximaler Lastwiderstand

$$R_{L\,max} = \frac{U_B + U_P}{I_L} \tag{37.91}$$

Bei konstanter Betriebsspannung der Schaltung lässt sich der Innenwiderstand dieser Konstantstromquelle berechnen.

$$\textit{Innenwiderstand} \quad r_i = R_S + r_{DS}(S \cdot R_S + 1) \tag{37.92}$$

Der Arbeitspunkt eines *IG-FET* wird nun in den ohmschen Bereich der Ausgangskennlinie gelegt und so die Drain-Source-Strecke *als spannungsgesteuerter Widerstand* betrieben. Mit Hilfe der Gate-Source-Spannung lässt sich dieser Widerstand zwischen Werten von einigen Ohm bis in den Megaohm-Bereich hinein nahezu leistungslos variieren. Da die I_D-U_{DS}-Kennlinien des FET aber nur für kleine Werte von U_{DS} annähernd linear verlaufen, sind die Anwendungsmöglichkeiten sehr eingeschränkt.

Wird ein Festwiderstand R_V nach Abb. 37.49 in Reihe zum Drain-Source-Widerstand eines J-FET geschaltet, entsteht ein Spannungsteiler, der sich leicht berechnen lässt.

$$\textit{Spannungsteiler} \quad U_{DS} = |U_{GS}| \cdot \frac{r_{DS}}{R_V + r_{DS}} \tag{37.93}$$

Der geringste Widerstand der Drain-Source-Strecke ergibt sich, wenn die Steuerspannung $U_{GS} = 0\,V$ ist. Damit ergibt sich eine interessante Kenngröße (mit β als Materialkonstante), der minimale Drain-Source-Widerstand.

minimaler Drain-Source-Widerstand

$$r_{DS(on)} = -\frac{1}{\beta \cdot U_P} \tag{37.94}$$

Der Widerstand der Drain-Source-Strecke lässt sich auch in der folgenden Form angeben:

Drain-Source-Widerstand

$$r_{DS} = \frac{r_{DS(on)}}{1 - \frac{U_{GS}}{U_P}} \tag{37.95}$$

Durch eine Rückkopplung zwischen Drain und Source wird die vorhandene Nichtlinearität weitgehend aufgehoben, wenn eine Spannung U_{GS} gewählt wird, die sich zur Hälfte aus der Steuerspannung U_{St} am Eingang und der Spannung U_{DS} zusammensetzt.

Eine entsprechende Schaltung mit einem selbstleitenden IG-FET zeigt Abb. 37.50.

Für die Drain-Source-Strecke ergibt sich R_{DS} als

$$R_{DS} = r_{DS} = \frac{r_{DS(on)}}{1 - \frac{U_{St}}{2 \cdot U_P}}$$

Damit errechnet sich der gesteuerte Widerstand R_{St}.

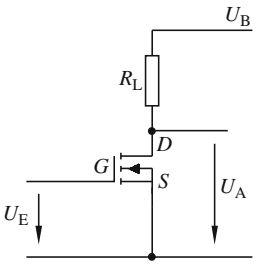


Abb. 37.51 Selbstsperrende N-Kanal-IG-FET als Schalter im Serienbetrieb

gesteuerter Widerstand

$$R_{St} = \frac{1}{\beta(\frac{U_{St}}{2} - U_P) + \frac{1-\frac{U_{St}}{U_{DS}}}{2R}} \approx r_{DS} \qquad (37.96)$$

Wählt man die Widerstände R sehr hochohmig, kann der Strom I vernachlässigt werden und der Näherungswert aus (37.96) wird gültig.

Im Vergleich zum bipolaren Transistor weist der FET beim Einsatz als Schalter einige Vorteile auf:

- nahezu leistungslose Steuerung aufgrund des hohen Eingangswiderstandes
- Drainstrom sehr gering im gesperrten Zustand
- für integrierte Schaltungstechnik aufgrund der einfachen Herstellung und geringen Abmessungen gut geeignet
- selbstsperrende IG-FET für digitale Logikschaltungen besonders geeignet, da die einzelnen Schalterstufen direkt gleichstrommäßig gekoppelt werden können
- gutes dynamisches Schaltverhalten, da nur Majoritätsträger am Strom beteiligt sind.

Die Schaltung nach Abb. 37.51 zeigt einen selbstsperrenden N-Kanal-IG-FET als Schalter im Serienbetrieb. Solange die Eingangs- oder Steuerspannung U_E kleiner ist als die Schwellenspannung $U_{GS(TO)}$, befindet sich der FET im gesperrten Zustand und auf der Arbeitsgeraden im Arbeitspunkt AP1. In diesem Zustand fließt nur der sehr geringe Drainstrom $I_{D(Off)}$, wobei der Gleichstromwiderstand $R_{DS(Off)}$ und der differentielle Widerstand $r_{DS(Off)}$ etwa den gleichen Wert (10^9–$10^{11}\,\Omega$) haben.

Überschreitet die Eingangsspannung die Schwellenspannung, so verschiebt sich der Arbeitspunkt an der Arbeitsgeraden entlang zu größeren Drainstromwerten. Bei einer ausreichenden Eingangsspannung stellt sich der Arbeitspunkt AP2 im ohmschen Bereich des Kennlinienfeldes nach Abb. 37.52 ein. Es fließt nun ein Drainstrom $I_{D(On)}$, da die Drain-Source-Strecke relativ niederohmig ist, wobei auch hier der Gleichstromwiderstand $R_{DS(Off)}$ und der differentielle Widerstand $r_{DS(Off)}$ etwa den gleichen Wert (10 bis einige $100\,\Omega$) haben.

Schalterstufen erhalten für die Integration besonders günstige Eigenschaften, wenn man als Lastwiderstand ebenfalls einen IG-FET verwendet, der gegensinnig zum eigentlichen Schalttransistor besteuert wird. Das wird sehr einfach, wenn man für Steuer- und Last-FET komplementäre Systeme verwendet (C-MOS-Technik). Wird der eigentliche Schalttransistor V2 nach

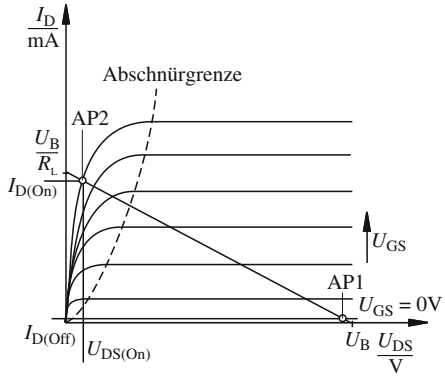


Abb. 37.52 Ausgangskennlinienfeld eines IG-FET mit Arbeitsgerade

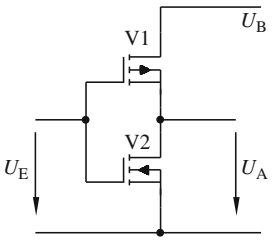


Abb. 37.53 Schalttransistor in C-MOS-Technik

Abb. 37.53 leitend, so wird der Lasttransistor V1 gesperrt. Der Lasttransistor V1 ist also bei Einschaltung sehr hochohmig, was im Hinblick auf einen geringen Leistungsverbrauch in integrierten Schaltkreisen vorteilhaft ist. Die Schaltung nach Abb. 37.53 stellt einen Inverter mit komplementären IG-FET dar.

37.3 Mehrstufige Verstärker

Obwohl integrierte Schaltkreise (IC = integrated circuits) die bisherigen diskreten Schaltungen und Transistoren immer mehr verdrängen, werden die Transistoren auch in Zukunft noch in vielen elektronischen Schaltungen eine erhebliche Bedeutung behalten.

Dieses gilt insbesondere auch für die Lösung von Anpassungsproblemen an Ein- und Ausgängen von integrierten Schaltkreisen sowie für die Leistungsverstärker.

Ein Verstärker besteht einerseits aus einer oder mehreren Vorstufen, die hauptsächlich zur Spannungsverstärkung und eventuell zur Signalbeeinflussung (zum Beispiel Klangregelung) dienen. Andererseits besteht er aus einem Leistungsverstärker, der mit einem Eingangssignal hoher Spannungsamplitude angesteuert wird und dessen Aufgabe es ist, vor allem eine Leistungs- oder Stromverstärkung vorzunehmen.

Der Leistungsverstärker selbst kann aus einer oder mehreren Treiberstufen und einer Endstufe bestehen.

Treiberstufen haben die Aufgabe, die zum Ansteuern der Endstufe notwendige Steuerleistung zur Verfügung zu stellen. Um

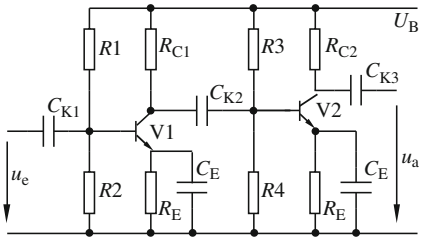


Abb. 37.54 Kapazitive Kopplung zweier Transistorstufen

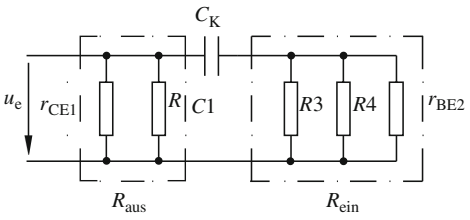


Abb. 37.55 Koppelkondensator in Wechselwirkung zwischen Ein- und Ausgangswiderstand

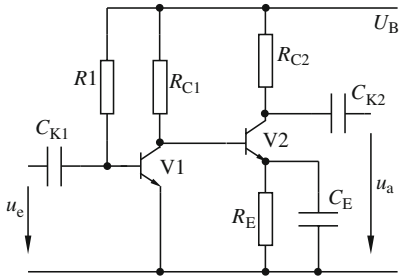


Abb. 37.56 Direkte Kopplung zweier Transistorstufen

eine *optimale* Leistungsübertragung zu gewährleisten, sollte zwischen Treiber- und Endstufe eine Leistungsanpassung vorliegen. Das bedeutet, dass der Eingangswiderstand der Endstufe gleich dem Ausgangswiderstand der Treiberstufe sein sollte.

Es nicht ohne weiteres möglich, zwei Verstärkerstufen miteinander zu koppeln, da die Gleichspannungspotentiale der miteinander zu koppelnden Schaltungspunkte nicht übereinstimmen.

Die Kopplung hat die Aufgabe, Signalspannungen vom Ausgang der einen Stufe möglichst unverfälscht auf den Eingang der nachfolgenden Stufe zu übertragen.

Dabei ist grundsätzlich zwischen **drei Kopplungsarten** zu unterscheiden: Übertragerkopplung, kapazitive Kopplung (RC-Kopplung) und galvanische Kopplung (Gleichstromkopplung).

Die *Übertragerkopplung*, also die Signalübertragung von einer Stufe zur nächsten mit Hilfe eines „Transformators", hat in der modernen Elektronik kaum noch eine Bedeutung. Das gilt auch für die Gleichstromkopplung in der diskreten Schaltungstechnik. Eine *Gegenkopplung* über zwei Stufen hinweg lässt sich jedoch nur bei der galvanischen Kopplung realisieren.

In Operationsverstärkern (OpAmp) ist die *galvanische* Kopplung wichtigstes Prinzip.

Die *kapazitive Kopplung* nach Abb. 37.54 ist wohl die problemloseste Kopplungsart. Aber während bei einer galvanischen Kopplung zwischen zwei Verstärkerstufen sowohl Gleichspannungs- als auch Wechselspannungssignale übertragen werden, können hier nur Wechselspannungen übertragen werden.

Die Kopplung der einzelnen Stufen erfolgt mit Hilfe eines Koppelkondensators C_{K2}, während die Stufen des Verstärkers gleichstrommäßig voneinander unabhängig sind. Daher können die *Arbeitspunkte* jeder Stufe getrennt voneinander eingestellt werden.

Die untere Grenzfrequenz des Verstärkers hängt wesentlich von dem RC-Glied (Hochpass) ab, das durch den Koppelkondensator und den Eingangswiderstand R_{ein} der nachfolgenden Stufe sowie den Ausgangswiderstand R_{aus} der vorausgehenden Stufe nach Abb. 37.55 gebildet wird.

Der Koppelkondensator C_K muss so bemessen sein, dass Wechselspannungen bei der unteren Grenzfrequenz noch ungehindert auf die nächste Stufe übertragen werden. Die Grenzfrequenz des Hochpasses ist gleich der unteren Grenzfrequenz f_{gu} des Verstärkers, wenn C_K die einzige beeinflussende Kapazität darstellt. Damit lässt sich die Kapazität des Koppelkondensators berechnen.

Koppelkondensator

$$C_K = \frac{1}{2\pi f_{gu}(R_{aus} + R_{ein})} \tag{37.97}$$

Die wirksamen Widerstände lassen sich nach Abb. 37.55 als Parallschaltung berechnen. Die Nennspannung dieses Koppelkondensators ergibt sich aus der Differenz der Ruhepotentiale des Kollektors von V1 und der Basis von V2.

Je niedriger die untere Grenzfrequenz sein soll, um so größer muss die Kapazität des Koppelkondensators sein.

Die untere Grenzfrequenz f'_{gu} eines mehrstufigen Verstärkers ist nicht identisch mit der unteren Grenzfrequenz f_{gu} der gleichartigen Teilverstärkerstufe. Für einen vielstufigen Verstärker mit n identischen Stufen gilt dann:

$$\textit{Grenzfrequenz} \quad f'_{gu} = f_{gu}(\sqrt{2})^{(n-1)} \tag{37.98}$$

Es findet auch eine Minderung der Verstärkung bei der oberen Grenzfrequenz f'_{go} statt, da sich die jeweiligen Querkapazitäten summieren. Das bedeutet, dass die Bandbreite B' des Gesamtverstärkers kleiner ist als die Bandbreite B der Teilverstärkerstufen.

Bei der RC-Kopplung mehrerer Stufen kann für bestimmte Frequenzen eine große Phasenverschiebung zwischen Ausgangs- und Eingangsspannung auftreten. Eine Gegenkopplung kann unter Umständen in eine Mitkopplung verkehrt werden. Wesentlich günstiger ist es, mehrere Transistoren galvanisch zu koppeln und alle Arbeitspunkte der Schaltung durch eine einzige Gleichspannungs-Gegenkopplung zu stabilisieren.

Diese direkte Kopplung nach Abb. 37.56 ist jedoch nur möglich, wenn die Potentiale des Kollektors von V1 und der Basis von V2 gleich sind.

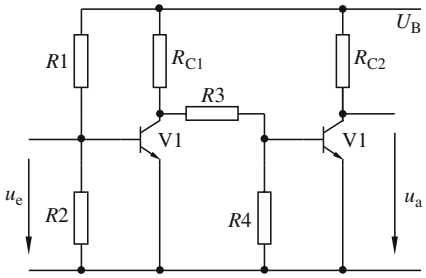


Abb. 37.57 Potentialanpassung durch Spannungsteiler

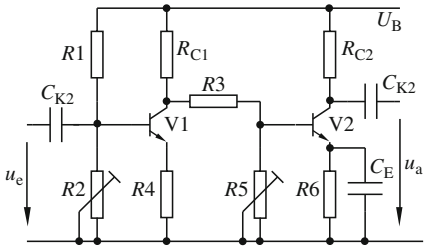


Abb. 37.58 Optimierung der Arbeitspunkte durch Potentiometer

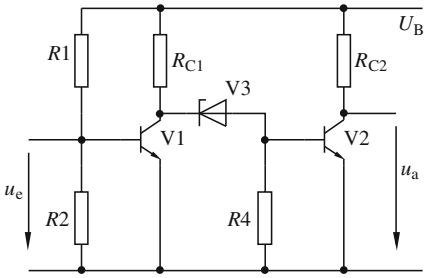


Abb. 37.59 Potentialanpassung durch Spannungsteiler mit Z-Diode

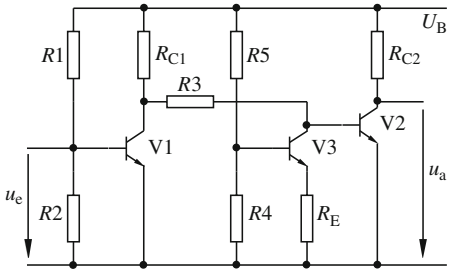


Abb. 37.60 Potentialanpassung durch Transistorkonstantstromquelle

Der Nachteil mehrstufiger Verstärker mit direkter Kopplung liegt darin, dass sich geringe Arbeitspunktverschiebungen der ersten Stufe verstärkt auf die nachfolgende(n) Stufe(n) auswirken. Temperaturänderungen oder Bauteiltolerenzen bei Ersatzbestückung können dafür verantwortlich sein.

Wenn die Potentiale des Kollektors von V1 und der Basis von V2 jedoch nicht gleich sind, so kann man das Basisruhepotential von V2 über einen von der Kollektor-Emitter-Spannung von V1 versorgten Spannungsteiler gewinnen.

$$\frac{U_{BE2}}{U_{CE1}} \approx \frac{R_4}{R_3 + R_4}$$

Nachteilig ist jedoch, dass auch die *Signalspannung* im Verhältnis der Spannungsteilerwiderstände R3 und R4 nach Abb. 37.57 *herabgesetzt* wird.

Damit lässt sich die Gesamtverstärkung näherungsweise berechnen.

Gesamtverstärkung

$$V_{u\,\mathrm{ges}} \approx V_{u1} \cdot \frac{\Delta U_{BE2}}{\Delta U_{CE1}} = V_{u1} \cdot \frac{R_4 \parallel r_{BE2}}{R_3 + R_4 \parallel r_{BE2}} \qquad (37.99)$$

Der Arbeitspunkt der 1. Stufe nach Abb. 37.58 wird mit Hilfe des Widerstandes R2 für den A-Betrieb eingestellt. Der Emitterwiderstand R4 sorgt für die thermische Arbeitspunktstabilisierung infolge der Stromgegenkopplung. Da der R4 nicht durch einen Emitterkondensator gebrückt ist, wirkt sich die Stromgegenkopplung verstärkungsmindernd auf die Wechselspannung aus. Arbeitspunktverschiebungen infolge Temperaturänderung werden an die 2. Stufe weitergegeben.

Zur Verbesserung der Aussteuerbarkeit der 2. Stufe wird deren Arbeitspunkt durch den Spannungsteiler aus R3 und R5 auf den gleichen Arbeitspunkt wie die 1. Stufe eingestellt. Der Verlust von Signalspannung am Widerstand R5 ist in dieser Variante nicht zu vermeiden.

Die Gesamtverstärkung ist relativ klein, weil die am Ausgang der 1. Stufe liegende Signalspannung u. a. durch den Spannungsteiler und den kleinen Eingangswiderstand der 2. Stufe stark herabgesetzt wird.

Die Herabsetzung der Signalspannung durch das Koppelglied kann vermieden werden, wenn man einen der beiden Spannungsteilerwiderstände R3 und R4 durch ein Bauelement ersetzt, dessen Wechselstromwiderstand (differentieller Widerstand) sich von seinem Gleichstromwiderstand wesentlich unterscheidet.

Es erfolgt wie gewünscht eine Herabsetzung des Gleichspannungspotentials, jedoch keine Herabsetzung der Signalspannung.

Nun bestehen zwei Möglichkeiten:

- man ersetzt R3 durch ein Bauelement mit niedrigem differentiellen Widerstand oder
- man ersetzt R4 durch ein Bauelement mit hohem differentiellen Widerstand.

In beiden Fällen ist das *Spannungsteilerverhältnis* wechselstrommäßig wesentlich günstiger als bei einer Kopplung mit ohmschem Spannungsteiler. Für die erste Möglichkeit bieten sich Dioden und Z-Dioden an, die nach Abb. 37.59 integriert werden.

Bei der zweiten Möglichkeit wird der Widerstand R4 durch eine Transistorkonstantstromquelle in Abb. 37.60 ersetzt.

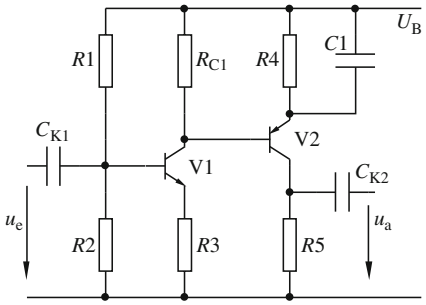


Abb. 37.61 Zweistufiger Verstärker mit komplementären Transistoren

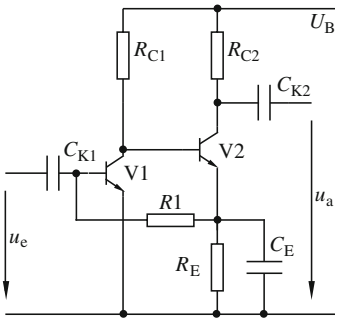


Abb. 37.62 Zweistufiger NF-Verstärker mit galvanischer Kopplung und Gegenkopplung

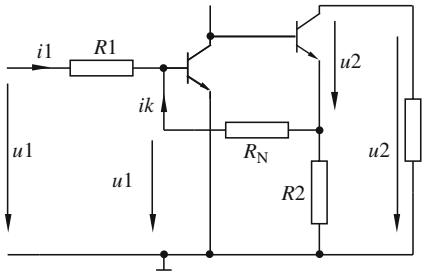


Abb. 37.63 Prinzip der Strom-Strom-Gegenkopplung

Es gilt

$$\Delta U_{CE1} \approx \Delta U_{BE2}$$

In beiden Fällen wird also die Signalspannung durch das Koppelglied nur unwesentlich herabgesetzt. Allerdings sind diese Schaltungen aufwendiger. Sie sollten wegen des besonders hohen Aufwandes für das Koppelglied nur bei hohen Anforderungen an geringes Rauschen verwendet werden.

Die genannten Nachteile lassen sich weitgehend vermeiden, wenn man den zweistufigen Verstärker mit komplementären Transistoren aufbaut, wie die Schaltung nach Abb. 37.61 zeigt.

Beide Transistoren arbeiten auch hier in Emitterschaltung. Mit den Widerständen R2 und R4 werden die Arbeitspunkte so eingestellt, dass beide Stufen im A-Betrieb arbeiten, so dass jeweils ein größtmöglicher Aussteuerbereich vorhanden ist. Mit dieser Schaltung lässt sich eine höhere Verstärkung erreichen.

Bei mehrstufigen Verstärkern wird neben der gleichstrommäßigen Gegenkopplung auch eine zusätzliche wechselstrommäßige Gegenkopplung über mehrere Stufen verwendet.

Eine Gegenkopplung liegt vor, wenn ein Teil der Signalspannung am Ausgang des Verstärkers gegenphasig, dass heißt mit einer Phasenverschiebung von 180° auf den Eingang des Verstärkers zurückgeführt wird. Solange diese Phasenbedingung eingehalten wird, kann die Rückkopplung über eine Stufe oder auch über mehrere Stufen erfolgen. Die Verstärkerkennlinie wird linearisiert. Dadurch werden die sonst auftretenden Verzerrungen verringert.

Bei der Schaltung nach Abb. 37.62 handelt es sich um einen zweistufigen NF-Verstärker mit galvanischer Kopplung beider Stufen. Für Wechselspannungen arbeiten die Transistoren V1 und V2 in Emitterschaltung ohne Gegenkopplung. Die Widerstände R1 und R_E dienen der Arbeitspunkteinstellung und -stabilisierung. R_E bewirkt eine Gleichstromgegenkopplung für den Transistor V2. Für Gleichspannungen arbeitet V2 bezüglich seines Emitterpotentials als Emitterfolger. Bei Vernachlässigung von U_{BE2} ist damit sein Emitter-DC-Potential gleich dem Kollektor-DC-Potential von V1.

Über den Widerstand R1 ist der Transistor V1 für DC voll gegengekoppelt. Aufgrund des Kondensators C_E ist der Verstärker für Wechselgrößen nicht gekoppelt.

Das Prinzip der Strom-Strom-Gegenkopplung lässt sich anhand des in Abb. 37.63 dargestellten Regelkreises unter Beachtung der Schaltung nach Abb. 37.62 nachvollziehen. Eine Erhöhung des Emitterstroms von V2 bewirkt eine Erhöhung des Spannungsabfalles an R_E, wodurch das Basisruhepotential von V1 größer wird. Dadurch wird der Basisstrom von V1 erhöht, was zu einer Erhöhung des Kollek- torstroms von V1 führt. Damit senkt sich das Basisruhepotential von V2; der Transistor V1 wird weniger leitend und der Emitterstrom von V2 verringert sich.

Bei galvanisch gekoppelten Verstärkerstufen sind die Transistoren direkt oder über ohmsche Widerstände miteinander verbunden. Alle Arbeitspunkte werden durch eine einzige Gleichspannungsgegenkopplung stabilisiert.

Durch eine wechselstrommäßige Gegenkopplung können die Eigenschaften des Verstärkers weiter verbessert werden. Bei geeigneter Dimensionierung kann zum Beispiel die Spannungsverstärkung weitgehend unabhängig von den Transistordaten und deren Streuung gemacht werden. Hierbei wird häufig eine Spannungs-Spannungs-Gegenkopplung verwendet. In Abb. 37.64 wird das Grundprinzip der Spannungs-Spannungs-Gegenkopplung dargestellt.

Die Ausgangswechselspannung u_2 wird über den Widerstand R1 auf den Emitter von V1 zurückgekoppelt. Das zur Emitterwechselspannung u_k phasengleiche Signal u_2 vergrößert u_k, so dass auch die Eingangswechselspannung u_1 größer werden muss. Die Gesamtverstärkung wird allerdings kleiner. Die Schaltung nach Abb. 37.65 gibt dieses Prinzip der Spannungs-Spannungs-Gegenkopplung wieder.

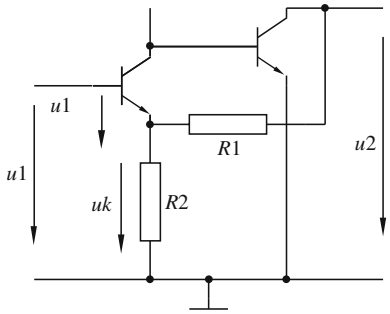


Abb. 37.64 Prinzip der Spannungs-Spannungs- Gegenkopplung

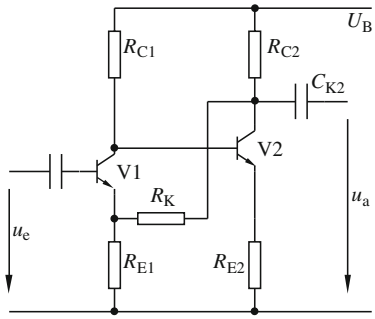


Abb. 37.65 Schaltung mit Anwendung der Spannungs-Spannungs-Gegenkopplung

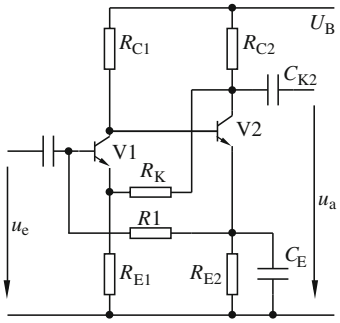


Abb. 37.66 Zweistufiger NF-Verstärker mit galvanischer Kopplung und doppelter Gegenkopplung

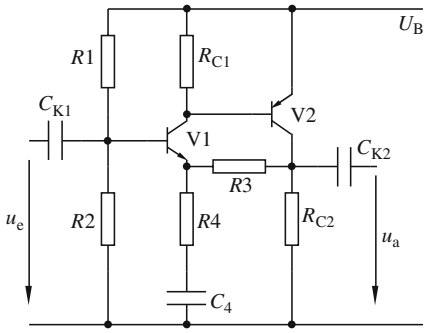


Abb. 37.67 Zweistufiger NF-Verstärker mit Komplementärtransistoren

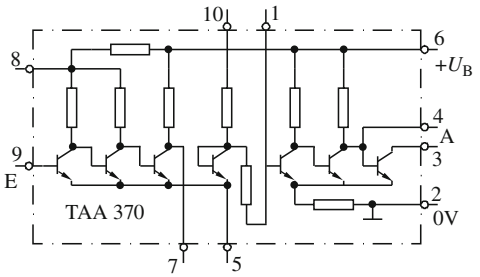


Abb. 37.68 Innenbeschaltung des IC TAA 370

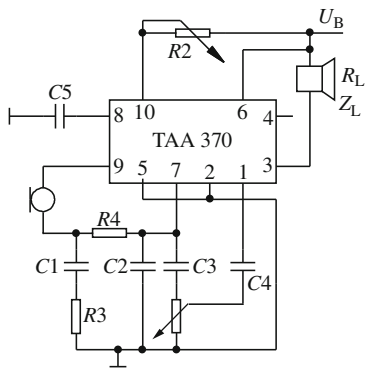


Abb. 37.69 Hörgeräteverstärker als Application des TAA 370

Kombiniert man beide Gegenkopplungsverfahren miteinander, ergibt sich die Schaltung nach Abb. 37.66. Der Emitter von V1 liegt nun nicht mehr auf Nullpotential. Die Gleichspannung an R_{E1} ist jedoch vernachlässigbar klein, wenn man diesen Widerstand genügend klein und R_K genügend groß wählt.

Diese beiden Widerstände sind für die AC-Gegenkopplung zuständig, während R1 und R_{E2} für die DC-Gegenkopplung über beide Stufen hinweg sorgen.

Die Gegenkopplung hat aber auch einen wesentlichen Einfluss auf die Bandbreite des Verstärkers. So wird die Bandbreite um so größer, je stärker die Verstärkung V_u durch die Gegenkopplung auf V_u' verkleinert wird. Dies ist besonders bei Breitbandverstärkern von Bedeutung.

Macht man nun die AC-Gegenkopplung sehr stark, $\beta_2 \cdot R_{C2}/R_K \gg 1$, so vereinfacht sich die Gleichung weiter.

$$\textit{Spannungsverstärkung} \quad V_u' \approx \frac{R_K}{R_{E1}} \tag{37.100}$$

In Anlehnung an die Schaltung nach Abb. 37.61 kann auch eine Schaltung nach Abb. 37.67 verwendet werden, in der Komplementärtransistoren miteinander verschaltet sind. Der Spannungsabfall am Widerstand R_{C1} stellt die Basis-Emitter-Spannung U_{BE2} für V2. Der Widerstand R3 stellt die DC-Gegenkopplung dar. Die Reihenschaltung aus R4 und C_4 verringert die Gegenkopplung für Wechselspannungen.

Diese Schaltungskombination ermöglicht oft eine einfachere Potentialanpassung.

NF-Verstärker werden häufig in integrierter Schaltungstechnik (IC) hergestellt. Es handelt sich dabei um galvanisch gekoppelte Verstärkerstufen, wobei die Gesamtverstärkung und der Frequenzgang häufig nur durch eine äußere Beschaltung beeinflusst werden. Ein Beispiel für einen solchen NF-Verstärker ist der IC TAA 370 in Abb. 37.68.

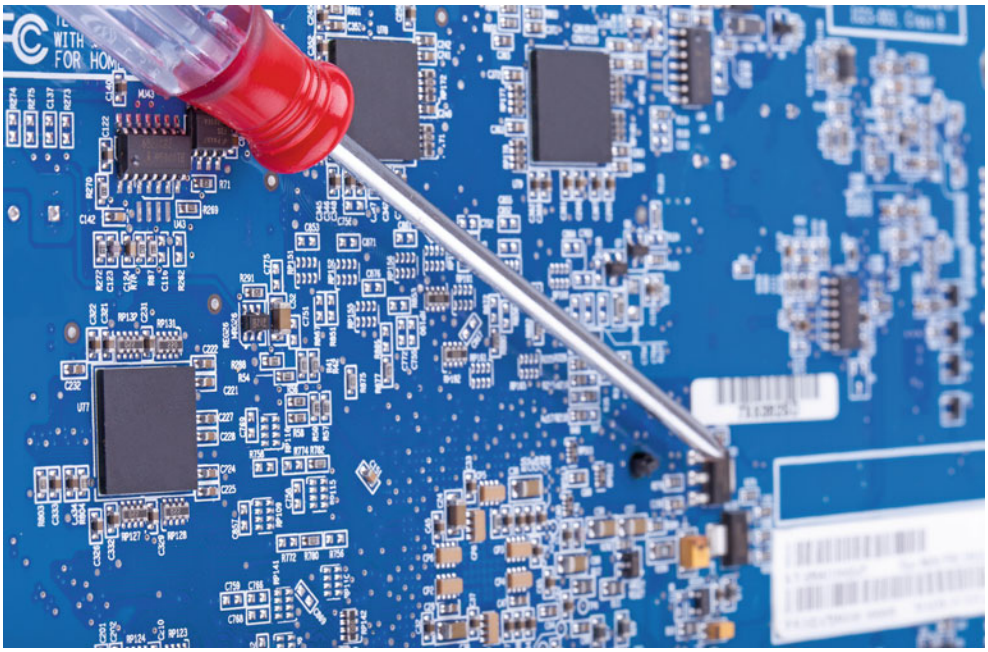
Als Schaltungsbeispiel (application) ist in Abb. 37.69 ein Hörgeräteverstärker dargestellt. Die Ausgangsleistung der Schaltung beträgt $P = 1{,}5\,\mathrm{mW}$, wobei sich die Verstärkung und damit die Lautstärke mit dem Potentiometer R1 einstellen lässt.

$$
\begin{aligned}
&U_B = +1{,}3\,\mathrm{V}, \quad R_1 = 25\,\mathrm{k\Omega}, \quad R_2 = 5\,\mathrm{k\Omega},\\
&R_3 = 1{,}5\,\mathrm{k\Omega}, \quad R_4 = 120\,\mathrm{k\Omega}, \quad C_1 = 0{,}47\,\mu\mathrm{F},\\
&C_2 = 22\,\mathrm{nF}, \quad C_3 = C_4 = 1\,\mu\mathrm{F}, \quad C_5 = 2{,}2\,\mu\mathrm{F},\\
&R_L = 100\,\Omega, \quad Z_L = 300\,\Omega
\end{aligned}
$$

Endstufen

38

Peter Döring



38.1 Betriebsarten 478
38.2 Schaltungen 478

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_38

38.1 Betriebsarten

Um möglichst große Leistungen an einen Verbraucher, wie zum Beispiel einen Lautsprecher, abgeben zu können, braucht man Endstufen. Das sind Verstärkerschaltungen, bei denen eine *möglichst große Signalleistung* im Vordergrund steht, während die Spannungsverstärkung nur eine Nebenrolle spielt.

Da große Leistungen selbstverständlich auch große Spannungen und Ströme bedeuten, werden die Transistorkennlinien in den Endstufen voll durchgesteuert. Eine nicht völlig lineare Kennlinie führt dabei zu Verzerrungen des Ausgangssignales.

Bei Leistungsverstärkern können verschiedene Arbeitspunkte auf der Arbeitsgeraden gewählt werden. Man unterscheidet dann Verstärker im A-Betrieb, Verstärker im B-Betrieb und Verstärker im AB-Betrieb.

Bisher wurde der Arbeitspunkt immer in die Mitte des geradlinigen Kennlinienfeldes gelegt. Diese Betriebsart nennt man *A-Betrieb*. Für die Arbeitspunkteinstellung ist die Vorspannung U_{BEA} erforderlich, aus der ein Basis-Ruhestrom I_{BA} beziehungsweise ein Kollektor-Ruhestrom I_{CA} folgt. Der Transistor wird dabei durch beide Halbwellen des Steuersignals ausgesteuert. Die Ruheströme fließen aber auch, wenn die Signalspannung Null ist.

Bei diesen im A-Betrieb arbeitenden Verstärkern werden beide Halbwellen einer sinusförmigen Signalspannung mit einem einzigen Transistor verstärkt. Man spricht deshalb auch von einem *Eintakt-Betrieb*.

Beim *B-Betrieb* wird der Transistor ohne Vorspannung U_{BE} betrieben. Der Basisstrom und der Kollektorstrom sind Null, wenn die Signalspannung Null ist. Daraus folgt, dass der Transistor nur eine Halbwelle einer Wechselspannung verstärken kann. Hierbei liegt der Arbeitspunkt im unteren Kennlinienknick der Steuerkennlinie. Die andere Halbwelle muss von einem zweiten Transistor verstärkt werden. Man spricht dann von einem *Gegentakt-Betrieb*, weil die zwei Transistoren abwechselnd, nämlich im Takt der positiven und negativen Halbwellen, im aktiven Bereich arbeiten. Nach der Verstärkung werden die beiden Halbwellen wieder zusammengefügt.

Da im B-Betrieb keine Ruheströme fließen, ist die Verlustleistung des Verstärkers wesentlich geringer als beim A-Betrieb, was kleinere Kühlkörper zur Folge hat und Energie sparen hilft.

Um nichtlineare Verzerrungen zu mildern, bildet man eine Synthese zwischen A- und B-Betrieb zum *AB-Betrieb*. Hierbei wird wie beim A-Betrieb durch eine Basis-Vorspannung U_{BE} ein Ruhestrom eingestellt, der aber wesentlich geringer ist als beim A-Betrieb.

Die Vorspannung U_{BEA} ist gerade so groß, dass sie die Schleusenspannung der Basis-Emitter-Diode kompensiert. Auch hier ist nur ein *Gegentakt-Betrieb* sinnvoll.

Sowohl Vorverstärker als auch Treiberstufen werden hauptsächlich als A-Verstärker in Emitter- oder Kollektorschaltung betrieben.

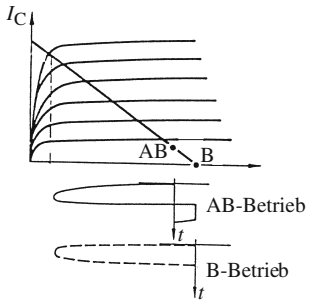


Abb. 38.1 Lage der Arbeitspunkte bei B- beziehungsweise AB-Betrieb im Kennlinienfeld

Bei großen Leistungen spielt der Wirkungsgrad eine erhebliche Rolle. Gute Endstufen erreichen einen Wirkungsgrad von nahezu 70 %. Die Verlustleistung der Transistoren wird im Transistor in Wärme umgewandelt und muss an die Umgebungsluft abgegeben werden. Wegen der relativ hohen Verlustleistung sind die Endtransistoren meistens auf Kühlkörper montiert.

38.2 Schaltungen

Durch den ständig fließenden großen Ruhestrom tritt bei A-Betrieb eines Leistungsverstärkers die größte Verlustleistung auf, wenn der Verstärker nicht ausgesteuert wird. Der Wirkungsgrad einer solchen Schaltung kann nicht größer als 50 % werden.

Für Leistungsverstärker werden im allgemeinen Verstärkerstufen in Kollektorschaltung (Emitterfolger) verwendet, da sie einen hohen Eingangswiderstand bei niedrigem Ausgangswiderstand aufweist. Eine solche Kombination ist zur Speisung niederohmiger Verbraucher wie Lautsprecher notwendig.

Die Schaltung nach Abb. 38.2 zeigt einen Emitterfolger mit gleich großen positiven und negativen Betriebsspannungen, wie er überwiegend bei Audio-Verstärkern verwendet wird.

Von Vorteil ist die einfache Arbeitspunkteinstellung (keine zusätzlichen Widerstände erforderlich) und der Fortfall der Koppelkapazitäten am Ein- und Ausgang. Der Lastwiderstand R_L steht symbolisch für den Lautsprecher.

Infolge der Arbeitspunkteinstellung für A-Betrieb, also im linearen Teil der Eingangskennlinie, ergibt sich ein Basis-

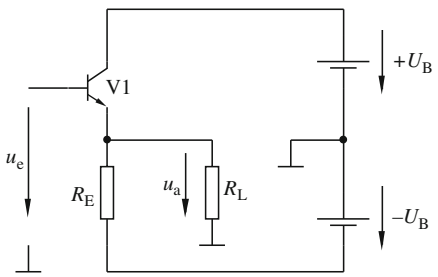


Abb. 38.2 Emitterfolger mit positiven und negativen Betriebsspannungen

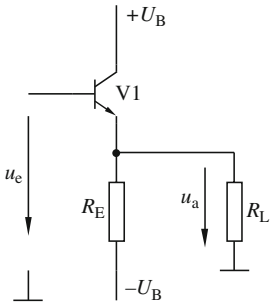


Abb. 38.3 Emitterfolger in der vereinfachten Form

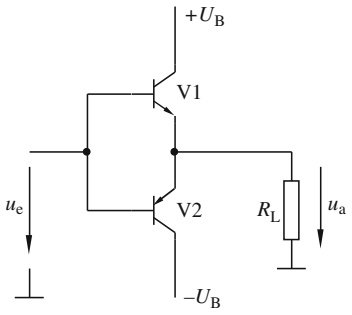


Abb. 38.4 Serien-Gegentakt-Verstärker

Ruhestrom I_{BA}, ein Kollektor-Ruhestrom I_{CA} und somit eine Kollektor-Emitter-Ruhespannung $U_{CEA} \approx U_B$. Damit lässt sich die Verlustleistung im Transistor berechnen.

Transistorverlustleistung

$$P_T = U_{CEA} \cdot I_{CA} \approx \frac{U_B^2}{R_E} \tag{38.1}$$

Bei $u_e = 0\,\text{V}$ ist die auftretende Verlustleistung gleichzeitig die maximale Transistorverlustleistung. Die bei $u_e \neq 0\,\text{V}$ auftretende Verlustleistung ist stets kleiner. Die von den Betriebsspannungsquellen gelieferte Gesamtleistung lässt sich ebenfalls näherungsweise berechnen.

$$\textit{Gesamtleistung} \quad P_{ges} \approx \frac{2U_B^2}{R_E} \tag{38.2}$$

Diese Leistung ist völlig unabhängig von der Aussteuerung des Transistors. Die an den Lastwiderstand abgegebene Ausgangsleistung hängt vom Widerstandsverhältnis R_E/R_L ab.

$$\textit{Ausgangsleistung} \quad P_L = \frac{U_B^2}{2} \cdot \frac{R_L}{2(R_L + R_E)^2} \tag{38.3}$$

Die Ausgangsleistung ist bei Leistungsanpassung ($R_E = R_L$) maximal.

$$\textit{maximale Ausgangsleistung} \quad P_{L\,max} = \frac{U_B^2}{8R_L} \tag{38.4}$$

Die Transistorverlustleistung P_T ist also 8mal so groß wie die maximale Ausgangsleistung. Die von den Betriebsspannungsquellen aufzubringende Leistung ist sogar 16mal so groß wie die maximale Ausgangsleistung.

Damit kann der Wirkungsgrad η dieser Schaltung berechnet werden.

$$\textit{Wirkungsgrad} \quad \eta = \frac{P_{L\,max}}{P_{ges}} \approx 6{,}25\,\% \tag{38.5}$$

Abb. 38.3 zeigt die Schaltung nach Abb. 38.2 in der üblichen vereinfachten Form. Durch die Krümmung der Kennlinien bei größeren Aussteuerungen entstehen Verzerrungen.

Die Nachteile des A-Betriebes sind: kleiner linearer Aussteuerbereich, geringe Ausgangsleistung, kleiner Wirkungsgrad.

Endstufen in A-Betrieb finden in der Praxis Verwendung als Gegentakt-A-Endstufen bei sehr hochwertigen Endstufen.

Im Ausgangskennlinienfeld nach Abb. 38.1 ist erkennbar, dass die Amplitude des Ausgangssignals bei B-Betrieb etwa die Größe der Betriebsspannung $U_{CE\,max} = U_B$ erreichen kann. Bei einer entsprechenden Ansteuerung kann jedoch nur eine Halbwelle des Signals verstärkt werden.

Dazu bedarf es eines neuen Schaltungskonzeptes. Man schaltet nach Abb. 38.4 zwei Transistoren gleichstrommäßig in Reihe (Serien-Gegentakt-Verstärker), wobei der eine ein npn- und der andere ein pnp-Transistor mit sonst identischen Daten (Komplementärtransistoren) ist. In dieser Schaltung wird während der positiven Halbwelle des Eingangssignals nur der npn-Transistor (V1) der Gegentakt-Endstufe leitend. Es fließt ein Kollektorstrom durch den Lastwiderstand mit der eingezeichneten Spannungsrichtung. Während der negativen Halbwelle der Eingangswechselspannung wird der pnp-Transistor (V2) leitend, und es fließt ein Strom in entgegengesetzter Richtung durch den Lastwiderstand. Die beiden einzeln verstärkten Halbwellen werden im Lastwiderstand wieder zu einem Gesamtsignal zusammengefügt.

Es fließt jedoch erst dann ein Kollektorstrom, wenn die Amplitude des Ansteuersignals größer ist als der Schwellwert der Basis-Emitterdiode der Transistoren. Das Ausgangssignal ist daher bei einem Gegentakt-B-Verstärker im Bereich des Nulldurchganges, kleine Eingangsspannungen vorausgesetzt, zwangsläufig verzerrt. Bei größeren Eingangswechselspannungen sind diese Übernahmeverzerrungen jedoch von untergeordneter Bedeutung.

Beim AB-Betrieb wird der Arbeitspunkt nach Abb. 38.1 etwas in Richtung des A-Arbeitspunktes verschoben. Um diesen Arbeitspunkt einzustellen, muss an die Basis-Emitterdioden eine *Basis-Vorspannung* U_{BE} gelegt werden, so dass beide Transistoren schon ohne Ansteuerung geringfügig leitend sind. Die Schaltung nach Abb. 38.5 zeigt die Prinzipschaltung eines Gegentaktverstärkers in AB-Betrieb mit den beiden Gleichspannungsquellen U_{BE}, die für die Basisvorspannung erforderlich sind.

Da der *Ruhestrom* nicht gleich Null ist, verschlechtert sich der Wirkungsgrad der Schaltung geringfügig. Liegt bei dieser Variante eine Ansteuerung mit einer kleinen Signalspannung vor,

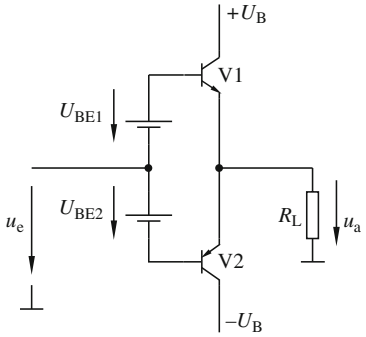


Abb. 38.5 Prinzipschaltung eines Gegentaktverstärkers in AB-Betrieb

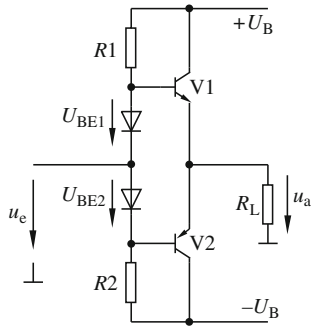


Abb. 38.6 Spannungsteiler mit integrierten Dioden

so ist die Verzerrung deutlich kleiner. Der Verstärkungsvorgang erfolgt in der gleichen Weise wie beim B-Betrieb.

Die maximale Ausgangswechselleistung, die mit einem Gegentakt-B-Verstärker erreicht werden kann, lässt sich mit folgender Überlegung bestimmen. Zwei Emitterfolger im Gegentakt-B-Betrieb haben eine doppelt so hohe Aussteuerbarkeit ($\pm U_B$ statt $\pm U_B/2$) und damit eine 4mal so große Ausgangsleistung wie ein Emitterfolger im A-Betrieb.

Die Ausgangsleistung ist bei Leistungsanpassung ($R_E = R_L$) maximal.

$$\textit{max. Ausgangsleistung} \qquad P_{L\max} = \frac{U_B^2}{2R_L} \tag{38.6}$$

$$\textit{Transistorverlustleistung} \qquad P_T = 0{,}1 \cdot \frac{U_B^2}{R_L} \tag{38.7}$$

Bei gleicher Ausgangsleistung ist die Verlustleistung eines Transistors bei Gegentakt-B-Betrieb ungefähr 40mal kleiner als bei A-Betrieb.

Die von den Betriebsspannungsquellen gelieferte Gesamtleistung lässt sich bei sinusförmiger Vollaussteuerung näherungsweise berechnen.

$$\textit{Gesamtleistung} \qquad P_{\text{ges}} = \frac{2U_B^2}{\pi \cdot R_L} \tag{38.8}$$

Damit kann der Wirkungsgrad h dieser Schaltung berechnet werden.

$$\textit{Wirkungsgrad} \qquad \eta = \frac{P_{L\max}}{P_{\text{ges}}} = 78{,}5\,\% \tag{38.9}$$

Vorteile von Leistungsverstärkern im AB- und B-Betrieb sind: kleiner Ruhestrom und kleine Verlustleistung, großer Wirkungsgrad ($\approx$ 70 %), großer Aussteuerbereich ($\approx \pm U_B$), große Ausgangswechselleistung.

Nachteile sind: Transistoren-Pärchen (Komplementär-Transistoren) erforderlich, gegenüber A-Betrieb größere Verzerrungen bei kleiner Aussteuerung. Will man ohne Komplementär-Transistoren arbeiten, wird am Eingang für einen Transistor eine Phasenumkehrstufe benötigt.

Die Vorspannung nach Abb. 38.5 lässt sich relativ einfach über einen Spannungsteiler aus den Betriebsspannungen U_B gewinnen. Die Schaltung nach Abb. 38.6 zeigt einen Spannungsteiler mit integrierten Dioden.

Da die Durchlasskennlinie einer Diode praktisch dieselbe Temperaturabhängigkeit wie die Eingangskennlinie eines Transistors aufweist, kann man mit dieser Variante auch die Temperaturdrift der Transistoren auffangen. Wenn die beiden Dioden sich in der gleichen Weise erwärmen wie die Transistoren (gemeinsame Montage auf einen Kühlkörper), nehmen die Spannungen U_{BE1} und U_{BE2} im selben Maße ab wie die Basis-Emitter-Spannungen der Transistoren. Die Folge ist, dass der Ruhestrom nahezu konstant bleibt.

Das eigentliche Problem beim Gegentakt-AB-Verstärker besteht in der Stabilisierung des Ruhestroms. Der Ruhestrom kann sich wegen der kleinen Lastwiderstände durch thermische Rückkopplung leicht aufschaukeln und zur Zerstörung des Transistors führen.

Die beiden Emitterwiderstände R_{E1} und R_{E2} in der Schaltung nach Abb. 38.7 bewirken eine Stromgegenkopplung, also eine Arbeitspunktstabilisierung. Die Widerstände sind durch die Dioden V5 und V6 gebrückt, die schon bei geringen Eingangssignalen leitend werden. Wäre dieses nicht der Fall, so würde die Wechselstromverlustleistung in den Emitterwiderständen unvertretbar groß werden. Der Kollektorruhestrom sollte an den Emitterwiderständen einen Spannungsabfall von ungefähr 0,2 bis 0,3 V hervorrufen (wenn Si-Dioden verwendet werden); die

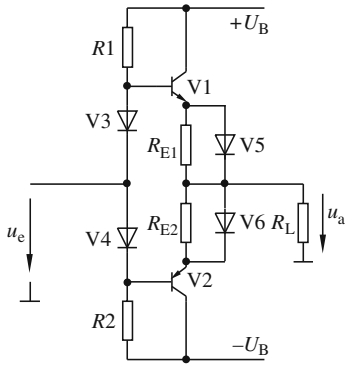


Abb. 38.7 Gegentakt-AB-Verstärker mit Emitterwiderständen zur Arbeitspunktstabilisierung

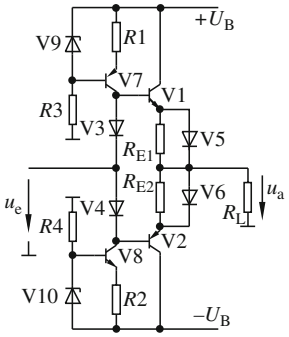


Abb. 38.8 Gegentakt-AB-Verstärker mit hohem Eingangswiderstand

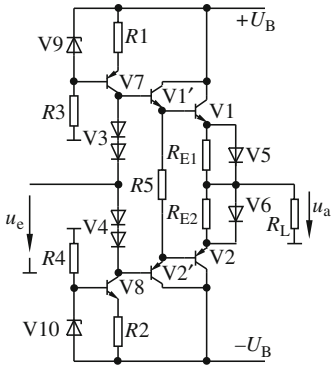


Abb. 38.9 Gegentakt-AB-Verstärker mit Darlington-Transistoren

Spannungsabfälle sollen bei $u_e = 0\,\mathrm{V}$ kleiner sein als die Schleusenspannungen der Dioden.

Wird die Eingangsspannung so groß, dass aufgrund des erhöhten Emitterstroms der Spannungsabfall an R_E den Wert von 0,7 V übersteigt, wird die entsprechende Diode leitend. Infolge des sehr niedrigen differentiellen Widerstandes der Diode fließt praktisch die *gesamte Wechselleistung* über die Diode, so dass an R_E kein wesentlicher Leistungsverlust entsteht und die Spannungsabfälle auf die Durchlassspannungen der Dioden begrenzt werden.

Diese Schaltung hat allerdings einen erheblichen Nachteil, da nur ein Teil des Signaleingangswechselstroms in die Basis der Transistoren gelangt, während der andere Teil über die Widerstände R1 und R2 ungenutzt abfließt. Die Folge ist eine relativ geringe Gesamtstromverstärkung der Stufe und damit eine geringe Leistungsverstärkung.

Ersetzt man die Widerstände R1 und R2 in der Schaltung nach Abb. 38.7 durch *Konstantstromquellen*, so erhöht man deutlich den Eingangswiderstand der Schaltung. Der differentielle Innenwiderstand der Quellen ist sehr hoch.

Die Schaltung nach Abb. 38.8 hat nun den hohen Eingangswiderstand eines Emitterfolgers ohne Eingangsspannungsteiler. Praktisch haben diese Konstantstromquellen einen unendlich hohen Innenwiderstand. Somit muss der Vorverstärker (Signalquelle) nur noch den für die Aussteuerung der Transisitoren erforderlichen Basisstrom liefern. Die Z-Dioden werden häufig durch die Reihenschaltung aus zwei bis drei in Durchlassrichtung betriebene Siliziumdioden ersetzt.

Die mit den bisher besprochenen Schaltungen erzielbaren Ausgangsströme werden von den Transistoren V1 und V2 oder deren Stromverstärkungen bestimmt.

Um eine höhere Stromverstärkung zu erreichen, verwendet man eine Darlington-Schaltung oder als Endtransistoren Darlington-Transistoren.

Die Transistoren V1′-V1 beziehungsweise V2′-V2 benötigen in der Schaltung nach Abb. 38.9 eine Basis-Emitter-Spannung, die doppelt so groß ist wie die in der Schaltung nach Abb. 38.8. Deshalb benutzt man hier jeweils zwei in Reihe geschaltete Dioden.

Durch den Widerstand R5 werden die Ruheströme von V1′ und V2′ stabilisiert. Auf diese Weise kann erreicht werden, dass die Basis-Ruheströme von V1 und V2 gleich sind, da ansonsten die Übernahmeverzerrungen nicht vernachlässigbar klein wären.

Manchmal ist es schwierig, für Verstärker mit höheren Ausgangsleistungen geeignete pnp/npn-Transistorpaare zu finden. Daraus folgt die Absicht, für bestimmte Anwendungen Transistoren vom gleichen Leitungstyp für V1 und V2 zu verwenden.

Dann wird eine Darlington-Schaltung mit komplementären Transistoren, wie die Schaltung nach Abb. 38.10 zeigt, ausgeführt. Die Transistoren V1 und V2 sind vom gleichen Leitungstyp (npn). Der Darlington-Transistor V1′-V1 wirkt wie der Transistor V1 nach Abb. 38.8 als npn-Transistor und der Darlington-Transistor V2′-V2 ebenfalls als npn-Transistor in der gleichen Funktion wie der Transistor V2 nach Abb. 38.8.

Die Schaltung nach Abb. 38.10 wird als *Quasi-Komplementär-Endstufe* bezeichnet. Der Darlington-Transistor V2′-V2 benötigt zur Vorspannungserzeugung lediglich die Diode V4. Die Widerstände R5 dienen der Ruhestromeinstellung der Transistoren V1′ und V2′.

Diese Gegentakt-Endstufen bilden die Grundlage für zahlreiche weitere Schaltungen. Die vorgestellten Endstufen arbeiten mit Emitterfolgern, die eine Spannungsverstärkung $V_u \approx 1$ aufweisen. Ziel ist, die Spannungsverstärkung zu erhöhen.

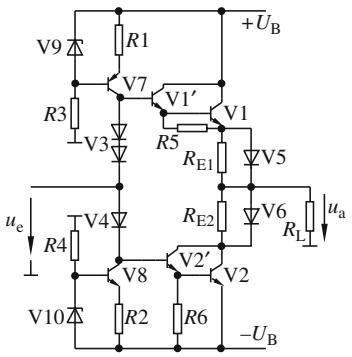


Abb. 38.10 Gegentakt-AB-Verstärker mit Endtransistoren vom gleichen Leitungstyp

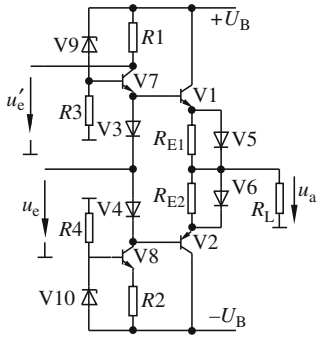


Abb. 38.11 Gegentakt-AB-Verstärker mit erhöhter Spannungsverstärkung

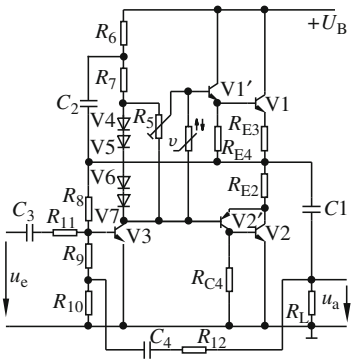


Abb. 38.13 Quasikomplementäre Endstufe mit unsymmetrischer Spannungsversorgung

Die Endtransistoren in der Schaltung nach Abb. 38.11 sind Darlington-Schaltungen entsprechend den Schaltungen nach Abb. 38.9 und 38.10. Der Übersichtlichkeit wegen sind sie als einfache Transistoren dargestellt. Legt man die Signalspannung u_e zwischen die Dioden V3 und V4, so ist die Schaltung identisch mit Abb. 38.8. Legt man jedoch die Signalspannung u'_e an den Emitter des Transistors V7, arbeitet dieser bezüglich der Signalspannung als Transistor in Basisschaltung und betrachtet die Stromquelle mit Transistor V8 als Arbeitswiderstand. Da eine Basisschaltung ein $V_u > 1$ hat, tritt am Kollektor von V7 nun eine Wechselspannung u_e auf, die größer ist als die Signalspannung u'_e.

Durch den Wechsel des Signaleinganges ist also eine Spannungsverstärkung erreicht worden. Zu beachten ist jedoch, dass bei Ansteuerung zwischen V3 und V4 das Eingangs-Ruhepotential Null ist. Bei Ansteuerung der Schaltung am Emitter des Transistors V7 liegt jedoch das Eingangs-Ruhepotential in der Nähe der Betriebsspannung $+U_B$ (genauer: auf $U_B - U_R$).

Ein Vorverstärker kann also nicht wahlweise angeschlossen werden, sondern sein Ausgangs-Ruhepotential muss angepasst werden. Die Schaltung nach Abb. 38.12 zeigt die Möglichkeit einer Ruhepotentialanpassung.

Der Vorverstärker liefert die Signalspannung u''_e an die Basisverstärkerstufe mit V11, wobei die Widerstände R1 und R6 eine Strom-Spannungs-Gegenkopplung darstellen. Die Wechselspannung u'_e ist entsprechend dem Widerstandsverhältnis R1/R2 größer als die vom Vorverstärkers gelieferte Signalspannung u''_e. Am Ausgang des Vorverstärkers ist das Eingangs-Ruhepotential wieder Null.

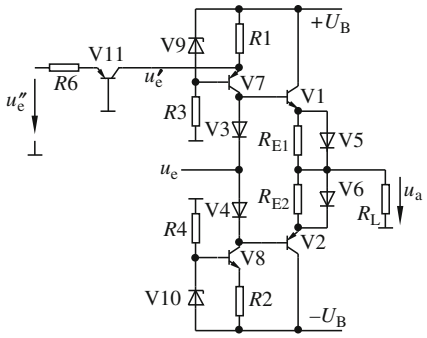


Abb. 38.12 Möglichkeit einer Ruhepotentialanpassung mit OP am Eingang

Gegentakt-Verstärker können auch mit unsymmetrischer Spannungsversorgung betrieben werden. Allerdings ist dann ein Koppelkondensator einzufügen. Um einen gleichstrommäßigen Kurzschluss zu vermeiden, muss der Lastwiderstand mittels des Kondensators mit der Endstufe verbunden sein.

Die Schaltung nach Abb. 38.13 zeigt eine quasikomplementäre Endstufe mit unsymmetrischer Spannungsversorgung. Die Transistoren V1 und V2 sind npn-Transistoren und werden durch das komplementäre Transistorenpaar V1′ und V2′ angesteuert. Der Transistor V3 arbeitet in Emitterschaltung und bewirkt eine Spannungsverstärkung. Mit dem Trimmer R5 wird der Ruhestrom der Transistoren V1 und V2 eingestellt.

Um eine zu starke wechselstrommäßige Belastung von V3 zu vermeiden und um die Spannungsverstärkung von V3 möglichst groß zu machen, ist ein Wechselspannungsbootstrap (C2) vorhanden. Der Kondensator C4 und der Widerstand R12 sind für eine Wechselspannungsgegenkopplung zur Einstellung einer definierten Spannungsverstärkung erforderlich.

Der Lastwiderstand R_L kann bei unsymmetrischer Spannungsversorgung einseitig am Plus- oder Minuspol der Versorgungsspannung liegen. Ist der Lastwiderstand R_L zum Beispiel am Minuspol angeschlossen, wird der Kondensator bei leitendem npn-Transistor etwa auf die halbe Betriebsspannung U_B aufgeladen. Während der Sperrphase des npn-Transistor wirkt der Kondensator als Spannungsquelle für den pnp-Transistor und treibt einen Strom durch ihn. Dadurch wird der Kondensator wieder etwas entladen.

Die Änderung der Kondensatorspannung muss durch die Wahl eines genügend großen Kondensators klein gehalten werden. Der Kondensator wird im Takt des Signals auf- und entladen. Der Kondensator erspart eine Spannungsquelle mit Mittelanzapfung, weil sich an ihm eine Spannung $U_B/2$ aufbaut, die als Speisespannung für den Transistor V2 wirkt. Von Nachteil ist, dass der Kondensator und der Lastwiderstand einen Hochpass bilden. Soll dessen untere Grenzfrequenz sehr niedrig sein, muss die Kapazität des Kondensators sehr groß sein.

Gegentakt-AB-Verstärker größerer Leistung und höherer Ansprüche bezüglich möglichst kleiner Verzerrungen sind bereits sehr aufwendig. Die Schaltung nach Abb. 38.14 realisiert einen hohen Eingangswiderstand durch einen Operationsverstärker in Elektrometerschaltung.

Die Spannungsverstärkung der Gesamtschaltung wird durch das Verhältnis der Widerstände R6 und R7 eingestellt. Da das Ausgangspotential des OP gleich Null ist, dient V7 zur Potentialverschiebung.

Aufgrund der Aufwendigkeit solcher Schaltungen haben sich inzwischen Gegentaktverstärker in integrierter Schaltungstechnik durchgesetzt.

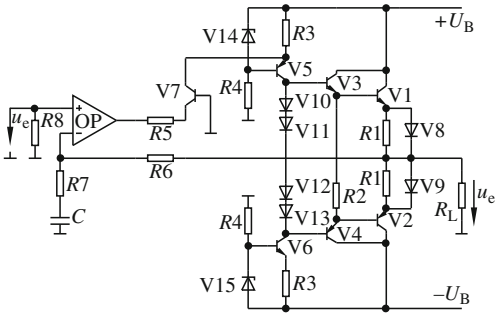
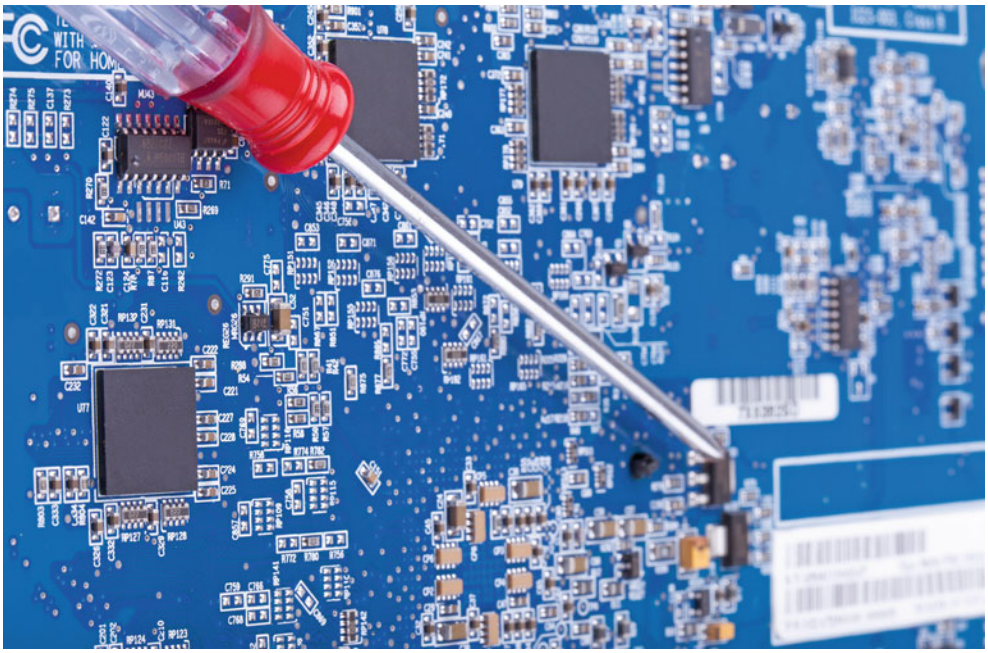


Abb. 38.14 Gegentakt-AB-Verstärker großer Leistung mit sehr hohem Eingangswiderstand

Operationsverstärker

39

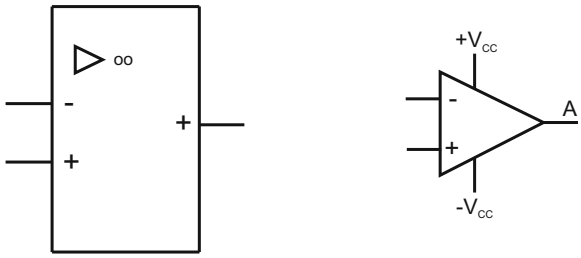
Peter Döring



39.1 Einführung 486
39.2 Differenzverstärker 486
39.3 Grundlagen des OP 487
39.4 Operationsverstärker als Verstärker 489

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_39



Schaltzeichen nach DIN EN 60617 — Übliches Schaltzeichen, nicht genormt; hier verwendet.

Abb. 39.1 Schaltzeichen eines Operationsverstärkers

39.1 Einführung

Eine wichtige *lineare, integrierte Schaltung* ist der *Operationsverstärker*, der in Kurzform mit OpAmp (operations-amplifier) oder OP bezeichnet wird. Diese Art von Verstärkern wurde fast ausschließlich als Rechenverstärker, also für Rechen-Operationen in Analog-Rechnern eingesetzt. Obwohl er aus einer Vielzahl von Transistoren, Dioden und Widerständen aufgebaut ist, kann er vom Anwender als ein einziges, kompaktes Verstärker-Bauelement betrachtet werden.

Der besondere Vorteil der OPs liegt darin, dass sich seine Eigenschaften durch einfache äußere Beschaltungen stark variieren lassen. OPs sind daher in immer größer werdendem Umfang in analogen Schaltungen, aber auch in Schaltungen der Digitaltechnik anwendbar und integrierbar.

Da dieser Abschnitt einführenden und erklärenden Charakter hat, wird hier auf die normgerechte Darstellung der Schaltzeichen an einigen Stellen verzichtet. Ebenso wird die Betriebsmittelkennzeichnung nach DIN nicht beachtet, da manche Bauteile trotz identischer Schaltzeichen unterschiedliche Benennungen hätten und dieser Umstand zu Missverständnissen führen kann (Abb. 39.1).

Beispiel für einen Widerstand:

Normaler Strombegrenzer	R
Heizwiderstand	E
Messwiderstand	B

Für Spulen (Induktivität) und Transformatoren werden die üblichen Schaltzeichen benutzt.

Für alle Halbleiterbauteile wird die übliche Kennung „V“ verwendet.

39.2 Differenzverstärker

Operationsverstärker (OP) haben einen Differenzverstärker mit Transistoren als Eingangsschaltung. Differenzverstärker sind dadurch gekennzeichnet, dass bei gleicher Ansteuerung der beiden Eingänge die Ausgangsspannung U_A (nahezu) Null ist; die Gleichtaktverstärkung ist also sehr klein. Das Gleichtaktsignal

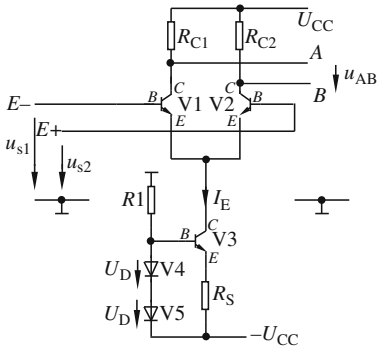


Abb. 39.2 Vollständiger Differenzverstärker

bewirkt jedoch in beiden Transistoren eine Verschiebung des Gleichstrom-Arbeitspunktes.

Die Kollektorströme ändern sich. In der Realität sind die Transistoren nicht völlig identisch, so dass das Gleichtaktsignal eine geringe Änderung des Ausgangssignals zur Folge hat. Differenzverstärker sind sehr unempfindlich gegenüber Temperatureinflüssen. Abb. 39.2 zeigt einen vollständigen Differenzverstärker. Bezogen auf den Ausgang A heißt der Eingang E^- invertierender Eingang und der Eingang E^+ nichtinvertierender Eingang.

Betrachtet man die Steuerspannung an beiden Eingängen der Schaltung, können unabhängig vom Aufbau und der Form dieser Signalspannungen bezogen auf die Wirkung im Differenzverstärker zwei Aussagen getroffen werden:

Jede Eingangsspannung an den Eingängen eines Differenzverstärkers lässt sich immer in zwei Komponenten zerlegen, nämlich in eine *Gleichtaktspannung* U_{Gl}, die sich als arithmetisches Mittel aus den beiden Eingangsspannungen ergibt und in eine *Differenzspannung* U_D.

$$\textit{Gleichtaktspannung} \qquad U_{GL} = \frac{u_{s1} + u_{s2}}{2} \tag{39.1}$$

$$\textit{Differenzspannung} \qquad U_D = u_{s2} - u_{s1} \tag{39.2}$$

Die Eingangsspannungsdifferenzen werden mit der hohen Differenzverstärkung v_d, gleiche Eingangsspannungen dagegen mit der wesentlich kleineren Gleichtaktverstärkung v_{Gl} verstärkt.

Das Verhältnis der beiden zueinander wird als Gleichtaktunterdrückung G bezeichnet und beschreibt die Güte eines Differenzverstärkers.

Gleichtaktunterdrückung

$$G = \frac{V_d}{V_{Gl}} = \frac{r_{BE} + 2R_E \cdot (\beta + 1)}{2r_{BE}} \approx \frac{\beta \cdot R_E}{r_{BE}} \qquad (39.3)$$

Meist wird die Gleichtaktunterdrückung G in dB angegeben.

Gleichtaktunterdrückung G in dB

$$C_{MRR}(\text{dB}) = 20 \cdot \log G \qquad \begin{array}{c|c|c|c} C_{MMR} & r_{BE} & R_E & \beta \\ \hline \text{dB} & \Omega & \Omega & 1 \end{array} \qquad (39.4)$$

Ersetzt man die Transistoren im Differenzverstärker durch Darlington-Transistoren, lässt sich der Differenz-Eingangswiderstand mindestens verdoppeln. Soll er noch größer werden, verwendet man als Eingangstransistoren FETs.

Die Endstufen in OPs bestehen im Wesentlichen aus Gegentakt-Endstufen oder Eintakt-Endstufen mit „open collector“.

39.3 Grundlagen des OP

Bei OPs mit Differenzeingang reagiert der Verstärker nur auf die Differenz der beiden Eingangssignale. Die *Differenzeingangsspannung* U_D ist die Spannung, die als resultierende Spannung am Eingang des Verstärkers liegt.

Differenzeingangsspannung

$$U_D = u_{s2} - u_{s1} \qquad (39.5)$$

Diese Spannung U_D wird nun mit der *Leerlaufverstärkung* V_0 (open loop gain) verstärkt an den Ausgang gebracht und ist als Ausgangsspannung U_A messbar.

$$\textit{Ausgangsspannung} \quad U_A = V_0 \cdot U_D \qquad (39.6)$$

Die Gleichungen geben den mathematischen Zusammenhang der elektrischen Größen unter der Voraussetzung wieder, dass in die Eingänge des OPs kein Eingangsstrom fließt.

Das *Ersatzschaltbild* eines OPs nach Abb. 39.3 „beschreibt“ elektrisch einen „realen OP“ mit den entsprechenden Ein- und Ausgangsgrößen.

Sollen die Eingangsströme Null sein, muss der Eingangswiderstand r_e des OPs unendlich groß sein. Die Ausgangsspannung U_A muss in der Folge ebenfalls unendlich groß sein können. Da diese Spannung an einem Lastwiderstand einen entsprechenden Strom zieht, muss der Ausgangsstrom bei konstanter Ausgangsspannung ebenfalls konstant sein. Das kann aber nur der Fall sein, wenn der Ausgangswiderstand r_a des OPs Null ist.

Deshalb muss die Leerlaufverstärkung V_0 ebenfalls unendlich groß sein. Nach Abb. 39.4 sind Ausgangsspannung U_A und Eingangsdifferenzspannung U_D nur proportional bis zu dem Wert

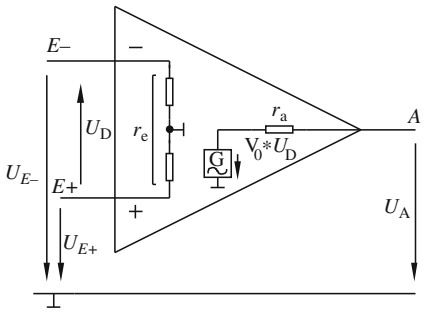


Abb. 39.3 Ersatzschaltbild eines Operationsverstärkers

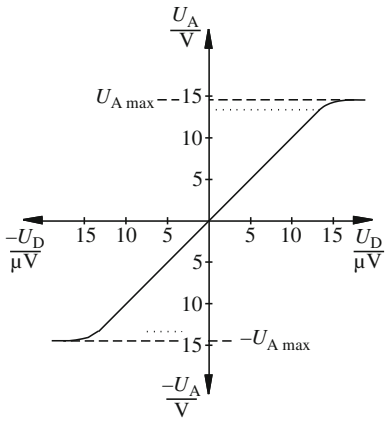


Abb. 39.4 Übertragungskennlinie eines OP

$\pm U_{A\,max}$, bei dem der OP in die Sättigung geht. Wegen der hohen Leerlaufverstärkung wird ein unbeschalteter OP bereits bei kleinen Eingangsspannungen in die Sättigung gesteuert.

So würde zum Beispiel bei $V_0 = 30.000$ eine Störspannung von nur 0,1 mV bereits eine Änderung der Ausgangsspannung von 3 V bewirken.

Die Daten eines „realen OPs“ können den Datenblättern der Hersteller entnommen werden. Die Daten eines „idealen“ und „realen“ OPs sind einander in der Tendenz gegenübergestellt (Tab. 39.1 und Tab. 39.2).

Die meisten Operationsverstärker werden in einer *Rückkopplungsschaltung* betrieben, das heißt, ein Teil der Ausgangsgröße (in der Regel ein Teil der Ausgangsspannung) wird auf den Eingang zurückgeführt. Bei der Rückkopplung unterscheidet man zwischen einer Mitkopplung und einer Gegenkopplung.

Bei der *Mitkopplung* wird die Ausgangsgröße gleichphasig auf den Eingang zurückgeführt, das heißt, das Eingangssignal wird in seiner Wirkung verstärkt. Für den Fall der reinen Mitkopplung der Schaltung nimmt die Amplitude der Ausgangsspannung so lange zu, bis die Sättigung des Verstärkers erreicht ist.

Bei der *Gegenkopplung* wird ein Teil der Ausgangsgröße gegenphasig auf den Eingang zurückgeführt und somit das Eingangssignal in seiner Wirkung geschwächt. Die Gesamtverstärkung nimmt ab.

Tab. 39.1 Operationsverstärker (Auswahl). Kenndaten bei $U_B = \pm 15\,\text{V}$ und $T_U = 25\,°\text{C}$

Typ		Bipolare Eingänge			FET-Eingänge	
		TBA 221 µA 741	TAA 521 709	TAA 761	LF 355	LF 357
Leerlaufverstärkung (open loop gain)	V_0	100 dB	93 dB	85 dB	80 dB	80 dB
Gleichtaktunterdrückung (CMRR)	G	90 dB	90 dB	79 dB	100 dB	100 dB
Offsetspannung (offset voltage)	U_o	±6 mV	±2 mV	±6 mV	± 3 mV	±3 mV
Eingangsruhestrom (bias current)	I_E	80 nA	300 nA	500 nA	30 pA	30 pA
Eingangswiderstand	r_e	2 MΩ	250 kΩ	200 kΩ	10^{12} Ω	10^{12} Ω
Ausgangswiderstand	r_a	75 Ω	150 Ω	o. K.	50 Ω	–
Ausgangsstrom (max.)	I_A	20 mA	10 mA	70 mA	25 mA	–
Ausgangsspannung (max.)	$U_{A\,\max}$	±13 V	±13 V	±13 V	±13 V	±13 V
Transitfrequenz	f_T	500 kHz	5 MHz	200 kHz	2,5 MHz	25 MHz

Tab. 39.2 Gegenüberstellung von idealem und realem OP

Charakteristische Eigenschaften	Idealer OP	Realer OP
Leerlaufverstärkung V_0	∞	10^3 bis 10^7
Eingangswiderstand r_e	∞	10^6 bis 10^{14} Ω
Ausgangswiderstand r_a	0 Ω	30 bis 100 Ω
Temperaturdrift	Nicht vorhanden	Von −50 °C bis +75 °C vernachlässigbar
Übertragungsbandbreite B	∞	10^4 bis 10^7 Hz
Aussteuerbereich $U_A = f(U_E)$	$-\infty$ bis $+\infty$	$-U_B$ bis $+U_B$

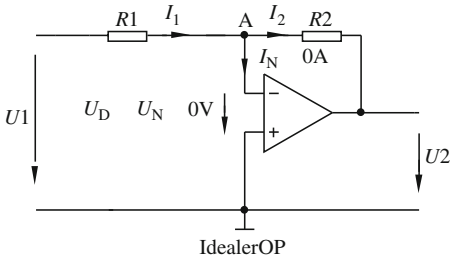


Abb. 39.5 Grundschaltung des idealen invertierenden OP

Die Grundschaltung des Inverters (invertierender OP) wird dargestellt in Abb. 39.5.

Wegen $V_0 = \infty$ gilt $\Rightarrow U_D = 0\,\text{V}$. Deshalb liegt der Punkt A auf Masse. Punkt A ist „*virtueller Nullpunkt*“, und es gilt:

$$\textit{Spannungsverstärkung} \quad V_u = -\frac{U_2}{U_1} = -\frac{R_2}{R_1} \tag{39.7}$$

Diese Gleichung gilt für den „idealen“ OP.

Die Verstärkung wird also nur durch das Widerstandsverhältnis und damit unabhängig von den spezifischen Kenngrößen des OP festgelegt.

Der Fehler bei Annahme eines „idealen“ OPs gegenüber korrekter Rechnung beträgt nur ungefähr 1 %. Unter Beachtung der Widerstandstoleranzen kann in der Praxis mit der Gleichung für den „idealen“ OP bei $V_0 \geq 60\,\text{dB}$ gearbeitet werden.

Bezogen auf einen an den Ausgang geschalteten Verbraucher wird der endliche Ausgangswiderstand r'_a des realen OP um den Faktor der *Schleifenverstärkung* g niederohmiger.

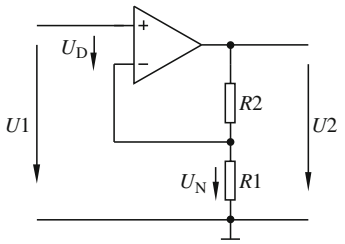


Abb. 39.6 Grundschaltung des idealen nichtinvertierenden OP

Ausgangswiderstand des beschalteten OP

$$r'_a = \frac{r_a}{V_0} + r_a \cdot \frac{V_u}{V_0} \tag{39.8}$$

Der Eingangswiderstand eines „realen“ OPs ist nicht, wie beim „idealen“ OP angenommen, unendlich groß. Darum gilt:

Eingangswiderstand des beschalteten OP

$$r'_e = R_1 \tag{39.9}$$

Beschaltet man einen OP gemäß Schaltung nach Abb. 39.6, so entsteht ein *nichtinvertierender Verstärker*.

Die Verstärkung des nichtinvertierenden OPs hängt nur von der äußeren Beschaltung ab und kann in weiten Grenzen unabhängig von der Leerlaufverstärkung frei festgelegt beziehungsweise eingestellt werden.

Gleichung (39.10) lässt erkennen, dass die Verstärkung V_u nicht kleiner als 1 werden kann, was beim invertierenden OP möglich ist.

$$\textit{Spannungsverstärkung} \quad V_u = 1 + \frac{R_2}{R_1} \tag{39.10}$$

Bei niedrigen Frequenzen ($f < 100\,\text{Hz}$) tritt als Eingangswiderstand r'_e der Schaltung der sehr hohe Gleichtaktwiderstand r_{Gl} des OPs auf.

Eingangswiderstand des beschalteten OP

$$r'_e \approx r_{\text{Gl}} \tag{39.11}$$

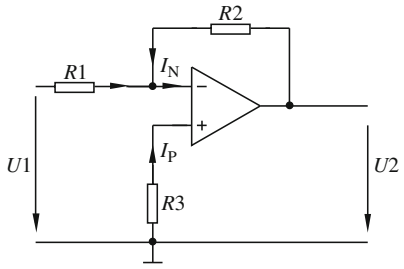


Abb. 39.7 Kompensation der Eingangsruheströme

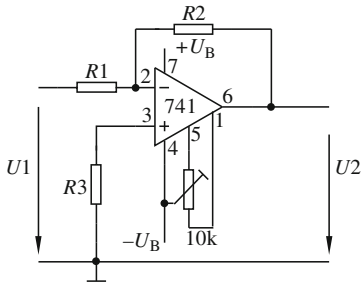


Abb. 39.8 Kompensation der Offset-Spannung

Ausgangswiderstand des beschalteten OP

$$r'_a = r_a \cdot \frac{V_u}{V_0} \tag{39.12}$$

Die beiden Eingänge E^+ und E^- sind eigentlich die Basisanschlüsse der Eingangstransistoren der Differenzverstärkerstufe. Es fließt daher ein Basisstrom, das heißt, in den OP fließen die *Eingangsruheströme* I_N und I_P. Obwohl diese Ströme sehr klein sind, können sie infolge der unterschiedlichen Spannungsabfälle an den äußeren Widerständen eine Spannungsdifferenz an den Eingängen des OPs bewirken. Diese Spannungsdifferenz wird verstärkt und würde ohne geeignete Maßnahmen als Störgröße am Ausgang des OPs auftreten.

Die Kompensation der Eingangsruheströme lässt sich durch den Widerstand R3 in der Schaltung nach Abb. 39.7 erreichen.

Dabei gilt:

$$R_3 = R_1 \parallel R_2 \tag{39.13}$$

Die identischen Spannungsabfälle an den beiden Eingängen verhindern eine Differenzspannung an den Eingängen.

Soll ein nichtinvertierender OP zur Anwendung kommen, umgeht man diese Schwierigkeit generell durch Verwendung eines OPs mit FET-Eingang. Bei diesen Typen können die Eingangsruheströme ganz vernachlässigt werden.

Beim realen OP sind die Transistoren der Differenzverstärkerstufe am Eingang natürlich nicht absolut identisch. Die interne Differenzspannung zwischen U_{BE1} und U_{BE2} wird als *Offset-Spannung* U_O bezeichnet. Diese Spannung ist die Ursache dafür, dass die Ausgangsspannung U_A am Ausgang ungleich Null ist, auch wenn die Spannung U_D am Eingang durch Kurzschließen und auf Masse legen der beiden Eingänge zu Null gemacht wurde.

Je nach OP-Typ gibt es verschiedene, vom Hersteller empfohlene Schaltungsmöglichkeiten zur *Kompensation* der Offset-Spannung U_O.

Besonders einfach ist die äußere Beschaltung für die Offset-Spannungskompensation beim Typ 741. Nach Abb. 39.8 ist lediglich ein 10 kΩ-Widerstand zwischen Pin 1 und Pin 5 anzuschließen und abzugleichen. Dieser Abgleich ist in der Weise vorzunehmen, dass bei an Masse gelegtem Eingang des OP die Ausgangsspannung gleich Null wird.

Die *Kompensation der Eingangsruheströme* erfolgt in Abb. 39.8 durch Widerstand R3 nach (39.13).

Unerwünschte und unvermeidbare interne Schalt- und Transistorkapazitäten ergeben zusammen mit den Widerständen in den OPs *Tiefpässe*. Ähnlich wie bei den Transistorverstärkerstufen führt das zu frequenzabhängigen Verstärkungsfaktoren, das heißt, V_u bleibt nicht konstant bis zu höchsten Frequenzen. Der Frequenzgang ist also nicht linear.

Auch die Phasendrehung zwischen Ein- und Ausgang ändert sich und kann aus einer Gegenkopplung eine Mitkopplung machen, so dass die OP-Schaltung „schwingt“. Der OP geht in den Oszillatorbetrieb über, er arbeitet als „Signalgenerator“. Diese unangenehme Nebenwirkung kann mit Hilfe eines RC-Gliedes, das nach Herstellerangaben zu dimensionieren ist, durch *Frequenzkompensation* vermieden werden. In zahlreichen OP-Typen wird die Frequenzkompensation bereits im Innern des OP vorgenommen.

39.4 Operationsverstärker als Verstärker

In elektronischen Schaltungen werden Verstärker mit unterschiedlichen Verstärkungsfaktoren und mit bestimmtem dynamischen Übertragungsverhalten benötigt.

Außerdem sind zusätzlich Schaltungen erforderlich, die diverse Rechenoperationen, wie zum Beispiel Addieren, Subtrahieren, Multiplizieren, Dividieren, Integrieren, Differenzieren ausführen können. Das gewünschte Verhalten kann man durch eine entsprechende Beschaltung eines OP erreichen.

39.4.1 Verstärker mit frequenzunabhängiger Gegenkopplung

Wenn bei OP-Verstärkern im Gegenkopplungszweig ausschließlich ohmsche Widerstände benutzt werden, ist die Gegenkopplung frequenzunabhängig.

Dadurch ist die Verstärkung V_u im zulässigen Arbeitsbereich des Verstärkers konstant und unabhängig von der Frequenz des Signals.

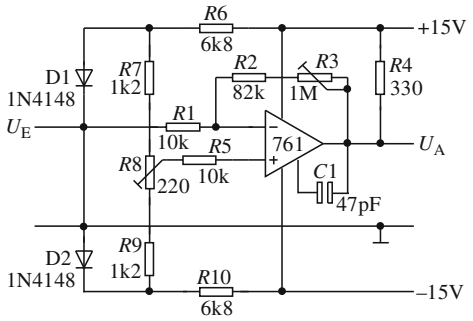


Abb. 39.9 Umkehrverstärker mit kontinuierlich einstellbarer Verstärkung

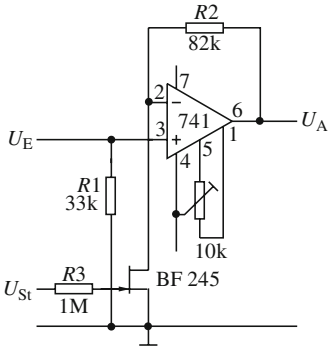


Abb. 39.10 Regelverstärker mit kontinuierlicher Einstellung der Verstärkung

Wird $R_2 < R_1$, so wird auch $V_u < 1$. Die Eingangsspannung U_E wird nicht verstärkt, sondern gedämpft.

Einen *Umkehrverstärker* mit kontinuierlich einstellbarer Verstärkung (Abgleich von Reglerschaltungen) zeigt die Schaltung nach Abb. 39.9. Die Verstärkung wird mit R1, R2 und R3 eingestellt. Offsetspannung und Offsetstrom werden mit R7, R8 und R9 sowie R5 eingestellt. Die Dioden D1 und D2 bilden mit R6 und R10 einen Massepunkt. Hier wird ein OP 761 mit open-collector-Ausgang eingesetzt.

Eine kontinuierliche Einstellung der Verstärkung kann mit Hilfe des Feldeffekttransistors in der Schaltung nach Abb. 39.10 vorgenommen werden. In diesem *Regelverstärker* wird der Vorteil ausgenutzt, dass beim Elektrometerverstärker eine Änderung der Rückkopplungswiderstände keinen Einfluss auf den Eingangswiderstand der Schaltung hat.

Der Eingangswiderstand beträgt hier $R_1 = 33\,\mathrm{k\Omega}$. Der normalerweise hohe Eingangswiderstand eines Elektrometerverstärkers wird in dieser Schaltung also nicht genutzt.

Die Verstärkungseinstellung erfolgt durch Änderung des Drain-Source-Widerstandes R_{DS} des FET, der in bestimmten Grenzen durch eine gesteuerte Gleichspannung variiert werden kann. Die Verstärkung dieses Regelverstärkers ist etwa im Bereich von $V_u = 1$ bis $V_u = 150$ veränderbar.

Die Elektrometerschaltung ist in Abb. 39.11 tragendes Element in dem sehr hochohmigen Spannungsmessgerät. Der Eingang dieser *Messschaltung* wird durch einen Spannungsteiler aus

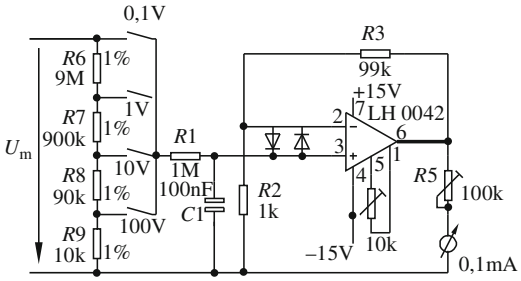


Abb. 39.11 Spannungsmessgerät mit OP in Elektrometerschaltung

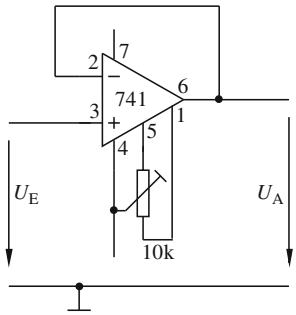


Abb. 39.12 Impedanzwandler mit OP

Messwiderständen gebildet. Er ist so dimensioniert, dass sich die Messbereiche 100 mV, 1 V, 10 V und 100 V ergeben.

Der Widerstand R1 dient zur Strombegrenzung in dem OP. Zusammen mit C1 bildet er einen Tief- pass, dessen Aufgabe es ist, Oberschwingungen und Spannungsspitzen wegzufiltern. Die antiparallelen Dioden zwischen den OP-Eingängen begrenzen die Eingangs-Differenz-Spannung des OP. Der Trimmer R5 ermöglicht einen Abgleich des Messinstrumentes.

Die Verstärkung V_u eines nicht-invertierenden OP-Verstärkers berechnet sich nach (39.10). Lässt man den Wert des Widerstandes R1 gegen ∞ gehen, so wird der Bruchwert zu Null. Das gleiche geschieht, wenn man $R_2 = 0\,\Omega$ setzt. Die Verstärkung der Schaltung nach Abb. 39.12 wird damit zu $V_u = 1$, das heißt, $U_E = U_A$. Es gelten (39.11) und (39.12).

Ausgangswiderstand

$$r_a' = \frac{r_a}{V_0} \quad \text{mit } V_u = 1 \tag{39.14}$$

Diese Schaltung bewirkt also eine *Impedanzwandlung*, da sie einen hohen Eingangswiderstand und einen sehr niedrigen Ausgangswiderstand hat.

Die Schaltung nach Abb. 39.13 realisiert eine *Spannungsquelle* mit sehr kleinem Innenwiderstand bei unsymmetrischer Spannungsversorgung des OP.

Die Referenzspannung liefert eine Z-Diode. Am Ausgang des OP steht diese Spannung nun mit dem sehr kleinen Ausgangswiderstand des Spannungsfolgers zur Verfügung, während der dynamische Innenwiderstand der Z-Diode Werte von ca. 10 bis 100 Ω haben kann.

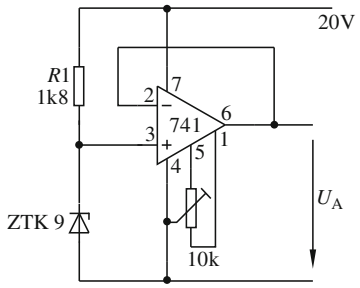


Abb. 39.13 OP als Spannungsquelle mit sehr kleinem Innenwiderstand

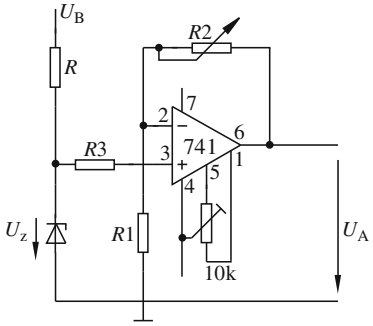


Abb. 39.14 Konstantspannungsquelle mit einstellbarer Ausgangsspannung

Deshalb kann der Ausgang der Schaltung erheblich stärker belastet werden als bei einer Stabilisierungsschaltung ausschließlich mit Z-Diode.

Um die Ausgangsspannung regel- oder einstellbar zu machen, kann auch die Prinzipschaltung in Abb. 39.14 einer *Konstantspannungsquelle* verwendet werden.

$$\textit{Ausgangsspannung} \quad U_A = \left(1 + \frac{R_2}{R_1}\right) \cdot U_Z \qquad (39.15)$$

Auch hier ist der Ausgangswiderstand durch die hohe Verstärkung des OP sehr klein und kann durch einen nachgeschalteten Spannungsfolger noch gemindert werden.

Der Operationsverstärker in der Schaltung nach Abb. 39.15 ist so beschaltet, dass die zwei Eingangsspannungen U_1 und U_2 elektrisch addiert und verstärkt werden.

Ausgangsspannung

$$-U_A = \frac{R}{R_1} \cdot U_1 + \frac{R}{R_2} \cdot U_2 \qquad (39.16)$$

Setzt man $R_1 = R_2 = R_3$, so vereinfacht sich die Gleichung zu

Ausgangsspannung

$$-U_A = \frac{R}{R_3} \cdot (U_1 + U_2) \qquad (39.17)$$

Das Minuszeichen in der Gleichung weist lediglich auf die Phasendrehung des invertierenden OP hin.

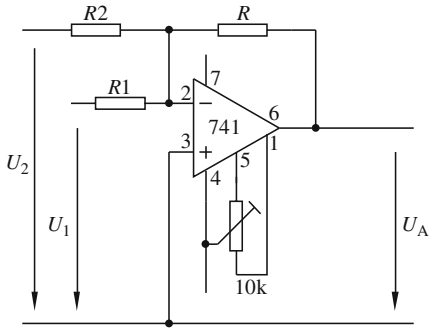


Abb. 39.15 OP als Addierer

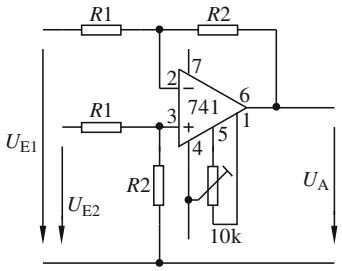


Abb. 39.16 OP als Subtrahierer

Der *Addierer* kann auf eine beliebige Anzahl n von Eingängen erweitert und nach folgender Gleichung berechnet werden:

Ausgangsspannung

$$-U_A = \frac{R}{R_1} \cdot U_1 + \frac{R}{R_2} \cdot U_2 + \ldots + \frac{R}{R_n} \cdot U \qquad (39.18)$$

Bei der Dimensionierung eines Addierer muss beachtet werden, dass die maximale Ausgangsspannung eines OP knapp 1 V unter der Versorgungsspannung des OP liegt.

Die Schaltung nach Abb. 39.16 ist ein *Subtrahierer*. Sie entspricht im Prinzip der Schaltung des Differenzverstärkers. Mit ihr wird die zwischen den beiden Eingängen wirksame Spannung verstärkt. Zur Vereinfachung, aber auch um eine Offset-Stromkompensation realisieren zu können, werden die Widerstände nach Abb. 39.16 gewählt. Werden beide Eingänge gleichzeitig angesteuert, so gilt:

Ausgangsspannung

$$U_A = \frac{R_2}{R_1} \cdot (U_{E2} - U_{E1}) \qquad (39.19)$$

Mit den gewählten Widerständen arbeitet der Subtrahierer korrekt, da die Eingangsspannungen mit dem gleichen Verstärkungsfaktor verstärkt und dann subtrahiert werden. Unsymmetrien in der Brücke können mit dem Potentiometer abgeglichen werden.

Im Zusammenhang mit Messbrücken (Brückenschaltung) nach Abb. 39.17 bietet der Subtrahierer die Möglichkeit, die massefreie Brückenspannung in eine massebezogene Spannung umzuwandeln, die dann weiterverarbeitet werden kann.

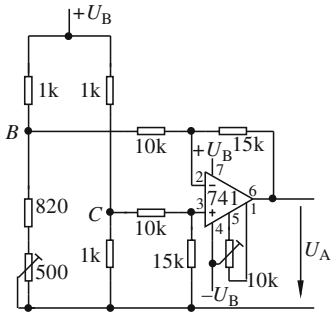


Abb. 39.17 Subtrahierer als Verstärker in Messbrücken

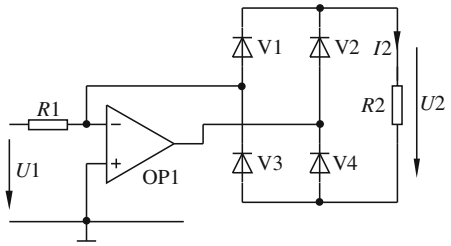


Abb. 39.18 OP als Gleichrichter mit massefreier Gleichspannung

Sollen Wechselspannungen mit kleiner Amplitude in einer normalen Brückengleichrichterschaltung B2 gleichgerichtet werden, ergibt sich wegen der Schwellenspannungen der Dioden der Nachteil, dass zwischen den positiven Halbwellen am Ausgang je nach Wert der Amplitude mehr oder weniger deutlich Lücken erkennbar sind.

Besonders in der Messtechnik sind Gleichrichterschaltungen erforderlich, die auch kleine Spannungen linear verarbeiten.

In der Schaltung nach Abb. 39.18, die einen OP mit Brückengleichrichter im Rückkopplungszweig zeigt, muss der *Gegenkopplungsstrom* wegen des Gleichrichters *stets in gleicher Richtung durch den Lastwiderstand* in der Brücke fließen.

Ein *Drehspulinstrument* als Lastwiderstand R2 im Querzweig der Brücke zeigt den arithmetischen Mittelwert dieses nichtlückenden Stroms an.

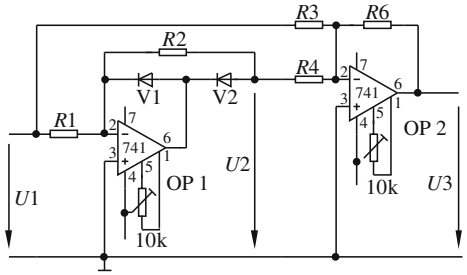
arithmetischer Mittelwert

$$I_{AV} = \frac{2 \cdot I_S}{\pi} = \frac{2 \cdot U_S}{R_1 \cdot \pi} \tag{39.20}$$

Diese Gleichrichterschaltung B2 ist zur Spannungsversorgung potentialfreier Lastwiderstände, wie zum Beispiel Anzeigeinstrumente geeignet.

Soll das gleichgerichtete Messsignal jedoch einen Lastwiderstand speisen, der einpolig an Masse liegt, so ist die obige Schaltung ungeeignet. Die Schaltung nach Abb. 39.19 besteht aus einem Einweggleichrichter mit OP und einem Addierer.

Wenn $R_1 = R_2 = R_3 = 2 \cdot R_4 = R_6 = R$ gewählt wird, dann ist $U_{S3} = U_{S1}$.



Zweiweggleichrichter-Schalter mit OPs

Abb. 39.19 Gleichrichter für kleine Spannungen mit OP (Zweiweggleichrichter)

Wird jedoch der Widerstand R6 anders gewählt, so gilt:

$$\textit{Ausgangsspannung} \quad U_{S3} = \frac{R_6}{R} \cdot U_{S1} \tag{39.21}$$

Der *Zweiweggleichrichter* nach Abb. 39.19 liefert eine pulsierende Gleichspannung wie eine normale B2-Schaltung.

Ein Drehspulinstrument am Ausgang zeigt wegen seiner mechanischen Trägheit einen konstanten (arithmetischen) Mittelwert an.

39.4.2 Verstärker mit frequenzabhängiger Gegenkopplung

Werden die ohmschen Widerstände, die bisher in den Schaltungen ausschließlich verwendet wurden, durch Kapazitäten oder Induktivitäten ersetzt, ist die Verstärkung V_u für Frequenzen unterhalb f_0 nicht mehr konstant.

Mit Wechselstromwiderständen ist der *Frequenzgang des beschalteten OP* definierbar als

Frequenzgang

$$F(j\omega) = \frac{\text{Rückführimpedanz}}{\text{Eingangsimpedanz}} \tag{39.22}$$

Die Darstellung des frequenzabhängigen Übertragungsverhaltens, aufgeteilt in den *Frequenzgang der Amplitudenverstärkung* in doppeltlogarithmischer Darstellung und den *Frequenzgang der Phasenverschiebung* linearer Darstellung, wird *Bode-Diagramm* genannt. Ein BodeDiagramm zeigt, wie ein Baustein ein sinusförmiges Eingangssignal frequenzabhängig an den Ausgang bringt.

Frequenzgang der Amplitudenverstärkung

$$F = \frac{u_a}{u_e} = f(\omega) \tag{39.23}$$

Die Amplitudenverstärkung wird meistens in Dezibel (dB) umgerechnet. *Dezibel* ist eine logarithmische Maßeinheit und

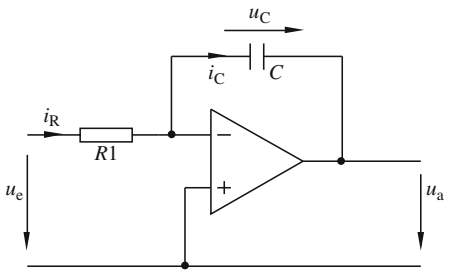


Abb. 39.20 Integrierer mit invertierendem OP

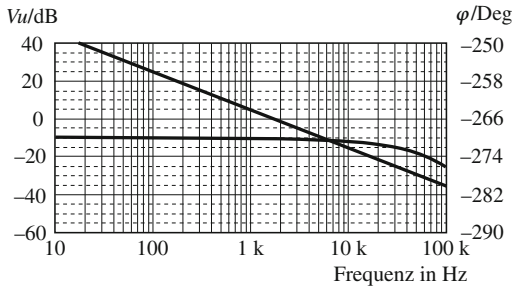


Abb. 39.21 Bodediagramm eines Integrierers

wurde ursprünglich als Maß für die Dämpfung eingeführt. Sie kennzeichnete ein Leistungsverhältnis.

Es gilt:

Leistungsverhältnis

$$\frac{P_a}{P_e} \text{ in dB} = 10 \cdot \lg \frac{P_a}{P_e} \text{ (dB)}$$
$$10\,dB = 1\,B \text{ (Bel)} \qquad (39.24)$$

Oft drückt man auch das Verhältnis zweier Spannungen oder Ströme in dB aus.

Spannungsverhältnis

$$\frac{U_a}{U_e} \text{ in dB} = 20 \cdot \lg \frac{U_a}{U_e} \text{ (dB)} \qquad (39.25)$$

Die Schaltung nach Abb. 39.20 zeigt einen *Integrator* mit invertierendem OP. Beim Integrator wird ein Teil der Ausgangsspannung über einen Kondensator auf den invertierenden Eingang des OPs zurückgekoppelt.

Das Bodediagramm eines Integrators zeigt Abb. 39.21, wobei die linke Achse die Verstärkung (gain) in dB und die rechte Achse die Phasenverschiebung (phase range) in Grad (deg) anzeigt (hier für $R_1 = 100\,\text{k}\Omega$ und $C = 1\,\text{nF}$).

Eine RC-Schaltung verursacht eine Phasenverschiebung der Ausgangsspannung um $-90°$. Zusammen mit der Phasenverschiebung des invertierenden OP von $-180°$ ergibt das die dargestellte Phasenverschiebung von $-270°$. Diese ist jedoch identisch mit einer Verschiebung von $+90°$.

Unter der Voraussetzung, dass der OP ideale Eigenschaften aufweist, gilt die Gleichung:

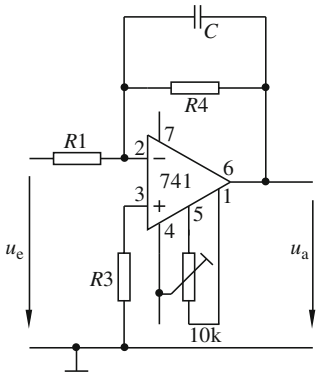


Abb. 39.22 Realer Invertierer mit invertierendem OP

Ausgangsspannung

$$u_a = -\frac{1}{R_1 \cdot C} \int_0^t u_e dt + U_{a0} \qquad (39.26)$$

U_{a0} ist die Spannung, die zu Beginn der Integration am Ausgang der Schaltung vorhanden ist. Das Produkt $R_1 \cdot C = T_I$ wird *Integrierzeit* genannt.

Ist die Eingangsspannung $u_e(t) = U_E = \text{const}$ (also eine Gleichspannung), so ist auch der Strom $I_R = -I_C = \text{const}$. Damit vereinfacht sich die Integralgleichung zur folgenden linearen Gleichung

$$\textit{Ausgangsspannung} \quad U_A = -\frac{U_E \cdot t}{R_1 \cdot C} \qquad (39.27)$$

Somit ist die Ausgangsspannung U_A eine *lineare Größe* und verläuft nicht, wie am Kondensator erwartet, nach einer e-Funktion. Sie kann je nach Polarität der Eingangsspannung nur noch linear ansteigen oder abfallen.

In der Praxis erweist sich eine Schaltung nach Abb. 39.20 als wenig zweckmäßig. Die endliche Gleichtaktunterdrückung, der Eingangsruhestrom und die Drift werden aufsummiert. Der Integrator würde immer in eine Endlage „laufen“.

Der Widerstand R3 in der Schaltung nach Abb. 39.22 dient zur Kompensation des Eingangsruhestroms. Der Widerstand R4 bestimmt die Gleichspannungsverstärkung V_u beziehungsweise stabilisiert den Gleichspannungsbetriebspunkt.

Elektrolytkondensatoren sollten wegen der relativ hohen Leckströme nicht verwendet werden. Durch den theoretisch unendlich großen Eingangswiderstand des OP entlädt sich der Kondensator fast nie.

Die Schaltung nach Abb. 39.23 zeigt die Prinzipschaltung eines *Differentiators* mit invertierendem OP. Betrachten wir den OP wieder als ideal, dass heißt, die Leerlaufverstärkung V_0 geht gegen Unendlich, so gilt

$$\textit{Ausgangsspannung} \quad -u_a = R_1 \cdot C \frac{du_e}{dt} \qquad (39.28)$$

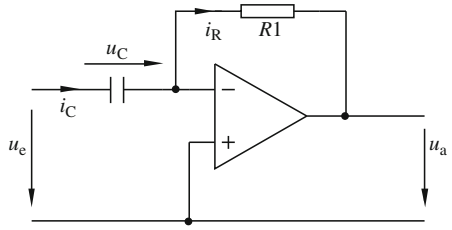


Abb. 39.23 Prinzipschaltung eines Differentiators

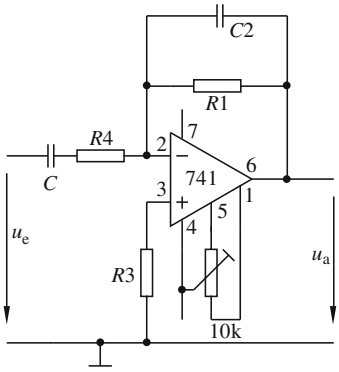


Abb. 39.24 Reale Schaltung eines Differentiators

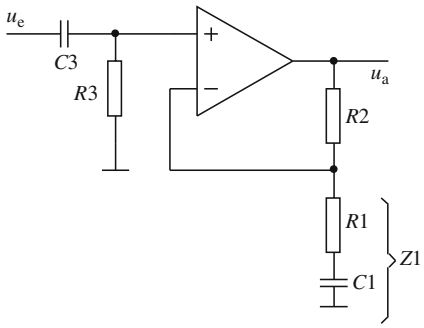


Abb. 39.25 OP als Wechselspannungsverstärker

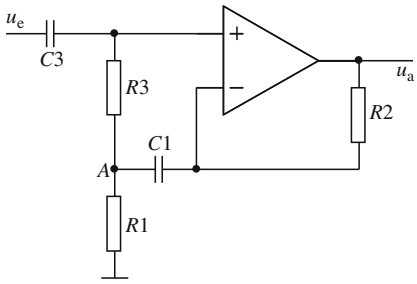


Abb. 39.26 NF-Bootstrap-Verstärker mit OP

Die Ausgangsgröße u_a des Differentiator hängt also von der zeitlichen Änderung der Eingangsspannung u_e ab. Eine konstante Eingangsspannung verursacht beim Differentiator also keine Ausgangsgröße.

Die Schaltung eines Differentiators nach Abb. 39.23 ist kaum zu verwenden, da Signale höherer Frequenzen mit der Leerlaufverstärkung V_0 verstärkt werden. Mit dem Kondensator C2 ($R_1 \cdot C_2 \approx R_4 \cdot C$ gewählt) in der Schaltung nach Abb. 39.24 kann das durch Störfrequenzen verursachte Rauschen des Differentiators herabgesetzt werden.

Die Kreisfrequenz ω sollte wegen $R_4 \cdot C$ (Eingangszeitkonstante) geschickt gewählt werden ($\omega \ll 1/R_4 \cdot C$).

Im Bode-Diagramm eines Differentiators ist der Amplitudengang eine Gerade, die mit 20 dB/Dekade ansteigt.

In der Schaltung nach Abb. 39.25 wird ein OP als Wechselspannungsverstärker verwendet. Aufgrund der vorhandenen Kondensatoren muss man hinsichtlich AC- und DC-Verstärkung/-Verhalten unterscheiden.

So trennt der Kondensator C3 den OP gleichspannungsmäßig von der Signalquelle, während der Widerstand R3 dafür sorgt, dass ein Ruhestrom fließen kann. Der Kondensator C1 macht die Gegenkopplung für DC voll wirksam.

Das bedeutet aber, dass Aus- und Eingang unter Berücksichtigung der Offset-Spannung auf dem gleichen Gleichspannungspotential liegen.

Eine Korrektur der Eingangsruheströme und ihrer Spannungsabfälle an R2 und R3 lässt sich leicht vornehmen, indem man $R_2 \approx R_3$ wählt.

Die Schaltung stellt insgesamt einen AC-Verstärker mit Hochpass-Verhalten dar. Die untere Grenzfrequenz f_u lässt sich analog zu CR-Gliedern berechnen als

$$\textit{untere Grenzfrequenz} \quad f_u = \frac{1}{2\pi \cdot R1 \cdot C1} \tag{39.29}$$

Der Eingangswiderstand der Schaltung ist praktisch gleich R3.

Hier ist nun eine Variante des Elektrometer-NF-Verstärkers hilfreich. Der NF-Bootstrap-Verstärker in der Schaltung nach Abb. 39.26 hat einen besonders hohen Eingangswiderstand.

Der Widerstand R3 liegt nicht an Masse, sondern am Punkt A, der eine Wechselspannung führt, die wegen $U_{D_} \approx 0\,\text{V}$ fast genau den gleichen Wert hat wie die Eingangswechselspannung u_e. Deshalb fließt durch den Widerstand R3 praktisch kein Wechselstrom. Der Widerstand R3 und damit der Eingangswiderstand dieser Schaltung werden dynamisch (also wechselstrommäßig) hochtransformiert. Die Ausgangs-Offsetspannung ist kleiner als in der oben erwähnten Schaltung.

39.4.3 OP als Leistungsverstärker

Bei Operationsverstärkern vom Typ 741 (Gegentaktausgang) liegt der zulässige Ausgangsstrom bei 20 mA, beim Typ 761 (open collector) liegt er bei 70 mA. Lieferbar sind von verschiedenen Herstellern auch Operationsverstärker, die Ausgangsleistungen von 25 W bei einem Ausgangsstrom von 3,5 A bereitstel-

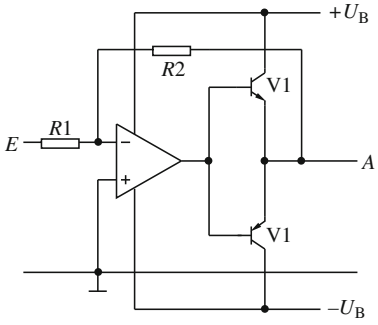


Abb. 39.27 OP-Verstärker mit Leistungsendstufe

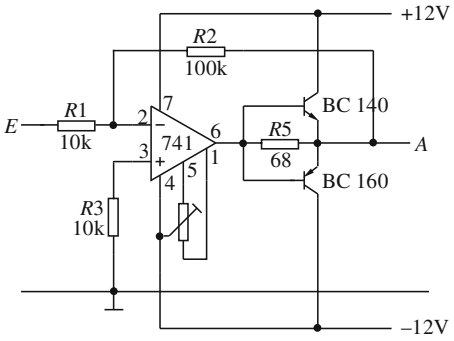


Abb. 39.28 OP mit Gegentakt-Emitterfolger im AB-Betrieb

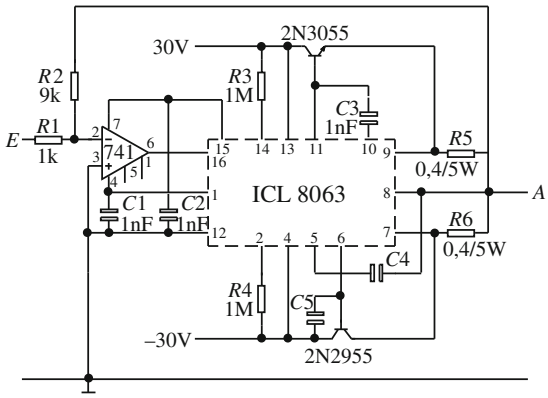


Abb. 39.29 Leistungs-OP ICL 8063 (Intersil)

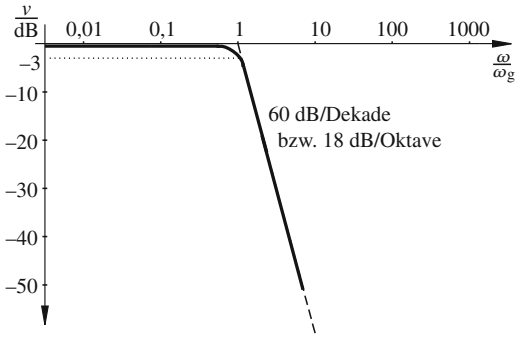


Abb. 39.30 Normiertes Bode-Diagramm eines aktuven Tiefpass-Filters

len. Zu ihrer Realisierung geben Datenblätter die häufig recht einfache externe Beschaltung an.

Bei zahlreichen Anwendungen sind noch höhere Leistungen erforderlich, die sich erreichen lassen, wenn dem OP eine Gegentakt-Leistungsendstufe nachgeschaltet wird. Diese Endstufen sind dann meistens mit externen Leistungstransistoren aufgebaut.

Bei der Schaltung eines OP-Verstärkers nach Abb. 39.27 ist als Leistungsendstufe ein komplementärer Emitterfolger für B-Betrieb angefügt. Der Gegenkopplungswiderstand R2 des invertierenden OP muss am Ausgang der Endstufe angeschlossen werden.

Die Übernahmeverzerrungen in der obigen Schaltung lassen sich mit der Schaltung nach Abb. 39.28 erheblich reduzieren. Dieser nachgeschaltete Gegentakt-Emitterfolger ermöglicht immerhin schon einen zulässigen Laststrom von ungefähr 200 mA.

Müssen die Übernahmeverzerrungen auf fast Null reduziert werden, erhöht sich der Schaltungsaufwand erheblich. Hier muss unter Umständen der Gegentakt-Emitterfolger im AB-Betrieb betrieben werden.

Um den Schaltungsaufwand zu reduzieren, wurden Ansteuerschaltungen in integrierter Schaltungstechnik entwickelt. Eine Möglichkeit ist der ICL 8063 nach Abb. 39.29, der für eine Ansteuerung mit dem OP-Typ 741 ausgelegt ist. Der Kurzschlussschutz für die beiden angesteuerten Leistungstransistoren ist im IC enthalten. Dem Ausgang der nebenstehenden Schaltung können Ströme bis ca. 3 A entnommen werden.

39.4.4 Aktive Filterschaltungen

Netzwerke aus ohmschen, induktiven und kapazitiven Widerständen, die eingesetzt werden, um den *Frequenzgang* gezielt zu beeinflussen, nennt man *Filter*.

Ideale Tiefpass-Filter sollen Nutzsignale mit $f < f_g$ unverfälscht übertragen und alle Störsignale mit $f > f_g$ unterdrücken. Dieses Ziel kann mit konstantem Übertragungsfaktor im Durchlassbereich ($f < f_g$) und mit starkem Amplitudenabfall oberhalb der Grenzfrequenz erreicht werden. Der Übergang vom Durchlassbereich zum Sperrbereich ($f > f_g$) sollte sehr steil erfolgen.

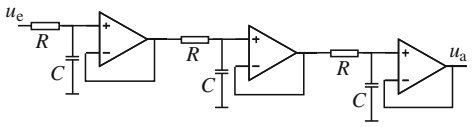
Durch das Einbringen eines Verstärkerelementes in den Filter erreicht man, dass aus der generellen Dämpfung des Signals eine Verstärkung mit $V_u \geq 1$ wird.

Es gibt sehr viele Schaltungsvarianten zu Aufbau und Ausführung von aktiven Filtern.

Man unterscheidet aktive Tief- und Hochpässe sowie selektive Filter wie aktive Bandsperren und Bandpässe.

Abb. 39.30 zeigt das normierte Bode-Diagramm eines aktiven Tiefpass-Filters 3. Ordnung.

Die Frequenz-Achse ist normiert, so dass der Durchlassbereich des TP bei $f/f_g = 1$ endet und Spannungen mit Frequenzen $f > f_g$ hier mit 60 dB/Dekade oder 18 dB/Oktave bedämpft werden.



Gauss-Filter 3. Ordnung

$f_0 = \frac{1}{2\pi RC}$ als Grenzfrequenz einer Filterstufe

Abb. 39.31 Tiefpass-Filter 3. Ordnung mit OPs

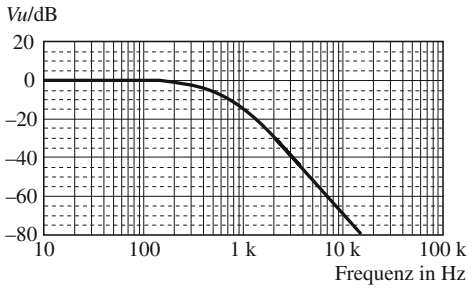


Abb. 39.32 Bode-Diagramm eines Gauß-Filters

Das normierte Bode-Diagramm eines aktiven Hochpass-Filters zeigt, dass der Durchlassbereich des HP bei $f/f_g = 1$ beginnt und Spannungen mit Frequenzen $f < f_g$ mit 60 dB/Dekade oder 18 dB/Oktave bedämpft werden. Bei der Grenzfrequenz f_g tritt eine Dämpfung/Verstärkung von 3 dB/Dekade/−3 dB/Dekade auf.

In der Schaltung nach Abb. 39.31 sind RC-Glieder mit OPs als Impedanzwandler zu einem Tiefpass 3. Ordnung zusammengeschaltet. Durch die direkte Gegenkopplung der OPs ist die DC-Verstärkung der Gesamtschaltung gleich eins. Jede der n Filterstufen hat bei der Frequenz f_0 eine Dämpfung von 3 dB/Dekade. Die gesamte Schaltung hat aber bei der Frequenz f_0 eine Dämpfung von $n \cdot 3$ dB/Dekade.

Der Tiefpass 3. Ordnung nach Abb. 39.31 hat eine Dämpfung von 3 dB/Dekade bei der Grenzfrequenz $f_g < f_0$. Oberhalb der Frequenz f_0 wird die Ausgangsspannung mit 60 dB/Dekade bedämpft. Diese Schaltung realisiert ein *Gauß-Filter*, da das Bode-Diagramm nach Abb. 39.32 die Gaußfunktion eines Tiefpass-Filters zeigt.

Charakteristikum eines Gauß-Filters ist der relativ flache Einsatz des Amplitudenganges oberhalb der Grenzfrequenz f_g. Der flache Einsatz ist aber gemessen am idealen Filter ein erheblicher Nachteil. Dem gegenüber steht der Vorteil dieses Filtertyps, dass bei sprungförmiger Änderung der Eingangsspannung die Ausgangsspannung nicht überschwingt.

Durch Vertauschen von R und C erhält man einen Hochpass 3. Ordnung. Gauß-Filter höherer Ordnung lassen sich so aufbauen.

Es gibt viele Methoden zum Aufbau aktiver Tief- und Hochpassfilter höherer Ordnung. Bei den meisten Filtern ist ihre Berechnung und Dimensionierung äußerst aufwendig.

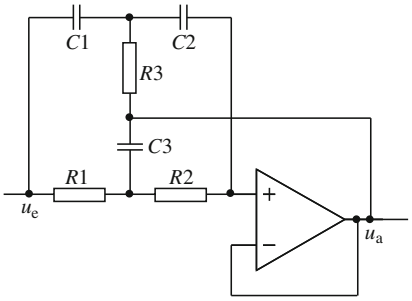


Abb. 39.33 Aktive Bandsperre mit OP

Um nur bestimmte Frequenzen zu verstärken oder zu unterdrücken, benötigt man *selektive Filter*. Mit einer Kombination von Hoch- und Tiefpass kann man selektive Filter aufbauen.

Die Schaltung nach Abb. 39.33 zeigt eine leicht berechenbare *Bandsperre*.

Wählt man $R_1 = R_2 = 2 \cdot R_3$ und $C_1 = C_2 = C_3/2$, so gilt für die Mittenfrequenz dieser Bandsperre folgende Gleichung:

$$\textit{Mittenfrequenz} \quad f = \frac{1}{2\pi \cdot R_1 \cdot C_1} \tag{39.30}$$

Dieses Filter überträgt Signale hoher und tiefer Frequenzen unverändert. In einem bestimmten Frequenzbereich werden jedoch die Signale stark unterdrückt. Bei der Grenzfrequenz (Resonanzfrequenz) hat das Filter einen hohen Übertragungswiderstand.

Das Bode-Diagramm nach Abb. 39.34 zeigt das normierte Bode-Diagramm einer Bandsperre. Es ist erkennbar. dass die obere Grenzfrequenz des Tiefpasses unterhalb der unteren Grenzfrequenz des Hochpasses liegt.

Zur Verstärkung bestimmter Frequenzen kann das Doppel-T-Filter in die Rückkopplung als Gegenkopplung des OP geschaltet werden.

Die Schaltung nach Abb. 39.35 zeigt einen aktiven *Bandpass*. Beim Bandpass werden Spannungen unterhalb und oberhalb

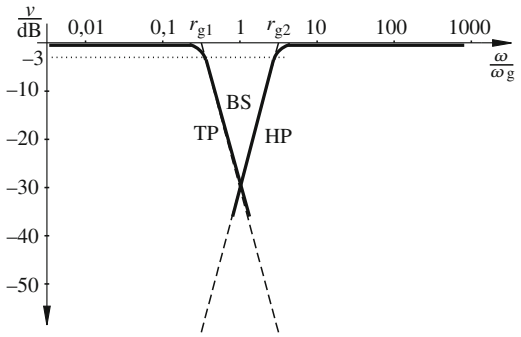


Abb. 39.34 Normiertes Bode-Diagramm einer Bandsperre

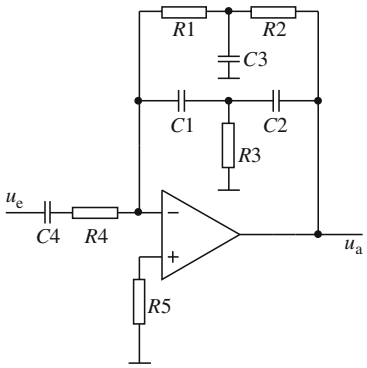


Abb. 39.35 Aktiver Bandpass

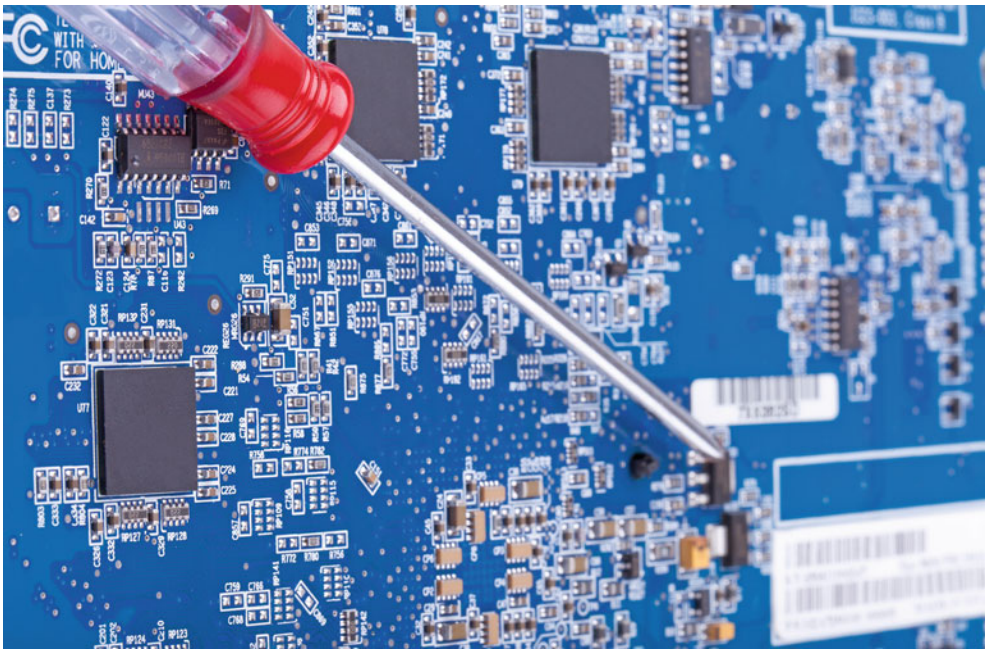
bestimmter Frequenzen bedämpft und das Frequenzband dazwischen unverändert übertragen.

In der Regel wird man versuchen, die Signale mit der favorisierten Frequenz zu verstärken.

Elektronische Schalter, Kippstufen

40

Peter Döring



40.1 Transistor als Schalter 500
40.2 Kippschaltungen mit Transistoren 503
40.3 Operationsverstärker als Schalter 507
40.4 Kippschaltungen mit Operationsverstärker 507
40.5 Zeitgeber 555 510
40.6 Trigger TCA 345 A 511

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_40

40.1 Transistor als Schalter

Wird ein Transistor als Schalter betrieben, so hat er zwei Arbeitspunkte, zwischen denen er im Betriebsfall je nach Betriebszustand wechselt. Der Übergang von einem Arbeitspunkt zum anderen wird durch ein entsprechendes Steuersignal am Eingang des Transistors erreicht.

Damit dieser Übergang möglichst schnell ablaufen kann, muss das Steuersignal einen *sprungartigen Verlauf* im Schaltzeitpunkt haben (Rechteckimpuls).

Abb. 40.1 zeigt einen Transistorschalter mit ohmschem Lastwiderstand im Serienbetrieb. Das Kennlinienfeld nach Abb. 40.2 zeigt die Anordnung der Arbeitspunkte A1 und A2 auf der Arbeitsgeraden. Die Arbeitsgerade ist die Kennlinie des Kollektorwiderstandes R_C, der in diesem Betriebsfall der Lastwiderstand R_{Last} ist.

Bei $I_B = 0\,\text{A}$ fließt der sehr geringe Kollektor-Emitter-Reststrom I_{CEO}. Damit ist $U_{\text{CE}} \approx U_B$. Der Arbeitspunkt A1 kann weiter nach unten verschoben werden, indem die Basis beispielsweise auf negatives Potential gelegt wird.

Steuert man den Transistor mit einem großen Basisstrom I_B an, fließt ein großer Kollektorstrom (Arbeitspunkt A2). Die Spannung U_{CE} nimmt einen minimalen Wert $U_{\text{CE min}} = U_{\text{CE sat}}$ bei $U_{\text{CB}} = 0\,\text{V}$ an.

Der Idealzustand eines Schalters mit $R = \infty$ (Arbeitspunkt A1) und $R = 0\,\Omega$ (Arbeitspunkt A2) wird von einem Transistorschalter nicht realisiert. Der Transistor erreicht seine Arbeitspunkte jeweils entlang der Arbeitsgeraden.

Je mehr ein Transistor übersteuert wird, umso sicherer wird der Arbeitspunkt A2 erreicht und gehalten. Als Folge des überhöhten Basisstroms ergibt sich ein weiter reduziertes $U_{\text{CE sat}}$. Den Grad der Übersteuerung gibt man als den Übersteuerungsfaktor $\ddot{u}$ an. Er ist das Verhältnis zwischen dem tatsächlich fließenden Basisstrom $I_{B\,\text{ist}}$ und dem Basisstrom $I_{B\,\text{soll}}$, der eigentlich nur erforderlich wäre, um $U_{\text{CB}} = 0\,\text{V}$ zu erreichen.

Im Schalterbetrieb wird ein Transistor meistens übersteuert. Je stärker die Übersteuerung, desto kürzer wird die Einschaltzeit t_{ein} des Transistors.

Die Einschaltzeit ergibt sich als Summe aus Verzögerungszeit t_d und Anstiegszeit t_r. Die Verzögerungszeit t_d(delay time) ergibt sich aus der Tatsache, dass die Sperrschichtkapazität C_{EB} am

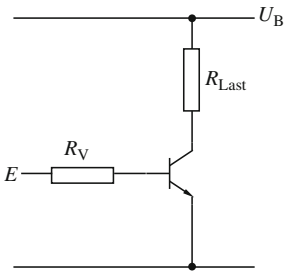


Abb. 40.1 Transistorschalter mit ohmscher Last

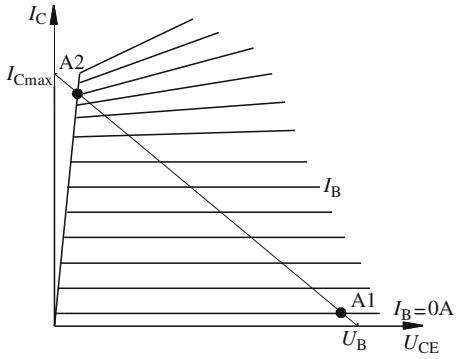


Abb. 40.2 Ausgangskennlinienfeld mit Arbeitsgerade und Arbeitspunkten

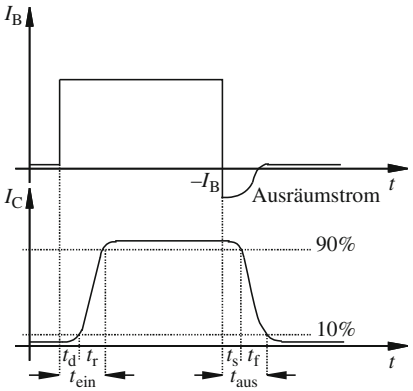


Abb. 40.3 Ein- und Ausschaltverhalten des Transistors

Anfang ungeladen ist und die Emitterdiode in Durchlassrichtung gepolt werden muss. Die Anstiegszeit t_r (rise time) ist eine definierte Größe, d. h., sie ist die Zeit, während der der Ausgangsstrom vom 0,1 auf den 0,9fachen Endwert ansteigt. Die Einschaltzeit t_{ein} wird mit zunehmendem Basisstrom kleiner.

Nachteilig ist jedoch, dass damit die Ausschaltzeit t_{aus} des Transistors größer wird. Die Basis-Emitter-Strecke wird durch den großen Basisstrom mit Ladungsträgern überschwemmt und muss beim Ausschalten erst von ihnen geräumt werden, um zu sperren. Diese Speicherzeit t_s (storage time) ergibt sich aus der Sperrschichtkapazität C_{CB}. Hinzu kommt die Abfallzeit t_f (fall time) als definierte Zeitgröße. Den Zusammenhang macht Abb. 40.3 deutlich.

Die Einschaltzeit lässt sich durch kleine Werte für R_B und große Übersteuerung verkürzen. Eine Schaltung zur Realisierung solcher Zustände zeigt Abb. 40.4. Der Basisvorwiderstand R_B wird durch die Teilwiderstände R_{B1} und R_{B2} ersetzt.

Im Einschaltmoment schließt der Kondensator C1 den Widerstand R_{B2} kurz, so dass der Basisstrom nur durch R_{B1} begrenzt wird und somit zur schnellen Übersteuerung führt. Ist der Kondensator geladen, begrenzt der Basisvorwiderstand $R_B$wieder den Basisstrom. Im Ausschaltmoment läuft der gleiche Prozess

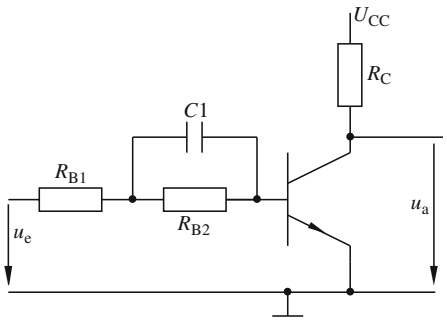


Abb. 40.4 Schaltung zur Verbesserung des Schaltverhaltens

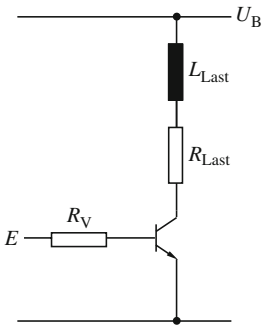


Abb. 40.5 Transistorschalter mit induktiv-ohmscher Last

mit umgekehrtem Vorzeichen.

$$\textit{Übersteuerungsfaktor} \quad \ddot{u} = \frac{I_{B\,\mathrm{ist}}}{I_{B\,\mathrm{soll}}} \tag{40.1}$$

mit

$$I_{B\,\mathrm{soll}} = \frac{I_C}{B} \tag{40.2}$$

gewählt: $\ddot{u} \approx 2 \ldots 10$

mit $R_C = R_{\mathrm{Last}}$:

$$R_C = \frac{U_B - U_{\mathrm{CE\,sat}}}{I_C} \qquad R_B = \frac{U_E - U_{\mathrm{BE}}}{I_B}$$

Überschlägig kann $R_B \approx 10 \cdot R_C$ gewählt werden.

Die *Sperrverluste* sind sehr klein gegenüber den Durchlassverlusten und können daher meistens vernachlässigt werden.

Verlustleistung im Sperrzustand

$$P_{\mathrm{Sperr}} = U_{\mathrm{CE}} \cdot I_{\mathrm{CBO}} \tag{40.3}$$

Die *Durchlassverluste* lassen sich messtechnisch und rechnerisch recht einfach bestimmen.

Verlustleistung im Durchlasszustand

$$P_D = U_{\mathrm{CE\,sat}} \cdot I_{\mathrm{CA2}} = P_V \tag{40.4}$$

Die *Umschaltverluste* lassen sich nicht so leicht bestimmen, weil die Strom- und Spannungswerte sich während des Umschaltens ständig ändern. Bei einem Schalter mit ohmscher Last sind Ein- und Ausschaltverluste beim Umschalten gleich groß, was bei kapazitiven und induktiven Lasten nicht der Fall ist. In der Praxis sind die wesentlich größeren Ausschaltverluste bei induktiven Lasten besonders zu beachten.

Die Berechnung der Umschaltverluste kann allenfalls überschlägig erfolgen.

Umschaltverluste

$$P_{\mathrm{Sch}} \approx U_B \cdot I_{C\,\max} \cdot \frac{(t_{\mathrm{ein}} + t_{\mathrm{aus}})}{T} \tag{40.5}$$

Bei Transistoren im Schalterbetrieb ergibt sich die gesamte Verlustleistung P_{ges} als die Summe der Teilverluste.

$$\textit{Verlustleistung} \quad P_{\mathrm{ges}} = P_{\mathrm{Sperr}} + P_D + P_{\mathrm{Sch}} \tag{40.6}$$

Bei *Leistungstransistoren* liegen die Werte für die Ein- und Ausschaltzeiten im Bereich von 1 . . . 3 ms.

Wenn die Lasten zudem noch in größeren Zeitabständen geschaltet werden, gilt folgende Vereinfachung:

Für $T \gg (t_{\mathrm{ein}} + t_{\mathrm{aus}})$ gilt

$$P_{\mathrm{ges}} \approx P_D = U_{\mathrm{CE\,sat}} \cdot I_{\mathrm{CA2}} = P_V \tag{40.7}$$

Für den *Impulsbetrieb* (im Gegensatz zum Dauerbetrieb) kann eine *mittlere Verlustleistung* definiert und berechnet werden.

mittlere Verlustleistung

$$P_m = P_V \cdot \nu = \frac{U^2}{R} \cdot \nu \tag{40.8}$$

$$\textit{Periodendauer} \quad T = t_i + t_p$$

$$\textit{Tastverhältnis} \quad \nu = \frac{t_i}{T} \tag{40.9}$$

In der Praxis werden Transistoren für den Betrieb als Leistungsschalter bezüglich ihrer Verlustleistung meistens recht großzügig ausgewählt. Es muss auf die korrekte Bemessung der Kühlkörper geachtet werden. Auch die Einhaltung der Grenzwerte $U_{\mathrm{CE\,max}}$ und $I_{C\,\max}$ des ausgewählten Transistors muss gewährleistet sein. Das ist besonders wichtig im Hinblick auf den kleinen Kaltwiderstand einer Glühlampe im Einschaltmoment oder auch wegen des hohen Anlaufstroms von Gleichstrommotoren.

Relais, Hubmagnete oder Motore haben einen überwiegend induktiven Anteil und können insgesamt durch die Ersatzschaltung einer Spule (Induktivität und ohmscher Spulenwiderstand in Reihe) elektrisch beschrieben werden.

In der Schaltung nach Abb. 40.5 ist der R_{Last} so gewählt, dass er mit dem in Abb. 40.1 übereinstimmt. Die Arbeitsgerade, die sich aus diesem Widerstand ergibt, ist im Kennlinienfeld gestrichelt eingetragen. Auf ihr liegen die *statischen Arbeitspunkte* A1 und A2.

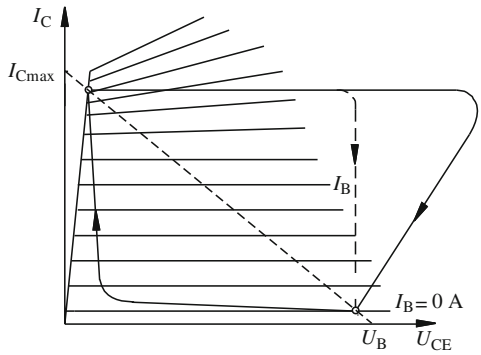


Abb. 40.6 Ausgangskennlinienfeld mit eingezeichnetem Verlauf des Arbeitspunktwechsels

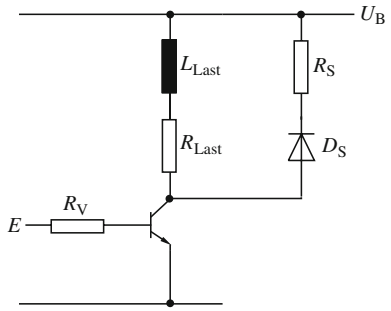


Abb. 40.7 Transistorschalter mit Freilaufdiode

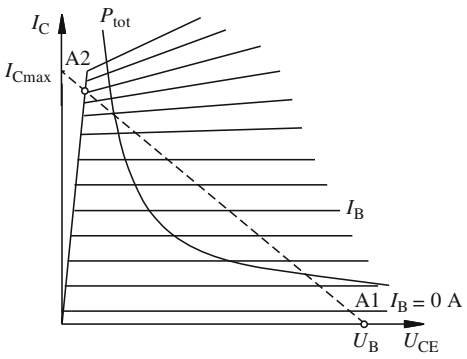


Abb. 40.8 Ausgangskennlinienfeld mit Arbeitsgerade und Verlusthyperbel

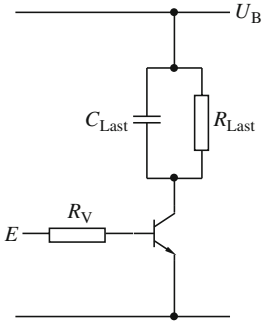


Abb. 40.9 Transistor mit Kondensator als Last

Aufgrund der Selbstinduktionsspannung kann der Kollektorstrom zunächst nur gering ansteigen, was zur Folge hat, dass sich der dynamische Arbeitspunkt dem Arbeitspunkt A2 von unten nähert. Diesen Vorgang zeigt das Kennlinienfeld nach Abb. 40.6.

Im Moment des Ausschaltens wird die in der Induktivität gespeicherte Energie frei und bewirkt eine hohe Selbstinduktionsspannung, die so gepolt ist, dass sie als positive Kollektorspannung wirkt. Die Selbstinduktionsspannung erhöht die Kollektor-Emitter-Spannung U_{CE} soweit, dass sie größer als die Betriebsspannung U_B werden kann. Sie kann Werte erreichen, durch die der Transistor zerstört wird ($U_{CE} > U_{CEO}$).

Der Umschaltvorgang verläuft entlang der dynamischen Umschaltkurve nach Abb. 40.6 und macht die Spannungsüberhöhung deutlich. Aus diesem Grund muss ein Transistorschalter im Serienbetrieb mit induktiven Lasten (Relais) vor Überspannungen im Ausschaltmoment geschützt werden.

Ein einfacher, aber wirksamer Schutz ist eine *Freilaufdiode*, wie sie in der Schaltung nach Abb. 40.7 zu sehen ist. Die Freilaufdiode liegt für die Betriebsspannung U_B in Sperrrichtung und ist somit unwirksam. Erst im Ausschaltmoment, also dann, wenn infolge der Induktionsspannung die Kollektorspannung größer als U_B wird, liegt die Diode in Durchlassrichtung. Der dann fließende Strom I_F baut die Energie des Magnetfeldes so rasch ab, dass eine gefährliche Überspannung nicht auftreten kann. Der Widerstand R_S begrenzt den Strom durch die Diode auf zulässige Werte und wandelt die magnetische Energie in Wärme um. Die Wirkung dieser Schutzmaßnahme ist im Kennlinienfeld nach Abb. 40.6 eingezeichnet.

Die Arbeitsgerade kann beim Schalterbetrieb des Transistors nach Abb. 40.8 durch die Verlustleistungshyperbel verlaufen, jedoch müssen die statischen Arbeitspunkte außerhalb liegen. Bei induktiver beziehungsweise kapazitiver Last muss darauf geachtet werden, dass der dynamische Arbeitspunkt nicht zu lange in der Hyperbel verweilt.

In der Schaltung nach Abb. 40.9 ist die Prinzipschaltung eines Transistorschalters mit einem Kondensator als Lastwiderstand dargestellt. Der Kondensator ist als „reale“ Kapazität dargestellt.

Ohne diesen Parallelwiderstand, der im Kennlinienfeld nach Abb. 40.10 als gestrichelte Arbeitsgerade dargestellt ist, wäre ein statischer Betrieb im Arbeitspunkt A2 nicht möglich.

Der Wert dieses R_{Last} ist mit dem Widerstandswert nach Abb. 40.1 identisch gewählt, um die Wirkung von C_{Last} anschaulicher darstellen zu können.

Qualitativer Verlauf und Richtung des Arbeitspunktverlaufs beim Ein- und Ausschalten sind im Kennlinienfeld nach Abb. 40.10 eingezeichnet. Je niedriger der Kapazitätswert ist, desto mehr nähert sich die Übergangskennlinie der gestrichelten Arbeitsgeraden.

In der Schaltung nach Abb. 40.11 wird ein Transistor als Schalter verwendet, um ein Leistungsrelais an den Ausgang eines Digitalgatters (hier NOT-Glied) anzupassen. Liegt der Ausgang des Gatters auf L-Potential, so ist der Transistor gesperrt; bei

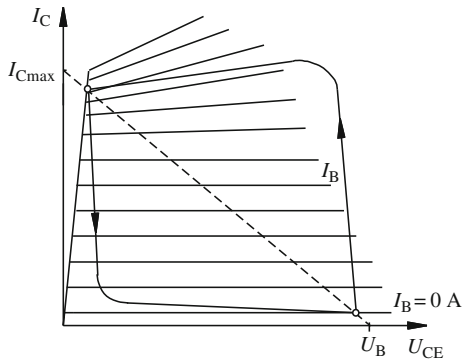


Abb. 40.10 Ausgangskennlinienfeld mit eingezeichnetem Verlauf des Arbeitspunktwechsels

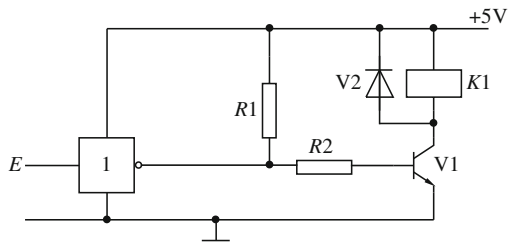


Abb. 40.11 Transistorschalter als „Interface"

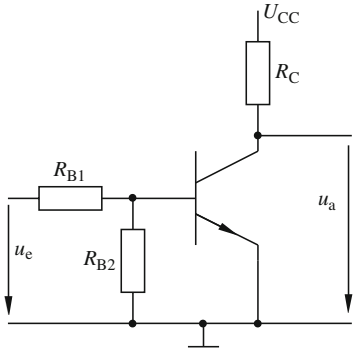


Abb. 40.12 Transistorschalter mit Unterstützung zum sicheren Sperren

H-Potential hingegen ist der Transistor leitend und damit das Relais im erregten Zustand.

Um zu verhindern, dass auch bei L-Potential der Transistor leitend wird, verwendet man häufig eine Schaltung nach Abb. 40.12. Hier reicht das L-Potential häufig nicht aus, um eine Basisvorspannung zu bewirken, die den Transistor leitend werden lässt. In der Signalverarbeitung wird der Transistorschalter meistens an seinem Ausgang durch die Eingangswiderstände nachfolgender Transistorschalter belastet. Diese Belastung kann sowohl gegen Masse als auch gegen die Betriebsspannung U_B auftreten.

40.2 Kippschaltungen mit Transistoren

Kippschaltungen bestehen hauptsächlich aus zwei Transistor-Schaltstufen, bei denen gleichzeitig eine Stufe leitend und die andere Stufe gesperrt ist.

Je nachdem, wie ein Kippvorgang ausgelöst oder eingeleitet wird, aber auch nach der Anzahl der stabilen Zustände einer Kippschaltung, unterscheidet man folgende Arten: bistabile Kippstufen, monostabile Kippstufen, astabile Kippstufen und Triggerschaltungen.

Die Art der Kippstufen hängt nach Abb. 40.13 von der technischen Ausführung der Koppelelemente Z1 und Z2 ab. Sind die Koppelelemente Z1 und Z2, wie hier im Bild dargestellt, ohmsche Widerstände R1 und R2, so ist diese Kippschaltung eine „bistabile Kippstufe".

Ist das Koppelelement Z1 ein ohmscher Widerstand R1, jedoch das Koppelelement Z2 durch das Koppelelement Z4 (C2 und R2) ersetzt, so ist diese Kippschaltung eine „monostabile Kippstufe".

Ist das Koppelelement Z2 weiterhin ein ohmscher Widerstand R2, wird jedoch das Koppelelement Z1 durch das Koppelelement Z3 (C1 und R1) ersetzt, so handelt es sich bei der Kippschaltung ebenfalls um eine „monostabile Kippstufe".

Werden in der Kippschaltung sowohl Koppelelement Z1 durch Koppelelement Z3 wie auch Koppelelement Z2 durch Koppelelement Z4 ersetzt, so handelt es sich bei dieser Kippschaltung um eine „astabile Kippstufe".

40.2.1 Bistabile Kippstufe

Eine *bistabile Kippstufe* nach Abb. 40.14, auch „Flip-Flop" (FF) genannt, besitzt zwei stabile Schaltzustände. Hier sind beide Transistorschalter identisch dimensioniert. Es ist erkennbar, dass der Kollektor von V1 direkt auf die Basis des Transistors V2 einwirkt wie auch der Kollektor von V2 direkt auf die Basis des Transistors V1 einwirkt.

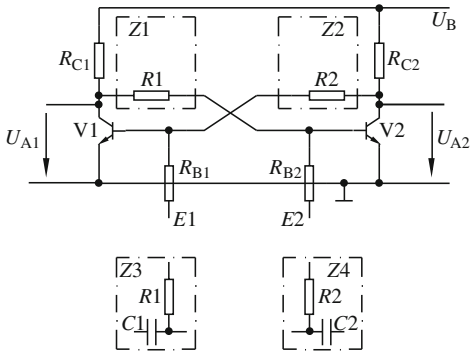


Abb. 40.13 Gemeinsame Kriterien und Synthese von Kippstufen

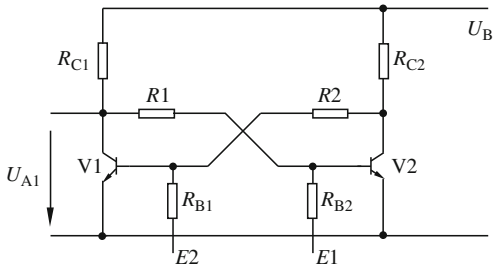


Abb. 40.14 Bistabile Kippstufe, auch „Flip-Flop" (FF)

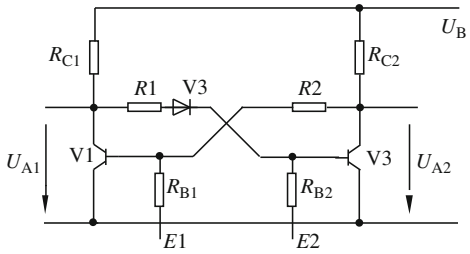


Abb. 40.15 Bistabile Kippstufe mit definiertem Grundzustand

Ist der Transistor V2 im Sperrzustand, so ist V1 leitend und damit $U_{A1} \approx U_{CE\,sat}$. Die Basis des Transistor V1 liegt über den Widerstand R_2 und R_{C2} an U_B. Somit liegt die Basis von V2 über R_1und R_{C1} an $U_{CE\,sat1}$ von V1 und ist sicher gesperrt. Je nachdem, welcher der beiden Transistoren als erster durchsteuert, stellt sich ein bestimmter, aber dann nicht von selbst veränderbarer Zustand an den Ausgängen ein.

An die Basisanschlüsse der Transistoren V1 und V2 nach Abb. 40.14 sind die Eingänge E1 und E2 angebunden. Geht man von obiger Darstellung aus, das heißt, V1 leitend und V2 gesperrt, und bringt man nun auf den Eingang E1 einen positiven Spannungsimpuls U_{E1}, so wird V2 leitend und V1 gesperrt und damit $U_{A1} \approx U_B$. Dies hat zur Folge, dass die Basis von V2 über R1 auf U_B gehalten wird und der Transistor V2 leitend bleibt.

Bringt man nun auf den Eingang E2 einen positiven Spannungsimpuls U_{E2}, so wird V1 leitend. Die beiden positiven Spannungsimpulse auf E1 und E2 haben also den ursprünglichen Zustand wiederhergestellt. Die Werte für U_{A1} und U_{A2} sind einander entgegengesetzt. Zum Umschalten in den jeweils entgegengesetzten Zustand ist nur ein kurzzeitiger Eingangsimpuls am jeweiligen Steuereingang erforderlich. Der dann eingenommene Zustand wird solange gespeichert, bis ein entsprechender Impuls am anderen Eingang einen neuen Zustand herbeiführt.

Aus diesem Grund können Flip-Flops (bistabile Kippstufen) in der Digitaltechnik als Informationsspeicher eingesetzt werden.

Beim Anlegen der Spannung U_B an ein FF tritt ein beliebiger, aber stabiler Ausgangszustand ein, der nur von den Parameterstreuungen der Transistoren abhängt.

Soll nach dem Anlegen der Spannung U_B ein definierter Ausgangszustand festliegen, so muss die Schaltung nach Abb. 40.14 entsprechend Abb. 40.15 modifiziert werden.

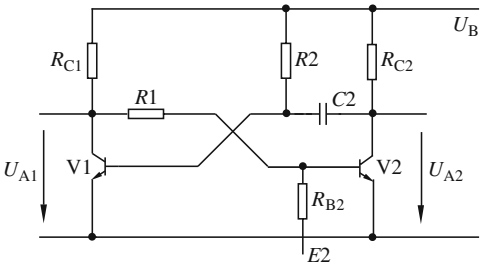


Abb. 40.16 Prinzipschaltung einer monostabilen Kippstufe

Die Diode V3 bewirkt in dieser Schaltung, dass der Transistor V1 immer vor dem Transistor V2 leitend wird und nach dem Einschalten der Versorgungsspannung somit $U_{A1} \approx U_{CE\,sat}$ ist.

Berechnet werden kann eine solche Schaltung prinzipiell nach folgenden Gleichungen (U_{BE} und U_{V3} vernachlässigt):

Kollektorwiderstand

$$R_{C1} = R_{C2} = \frac{U_B - U_{CE\,sat}}{I_C} \approx \frac{U_B}{I_C} \tag{40.10}$$

Vorwiderstand

$$R_1 = R_2 = \frac{1}{\ddot{u}} \cdot R_C \cdot B \tag{40.11}$$

Durch weitere Maßnahmen, wie zum Beispiel Parallelschaltung von Kondensatoren zu den Widerständen R1 und R2, kann die *Flankensteilheit* der Ausgangsspannungen verbessert werden.

Parallelkondensatoren

$$C_K \leq \frac{R_C + R_B}{4{,}6 \cdot f \cdot R_C \cdot R_B} \tag{40.12}$$

mit der Frequenz der Eingangssignale $f = \frac{f_{ein}}{2}$.

40.2.2 Monostabile Kippstufe

Es gibt zwei Versionen der monostabilen Kippstufe (Abb. 40.13). Sie werden mit Kennzeichnung der Eingänge E1 und E2 vorgestellt. Ein positives Signal auf den Eingang E2 lässt das Monoflop nach Abb. 40.16 vom stabilen Zustand in den nichtstabilen Zustand kippen. Nach einer von der Schaltungsdimensionierung abhängigen Zeit t_i kippt er von selbst in die stabile Schaltstellung zurück und kann nach einer bestimmten Zeit t_{rec} erneut gesetzt werden.

Beim Anlegen der Versorgungsspannung U_B wird der Transistor V1 leitend, weil sein Basisstrom über RC2 und R2 größer ist als derjenige nach V2. Der Kondensator C2 wird über R_{C2} geladen, das heißt, die Spannung U_{C2} am Kondensator ist am Ende des Ladevorganges gleich der Spannung U_B.

Es fließt kein Strom mehr über C2 in die Basis des Transistors. Lediglich der normale Basisstrom I_B fließt über R2 zu. Die Spannung U_{BE} am Transistor V1 beträgt ca. 0,7 V.

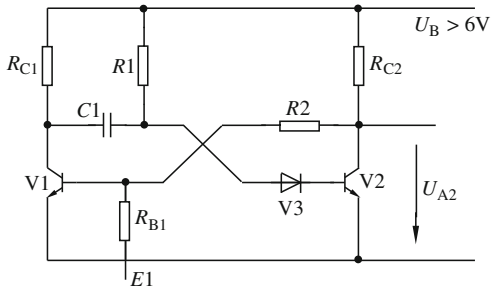


Abb. 40.17 Verbesserte monostabile Kippstufe

Ein positives Signal auf den Eingang E2 lässt den Transistor V2 leitend werden, wodurch die Spannung an der Basis von V1 auf einen Wert von $U_{BE} = -U_B$ heruntergezogen wird, da der positive Anschluss von C2 auf Masse gelegt wird. Der Transistor V1 sperrt infolgedessen. Der Kondensator C2 beginnt nun damit, sich über den Widerstand R2 auf die Batteriespannung $+U_B$ umzuladen. Wenn sich der Kondensator auf $U_{BE} = +0{,}7\,\text{V}$ umgeladen hat, wird der Transistor V1 wieder leitend und damit $U_{A1} \approx 0\,\text{V}$.

$$\textit{Schaltzeit} \quad t_i = R_2 \cdot C_2 \cdot \ln 2 \tag{40.13}$$

Der Kondensator C2 muss sich wieder in den normalen Zustand umladen, also von ca. 0 V auf $U_{C2} = +U_B$. Man spricht hier von der

$$\textit{Erholzeit} \quad t_{rec} = 5 \cdot R_{C2} \cdot C_2 \tag{40.14}$$

Die nächste Triggerung (positives Signal an E2) des Monoflops sollte nicht vor Ablauf der Erholzeit erfolgen, da sonst C2 nicht voll aufgeladen wird und die Schaltzeit „unkalkulierbar" verkürzt wird. Mit handelsüblichen Bauteilen lassen sich Impulszeiten von ca. 10 ms bis maximal 10 s erreichen. Bei Verwendung von FET können längere Zeiten erreicht werden.

Wegen $t_i \approx t_{rec}$ ist für t_i durch C2 eine untere Grenze gegeben, die praktisch bei $t_i \approx \frac{T}{10}$ liegt.

Im Moment der Triggerung liegt die volle negative Batteriespannung U_B an der Basis-Emitter-Strecke des Transistors und beansprucht sie in Sperrrichtung. Diese BE-Strecken haben aber schon bei Spannungen um 6 V ihre Durchbruchspannungen, wodurch dieser pn-Übergang stark gefährdet ist. Die Zeit t_i wird durch die schnelle Entladung des Kondensators stark verkürzt. Abhilfe schafft hier die Diode V3 nach Abb. 40.17.

40.2.3 Astabile Kippstufe

Im Prinzip ist eine astabile Kippstufe nach Abb. 40.18 eine geschickte Kombination zweier Monoflops ohne Eingänge. Beide Monoflops sind so geschaltet, dass sie sich jeweils nach Ablauf der durch die Zeitglieder bestimmten Schaltzeiten gegenseitig triggern (schalten). Da beide Schalter nur noch über Kondensatoren wechselstrommäßig miteinander gekoppelt sind, hat die

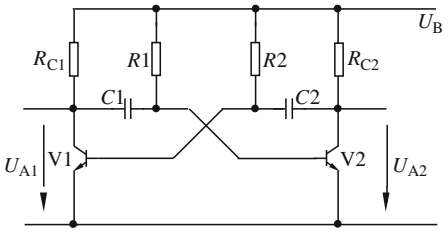


Abb. 40.18 Prinzipschaltung einer astabilen Kippstufe

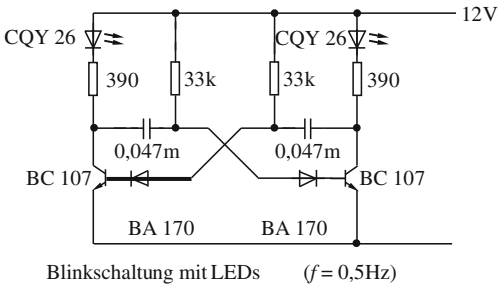


Abb. 40.19 Blinkschaltung mit LEDs

Schaltung zwei instabile Zustände. Sie kippt daher fortlaufend von einem instabilen Zustand in den anderen und erzeugt dabei eine Ausgangsspannung mit einem kontinuierlichen, rechteckförmigen Verlauf (Rechteckgenerator).

Gleichstrommäßig sind die beiden Transistorschalter so ausgelegt, dass sie bei fehlenden Kondensatoren voll leiten würden. Die Frequenz der Ausgangsspannung sowie das Impuls-Pausen-Verhältnis hängen von den Zeitkonstanten der beiden RC-Glieder $R_2 \cdot C_2$ beziehungsweise $R_1 \cdot C_1$ ab.

$$\textit{Periodendauer} \quad T = (R_1 \cdot C_1 + R_2 \cdot C_2) \cdot \ln 2 \tag{40.15}$$

Die Blinkschaltung in Abb. 40.19 ist in dieser Form verwendbar. Die Dioden vor der Basis der Transistoren haben die gleiche Funktion wie in Abb. 40.17 beschrieben.

40.2.4 Triggerschaltungen

Als Schmitt-Trigger wird eine Kippschaltung bezeichnet, die beim Überschreiten einer bestimmten Eingangsspannung U_{E1} vom Ruhezustand in den Arbeitszustand kippt und erst beim Unterschreiten einer bestimmten Eingangsspannung U_{E2} wieder in den Ruhezustand zurückkippt. Schmitt-Trigger nach Abb. 40.20 werden häufig als Schwellenwertschalter eingesetzt.

Wie bei jeder Kippschaltung erfolgt das Kippen stets sprunghaft, so dass auch bei einem beliebigen Verlauf der Eingangsspannung die Ausgangsspannung eines Schmitt-Triggers rechteckförmig ist und nur die beiden Spannungswerte $U_{A\,min}$ und $U_{A\,max}$ erscheinen, also ein digitales Signal vorliegt. Da das Kippen meistens schnell erfolgt, ergeben sich steile Flanken. Das Eingangssignal ist ein analoges Signal; der Schmitt-Trigger kann also als Analog-Digital-Wandler gelten.

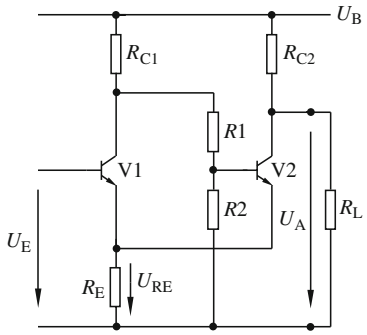
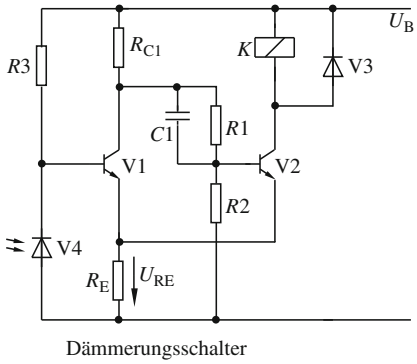


Abb. 40.20 Schmitt-Trigger als Schwellenwertschalter



Dämmerungsschalter

Abb. 40.22 Lichtabhängiger Schwellenwertschalter

Über den gemeinsamen Emitter-Widerstand R_E liegt eine gleichstrommäßige Mitkopplung zwischen den beiden Transistoren vor. Der Spannungsteiler aus R_1 und R_2 muss so dimensioniert sein, dass V2 mit der sich einstellenden Emitterspannung U_{RE} voll durchgesteuert ist. Die Ausgangsspannung $U_{A\,\min}$ ist die Summe aus $U_{RE} + U_{CE\,sat}$. Der Schmitt-Trigger befindet sich jetzt im Ruhezustand.

Er bleibt im Ruhezustand, solange die Eingangsspannung $U_E < U_{E1}$ ist, wobei U_{E1} die obere Schwellspannung des Schmitt-Triggers darstellt. Die Mitkopplung bewirkt, dass der Schmitt-Trigger bei einer Erhöhung von U_E über U_{E1} hinaus sehr schnell vom Ruhezustand in den Arbeitszustand kippt, wobei V1 leitend und V2 gesperrt ($U_{A\,max} = U_B$) ist. Dieser Arbeitszustand bleibt solange erhalten, wie $U_E > U_{E2}$ ist. Wird die Eingangsspannung $U_E < U_{E2}$, kippt der Schmitt-Trigger in den Ruhezustand zurück.

Die Spannungsdifferenz zwischen U_{E1} und U_{E2} wird als *Schalthysterese* des Schmitt-Triggers bezeichnet. Der Zusammenhang zwischen Ein- und Ausgangsspannung wird in Abb. 40.21 in Form der Übertragungskennlinie dargestellt. Die Flankensteilheit kann durch Parallelschalten eines Kondensators zu R_1 verbessert werden. Sprunghafte Änderungen des Kollektorpotentials von Transistor V1 wirken sich so schneller auf V2 aus.

Bei dem *lichtabhängigen Schwellenwertschalter* nach Abb. 40.22 wird der Schaltvorgang bei einer bestimmten Beleuchtungsstärke ausgelöst.

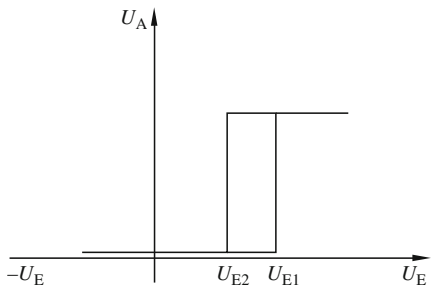


Abb. 40.21 Übertragungskennlinie des Schmitt-Triggers

Solange auf die Fotodiode V4 kein Licht fällt, ist diese hochohmig, V1 leitend und damit V2 gesperrt. Fällt ausreichend Licht auf V4, kippt der Trigger und das Relais *K* zieht an.

Als Hilfe zur Berechnung eines Schmitt-Triggers können folgende Überlegungen dienen:

Ruhelage 1

V1 gesperrt, V2 leitend mit $U_E < U_{E1}$

Einschaltspannung

$$U_{E1} = U_{RE2} + U_{BE1} = I_{C2} \cdot R_E + U_{BE1} \tag{40.16}$$

Kollektorstrom

$$I_{C2} = \frac{U_B - U_{CE\,sat}}{R_{C2} + R_E} \tag{40.17}$$

Ausgangsspannung

$$U_A = U_{RE2} + U_{CE\,sat2} \tag{40.18}$$

Ruhelage 2

V1 leitend, V2 gesperrt mit $U_E > U_{E2}$

Ausschaltspannung

$$U_{E2} = U_{RE1} + U_{BE1} = I_{C1} \cdot R_E + U_{BE1} \tag{40.19}$$

Kollektorstrom

$$I_{C1} = \frac{U_B - U_{CE\,sat1}}{R_{C1} + R_E} \tag{40.20}$$

Ausgangsspannung

$$U_A = U_B \quad \text{(Leerlauf)} \tag{40.21}$$

Hilfreich zur Berechnung ist die Überlegung, dass beim Ausschalten für den Ausschaltmoment $I_{C1} = I_{C2}$ sein muss!

$$\textit{Hysteresespannung} \quad \Delta U_E = U_{E1} - U_{E2} \tag{40.22}$$

40.3 Operationsverstärker als Schalter

Wird ein *OP ohne Gegenkopplung* betrieben, so ist seine Leerlaufverstärkung V_0 voll wirksam. Legt man den OP nach Abb. 40.23 mit dem invertierenden Eingang auf Masse und an den nichtinvertierenden Eingang eine variable Spannung, so ist diese Eingangsspannung U_E auch gleichzeitig die *Differenzspannung* U_D, die mit V_0 verstärkt als Ausgangsspannung U_A an den Ausgang des OP gelangt.

Durch die sehr hohe Leerlaufverstärkung (beim idealen OP ist $V_0 = \infty$) gelangt der OP schnell in die Sättigung, das heißt die Ausgangsspannung erreicht schnell die positive oder negative Aussteuergrenze des jeweiligen OP nach Abb. 40.24.

Beispiel

Die Aussteuergrenze eines OP soll $U_A = \pm 12\,\mathrm{V}$ und die Leerlaufverstärkung $V_0 = 10^5$ betragen.

Es gilt: $U_A = U_D \cdot V_0 \Rightarrow U_D = U_A/V_0 = \pm 12\,\mathrm{V}/10^5 = \pm 0{,}12\,\mathrm{mV}$. ◀

Demnach reicht eine Eingangsspannung von $\pm 0{,}12\,\mathrm{mV}$ aus, um den OP an die jeweilige Aussteuergrenze zu bringen. Der Signalwechsel am Ausgang findet also theoretisch immer beim Polaritätswechsel der Eingangsspannung statt. Der OP realisiert einen Komparator ohne Hysterese (Komparator = Spannungs-Vergleicher; to compare: engl. = vergleichen).

Natürlich ändert sich die Ausgangsspannung nicht in unendlich kurzer Zeit. Die Anstiegsgeschwindigkeit (slew rate) der Ausgangsspannung des OP 741 liegt zum Beispiel bei $1\,\mathrm{V/ms}$, das heißt, der Wechsel der Ausgangsspannung von der positiven zur negativen Aussteuergrenze (und umgekehrt) dauert im obigen Beispiel ca. 24 µs. Außerdem muss eine zusätzliche „Erholzeit" infolge der Übersteuerung des OP berücksichtigt werden.

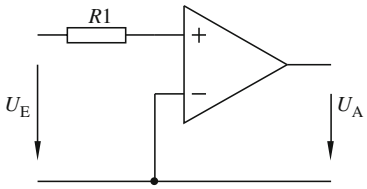


Abb. 40.23 Nichtinvertierender OP als Schalter

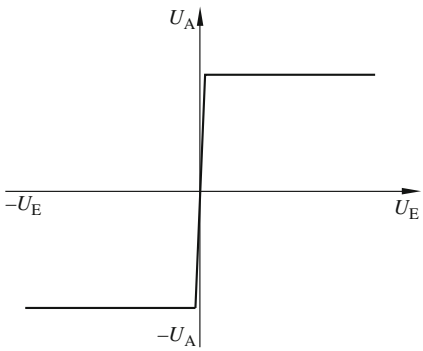


Abb. 40.24 Übertragungskennlinie eines Komparators nach Abb. 40.23

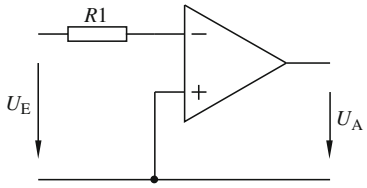


Abb. 40.25 Schaltung eines invertierenden Komparators

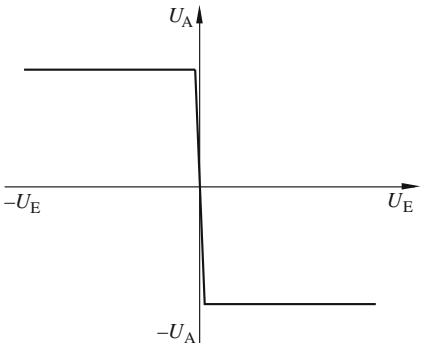


Abb. 40.26 Übertragungskennlinie eines invertierenden Komparators

Legt man nach Abb. 40.25 den nichtinvertierten Eingang auf Masse und an den invertierten Eingang die Eingangsspannung nach nebenstehender Schaltung, so kehren sich die oben dargestellten Überlegungen um (siehe Abb. 40.26).

Der *Schaltpunkt* dieser Schwellwertschalter liegt also bei einer Spannung U_E nahe 0 V.

40.4 Kippschaltungen mit Operationsverstärker

40.4.1 Triggerschaltungen mit Operationsverstärker

Ein Schwellwertschalter ohne Hysterese ist in der Praxis kaum verwendbar, denn die dem eigentlichen Signal überlagerten Störspannungen führen zum mehrfachen Umschalten.

Durch Mitkopplung lassen sich Schaltungen realisieren, die diese Hysterese aufweisen. In der Schaltung nach Abb. 40.27 wird die Spannung am Widerstand R2 mitgekoppelt. Weil die Eingangsspannung auf den invertierenden Eingang geführt wird, heißt diese Schaltung „Invertierender Komparator mit Hysterese".

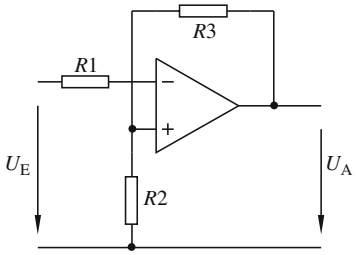


Abb. 40.27 Schwellwertschalter mit Hysterese

Die Ausgangsspannung kann nur zwei definierte Werte annehmen, nämlich $U_A \approx \pm U_B$.

$$\textit{Mitkopplungsspannung} \quad U_{R2} = U_B \cdot \frac{R_2}{R_2 + R_3} \qquad (40.23)$$

Wenn das Eingangssignal U_E den Wert von U_{R2} erreicht, kippt die Ausgangsspannung von $+U_B$ nach $-U_B$ um. Der Grund liegt darin, dass die Differenzspannung U_D ihren Nullpunkt durchläuft. Nun liegt der nichtinvertierende Eingang wegen des Spannungsteilers aus R2 und R3 auf einem negativen Potential, nämlich auf $-U_{R2}$.

Wenn das Eingangssignal den Wert von $-U_{R2}$ erreicht, kippt die Ausgangsspannung von $-U_B$ nach $+U_B$ um. Die rückgekoppelte Spannung U_{R2} des Spannungsteilers stellt also die Ein- und die Ausschaltschwelle des Komparators dar.

Das Übertragungsverhalten des Komparators ist in Abb. 40.28 dargestellt. Die Schaltpunkte liegen symmetrisch um das Nullpotential. Der Betrag der Schalthysterese ist also die Summe aus U_{E1} und U_{E2} und lässt sich berechnen als

$$\textit{Hysteresespannung} \quad U_{\mathrm{Hy}} = 2 \cdot U_B \cdot \frac{R_2}{R_2 + R_3} \qquad (40.24)$$

Sollen die Schaltpunkte nicht symmetrisch um das Nullpotential liegen, muss der jeweils benutzte Eingang auf den gewünschten Spannungswert (Referenzspannung) mit umgekehrtem Vorzeichen gelegt (vorgespannt) werden.

Die Ausgangsspannung U_A ist wiederum die positive oder negative Aussteuergrenze des OP.

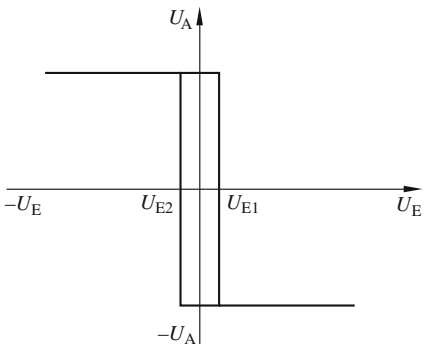


Abb. 40.28 Übertragungsverhalten eines Komparators mit Hysterese

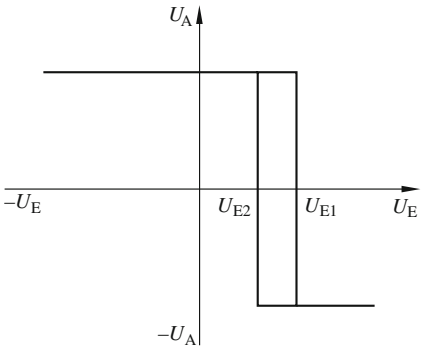


Abb. 40.29 Übertragungskennlinie eines Komparators einstellbarer Schaltschwelle

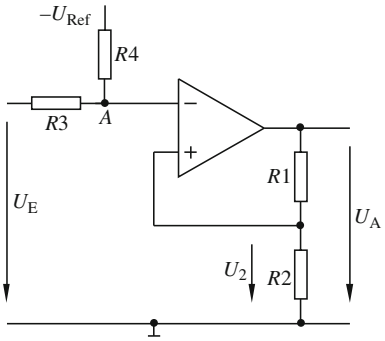


Abb. 40.30 Komparator mit frei wählbaren Schaltpunkten

Die Übertragungskennlinie nach Abb. 40.29 zeigt das prinzipielle Schaltverhalten. Der OP in der Schaltung nach Abb. 40.30 schaltet um, wenn das Potential am Punkt A den Wert Null durchläuft. Mit den Widerständen R1 und R2 sind die *Ein- und Ausschaltpunkte* im Zusammenhang mit R3 und R4 einstellbar. In dieser Schaltung sind die Schaltpunkte völlig frei wählbar.

Beispiel

Ein invertierender Komparator (mit idealem OP) soll für die Schaltpunkte $U_{E1} = 10\,\mathrm{V}$ und $U_{E2} = 8\,\mathrm{V}$ berechnet werden. Die Aussteuergrenzen des OP liegen bei $U_A = \pm 12\,\mathrm{V}$. $R_3 = R_4 = 100\,\mathrm{k\Omega}$ und $R_2 = 33\,\mathrm{k\Omega}$.

Zu bestimmen sind die Referenzspannung, die Mitkopplungsspannung U_2 und der Wert von R2.

Lösung:

In zwei Schaltbilder sind die Potentiale unmittelbar vor dem Signalwechsel von U_A einzutragen. Daraus kann man erkennen:

Die Hysterese beträgt 2 V. Die Spannung am Punkt A verändert sich nur um 1 V infolge der Spannungsteilung durch R3 und R4, wenn sich die Eingangsspannung von 10 auf 8 V ändert.

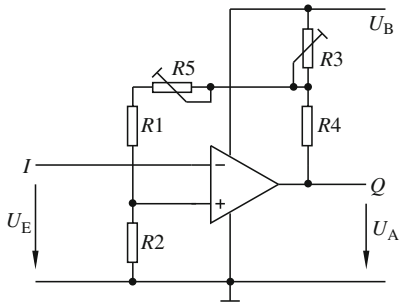


Abb. 40.31 Komparator mit einstellbarer Schaltschwelle und einstellbarer Hysterese

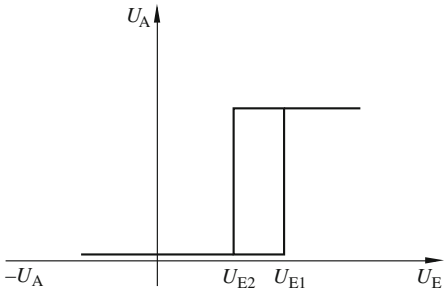


Abb. 40.32 Übertragungskennlinie zum Komparator nach Abb. 40.31

Dieselbe Spannungsänderung muss durch die Mitkopplungsspannung U_2 erfolgen; somit muss $U_2 = 0{,}5\,\text{V}$ sein.

Damit kann R2 als Teilerwiderstand berechnet werden. Wegen $R_3 = R_4$ und theoretisch unendlich großem OP-Eingangswiderstand ist $U_{R3} = U_{R4}$.

Mit $U_D = 0\,\text{V}$ gilt im Schaltaugenblick: $U_{R3} = 10\,\text{V} - 0{,}5\,\text{V} = 9{,}5\,\text{V}$. Damit ergibt sich als Referenzspannung $U_{\text{Ref}} = -9\,\text{V}$! ◀

Die Schaltung nach Abb. 40.31 zeigt einen invertierenden Komparator, bei dem *Schaltschwelle und Hysterese* mit Hilfe der Potentiometer R5 und R3 eingestellt werden. Der OP liegt an einer unsymmetrischen Versorgungsspannung, so dass die Ausgangsspannung U_A entsprechend der Übertragungskennlinie nach Abb. 40.32 maximal zwischen 0 V und U_B springen kann.

40.4.2 Astabile Kippstufe mit Operationsverstärker

Beschaltet man den Eingang eines invertierenden Komparators (mit Hysterese) nach Abb. 40.33 mit einem RC-Glied, das seine Spannung vom Ausgang des OP bezieht, so erhält man eine *astabile Kippstufe*.

Die Mitkopplungsspannung U_2 hat dieselbe Polarität wie die Ausgangsspannung und wird durch den Spannungsteiler erstellt.

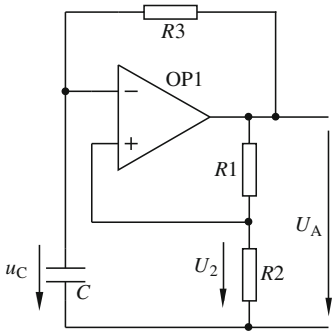


Abb. 40.33 Astabile Kippstufe mit OP

Der Kondensator lädt sich über den Widerstand R3 auf die Spannung U_A auf. Erreicht die Kondensatorspannung u_c den Wert von U_2, so kippt der Ausgang, und der Kondensator lädt sich in die andere Richtung um. Dieser Vorgang wiederholt sich kontinuierlich.

$$\textit{Impulsdauer} \quad t_i = R_3 \cdot C_1 \cdot \ln \frac{U_A + U_2}{U_A - U_2} \tag{40.25}$$

Da $t_i = t_p$ ist und somit $T = 2 \cdot t$ und außerdem U_2 ersetzbar ist, erhält man für die

Periodendauer

$$T = 2 \cdot R_3 \cdot C_1 \cdot \ln\left(1 + 2 \cdot \frac{R_2}{R_1}\right) \tag{40.26}$$

Mit $f = 1/T$ lässt sich die *Taktfrequenz der astabilen Kippstufe* berechnen. Die Ausgangsspannung ist eine Rechteckspannung mit positiver und negativer Amplitude (Rechteckgenerator).

40.4.3 Monostabile Kippstufe mit Operationsverstärker

Schaltet man parallel zum Kondensator C1 nach Abb. 40.34 eine Diode V1, dann sorgt die Schleusenspannung U_S der Diode dafür, dass der Kondensator sich nur in einer Richtung bis zur Mitkopplungsspannung U_2 laden kann (Voraussetzung: $U_2 > U_S$ gewählt).

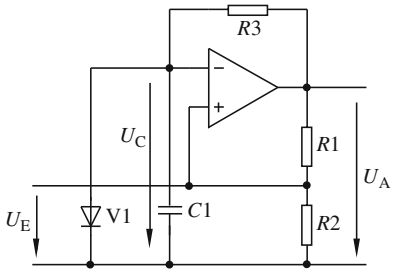


Abb. 40.34 Monostabile Kippstufe mit OP

Teil V

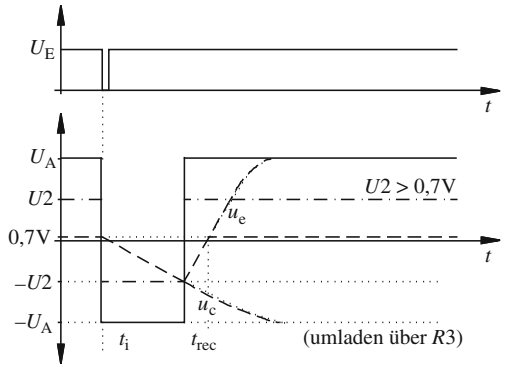


Abb. 40.35 Liniendiagramm einer monostabilen Kippstufe

Die Schaltung kann nur nach einem der beiden Ladevorgänge kippen. Es handelt sich also um eine *monostabile Kippstufe*.

Die Ausgangsspannung wird für eine bestimmte Zeit t_i auf einem bestimmten Pegel gehalten. (*Timer*, *Zeitglied*).

Die *Polarität* der Ausgangsspannung wird von der *Richtungslage der Diode V1* bestimmt.

$$\textit{Impulsdauer} \quad t_i = R_3 \cdot C_1 \cdot \ln \frac{U_A + 0{,}7\,\mathrm{V}}{U_A - U_2} \qquad (40.27)$$

Abgerufen wird hier die Zeit t_i, indem man für einen Moment den nichtinvertierenden Eingang des OP negativer als den invertierenden Eingang macht. Dadurch kippt der OP am Ausgang auf $-U_A$, der nichtinvertierende Eingang wird somit auf $-U_2$ gelegt, und der Kondensator lädt sich nach Abb. 40.35 um. Wird $u_C = -U_2$, so kippt der OP erneut.

Der Kondensator lädt sich aber jetzt nur auf die Schleusenspannung U_S der Diode, so dass ein weiterer Kippvorgang nicht ausgelöst wird. Wird die Diode umgepolt, so ist der Ablauf genau umgekehrt zu verfolgen. Die Zeit t_{rec}, die der Kondensator benötigt, um sich von $-U_2$ auf U_S aufzuladen, sollte bis zum nächsten Abrufen des Zeitgliedes verstreichen, da sonst t_i nicht immer gleich ist.

40.4.4 Bistabile Kippstufe mit Operationsverstärker

Eine *bistabile Kippstufe* mit OP stellt die Schaltung nach Abb. 40.36 dar.

Die Wirkungsweise der Schaltung ergibt sich aus dem Aufbau. Der Schaltungsaufwand ist sehr gering gegenüber dem Aufwand, der betrieben werden muss bei der Realisierung einer solchen Schaltung mit Transistoren. Auch die Schaltschnelligkeit des OP ist deutlich höher.

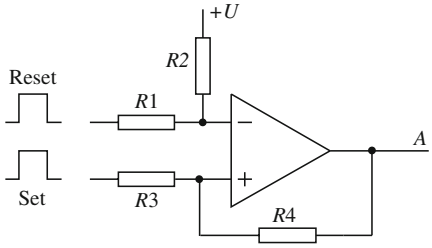


Abb. 40.36 Bistabile Kippstufe mit OP

40.5 Zeitgeber 555

Monostabile und astabile Kippschaltungen lassen sich für zahlreiche Anwendungen mit einer *integrierten Standardschaltung* realisieren, die als *Zeitgeber* mit der Typenbezeichnung IC 555 (Abb. 40.37) bekannt ist.

Der Timer-IC 555 ist in der Schaltung nach Abb. 40.38 als *monostabile Kippstufe* beschaltet. Die Verweilzeit t_i hängt nur von dem zeitbestimmenden RC-Glied und nicht von der Betriebsspannung des IC ab. Pin 2 ist der „Set"-Eingang, der auf die negative Flanke des Set-Signals reagiert.

An Pin 5 ist stets ein Kondensator $C_B = 10\,\mathrm{nF}$ gegen Masse zu schalten. Bei einer Versorgungsspannung von $V_{CC} = +5\,\mathrm{V}$ ist der Ausgang TTL-kompatibel; während der Verweilzeit t_i ist $U_a \approx V_{CC}$. Die Zeit kann zwischen 1 ms und ca. 15 min liegen.

$$\textit{Verweilzeit} \quad t_i = R_1 \cdot C \cdot \ln 3 \qquad (40.28)$$

Die einfache monostabile Kippstufe ist in der Schaltung nach Abb. 40.39 zu einer *nachtriggerbaren monostabilen Kippstufe* erweitert worden. Gibt man innerhalb der Verweilzeit t_i einen

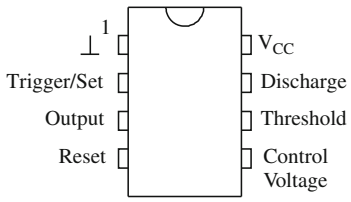


Abb. 40.37 Belegungsplan eines IC 555

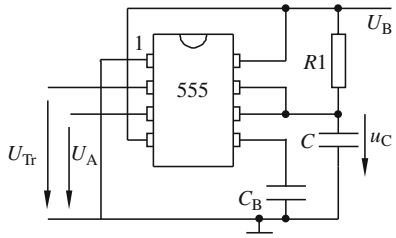


Abb. 40.38 Timer-IC 555 als monostabile Kippstufe

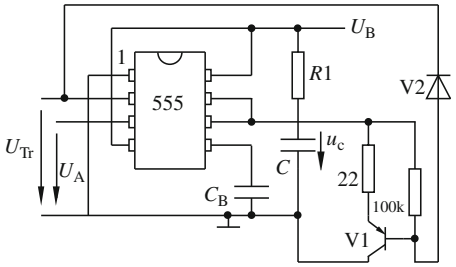


Abb. 40.39 Nachtriggerbarer Timer mit dem IC 555

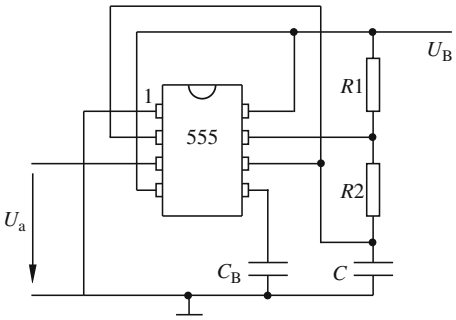


Abb. 40.40 Astabile Schaltung mit dem IC 555

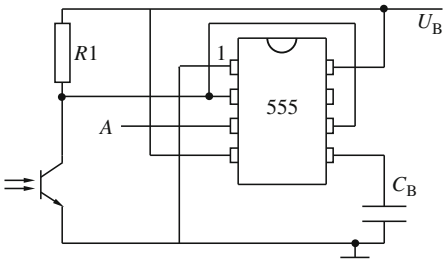


Abb. 40.41 Lichtabhängiger Trigger mit dem IC 555

erneuten Impuls mit negativer Flanke auf den Trigger (Pin 2), so wird dadurch das Rückkippen der KS verzögert. Bei einer schnellen Folge von Impulsen kann dadurch das Rückkippen der KS beliebig verzögert werden.

Das „Nachtriggern“ wird solange fortgesetzt, bis eine ausreichend lange Impulslücke auftritt. Die *Verweilzeit* t_i' nach dem letzten Impuls hängt allerdings auch von V_{CC} ab.

Der IC 555 ist auch als astabile Kippstufe einsetzbar. Während der Impulsdauer t_i ist die Ausgangsspannung $U_a \approx V_{CC}$. Impulsdauer t_i und Pausendauer t_p sind unabhängig voneinander einstellbar.

Für die astabile Schaltung nach Abb. 40.40 gelten folgende Regeln:

$$\textit{Impulsdauer} \qquad t_i = (R_1 + R_2) \cdot C \cdot \ln 2 \qquad (40.29)$$

$$\textit{Pausendauer} \qquad t_p = R_2 \cdot C \cdot \ln 2 \qquad (40.30)$$

Als *Trigger* wird der IC 555 in der Schaltung nach Abb. 40.41 verwendet. Fällt Licht auf den Fototransistor, wird er niederohmig, und die abfallende Flanke triggert den IC 555, der aufgrund der Ausgangsspannung $U_a \approx V_{CC}$ und des Ausgangsstroms (max. 200 mA) unmittelbar ein Relais schalten könnte.

40.6 Trigger TCA 345 A

Neben einer Vielzahl an Triggern, z. B. aus der TTL-Typenreihe 74xx und der CMOS-Typenreihe 40xx, gibt es auch andere Trigger-IC-Typen. Der TCA 345A nach Abb. 40.42 ist einfach in eine Schaltung zu integrieren und durch die Auswahl der Versorgungsspannung U_B leicht entsprechend dem zugehörigen Diagramm auf bestimmte Schwellwerte und zugehörige Hysteresespannungen einzustellen.

Die Schaltung nach Abb. 40.43 zeigt einen Dämmerungsschalter mit dem LDR RPY 63 und dem TCA 345 A.

Die Schaltung wird mit dem Poti R2 so abgeglichen, dass bei Tageslicht (LDR niederohmig) die Eingangsspannung des TCA größer als U_{Ein} ist, so dass der Ausgang (Pin 4) auf 10 V liegt. Dadurch wird ein Schalten des Relais verhindert. Nach Einbruch der Dunkelheit unterschreitet die Eingangsspannung den Wert von U_{Aus} und das Relais zieht an.

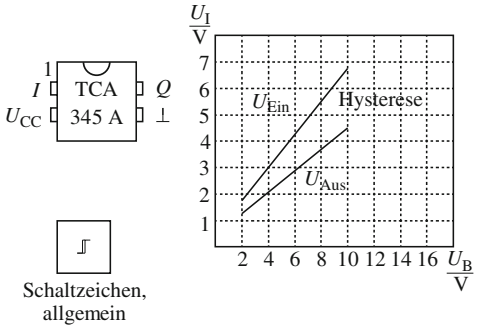


Abb. 40.42 TCA 345 mit Belegungsplan und Übertragungskennlinie

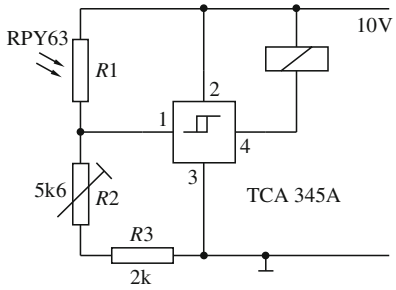
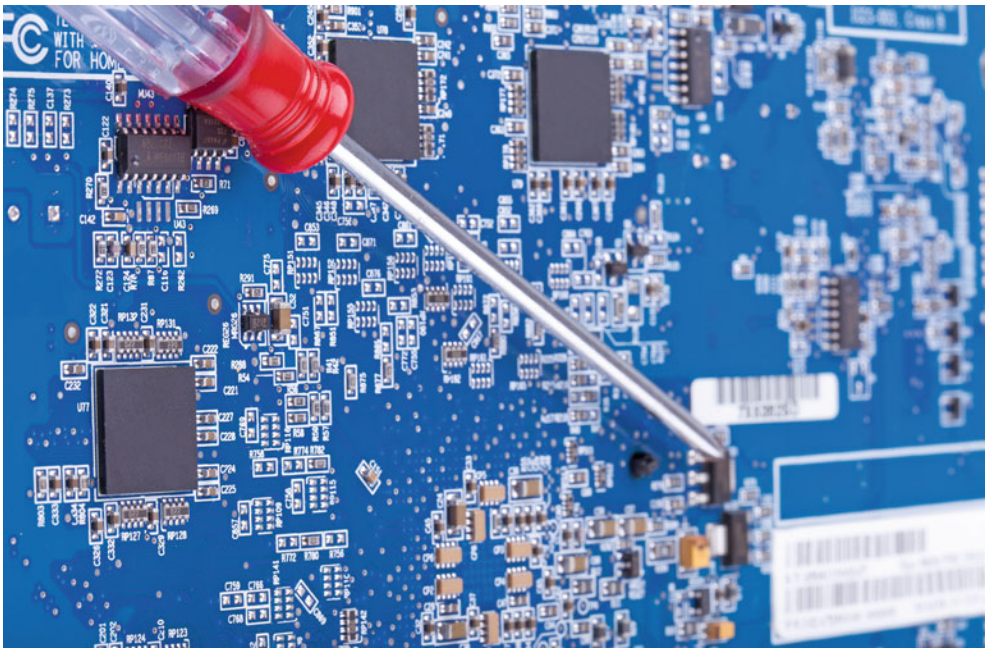


Abb. 40.43 TCA 345-Applikation als Dämmerungsschalter

Oszillatoren

41

Peter Döring



41.1 Allgemeines 514
41.2 Sinusgeneratoren (RC-Oszillatoren) 514
41.3 Funktionsgeneratoren 516

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_41

41.1 Allgemeines

Oszillatoren sind *elektronische Schwingungserzeuger*. Sie dienen zur Erzeugung von *ungedämpften sinusförmigen Schwingungen*. Einsatzgebiete liegen vor in Sendern (Erzeugung der Hochfrequenz), Empfängern (Bildung der Zwischenfrequenz), Messgeräten (Sinus-, Rechteckgeneratoren) und elektronischen Musikinstrumenten (Synthesizer).

Oszillatoren bestehen aus einem Verstärker und einem *Rückkopplungsnetzwerk*, das meistens eine *Mitkopplung* darstellt. Diese Rückkopplung muss bestimmte Bedingungen erfüllen, damit Sinusschwingungen entstehen.

Das Rückkopplungsnetzwerk besteht aus einem Resonator und einer Anpassungsschaltung. Unter einem Resonator ist ein schwingungsfähiges Netzwerk, zum Beispiel ein Schwingkreis oder ein Schwingquarz, zu verstehen, das mit seiner Eigenfrequenz die Schwingfrequenz des Oszillators bestimmt. Da ein Resonator normalerweise verlustbehaftet ist, würde nach dem Anschwingen die Ausgangsamplitude exponentiell abnehmen, so dass eine gedämpfte Schwingung statt der gewünschten ungedämpften Schwingung zustande käme. Die Verluste der Schaltung müssen also durch einen Verstärker kompensiert werden.

Deshalb wird ein Teil der Ausgangsgröße (Spannung oder Strom) auf den Resonator zurückgeführt, so dass die Schwingung des Resonators erneut angestoßen wird (Selbsterregung).

Eine Selbsterregung kann nur bei einer Phasengleichheit von Eingangs- und Ausgangsgröße entstehen, das heißt, es muss eine Mitkopplung vorliegen (Phasenbedingung).

Eine Mitkopplung hat als generelle Tendenz die Übersteuerung des Verstärkers zur Folge und liefert keine sinusförmige Ausgangsspannung mehr. Es muss also eine Amplitudenbegrenzung vorgenommen werden.

Bildet man das Verhältnis von Ausgangsspannung u_a zur Eingangsspannung u_e und nennt diese Verstärkung v, so kann man auch das Verhältnis von rückgekoppelter Spannung u_r zur Ausgangsspannung u_a bilden und nennt dieses Verhältnis Kopplungsfaktor K.

$$\textit{Grundbedingung} \quad v = \frac{u_a}{u_e} \quad u_r = K \cdot u_a \tag{41.1}$$

Das *Produkt* $K \cdot v$ bezeichnet man als *Ringverstärkung*. Die Ringverstärkung muss mindestens gleich eins sein ($K \cdot v = 1$), damit die Selbsterregung einsetzen kann (Amplitudenbedingung). Dieser Mindestwert reicht aus, um eine bestehende Schwingung zu erhalten (ungedämpfte Schwingung).

Ist die Ringverstärkung $K \cdot v < 1$, kommt keine Schwingung zustande oder eine vorhandene Schwingung „reißt“ ab.

Ist die Ringverstärkung $K \cdot v > 1$, schaukelt sich die Amplitude der Schwingung auf, und es treten durch Übersteuerung Verzerrungen auf.

Soll ein Oszillator anschwingen, muss die Ringverstärkung $K \cdot v > 1$ sein; nach dem Anschwingen jedoch muss sich der Verstärkungs- oder Kopplungsfaktor mit wachsender Amplitude „automatisch“ auf $K \cdot v = 1$ verringern.

Sind die Schwingungsbedingungen nur für eine Frequenz erfüllt, schwingt der Verstärker nur auf dieser einen Frequenz; die Schwingung ist sinusförmig.

41.2 Sinusgeneratoren (RC-Oszillatoren)

Bei sehr hohen Ansprüchen an die Genauigkeit der erzeugten Frequenz wird als frequenzbestimmendes Bauteil anstelle des Schwingkreises ein Schwingquarz eingesetzt. Die maximale Abweichung seiner Eigenschwingungen beträgt etwa 0,0001 %.

RC-Oszillatoren werden vorzugsweise für den Frequenzbereich von 0,01 Hz bis 500 kHz verwendet, da LC-Oszillatoren in diesem Bereich nur durch „unhandlich“ große Induktivitäten (schlechte Güte, relativ hohes Gewicht) und Kapazitäten realisiert werden können.

Man unterscheidet zwei Arten von RC-Oszillatoren: Phasenschieber- und Wien-Brücken-Oszillatoren.

Das Prinzip des Phasenschieberoszillators lässt sich nach Abb. 41.1 erläutern. Die Ausgangsspannung u_a einer Emitterschaltung ist gegenüber der Eingangsspannung um 180° phasenverschoben. Um die Phasenbedingung (Mitkopplung) zu erfüllen, muss das Rückkopplungsnetzwerk ebenfalls eine Phasenverschiebung von 180° zwischen der Ausgangsspannung des Verstärkers und der rückgekoppelten Teilspannung erzeugen.

Mindestens drei in Reihe geschaltete RC-Glieder sind erforderlich, um diese Phasenverschiebung zu erzeugen, da jedes Glied eine Phasenverschiebung von weniger als 90° bewirkt. Im Allgemeinen werden alle Kondensatoren und Widerstände der Phasenschieberkette gleich groß gemacht. Eine Selbsterregung tritt ein, wenn die am Poti R_C abgegriffene Rückkopplungsspannung groß genug ist.

Der Oszillator schwingt mit der Frequenz f_0, für die diese Phasenschieberkette insgesamt genau eine Phasenverschiebung von 180° bewirkt. Die Phasenverschiebung der einzelnen RC-Glieder ist nicht gleich groß, da jeweils das folgende Glied das vorausgehende belastet.

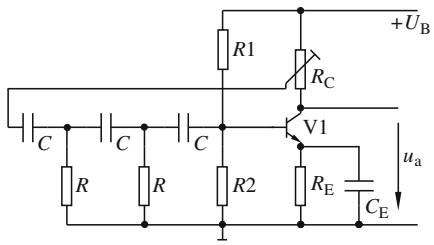


Abb. 41.1 Prinzip des Phasenschieberoszillators

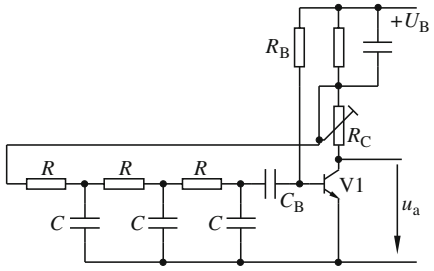


Abb. 41.2 Phasenschieberoszillators

Frequenz der Ausgangsspannung

$$f_0 = \frac{1}{\sqrt{6} \cdot 2\pi \cdot R \cdot C} \tag{41.2}$$

Die Widerstände R1, R2 und r_{BE} ergeben sich durch den Verstärker. Deshalb sind die Widerstände R danach zu dimensionieren.

$$\textit{Widerstandsbestimmung} \qquad R = R_1 \parallel R_2 \parallel r_{BE} \tag{41.3}$$

$$\textit{Rückkopplungsfaktor} \qquad K = \frac{1}{29} \tag{41.4}$$

Um die Amplitudenbedingung zu erfüllen, muss der Verstärkungsfaktor des Verstärkers generell $\rightarrow v \geq 29$ sein.

Die Kondensatoren und Widerstände der Phasenschieberkette können auch gemäß der Schaltung nach Abb. 41.2 vertauscht werden, ohne dass sich die Wirkungsweise grundsätzlich ändert. Die Amplitudenbegrenzung erfolgt hier durch eine Verstärkungsregelung. Bei großer Amplitude lädt sich der Kondensator C_B aufgrund der Gleichrichterwirkung der Basis-Emitter-Strecke auf. Dadurch kommt es zu einer Arbeitspunktverschiebung des Transistors, wodurch die Verstärkung geringer wird, bis die Ringverstärkung $K \cdot v = 1$ ($v = 29$) ist.

Bei Phasenschieberoszillatoren ist die Amplitudenstabilisierung sehr kritisch, da die Abstimmelemente R oder C synchron veränderbar sein müssen. Die Frequenzkonstanz dieser Schaltung genügt selten den Ansprüchen moderner Schaltungen. Aus diesen Gründen werden Phasenschieberoszillatoren in der Praxis wenig angewendet.

Eine wesentliche Verbesserung ergibt sich bei Verwendung einer *Wien-Brücke* als Rückkopplungsnetzwerk. Die Wien-Brücke nach Abb. 41.3 ist ein frequenzabhängiger Spannungsteiler. Die Schaltung besteht aus der Reihenschaltung eines RC-Tiefpasses mit einem RC-Hochpass, wobei die Ausgangsspannung u_a am Hochpass abgegriffen wird.

Bildet man mathematisch das Verhältnis von Ausgangsspannung u_a zur Eingangsspannung u_e, ergibt sich für eine bestimmte Frequenz eine maximale Ausgangsspannung und ein Phasenverschiebungswinkel von 0° zwischen Aus- und Eingangsspannung.

Daraus ergibt sich bei $\frac{u_e}{u_a} = 3 + j\frac{(\omega RC)^2 - 1}{\omega RC}$ und Trennung von Real- und Imaginärteil die

$$\textit{Resonanzfrequenz} \qquad f_0 = \frac{1}{2\pi \cdot RC} \tag{41.5}$$

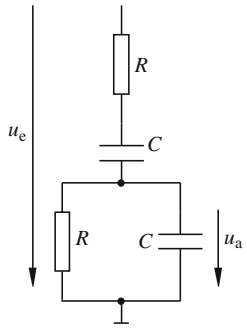


Abb. 41.3 Wien-Brücke als Rückkopplungsnetzwerk

und die maximale

$$\textit{Ausgangsspannung bei } f_0 \qquad U_{a\,\max} = \frac{1}{3} U_e \tag{41.6}$$

Wird ein Oszillator mit der Wien-Brücke aufgebaut, darf der Verstärker zwischen Aus- und Eingangsspannung keine Phasenverschiebung haben. Es müssen also zwei Emitterschaltungen in Reihe verwendet werden. Bei Verwendung eines Operationsverstärkers muss die Rückkopplung auf den nichtinvertierenden Eingang erfolgen. Die Schaltung nach Abb. 41.4 zeigt einen Wien-Brücken-Oszillator mit OP, bei dem die rückgekoppelte Spannung u_r auf den nichtinvertierenden Eingang gelegt ist. Die Ausgangsspannung u_a der Schaltung ist hier die Eingangsspannung der Wien-Brücke.

Die Verstärkung V_u des OP kann mit den Widerständen R1 und R2 eingestellt werden und muss, da der Kopplungsfaktor nach (41.6) bekannt ist, zum Anschwingen größer als 3 sein.

Mit $R_2 = 22\,\text{k}\Omega$ und $R_1 = 10\,\text{k}\Omega$ ergibt sich eine Verstärkung $V_u = 3{,}2$, die zum Anschwingen ausreicht. Die gegeneinander geschalteten Z-Dioden parallel zum Widerstand R2 dienen hier zur Amplitudenbegrenzung, führen also die Verstärkung auf $V_u = 3$ zurück.

Bessere Eigenschaften als die Wien-Brücke zeigt ihre Weiterentwicklung zur Wien-Robinson-Brücke als Rückkopplungsnetzwerk nach Abb. 41.5. Der Wien-Brücke ist ein fester, frequenzunabhängiger ohmscher Spannungsteiler parallel ge-

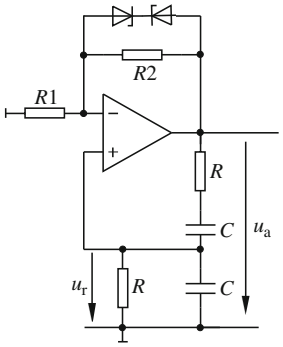


Abb. 41.4 Wien-Brücken-Oszillator mit OP

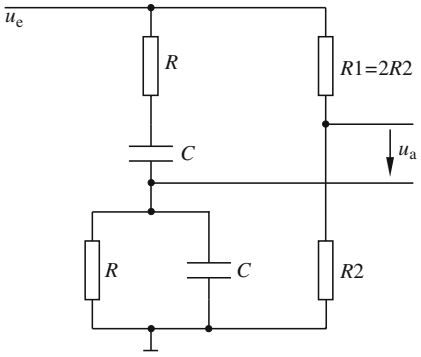


Abb. 41.5 Wien-Robinson-Brücke als Rückkopplungsnetzwerk

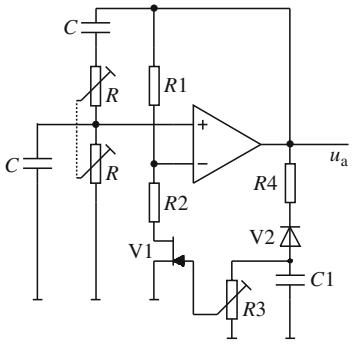


Abb. 41.6 Wien-Robinson-Oszillator mit OP

schaltet, dessen Widerstände R1 und R2 im Verhältnis 2 : 1 stehen. Im Resonanzfalle ist die Ausgangsspannung dieser Brücke $u_a = 0\,\text{V}$. Das Verhalten der Wien-Robinson-Brücke kommt dem eines idealen Schwingkreises wesentlich näher.

Die Wien-Robinson-Brücke muss etwas verstimmt betrieben werden, da ihre Ausgangsspannung im Resonanzfall gleich Null ist, also auch die Rückkopplungsspannung gleich Null wäre. Die Verstärkerstufe muss eine relativ große Verstärkung haben. Die Schaltung nach Abb. 41.6 zeigt einen einfachen Wien-Robinson-Oszillator.

Der Widerstand R2 nach Abb. 41.5 ist hier ersetzt durch die Reihenschaltung des Widerstandes R2 mit dem FET V1. Das Netzwerk aus Diode V2, Kondensator C1 und Poti R3 bewirkt eine Stabilisierung der Ausgangsamplitude.

Die Frequenz des Oszillators lässt sich durch Verstellen der beiden Widerstände R verändern, die allerdings Doppelpotentiometer mit guten Gleichlaufeigenschaften sein müssen.

41.3 Funktionsgeneratoren

Als Funktionsgeneratoren werden Signalgeneratoren bezeichnet, die mindestens zwei verschiedene Ausgangsspannungen mit jeweils unterschiedlichem Verlauf erzeugen. Kompakte Geräte dieser Art werden in breit gefächerten Qualitäts- und

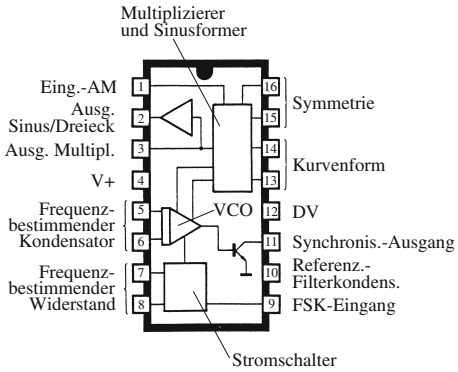


Abb. 41.7 Pin-Belegung des integrierten Funktionsgenerators XR 2206

Ausstattungskategorien hergestellt. Die Signale können in den verschiedensten Frequenzbereichen, mit unterschiedlichen Amplituden und in variablen Anstiegs- oder Abfallzeiten bei den Sägezahn- und Dreiecksignalen Verwendung finden.

Eine Dreieckspannung kann mit Hilfe eines Operationsverstärkers erzeugt werden, der als Integrierer geschaltet ist und eine Rechteckspannung als Eingangsspannung hat.

Sägezahngeneratoren erzeugen eine sägezahnförmige Ausgangsspannung. Sie kann auf vielfältige Art erzeugt werden. Die Ladekurve von Kondensatoren, die mit einer Stromquelle, mit einem Unijunction-Transistor UJT oder mit einem OP geladen werden, können zu diesem Zweck benutzt werden.

Rechteckgeneratoren können mit monostabilen Kippstufen, mit Schmitt-Triggern oder mit Operationsverstärkern hergestellt werden.

Wird ein Funktionsgenerator gesucht, der alle diese Signalspannungen erzeugen soll, können integrierte Bauelemente wie zum Beispiel der IC 8038 oder der XR 2206 verwendet werden.

Der XR 2206 ist ein monolithisch integrierter Funktionsgenerator, der Sinus-, Rechteck-, Dreieck-, Rampen- und Pulsfrequenzen von hoher Stabilität liefert. Die Ausgangssignale können durch eine externe Spannung sowohl amplituden- als auch frequenzmoduliert werden. Allein durch die äußere Beschaltung kann die Arbeitsfrequenz im Bereich von 0,01 Hz bis über 1 MHz gewählt werden. Die Pin-Belegung zeigt Abb. 41.7.

Die Frequenz f_0 der Schwingungen hängt von dem Kondensator C_T zwischen den Pins 5 und 6 und dem Widerstand R_T an Pin 7 beziehungsweise Pin 8 ab. Für Widerstandswerte von 4 bis 200 kΩ ist die Temperaturstabilität optimal; für den Kondensator werden Werte zwischen 1 nF und 100 mF empfohlen.

Die Schaltung nach Abb. 41.8 zeigt die Standardbeschaltung zur Erzeugung von Sinussignalen, wobei der Schalter S1 geschlossen sein muss. Das Poti an Pin 7 dient der Frequenzeinstellung. Wird der Schalter S1 geöffnet, werden dreieckförmige Signale erzeugt. Ihre Amplitude ist etwa doppelt so groß wie die der sinusförmigen Spannung.

Der XR 2206 kann nach Abb. 41.9 mit zwei verschiedenen Widerständen an den Pins 7 und 8 betrieben werden. Je nach dem

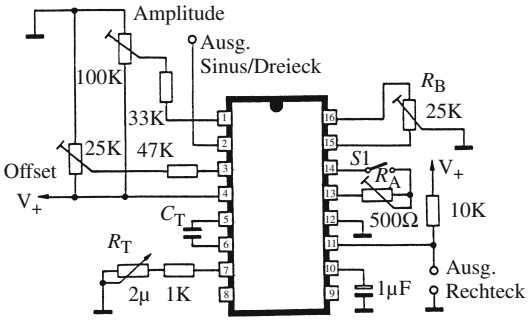


Abb. 41.8 Schaltung zur Erzeugung von Sinussignalen

Pegel des logischen Signals an Pin 9 ist entweder der Widerstand R1 oder der Widerstand R2 wirksam.

Die Schaltung nach Abb. 41.10 erzeugt Impuls- und Sägezahnspannungen. Impulsbreite und Tastverhältnis beziehungsweise Anstiegs- und Abfallzeit können durch Wahl der Widerstände R1 und R2 zwischen 1 und 99 % gewählt werden. Ihre Werte sollten zwischen 1 kΩ und 2 MΩ liegen.

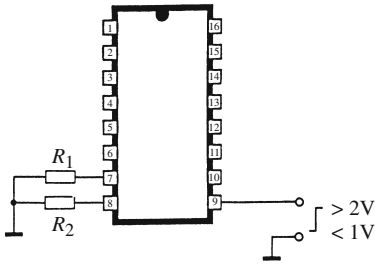


Abb. 41.9 Widerstandsauswahl durch Pegelbeschaltung

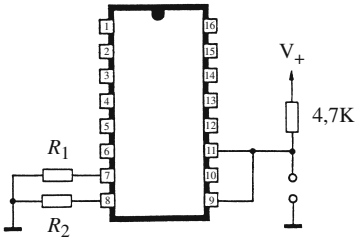
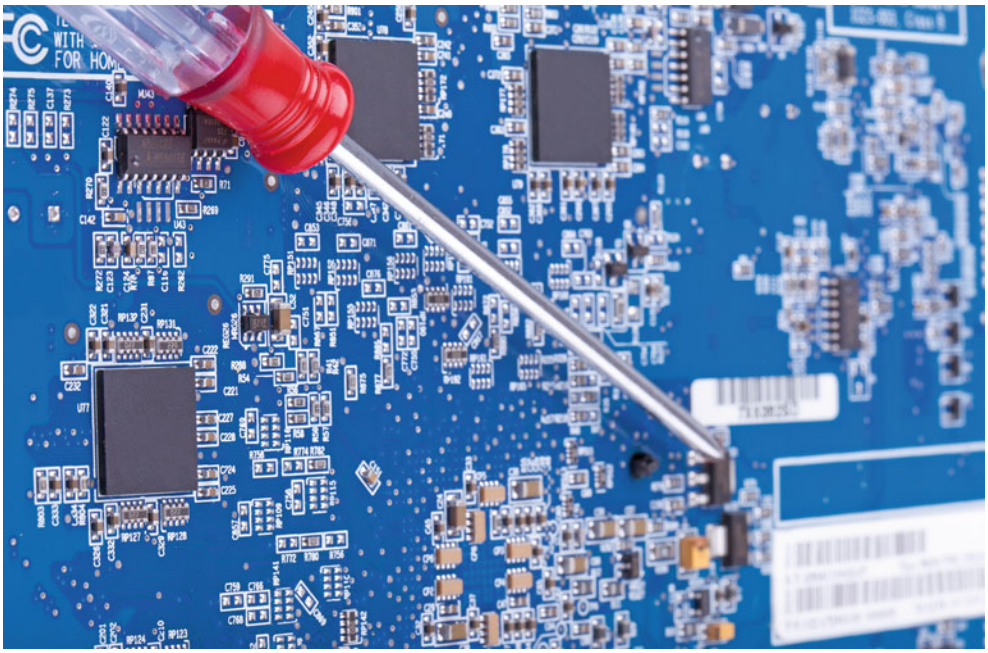


Abb. 41.10 Impuls- und Sägezahnspannungserzeugung

Schaltungstechniken

42

Peter Döring



42.1 Integrierte Schaltungen 520
42.2 SMD-Technik 522

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_42

42.1 Integrierte Schaltungen

Integrierte Schaltungen (IC = **i**ntegrated **c**ircuit) werden in weiter zunehmender Zahl auf den Markt gebracht. Niedrige Herstellungskosten bei großen Stückzahlen, hohe Zuverlässigkeit, geringer Platzbedarf und hohe Arbeitsgeschwindigkeiten sind die wesentlichen Vorteile der integrierten Schaltungen im Vergleich zu Schaltungen mit diskreten Bauelementen.

Die Bauelemente eines IC und ihre Verdrahtung werden in einem gemeinsamen Fertigungsprozess auf einem *einkristallinen Halbleiterplättchen* aus Silizium (Chip) hergestellt. Diese Chips nach Abb. 42.1 sind nur wenige Quadratmillimeter groß und werden in ein Gehäuse eingebaut, wobei die Chipanschlusskontakte durch sehr dünne Drähte mit den „Pins“ des Gehäuses verbunden werden. Aus Kostengründen werden möglichst viele Chips auf einer Siliziumscheibe (wafer) untergebracht.

Die Komplexität der Schaltungen auf den Chips steigt stetig mit weiter voranschreitender Integrationsdichte, also der Anzahl der Transistorfunktionen je Chip. Möglich wird das durch die bereits in Abschn. 33.4 beschriebene *Planartechnik*.

Bei den *monolythisch integrierten Schaltungen* sind alle zum Funktionieren erforderlichen Bauelemente der Schaltung, die auf einem konventionellen Vorbild mit diskreten Bauelementen beruht, in einen Halbleiterkristall hineindotiert, auch Bauelemente wie Kondensatoren, Dioden, Transistoren oder Widerstände nach Abb. 42.2. Neben bipolaren Transistoren werden auch MOS-FET in Planartechnik (Abschn. 35.2) in monolythisch integrierten Schaltungen eingebracht. Im Kristall müssen die Einzelbauelemente durch Isolierschichten voneinander getrennt sein.

Überwiegend werden npn-Transistoren in den IC realisiert, da sie bessere elektrische Eigenschaften haben und einfacher herstellbar sind als pnp-Transistoren. Die besondere Anordnung der einzelnen dotierten Bereiche ergibt sich aus technologischen Notwendigkeiten. Die Kontakte können nur auf der Siliziumoberfläche angebracht werden, und die Transistoren sind von p-dotierten Gebieten zur gegenseitigen Isolierung umgeben, so dass sich immer in Sperrrichtung liegende pn-Übergänge ergeben.

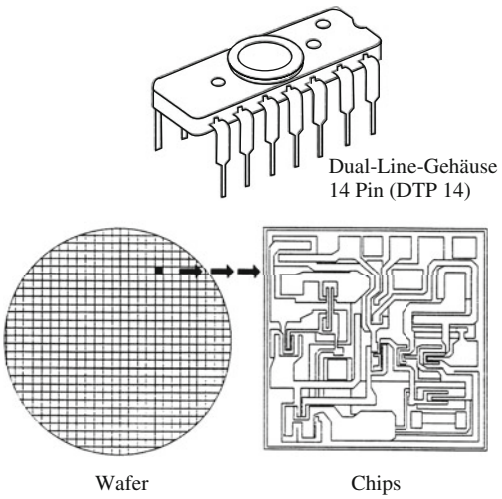


Abb. 42.1 Wafer, Chip und Bauform eines IC

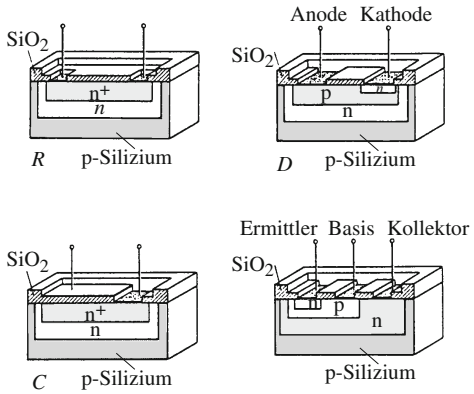


Abb. 42.2 Schichtenfolge der Bauelemente

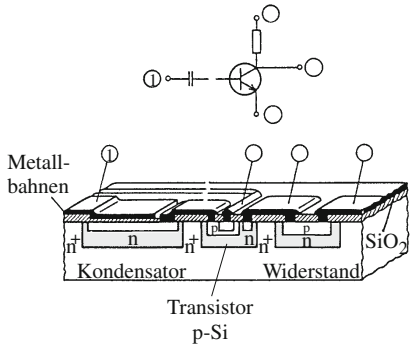


Abb. 42.3 Monolythischer IC und konventionelles Vorbild

Abb. 42.3 zeigt eine einfache Schaltung mit diskreten Bauelementen und ihre Realisierung in monolythisch integrierter Technik.

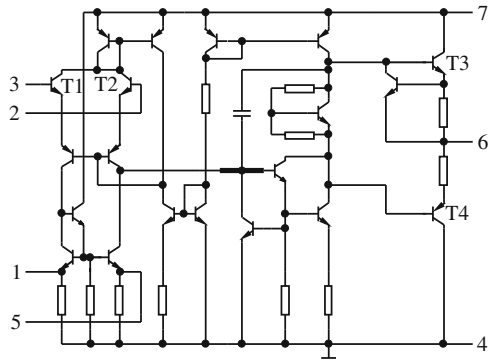
Als *Isolationsschicht* zwischen der Siliziumoberfläche und den Leiterbahnen dient eine SiO_2-Schicht, die durch thermische Oxydation der Siliziumoberfläche erzeugt wird. Weitere Leiterbahnschichten können auf ähnliche Weise darübergelegt werden, während Löcher in der SiO_2-Schicht zum Anschluss von Kontakten eingebracht werden. Die Leiterbahnen bestehen im Wesentlichen aus Aluminium.

Auch bei den integrierten Schaltungen unterscheidet man zwischen *Bipolarschaltungen* und *MOS- Schaltungen* und ferner zwischen *integrierten Analog- und Digitalschaltungen*. Die Herstellungsverfahren sind zwar weitgehend identisch, aber während man bei Analogschaltungen mit einigen hundert Einzelbauelementen auf einem Chip auskommt, werden bei Digitalschaltungen einige hunderttausend Einzelbauelemente benötigt. Die Frage, ob *bipolare* oder *unipolare Systeme* günstiger sind, lässt sich nur über Nebenbedingungen klären, wie zum Beispiel die Integrationsdichte und die beabsichtigte Anwendung (Tab. 42.1).

Integrierte Schaltkreise in *Bipolar-Technik* kommen als *Analog- und Digitalschaltungen* zur Anwendung, im ersten Fall als Ver-

Tab. 42.1 Integrationsgrade (Fa. Siemens)

Integrationsgrad	Anzahl der Funktionen	Mittlere Anzahl der Transistoren	System	Mittlerer Flächenbedarf in mm²
SSI (small scale integration)	2–20	100	Bipolar digital + analog	3
MSI (medium scale integration)	20–100	500	Bipolar digital + analog unipolar digital MOS-IC	8
LSI (large scale integration)	100–50.000	< 100.000	Bipolar digital + analog unipolar digital MOS-IC	20
VLSI (very large scale integration)	> 50.000	> 100.000	Unipolar digital CMOS-IC	>30



Innenschaltungeines TL x741x

Abb. 42.4 Innenbeschaltung eines Operationsverstärkers

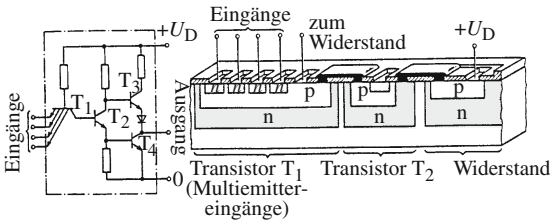


Abb. 42.5 NAND-Gatter in TTL-Technik (Fa. Siemens)

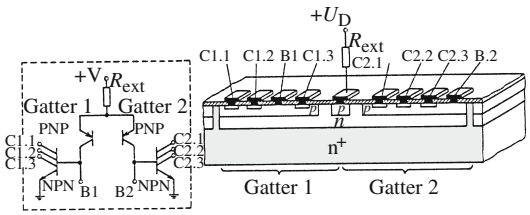


Abb. 42.6 NAND-Gatter in I²L-Technik (Fa. Siemens)

stärker (siehe Kap. 37) und im zweiten Fall als Schalter (siehe Kap. 40). Als Vorteile gegenüber der MOS-Technik ist hier die Fähigkeit zu nennen, große Ströme zu treiben und die kurze Schaltzeit beziehungsweise die hohe Grenzfrequenz der bipolaren Transistoren.

Integrierte monolythische Schaltungen für analoge Anwendungen enthalten überwiegend widerstandsgekoppelte Verstärker ohne Koppelkondensatoren. Analogschaltungen setzen sich meist aus verhältnismäßig einfachen Grundschaltungen zusammen, wie sie in Kap. 37 besprochen werden. Differenzverstärker werden aufgrund der Temperaturkompensation häufig verwendet. Die tatsächliche Innenbeschaltung ist für den Anwender nicht bedeutsam, jedoch wird der innere Funktionsablauf und die Wirkungsweise der Gesamtschaltung durch Blockschaltbilder beschrieben. Die Komplexität der Innenbeschaltung eines einfachen Operationsverstärkers (Kap. 39) zeigt die Schaltung nach Abb. 42.4.

Andere integrierte Analogschaltungen dienen Regelungszwecken, in der Energieelektronik zum Zünden von Thyristoren und Triacs oder sind für zahlreiche Anwendungen in Rundfunk- und Fernsehgeräten entworfen worden.

Spannungsregler (TDB 78xx), Zeitgeberschaltungen (TDB 555), Drehzahlregler (TCA 955) und viele mehr sind in der verfügbaren Bestellpalette enthalten.

Die integrierten Digitalschaltungen unterscheidet man nach der Ausführungsform der Grundgatter und spricht von *Logikfamilien*. Von den wichtigsten bipolaren Familien (TTL = transistor transistor logic; ECL = emitter coupled logic; I²L = integrated injection logic) werden zwei kurz behandelt (siehe Teil VII).

Abb. 42.5 stellt ein NAND-Gatter in TTL-Ausführung dar. Die Eingänge des Gatters sind als mehrere Emitter eines Transistors in der integrierten Schaltung erkennbar. Die einzelnen Bauelemente sind durch p-dotiertes Silizium voneinander getrennt.

Eine Weiterentwicklung dieser TTL-Technik zu höherer Integrationsdichte und verbesserten Schaltzeiten erbrachte die I²L-Technik. Über die zusätzlich integrierten *Komplementärtransistoren* nach Abb. 42.6 bekommen die eigentlichen Transistoren einen konstanten Strom, so dass die Schaltung frei von größeren Stromschwankungen bleibt. Die rasche Änderung des differentiellen Innenwiderstandes der Konstantstromquelle ergibt höhere Arbeitsgeschwindigkeiten. Wesentlichster Vorteil der I²L-Technik ist jedoch, dass ungefähr 11mal soviele Gatter auf einer bestimmten Fläche integriert werden können als bei der TTL-Technik.

Integrierte Schaltkreise in MOS-Technik bringen den großen Vorteil mit, dass ein MOS-Transistor fast leistungslos gesteuert wird (siehe Abschn. 35.2). Der Stromkanal mit den Source- und Drain-Gebieten ist von den umgebenden p-Gebieten elektrisch isoliert. Da dieses auch beim Gate der Fall ist, stellt ein MOS-FET ein selbstisolierendes Bauelement dar, das nicht durch zusätzliche Maßnahmen im Chip von anderen Bauelementen isoliert werden muss. Dieser und weitere Effekte sorgen dafür, das MOS-Schaltungen erheblich weniger Platz beanspruchen als gleichartige bipolare Schaltkreise. Die CMOS-Technik (C = complementary) ist heute die dominierende Technik bei

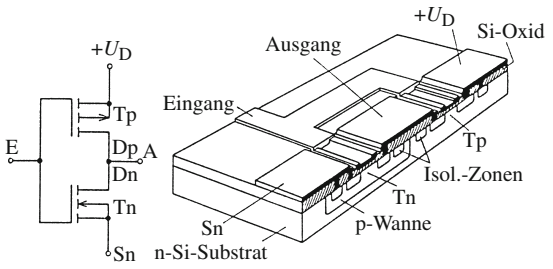


Abb. 42.7 Inverter in CMOS-Technik (Fa. Siemens)

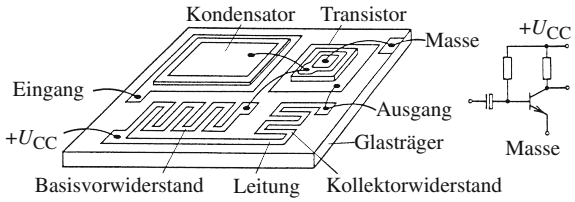


Abb. 42.8 Hybrid-Dünnschicht-Schaltkreis und realisierte Schaltung

den integrierten digitalen Schaltkreisen, weil sie leistungsarm ist und sehr hohe Integrationsdichten ermöglicht. Im Abb. 42.7 ist erkennbar, dass eine Aluminiumschicht den gesamten Chip überzieht. Da direkte Kontaktierungen in den meisten Fällen nicht notwendig sind (isoliertes Gate), kann auf ausgeprägte Leiterbahnen verzichtet werden. Die Schaltung in Abb. 42.7 zeigt einen Inverter in CMOS-Technik.

Die Aufteilung von integrierten Schaltungen in Standard-Schaltungen und in kundenspezifische Schaltungen ist fließend. Als Standard-Schaltungen werden IC bezeichnet, die jedem Anwender zugänglich sind und ohne weitere Abänderung verwendet werden können.

Kundenspezifische Schaltungen sind IC, die im Auftrag von Anwendern auf der Grundlage spezieller technischer Vorgaben und Konzepte realisiert werden.

Integrierte Schaltkreise in *Hybrid-Technik* (hybrid = gemischt) werden bei geringen Stückzahlen für kundenspezifische Schaltungen eingesetzt. Wenn eine Schaltung zu viele passive Schaltelemente (Kondensatoren) enthält, so dass eine monolythisch integrierte Schaltung nicht angezeigt erscheint, setzt man diese Technologie ebenfalls ein. Dabei werden die passiven Bauelemente auf eine isolierende Trägerplatte aus Glas oder Keramik aufgebracht. Sie existieren als Schicht auf der Oberfläche, Widerstände, zum Beispiel in Mäanderform. Transistoren und integrierte Halbleiterschaltkreise werden in Form von Chips auf die Trägerplatte gebracht und durch dünne Gold- und Silberdrähte mit den anderen Teilen der Schaltung verbunden. Bei der Gesamtherstellung ist zwischen der Dickschicht- und Dünnschichttechnologie zu unterscheiden. Erstere benutzt man für kleinere Serien, deren Schaltungsfunktionen flexibel sein sollen, letztere für große Stückzahlen mit festen Schaltungsfunktionen. Dünnschichtschaltungen lassen sich für fast alle Anwendungsbereiche bis hin zur Höchstfrequenztechnik herstellen. Abb. 42.8 zeigt als Beispiel einen Hybrid-Dünnschicht-Schaltkreis, der die dargestellte konventionelle Schaltung nachbildet.

42.2 SMD-Technik

Auch bei den Schaltkreisen mit diskreten Bauelementen geht der Trend zu kleineren Schaltungsaufbauten mit höherer Bestückungsdichte und zur Automatisierung der Fertigung. In gedanklicher Fortentwicklung der Hybrid-Technologie kommt man von der Einsteck- zur Oberflächenmontage. Dazu mussten neue Gehäuseformen entwickelt werden.

Oberflächenmontierbare Bauelemente nennt man SMD (**s**urface **m**ounted **d**evices). Die Fertigungstechniken zur Verarbeitung solcher Bauelemente nennt man SMT (**s**urface **m**ounted **t**echnology).

Die oberflächenmontierbaren Bauelemente werden auf der gleichen Seite der Platine aufgebracht, auf der sich die Leiterbahnen befinden. Deshalb lassen sie sich auf beiden Platinenseiten montieren (Abb. 42.9), was bei der Einsteckmontage nur mit großem Aufwand zu bewerkstelligen ist. Die Bauelemente haben keine oder nur sehr kurze Anschlussdrähte und sind in ihren Abmessungen relativ klein. Die Platine muss nicht mit Bohrlöchern für die Bestückung versehen werden. Es sind alle Leiterplattenmaterialien geeignet; sowohl Keramiksubstrate als auch flexible Träger werden eingesetzt. Die Leiterplatten können verkleinert werden, und es ergeben sich verbesserte Hochfrequenzeigenschaften.

Als Gesamtvorteile werden genannt: Miniaturisierung und Rationalisierung bei höherer und gleichmäßigerer Qualität sowie Zuverlässigkeit.

Allerdings werden sehr hohe Anforderungen an die Positioniergenauigkeit und Zuverlässigkeit der Bestückungsautomaten und die Löt-(Klebe-)verfahren gestellt.

Die neuentwickelten Gehäuseformen müssen den Anforderungen hinsichtlich geringer Abmessungen, geeigneter Lötanschlüsse, hoher Lötwärmebeständigkeit, automatengerechter Verpackung und automatischer Bestückbarkeit gerecht werden. Abb. 42.10 zeigt einige bei der Fa. Siemens gefertigte SMD-Gehäuseformen. Nahezu alle diskreten aktiven und passiven Bauelemente sowie zahlreiche integrierte Schaltungen gibt es mittlerweile in SMD-Gehäusen.

Als Verpackungsart setzt sich zunehmend der Gurt gegenüber dem Magazin durch. Einige Magazinformen zeigt Abb. 42.11.

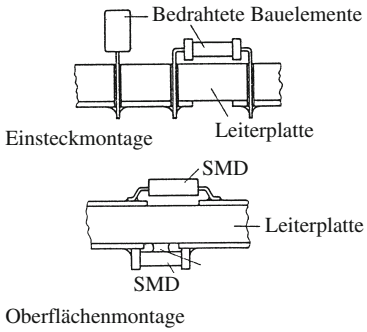


Abb. 42.9 Einsteck- und Oberflächenmontage



Abb. 42.10 Gebräuchliche SMD-Gehäuse

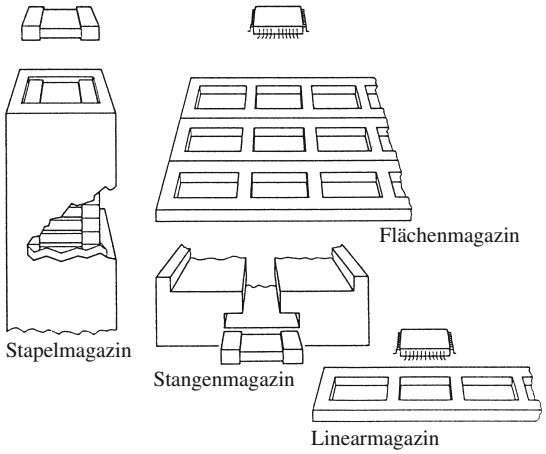


Abb. 42.11 Gebräuchliche Magazintypen

Für die SMD-Technik mussten neue Layoutregeln entwickelt werden, in denen berücksichtigt wird, wie die Platine gefertigt, geprüft, repariert und gewartet werden soll.

Bei den geringen Abmessungen der Bauteile und der großen Packungsdichte wirken sich Löt- und Klebefehler oder Unregelmäßigkeiten im Lötvorgang besonders gravierend aus. Hier müssen Lötverfahren und Klebeverfahren sorgfältig ausgesucht werden.

Da zum Beispiel beim Wellenlöten die Platine mit dem Bauteil kopfunter liegt, muss der Klebepunkt das Bauteil sicher fixieren. Liegt zuviel Kleber unter dem Bauteil, verschiebt es sich eventuell. Liegt zuwenig Klebemittel an, hält das Bauteil beim Lötvorgang die Position nicht ein, zumal die SMD-Bauteile vom Lötmittel voll umflutet werden und der hohen Temperatur des Bades ausgesetzt sind. Es muss also nach dem Grundsatz „so wenig wie möglich, aber so viel wie nötig“ verfahren werden.

Bei Verwendung von Keramiksubstraten mit einem Reflow-Lötvorgang kann auf Kleber ganz verzichtet werden. Hier wird zunächst Lötpaste auf die Lötflächen aufgebracht, anschließend werden die Lötanschlüsse der Bauelemente in die Paste gedrückt und schließlich die Paste durch Wärmeeinwirkung aufgeschmolzen.

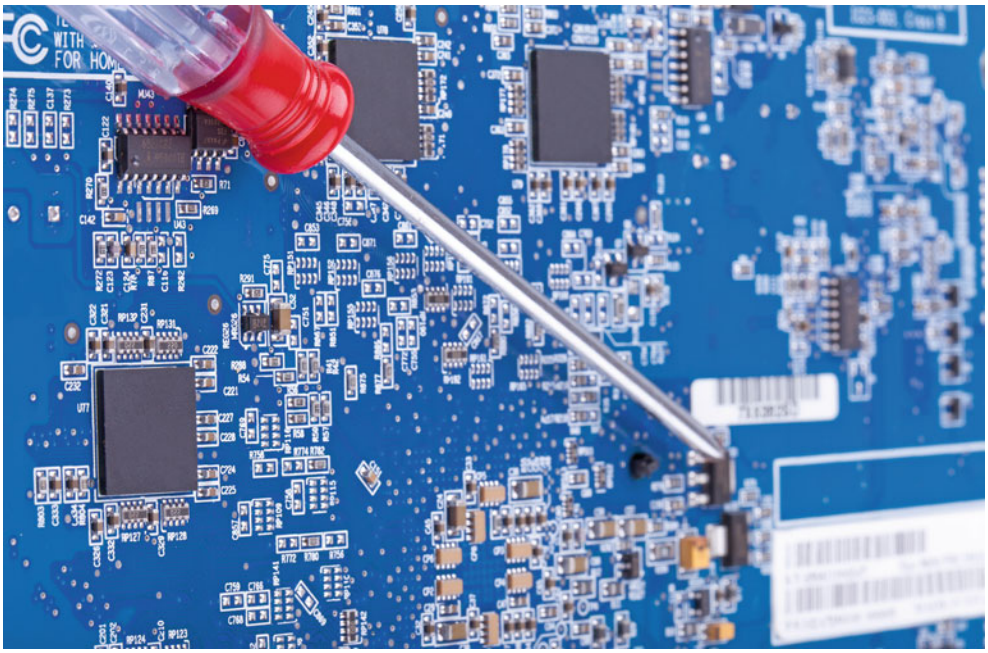
Bestückungsautomaten haben die Probleme der Kleberdosierung und der Lötungen praktisch gelöst. Zusätzliche Anforderungen an den Bestückungsautomaten müssen von Fall zu Fall entschieden und gelöst werden. So sollten möglichst alle Bauelemente mit einem Bestückkopf bestückbar und eine genaue Platzierung der Bauelemente gewährleistet sein. Auch sollte der Bestückungsautomat keine Auswirkung auf das Layout haben.

Welche Bestückungsautomaten eingesetzt werden, hängt vor allem von der Zahl der zu bestückenden Platinen und der Häufigkeit des Platinenwechsels ab. Bei der Pick-and-Place-Methode wird ein Bauelement nach dem anderen aufgesetzt, bei der Simultan-Methode dagegen werden mehrere Bauelemente gleichzeitig erfasst und auf der Platine gleichzeitig abgesetzt, was hohe Stückzahlen ermöglicht. Bei einem Platinenwechsel werden dann aber hohe Rüstzeiten erforderlich.

Optoelektronik

43

Peter Döring



43.1 Grundsätzliche Überlegungen 526
43.2 Optoelektronische Bauelemente 526
43.3 Anzeigeeinheiten 533
43.4 Signalübertragung mit Optokoppler 535
43.5 Faseroptische Übertragungsmittel 535

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_43

43.1 Grundsätzliche Überlegungen

Alle Halbleitermaterialien werden bei Energiezufuhr von außen in Form von Wärme oder Licht niederohmiger, da neue Ladungsträgerpaare gebildet werden, die die Eigenleitfähigkeit erhöhen. Die Zahl der durch das auftreffende Licht freigesetzten Elektronen wird umso größer, je größer die Beleuchtungsstärke ist, weil erhöhte Lichteinstrahlung eine Energiezufuhr bedeutet. Dieser Vorgang wird als „Innerer foto- elektrischer Effekt“ bezeichnet.

Trifft Licht auf eine pn-Sperrschicht, werden infolge der Energiezufuhr Kristallbindungen aufgerissen. Es entstehen bewegliche Ladungsträger-Paare, die in- folge des vorhandenen elektrischen Feldes sofort abfließen. Dabei wandern die Löcher in die p-Schicht und die Elektronen in die n-Schicht. Ohne angelegte äußere Spannung wird die p-Schicht zum Pluspol und die n-Schicht zum Minuspol einer Spannungsquelle.

Licht ist physikalisch gesehen eine elektromagnetische Strahlung in einem bestimmten Frequenzbereich. Dabei wird sichtbares Licht von nichtsichtbarem Licht unterschieden. Das natürliche weiße Sonnenlicht ist eine Mischung von elektromagnetischen Schwingungen der verschiedensten Wellenlängen. Wellenlänge und Frequenz sind zueinander umgekehrt proportional. Die Energie der Lichtstrahlung ist der Frequenz proportional.

Die Farbanteile des Lichtes werden als Spektralfarben bezeichnet. Das unsichtbare Infrarotlicht (IR) hat eine Wellenlänge λ von 780 bis etwa 1000 nm. Rotes, sichtbares Licht (780 bis 630 nm) schließt sich an, geht in orangefarbenes Licht (630 bis 590 nm) über, während gelbes Licht (590 bis 560 nm) den farblichen Übergang zum Grünbereich (560 bis 490 nm) bildet. Über den Anteil an blauem Licht (490 bis 440 nm) und violetter Spektralfarbe (440 bis 380 nm) endet der sichtbare Teil und geht in den nichtsichtbaren Teil (380 bis 10 nm), dem ultravioletten Licht (UV) über.

43.2 Optoelektronische Bauelemente

Bauelemente zur Umwandlung elektrischer Größen in optische Strahlung und umgekehrt werden als *optoelektronische Bauelemente* bezeichnet. Man unterscheidet dabei zwischen lichtemittierenden (lichtabstrahlenden) und lichtabsorbierenden (lichtaufnehmenden) Bauelementen.

Bei einigen, im sichtbaren Licht arbeitenden Fotohalbleitern führt auch die vom Auge nicht wahrgenommene infrarote Strahlung zu einer Änderung der Leitfähigkeit. Hier werden für bestimmte Anwendungen optische Filter verwendet, um deren Einflüsse gering zu halten.

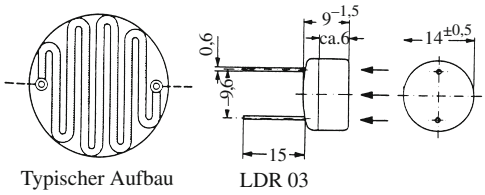


Abb. 43.1 Typischer Aufbau und Bauform eines LDR

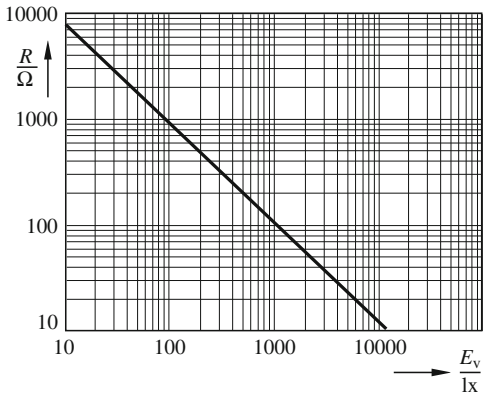


Abb. 43.2 Kennlinie des LDR 03

43.2.1 Fotowiderstand (LDR – light dependent resistor)

Fotowiderstände bestehen aus Halbleiter-Mischkristallen als Basismaterial. LDR können sowohl an Gleichspannung wie auch an Wechselspannung betrieben werden, da sie ohne pn-Sperrschicht sind. Der Widerstandswert von Fotowiderständen wird mit zunehmender Beleuchtungsstärke kleiner. Ein LDR hat bei einer bestimmten Lichtwellenlänge seine größte Empfindlichkeit.

Für die Herstellung von LDR, deren spektrale Empfindlichkeit im Bereich des sichtbaren Lichtes liegt, werden als Halbleitermaterialien Cadmiumsulfid (CdS) und Cadmiumselenid (CdSe) verwendet.

Die spektrale Empfindlichkeit von LDR aus Bleisulfid (PbS) und Indiumantimonid (InSb) liegt dagegen im Infrarotbereich; die spektrale Empfindlichkeit von Germanium und Silizium liegt zwischen 800 und 1600 nm.

Fotowiderstände haben die höchste Lichtempfindlichkeit unter den fotoelektronischen Halbleiterbauelementen. Abb. 43.1 zeigt den typischen mäanderförmigen Aufbau eines LDR.

Die wichtigsten Kennwerte von LDR sind der Dunkelwiderstand und der Hellwiderstand, der in den Datenblättern meistens für eine Beleuchtungsstärke von 100 Lux angegeben wird.

Die Kennlinie des LDR 03 zeigt Abb. 43.2. Seine maximale Versorgungsspannung beträgt $U_B = 150\,\text{V}$ und seine maximale Verlustleistung $P_{\text{tot}} = 100\,\text{mW}$.

Auf Erfahrung bauen.
Innovationskraft nutzen.
Sensorik voranbringen.

Optimale Lösungen für Ihre Anforderungen: Mit den industriellen Sensoren und Systemen von Pepperl+Fuchs profitieren Sie von modernsten Technologien und zukunftsweisenden Innovationen – insbesondere auf dem Weg zu vollständig vernetzten Produktionsprozessen für die Anwendungen der Zukunft.

www.pepperl-fuchs.com



Your automation, our passion.

PEPPERL+FUCHS

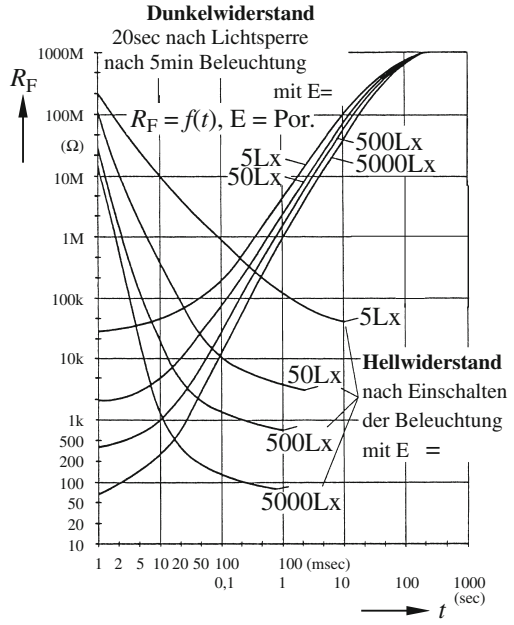


Abb. 43.3 Einstellträgheit eines Silizium-LDR

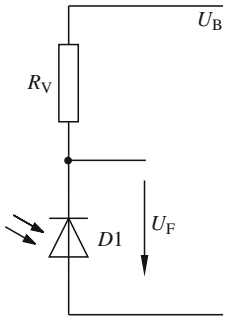


Abb. 43.4 Fotodiode mit Vorwiderstand

Dunkelwiderstand R_0 = Widerstandswert nach 1 Minute völliger Abdunkelung; $R_0 > 10\,\text{M}\Omega$.

Hellwiderstand R_H = Widerstandswert bei 100 Lux oder 1000 Lux; $R_{H100} = 500\,\Omega \ldots 50\,\text{k}\Omega$.

Beim praktischen Einsatz von Fotowiderständen muss beachtet werden, dass der Widerstandswert einer Änderung der Beleuchtungsstärke mit einer relativ hohen Trägheit folgt. Abb. 43.3 zeigt prinzipiell die Einstellträgheit eines Silizium-LDR.

Fotowiderstände sind daher nicht besonders gut für einen Einsatz geeignet, bei dem schnelle Änderungen der Beleuchtungsstärke erfasst werden müssen. Bei $E = 50\,\text{Lx}$ benötigt dieser LDR eine Zeit von 100 ms, um seinen Widerstandswert R_F von $100\,\text{M}\Omega$ auf $10\,\text{k}\Omega$ zu mindern; bei $E = 500\,\text{Lx}$ noch eine Zeit von 15 ms.

Fotodioden werden in Sperrrichtung an einer äußeren Spannung betrieben; damit ist zu ihrem Betrieb nach Abb. 43.4 ein Vorwiderstand und eine Betriebsspannung erforderlich.

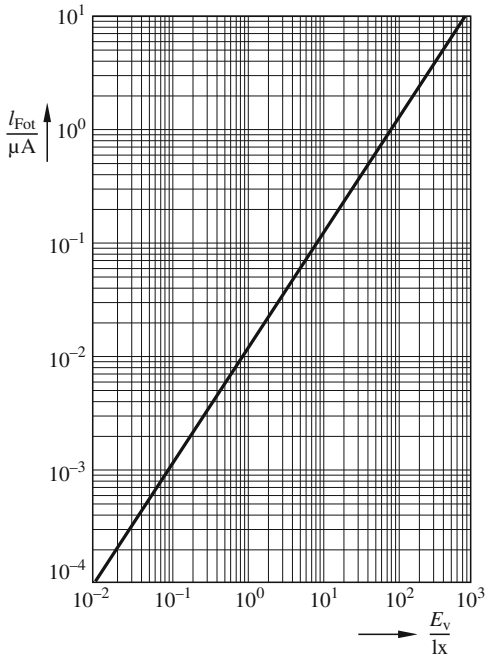


Abb. 43.5 Zusammenhang zwischen Fotostrom und Beleuchtungsstärke

43.2.2 Fotodiode und Fotoelement

Ohne Beleuchtung fließt durch die pn-Sperrschicht einer Fotodiode wie bei jeder normalen Diode ein Sperrstrom, der bei den Fotodioden meistens als Dunkelstrom I_{R0} bezeichnet wird. Als Folge der Beleuchtung tritt ein zusätzlicher Fotostrom I_{Fot} auf, der linear mit der Beleuchtungsstärke ansteigt. Abb. 43.5 zeigt den Zusammenhang zwischen Fotostrom und Beleuchtungsstärke.

Wie jedes Halbleiter-Bauelement hat auch die Fotodiode eine deutliche Temperaturabhängigkeit (Abb. 43.6). Besser zu erkennen ist diese Abhängigkeit im Abb. 43.7, deren Kennlinie das Verhältnis normiert darstellt. Bei einer Änderung der Temperatur kann der Korrekturfaktor abgelesen werden, mit dem der bezogene Fotostrom multipliziert wird.

Wie alle lichtempfindlichen Bauelemente haben auch die Fotodioden eine spektrale Empfindlichkeit. Für die Fotodiode BPW 32 ist die relative spektrale Empfindlichkeit in Abb. 43.8 dargestellt. Die größte Empfindlichkeit dieser Fotodiode liegt bei einer Wellenlänge von $\lambda \sim 800\,\text{nm}$, was etwa einer Farbe zwischen Dunkelrot und Infrarot entspricht.

Dies ist auch bei den meisten anderen Typen von Silizium-Fotodioden der Fall. Zu unterscheiden ist zwischen den pn-Fotodioden und den pin-Fotodioden.

Der großflächige pn-Übergang bei den Fotodioden hat eine große Sperrschichtkapazität zur Folge. Daher liegen die Schaltzeiten von pn-Fotodioden im Bereich von Mikrosekunden.

Um die Sperrschichtkapazität zu verkleinern, wurden die pin-Fotodioden entwickelt, deren Schaltzeiten im Nanosekunden-

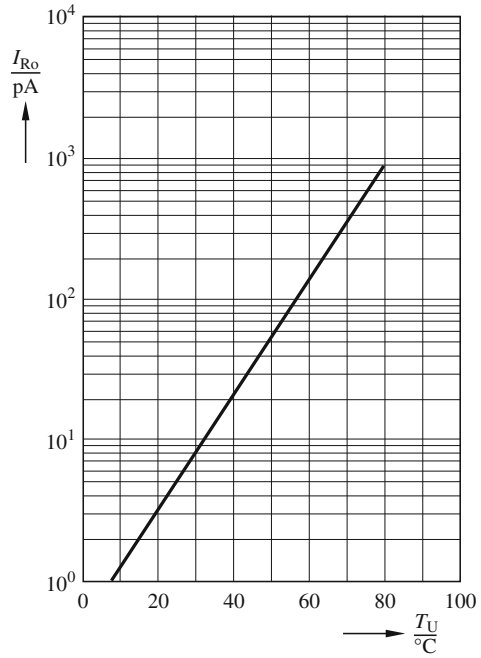


Abb. 43.6 Temperaturabhängigkeit einer Fotodiode

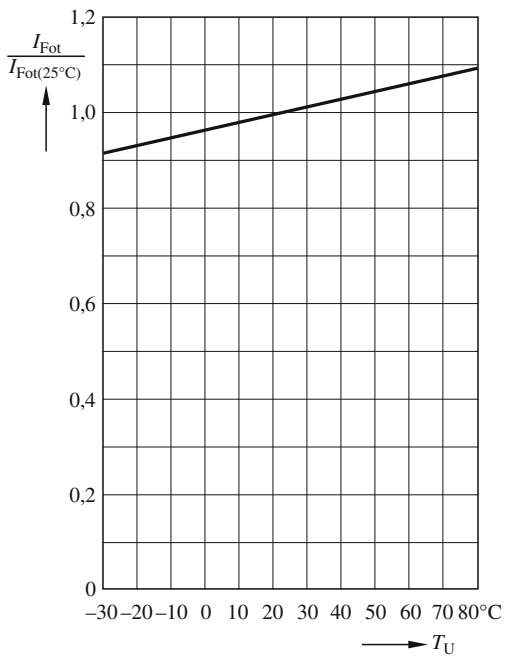


Abb. 43.7 Normierte Darstellung der Temperaturabhängigkeit einer Fotodiode

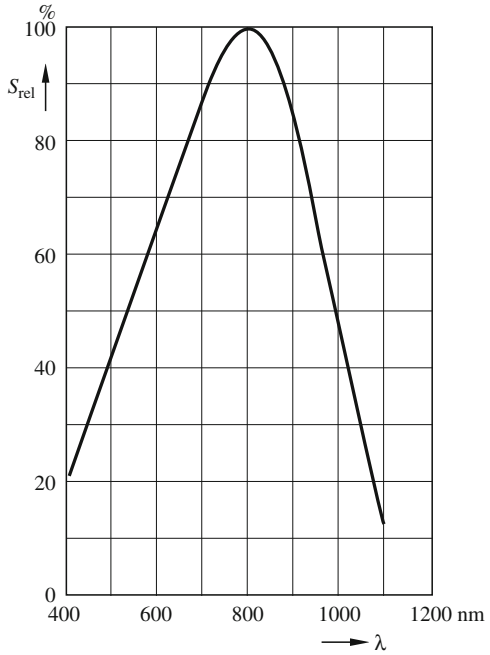


Abb. 43.8 Relative spektrale Empfindlichkeit der Fotodiode BPW 32

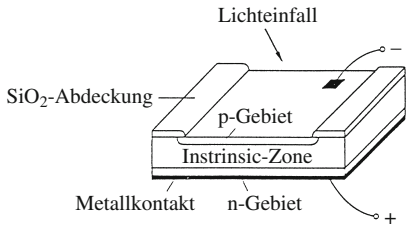


Abb. 43.9 Technologischer Aufbau einer pin-Fotodiode

Bereich liegen (Abschn. 33.4.4). Abb. 43.9 zeigt den technologischen Aufbau von pin-Fotodioden, bei denen sich zwischen den sehr dünnen p- und n-Schichten eine breite Intrinsic-Schicht befindet. Wegen der daraus resultierenden hohen Feldstärke in dieser Schicht werden die bei Beleuchtung erzeugten Ladungsträger-Paare viel schneller als bei den pn-Fotodioden in die p- beziehungsweise n-Schicht abgesaugt. Diese größere Beweglichkeit der Ladungsträger verbessert das Schaltverhalten. Infolge der relativ breiten i-Schicht haben pin-Fotodioden wesentlich höhere Sperrspannungen. Sie liegen bei $U_R \sim 50$ bis 100 V. Die Fotoströme sind bei den pin-Fotodioden nur etwa halb so groß wie bei den pn-Fotodioden.

In den letzten Jahren haben pin-Fotodioden eine steigende Bedeutung erlangt. Sie werden zum Beispiel aufgrund ihrer hohen Empfindlichkeit im Infrarotbereich und wegen ihrer kurzen Schaltzeiten bei der Fernsteuerung mit moduliertem Infrarotlicht eingesetzt.

Fotodioden können wegen des gleichartigen technologischen Aufbaues und des gleichen Funktionsprinzips auch als Fotoelemente betrieben werden. Sie arbeiten dann im IV. statt im III. Quadranten des Kennlinienfeldes (Abb. 43.10). Infolge der Ladungstrennung in der pn-Schicht durch das einfallende Licht und mit Unterstützung der Diffusionsspannung bildet sich eine Spannung in Durchlassrichtung, das heißt, es ist keine äußere Betriebsspannung notwendig.

Legt man einen Widerstand an die Kontaktierung der p- und n-Schichten, fließt ein Strom, der in Bezug auf die Spannung

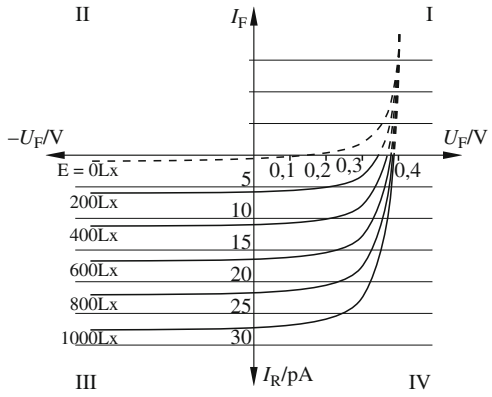


Abb. 43.10 Vier-Quadranten-Kennlinienfeld einer Fotodiode

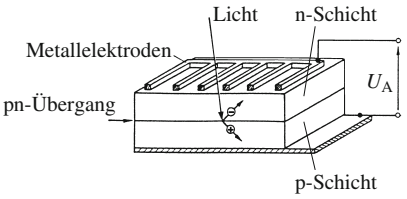


Abb. 43.11 Schnitt durch ein Fotoelement

negativ zu sehen ist. Damit hat man eine Spannungsquelle, die Licht direkt in elektrische Energie umwandelt (Fotoelement, Solarzelle). Fotoelement und Solarzelle unterscheiden sich nur dadurch, dass eine Solarzelle zur Erzeugung höherer Leistungen vorgesehen ist.

Eine typische Solarzelle ist 10 cm × 10 cm groß und besteht aus kristallinem Silizium. Der interne Aufbau und die Kontaktierung der Anschlüsse ist in Abb. 43.11 erkennbar. Die Oberfläche ist mit einer „Antireflex-Schicht„ (schwarzblaue Oberfläche) versehen, damit möglichst viel Licht eindringen kann. Zur Erzeugung höherer Leistungen werden solche Zellen parallel und in Reihe zusammengeschaltet.

Hundert solcher Solarzellen erbringen eine elektrische Leistung von 100 W unter der Bedingung, dass die Strahlungsleistung der Sonne 1000 W/m^2 beträgt. Der Wirkungsgrad liegt folglich bei 10 %. Zur Energieumwandlung trägt bei Verwendung von Silizium nicht nur das sichtbare Licht bei, sondern nach Abb. 43.12 auch Licht mit höherer Wellenlänge.

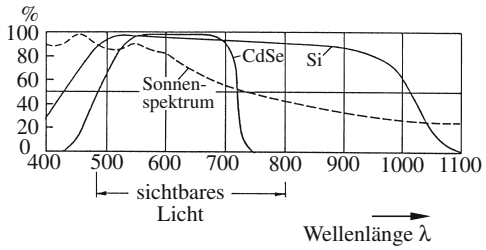


Abb. 43.12 Strahlungsspektrum des Sonnenlichtes und Spektralempfindlichkeit von Solarelementen

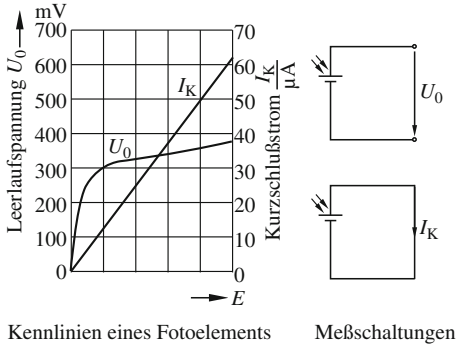


Abb. 43.13 Kennlinie und Messschaltung eines Fotoelementes

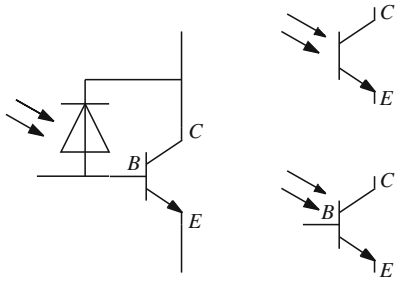


Abb. 43.14 Ersatzschaltbild und Schaltzeichen von Fototransistoren

Die Kennwerte von Fotoelementen, wie Leerlaufspannung U_0 und Kurzschlussstrom I_K lassen sich mit den Messschaltungen nach Abb. 43.13 ermitteln.

Solarzellen werden in zunehmendem Maße für die Energieversorgung ortsfester Verbraucher mit niedrigem Verbrauch, wie zum Beispiel Leuchtbojen, Sendeanlagen, Parkscheinautomaten in Großstädten, kleinere Wochenendhäuser, verwendet.

43.2.3 Fototransistoren

Wie jeder Transistor enthält auch der Fototransistor zwei pn-Übergänge, die lichtempfindlich sind, weshalb normale Transistoren in lichtundurchlässige Gehäuse gegossen werden. Hier wird die Kollektor-Basis-Sperrschicht als lichtempfindliche Schicht benutzt. Die Wirkungsweise und damit das Ersatzschaltbild nach Abb. 43.14 entspricht einer Fotodiode parallel zur CB-Strecke. Der durch die freigesetzten Ladungsträger hervorgerufene Strom wirkt wie ein Basisstrom. Die Lichtempfindlichkeit ist um den Verstärkungsfaktor B des Transistors größer als die der Fotodiode. Der Basisanschluss kann herausgeführt sein, was die Einstellung eines Arbeitspunktes erleichtert und die Steuermöglichkeiten vergrößert.

Das Ausgangskennlinienfeld des Transistors nach Abb. 43.15 hat nicht mehr den Basisstrom I_B als Parameter, sondern die Beleuchtungsstärke E. Bei $E = 0$ Lx fließt praktisch kein Kollektorstrom I_C. Mit größer werdender Beleuchtungsstärke steigt der Strom an. Die I_C-Achse im Kennlinienfeld ist meistens lo-

Tab. 43.1 Übliche Zusammensetzung von LED's und IRED's

Werkstoff	SiC	GaP	GaP	GaAsP	GaAs	GaAsP	GaAs	GaAs
Schleusenspannung		2,7 V	2,4 V	2,2 V		1,6 V	1,4 V	
Dotierung	SiC	Stark mit N	Schwach mit N	Schwach mit N	Mit P	Zn + O	Zn	Si
Wellenlänge λ (nm)	480	565	590	625	650	700	900	930
Farbe	Blau	Grün	Gelb	Orange	Hellrot	Rot	Infrarot	Infrarot

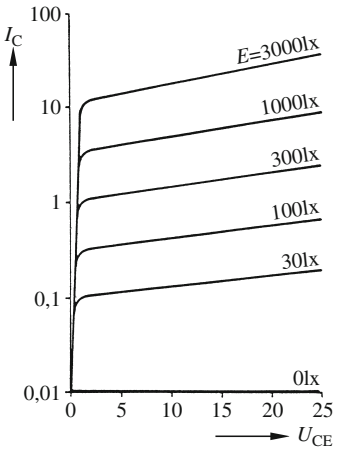


Abb. 43.15 Kennlinienfeld eines Fototransistors

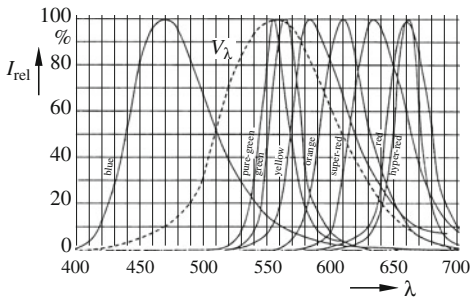


Abb. 43.16 Spektralkennlinien und Strahlungsmaxima einiger LED's

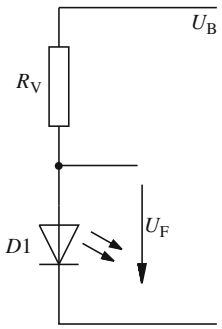


Abb. 43.17 LED mit Vorwiderstand

garithmisch eingeteilt, da sich der Kollektorstrom um mehrere Zehnerpotenzen ändert.

Die Schaltgeschwindigkeiten von Fototransistoren liegen zwischen 2 und 100 ms und sind damit kleiner als die der Fotodioden. Die Schaltgeschwindigkeit ist umso niedriger, je kleiner der Lastwiderstand und je größer die Amplitude des Lichtimpulses ist. Um die Lichtempfindlichkeit weiter zu erhöhen, kann der Transistor auch als Darlington-Fototransistor ausgeführt werden.

43.2.4 Lumineszenzdioden und Flüssigkristalle

In den *lichtemittierenden Fotohalbleitern* wird elektrische Energie in Strahlungsenergie umgewandelt. Das geschieht im Bereich einer dünnen pn-Sperrschicht. Hierbei wandern etwa gleich viele Elektronen von der n-Schicht in die p-Schicht wie Löcher von der p-Schicht in die n-Schicht. Die n-Schicht ist jedoch deutlich stärker dotiert als die p-Schicht. Dies führt dazu, dass der Strom durch die Sperrschicht fast vollständig ein Elektronenstrom ist. Die in die p-Schicht gelangenden Elektronen rekombinieren mit den dort vorhandenen Löchern. Dabei wird Energie frei, die je nach Ausgangsmaterial der Diode als sichtbares Licht oder als Infrarotstrahlung nach außen tritt.

Technisch benutzt wird dieser Effekt in der Leuchtdiode, die auch als Lumineszenzdiode (LED = **l**ight **e**mitting **d**iode) bezeichnet wird. LED's werden grundsätzlich in Durchlassrichtung betrieben.

Grundmaterial für Leuchtdioden sind *Gallium-Verbindungen* mit unterschiedlichen Dotierungen. Für den Bereich des sichtbaren Lichtes werden grün-, gelb-, orange-, rot- und blauleuchtende LED's geliefert. Für den nichtsichtbaren Infrarotbereich werden verschiedene IRED (**i**nfra**r**ed **e**mitting **d**iode) angeboten. In allen Fällen erstreckt sich das erzeugte Lichtspektrum jeweils nur über einen schmalen Bereich (monochromatische Leuchtquellen). Abb. 43.16 zeigt die Strahlungsmaxima und die Spektralkennlinien einiger LED's. Tab. 43.1 gibt eine Übersicht über die Zusammensetzung einiger LED's mit den zugehörigen Gallium-Verbindungen und Dotierungsstoffen. Als Fremdatome dienen Zinkdampf (Zn + O), Stickstoff (N), Phosphor (P) oder Silizium (Si).

In der Prinzipschaltung nach Abb. 43.17 ist der zur Strombegrenzung und Spannungseinstellung erforderliche Widerstand R_V enthalten. Die erzeugte Lichtstärke I_V wird bei LED's für den sichtbaren Lichtbereich meistens in Millicandela (mcd) angegeben. Sie hängt nahezu linear von der Größe des Durchlassstroms I_F ab. Für den praktischen Betrieb von LED's ist der Zusammenhang zwischen I_F und U_F von Bedeutung. Kennwerte einer Leuchtdiode sind die Leuchtfläche, die Strahlungsleistung (Lichtstrom) und die Lichtstärke (Helligkeit). Die Schleusenspannung ist aufgrund anderer Ausgangsmaterialien höher als bei Silizium-Dioden.

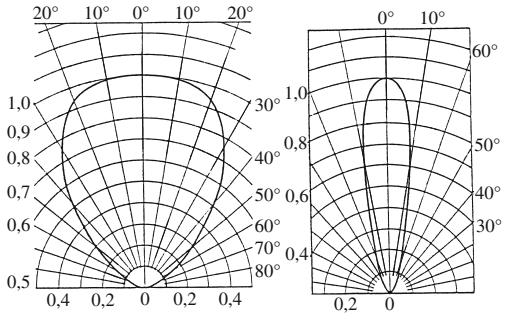


Abb. 43.18 Richtcharakteristiken von LED's (Erläuterung siehe Text)

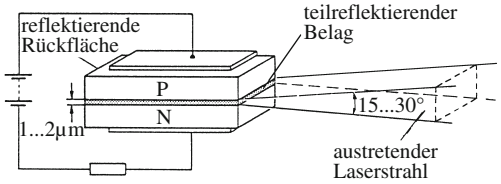


Abb. 43.19 Prinzipieller Aufbau einer Laserdiode

Grenzwerte:

$$U_{R\,\max} = 5\,\mathrm{V}, \qquad I_{F\,\max} = 50\,\mathrm{mA},$$
$$P_{\mathrm{tot}} = 100\,\mathrm{mW}, \quad \vartheta_{J\,\max} = 100\,^{\circ}\mathrm{C}$$

Durch entsprechende Form der aufgesetzten Kunststoffkörper ist es möglich, den LED's unterschiedliche Richtcharakteristiken zu geben. In Abb. 43.18 sind die Richtcharakteristiken für zwei Standardausführungen mit Öffnungswinkeln von 60° (breit) und 25° (gebündelt) angegeben.

Die Lebensdauer von LED's und IRED's liegt bei normalen Betriebsbedingungen bei 100.000 h. Geringer Spannungs- und Strombedarf (je nach Typ 5, 10 oder 20 mA), kleine Abmessungen, einfache Montage und hohe Packungsdichte geben ihnen einen sehr breiten Anwendungsbereich.

Da LED's Schaltzeiten von 5 bis 20 ns haben, können sie auch zur Abstrahlung von sich sehr schnell ändernden Lichtsignalen verwendet werden, zum Beispiel in Optokopplern (siehe Abschn. 43.4).

Bei den Laser-Dioden (engl.: **l**ight **a**mplification by **s**timulated **e**mission of **r**adiation) wird das im pn-Übergang erzeugte monochrome Licht im Inneren des Kristalls an den inneren Flächen verspiegelt und tritt an der Stirnfläche mit relativ schmalem Austrittswinkel, aber großer Lichtstärke aus. Abb. 43.19 zeigt den grundsätzlichen Aufbau einer Laserdiode als Kantenstrahler. Wird das Licht impulsartig abgestrahlt, sind mit diesen Bauelementen Leistungen bis ungefähr 100 W möglich. Sie eignen sich zur Nachrichten- und Datenübermittlung in Lichtwellenleitern. Zur Abtastung von CD-Plattenspielern und als Lesestift in Scannergeräten werden sie ebenfalls benutzt. Laser höherer Leistung sind zum Beispiel aus der Medizin (optisches Skalpell) nicht mehr wegzudenken.

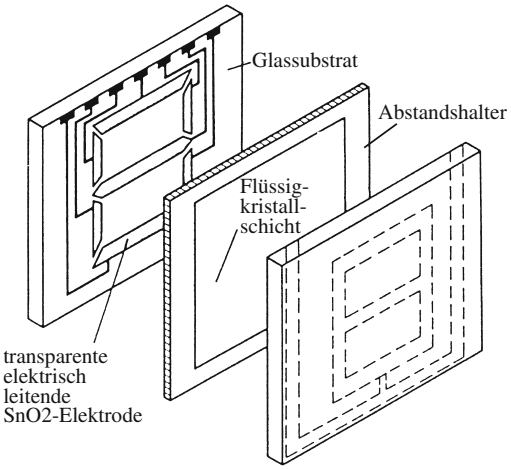


Abb. 43.20 Prinzipaufbau einer LCD-Anzeige

Eine interessante und zukunftsträchtige Variante zur Anzeige von Informationen stellen die Flüssigkristalle (engl.: liquid crystal) dar. Flüssigkristalle sind glasklare Flüssigkeiten, deren Moleküle einen regelmäßigen einkristallinen Aufbau aufweisen. Sie befinden sich in einem speziellen Aggregatzustand, in dem Stoffe aus dem flüssigen in den festen Zustand übergehen. Flüssigkristallwerkstoffe zeigen bei Einwirkung eines elektrischen Feldes Veränderungen ihrer Kristallstruktur.

Je nach Grundsubstanz gehen sie entweder vom durchsichtigen in den weitgehend undurchsichtigen Zustand über (wird meist verwendet) oder umgekehrt. Dieses geschieht dadurch, dass ihre Moleküle sich in bestimmter Weise ausrichten.

Nach Abschalten des elektrischen Feldes stellt sich der ursprüngliche Zustand wieder ein. Flüssigkristalle leuchten nicht. Einfallendes Fremdlicht wird an den undurchsichtigen Bereichen reflektiert und macht damit die flächenmäßige Form der Elektroden des elektrischen Feldes sichtbar. Mit Unterstützung von Polarisationsfiltern und geschickt angebrachten Lichtquellen kann der Prinzipaufbau nach Abb. 43.20 vollendet werden. LCD-Anzeigen (**l**iquid **c**rystal **d**isplay) werden in Uhren, Taschenrechnern und zahlreichen anderen Geräten verwendet. Bei wirksamen Spannungen von 1,5 bis 3 V und einem Strom in der Größenordnung von 1 µA werden Anzeigen mit sehr geringen elektrischen Leistungsanforderungen erreicht.

43.3 Anzeigeeinheiten

Abb. 43.21 zeigt eine einfache Schaltung zur Kontrollanzeige der Betriebsspannung mit einer LED. Eine Umpolung der LED macht die Schaltung zur Anzeige einer negativen Betriebsspannung fähig. Soll eine solche Anzeige für Wechselspannung installiert werden, muss die LED vor der hohen Sperrspannung geschützt werden.

Eine antiparallel geschaltete normale Diode übernimmt diesen Schutz, da sie eine Schleusenspannung von 0,7 V hat und darum die Sperrspannung an der LED unter 5 V hält. Die beiden

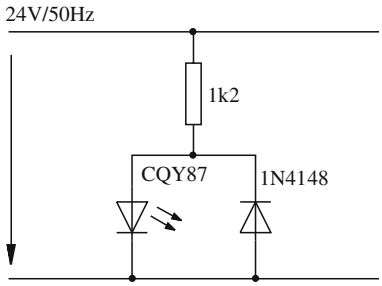


Abb. 43.21 Betriebsspannungsanzeige

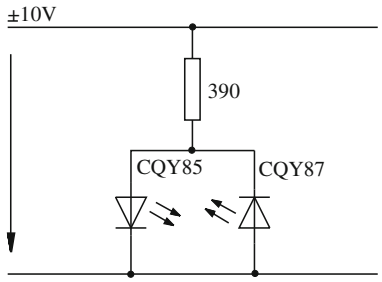


Abb. 43.22 Farbige Polaritätsanzeige mit LED's

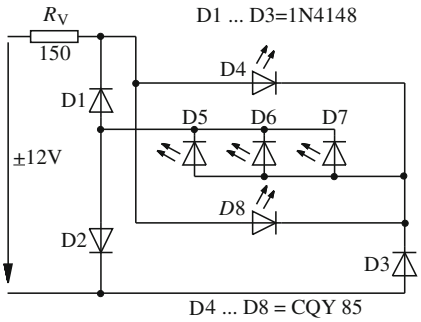


Abb. 43.23 Polaritätsanzeige mit Symbolen

LED's in Abb. 43.22 haben unterschiedliche Farbabstrahlungen, übernehmen den gegenseitigen Schutz und zeigen die jeweilige Polarität farblich verschieden an.

Abb. 43.23 zeigt eine Schaltung, bei der LED's so geschaltet sind, dass sie die Polarität in Symbolen anzeigt. Ist zum Beispiel die Betriebsspannung positiv, so sind die Dioden D1 und D3 in Sperrrichtung geschaltet. Auf diese Weise kann der Strom nur den Weg über alle LED's und die Diode D2 nehmen. Ist dagegen die Betriebsspannung negativ, so fließt der Strom über die dann in Durchlassrichtung liegenden Dioden D1 und D3 sowie über die LED's D5 bis D7.

Zur Darstellung einer beliebigen Dezimalziffer kann man 7-Segment-Anzeigen nach Abb. 43.24 verwenden. Für jedes Segment ist ein Anschluss (a bis g) nach außen geführt, im vorliegenden Bild zusätzlich der Anschluss für die gemeinsame Kathode. Bei manchen 7-Segment-Anzeigen wird die gemeinsame Anode herausgeführt, so dass die Anzeige mit negativen Spannungen angesteuert werden kann. In den Segmenten sind

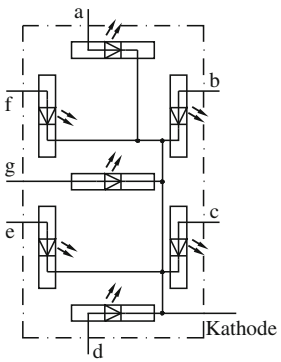


Abb. 43.24 7-Segment-Anzeige mit Leuchtdioden

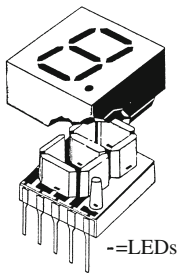


Abb. 43.25 7-Segment-Anzeige mit Dezimalpunkt und LED's

LED's angeordnet, deren Licht über Lichtleiter nach außen geführt wird. Abb. 43.25 zeigt eine 7-Segment-Anzeige in ihrer mechanischen Ausführung.

Erhöht man die Anzahl der Segmente auf sechzehn, können außer Ziffern auch Buchstaben und Sonderzeichen dargestellt werden. Diese Sechzehnsegment-Anzeigen nach Abb. 43.26 bezeichnet man als alphanumerische Anzeigeeinheit. Diese Anzeigen werden auch mehrstellig angeboten und können zu langen Zeilen erweitert werden.

Bei der Realisierung solcher Anzeigen mit LED's stößt man auf ein gewichtiges Problem. Diese Dioden haben eine Stromaufnahme von ca. 20 mA und vervielfachen sich mit der Anzahl der

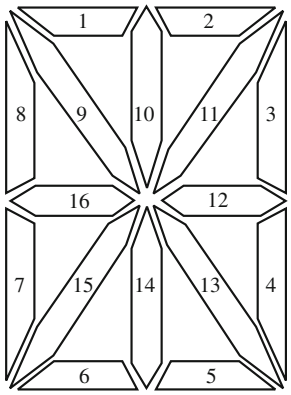


Abb. 43.26 Darstellung der Segmente einer alphanumerischen Anzeigeeinheit

verwendeten LED's. Mehrstellige Anzeigen lassen die Stromversorgung für kleine Geräte nahezu unlösbar werden.

Baut man sowohl die 7-Segment-Anzeigen wie auch die alphanumerischen Anzeigen mit Flüssigkristallen auf, lassen sich sehr komplexe Anzeigeeinheiten mit zahlreichen Zeilen und Stellen aufbauen, die infolge ihrer geringen Leistungsaufnahme direkt an digitalen IC mit hoher Integrationsdichte betrieben werden können.

43.4 Signalübertragung mit Optokoppler

Optokoppler sind *optoelektronische Koppelelemente* zur Signalübertragung bei galvanischer Trennung von Ein- und Ausgang. Als Sender dient eine IR-Diode, die über einen Lichtleiter direkt auf einen Empfänger strahlt. Als Empfänger dienen überwiegend Fotodioden, Fototransistoren mit und ohne herausgezogene Basis und Fototriacs. Die Schaltungen nach Abb. 43.27 zeigen den prinzipiellen Aufbau einiger Optokoppler. Zwischen Eingangs- und Ausgangsseite dürfen, je nach Bauform, Potentialdifferenzen bis zu einigen kV bestehen.

In Optokopplern werden elektrische Signale in optische Signale umgewandelt und über eine Isolationsstrecke übertragen. Im Anschluss daran wird das optische Signal wieder in ein elektrisches Signal umgewandelt. Der Optokoppler ist durch das System der Signalübertragung absolut rückwirkungsfrei.

Die wichtigsten Kenngrößen sind das *Stromübertragungsverhältnis* CTR (**c**urrent **t**ransfer **r**atio), dass das Verhältnis von Ausgangsstrom zu Eingangsstrom beschreibt, und die *Grenzfrequenz* f_g, bei der der AC-CTR-Wert auf 50 % des DC-CTR-Wertes abgesunken ist. Abb. 43.28 zeigt einige Optokoppler der Fa. Siemens mit Pin-Belegung, Bauform und innerer Schaltung.

Bei Optokopplern mit Fotodioden ergibt sich ein CTR-Wert von ca. 1 %, bei Fototransistoren kann der CTR-Wert 100 % betragen. Werden im Optokoppler Darlington-Transistoren verwendet, steigt der CTR-Wert auf bis zu 500 %. Manche Optokoppler verfügen über eine Grenzfrequenz nahe bei 10 MHz. Bei Optokopplern mit einem Fototransistor als Empfänger ergeben sich Schaltzeiten von ungefähr 3 µs.

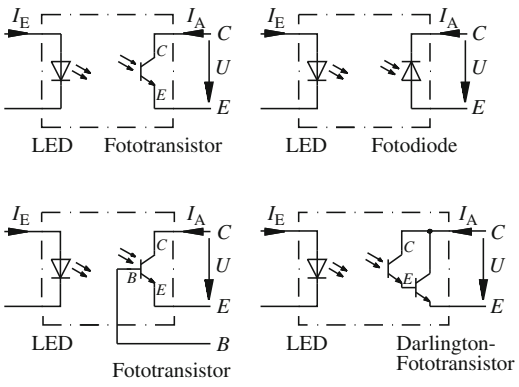


Abb. 43.27 Prinzipschaltbilder einiger Optokoppler

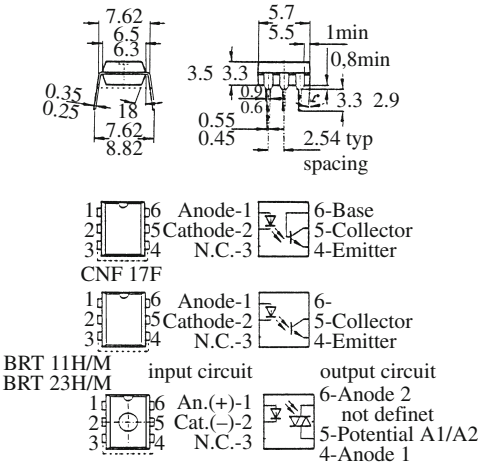


Abb. 43.28 Optokopplertypen (Fa. Siemens)

Bei LED's als Sender macht sich eine alterungsbedingte Abnahme ihrer Strahlungsleistung über einen längeren Zeitraum durch eine Verringerung des CTR-Wertes bemerkbar. Hohe Ströme und/oder hohe Umgebungstemperaturen sind der Grund dafür und sollten vermieden werden.

Optokoppler lassen sich sehr günstig in Verbindung mit Digitalschaltungen verwenden. Sie werden in *Interface-Schaltungen* verwendet, um zum Beispiel eine *Potentialtrennung* zwischen der Zentraleinheit eines Computers und seinen Peripheriegeräten herbeizuführen.

Wird die optische Kopplung zwischen Lichtsender und Lichtempfänger nicht im Inneren des Bauelementes, sondern über äußere Reflexstellen vorgenommen, spricht man von Lichtschranken. Nur wenn das Licht an einer geeigneten, dafür vorgesehenen Stelle reflektiert wird, gelangt es an den Empfänger. Derartige Bauelemente werden auch *Reflexsensoren* genannt.

In die Gruppe der Lichtschranken gehören auch die *Gabellichtschranken*, bei denen Lichtsender und -empfänger in getrennten Holmen eines U-förmigen Gehäuses integriert sind. Der *Lichtstrahl* kann durch einen dazwischengeschobenen Gegenstand unterbrochen werden.

Gabellichtschranken werden zur Prozessüberwachung verwendet, aber auch in Alarmanlagen, Positions- meldern, zur Drehzahlüberwachung oder zur Informationsübernahme in Belegelesern (Scanner).

43.5 Faseroptische Übertragungsmittel

Die *optische Nachrichtenübertragung* über Lichtwellenleiter (LWL) gewinnt ständig an Bedeutung. Um hohe Frequenzen realisieren zu können, wird als Strahlungsquelle eine IR- oder eine Laserdiode und als Empfänger eine Fotodiode verwendet.

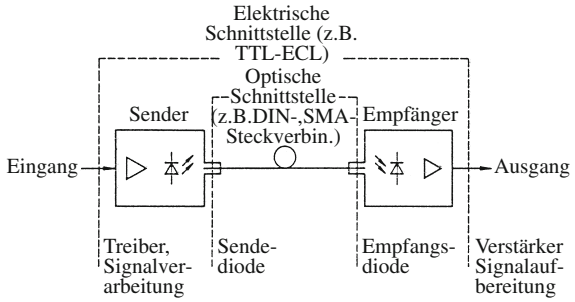


Abb. 43.29 LWL-Übertragungsstrecke (Fa. Siemens)

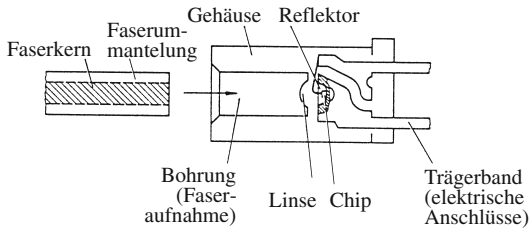


Abb. 43.30 Empfangsdioden für Plastik-LWL (Fa. Siemens)

Medien für die optische Übertragung sind Glas oder Plastik. Zur Nachrichtenübertragung mit hohen Übertragungsraten bei großen Entfernungen werden bevorzugt Glasfasern verwendet. Plastikfasern dagegen sind für niedrige Übertragungsraten im lokalen Bereich und zur Lösung vielfältiger Anwendungen in der Steuer- und Regelungstechnik verwendbar.

Bedingt durch die gegenwärtige Fasertechnologie werden die aktiven LWL-Bauteile für die Wellenlängen um 850 nm ausgelegt. Vorteile dieser Technik liegen in der Unempfindlichkeit gegenüber elektromagnetischen Störungen und in der eindeutigen galvanischen Trennung zwischen Sender und Empfänger.

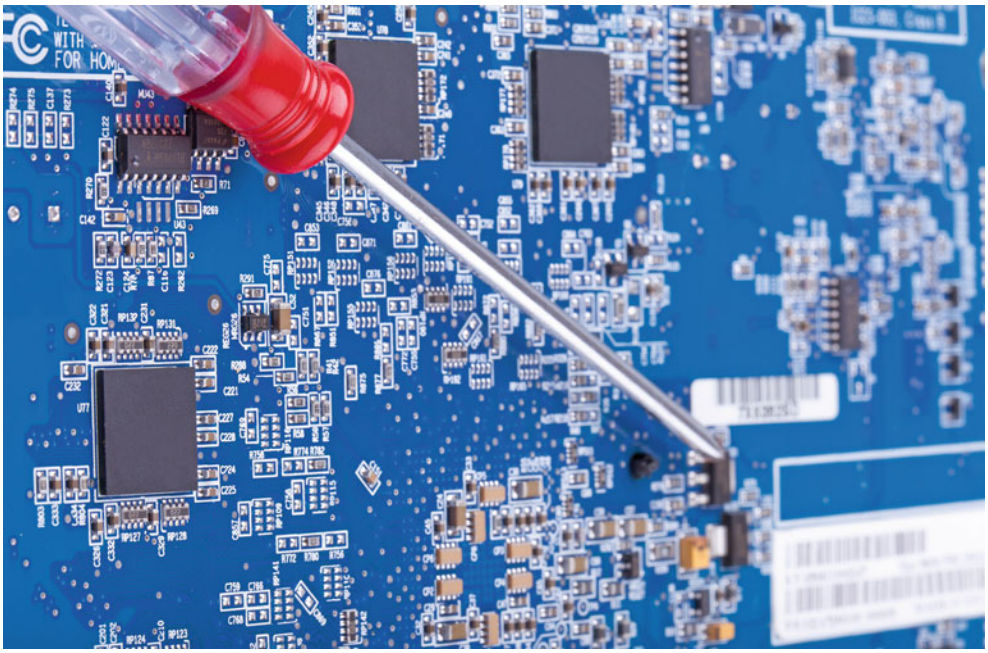
Abb. 43.29 zeigt das Schema eines optischen Übertragungssystems für den lokalen Bereich, hier eine Schnittstelle zwischen einer TTL-Technologie und einer ECL-Technologie.

Die Anbindung der LWL an die LWL-Bauteile und die optomechanische Verbindung der LWL miteinander sind die eigentlichen Probleme dieser Technologie, die sich aber durch hohe Präzision bei der Montage bewältigen lassen (Abb. 43.30).

Analog-Digital-Wandler

44

Peter Döring



44.1 Grundlagen 538
44.2 Spannungs-Frequenz-Wandler 538
44.3 Sägezahnverfahren 538
44.4 Dual-Slope-Verfahren 539
44.5 Flash-Wandler 539
44.6 Wandler nach dem Wägeverfahren 540
44.7 Integrierte Wandler 540

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_44

44.1 Grundlagen

Häufig werden physikalische Größen (Temperatur, Druck, Längen, Drehzahl, u. a.) dezentral erfasst und zentral ausgewertet, dargestellt und bearbeitet. Mit Hilfe entsprechender Sensoren erhält man die Messgröße meist in analoger Form als Stromstärke oder als Spannung. Ein digitales Signal lässt sich aber besser über größere Entfernungen ohne Signalwertfälschung übertragen. Außerdem kann mit bestimmten Verfahren der Nachrichten- und Datentechnik (Multiplexverfahren) eine Übertragungsleitung mehrfach ausgenutzt werden.

Um vorhandene analoge Signale in verwertbare digitale Signale umzuwandeln, benötigt man Analog-Digital-Wandler.

Eine digitale Messung lässt sich als Zählvorgang begreifen. Also kann man die analoge Größe in eine Impulsreihe mit bestimmter Frequenz umformen und diese Impulse dann in einer bestimmten Zeiteinheit zählen.

44.2 Spannungs-Frequenz-Wandler

Zur Umformung einer Signalspannung in eine der Spannung proportionale Frequenz verwendet man einen Spannungs-Frequenz-Umsetzer (-Wandler) nach Abb. 44.1. Der Operationsverstärker N1 ist hier als Integrator geschaltet, der die Spannung u_e über die Zeit integriert und damit eine negative Sägezahnspannung u_a bewirkt. Je größer die Eingangsspannung ist, desto steiler ist der Anstieg der Ausgangsspannung.

Die Spannung U_e vergleicht der Operationsverstärker N2 (Komparator) mit der negativen Spannung U_V des Spannungsteilers aus R2 und R3. Sobald u_a den Wert U_V erreicht, kippt der Komparator N2 und steuert den Transistor V1 durch. Der leitende Transistor hebt die Vergleichsspannung auf nahezu 0 V an. Gleichzeitig entlädt sich der Kondensator C1 über den Transistor, bis die Ausgangsspannung $u_a = 0\,\text{V}$ wird.

Dieser Vorgang führt dazu, dass der Komparator erneut seine Ausgangsspannung umkehrt und nun den Transistor sperrt. Damit beginnt der Integrationsvorgang der Messgröße u_e erneut.

Am Ausgang der Schaltung erhält man Impulse, deren Frequenz vom Wert der Eingangsspannung abhängt. Die Frequenz f ist proportional zur Spannung u_e.

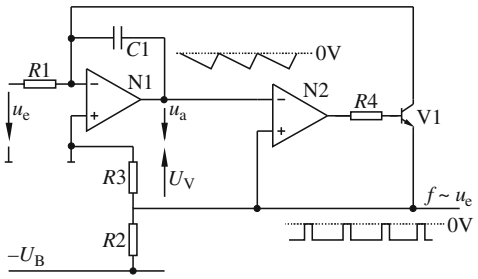


Abb. 44.1 Spannungs-Frequenz-Umsetzer

Da sich der Kondensator nur in einer endlichen Zeit entlädt, kommt hier ein systematischer Messfehler zum Tragen.

$$\textit{Erzeugte Frequenz} \quad f \approx \frac{u_e}{U_V \cdot R_1 \cdot C_1} \tag{44.1}$$

unter der Bedingung für die Rückstellzeit

$$t_R \quad \Rightarrow \quad \frac{\Delta u_a \cdot R_2 \cdot C_1}{U_B} \ll R_1 \cdot C_1$$

Schaltet man hinter diesen Umsetzer einen digitalen Frequenzmesser, so wird die Gesamtschaltung zu einem digitalen Spannungsmesser.

Für $R1 = 100\,\text{k}\Omega$, $R2 = 700\,\Omega$, $R3 = 800\,\Omega$, $C1 = 1{,}25\,\text{nF}$, $U_B = -15\,\text{V}$ und $u_e = 10\,\text{V}$ ergibt sich nach (44.1) eine Frequenz von 10 kHz. Da die Rückstellzeit nicht berücksichtigt ist, liegt der tatsächliche Wert niedriger. Der Spitzenwert von u_a und U_V liegt bei $-8\,\text{V}$.

44.3 Sägezahnverfahren

Beim Sägezahnverfahren wird die analoge Spannung u_e mit einer linear ansteigenden Sägezahnspannung nach Abb. 44.2 verglichen. Die Komparatoren N1 und N2 vergleichen die Messgröße und die Sägezahnspannung miteinander. Sobald die Sägezahnspannung $\geq 0\,\text{V}$ wird, beginnt der Vergleich, denn nun liegt der eine Eingang des EXCLUSIV-ODER-Gatters (N3) auf „H“, das UND-Gatter (N4) ist vorbereitet, und die Impulse des Rechteckgenerators (Oszillators) gelangen an den Ausgang A der Schaltung.

Wenn die Sägezahnspannung den Wert der Messgröße u_e erreicht, kippt der Komparator N1 am Ausgang von „L“ auf „H“. Der Ausgang des EXCLUSIV-ODER-Gatters (N3) verändert sich auf „L“, und das UND-Gatter (N4) sperrt weiteren Impulsen des Rechteckgenerators den Durchgang zum Ausgang.

Die beiden Schaltpunkte der Komparatoren N1 und N2 begrenzen also die Zeitdauer Δt, während der die Impulse gezählt werden. Bei linearer Sägezahnspannung ist Δt der Messgröße u_e direkt proportional, und die Anzahl der Impulse entspricht der Höhe der Eingangsspannung. Nach jeder Messung muss der Zähler auf Null zurückgesetzt werden. Das letzte Zählergebnis kann mit einem Anzeigespeicher bis zur Vorlage des nächsten Zählergebnisses festgehalten werden, um eine ruhende Anzeige zu ermöglichen.

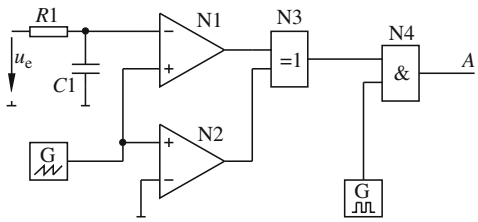


Abb. 44.2 Schaltung zum Sägezahnverfahren

Es lassen sich Messgenauigkeiten von etwa 0,1 % des Endwertes erreichen. Um Oberwellen der Messgröße aus dem Messergebnis herauszuhalten, ist ein Tiefpassfilter am Eingang vorteilhaft.

44.4 Dual-Slope-Verfahren

Beim „Doppelintegrationsverfahren“ (dual-slope-converter) erfolgt die Umwandlung in zwei aufeinander folgenden Schritten.

Die Messgröße u_e wird während der konstanten vorgegebenen Zeit T1 integriert. Die Ausgangsspannung u_a (Sägezahnspannung) ändert sich entsprechend (44.2).

Ausgangsspannung

$$u_a(t) = -\frac{1}{R_1 \cdot C_1} \cdot u_e \cdot T_1 \qquad (44.2)$$

Bei konstanter Integrationszeit wird der Höchstwert der Ausgangsspannung u_a durch den Wert der Eingangsspannung u_e bestimmt.

Im zweiten Schritt wird der Kondensator C1 über die Konstantspannungsquelle entladen. Die Ausgangsspannung u_a des Integrators N1 in der Schaltung nach Abb. 44.3 ändert sich linear, bis der Komparator N2 den Spannungswert Null feststellt. Die Entladezeit T2 richtet sich ausschließlich nach der vorher im 1. Schritt aufgenommenen Ladungsmenge, also nach dem Höchstwert von u_a.

Diese Entladezeit wird genutzt, um die Digitalisierung der Messgröße u_e zu bewirken, indem der Oszillator N4 während der Entladezeit T2 seine Impulse auf den Ausgang gibt.

Die Zählzeit T2 und die konstante Oszillatorfrequenz bestimmen die Anzahl der Impulse.

$$\textit{Zählzeit} \quad T_2 = \frac{u_e}{U_{\text{ref}}} \cdot T_1 \qquad (44.3)$$

Der Steuerlogik-Baustein bildet die Integrationszeit durch Frequenzteilung aus dem Oszillator N4, hält den FET-Transistor V1 für die Integrationszeit T1 leitend und den FET-Transistor V2 gesperrt. Nach Ablauf der Integrationszeit T1 gibt die Steuerlogik ein „H“-Signal auf das UND-Gatter N3 und schaltet die Transistoren um, so dass sich der Kondensator über die Konstantspannungsquelle U_{ref} entlädt. Während dieser Zeit T2 gibt auch der Komparator N2 ein „H“-Signal an das UND-Gatter N3, so dass die Impulse des Oszillators an den Ausgang gelangen.

Der Zählvorgang beziehungsweise die Zeit T2 ist beendet, sobald sich der Kondensator auf $u_a = 0\,\text{V}$ entladen hat und der Komparator N2 ein „L“-Signal an das UND-Gatter N3 gibt, womit er weitere Impulse unterdrückt.

Es lassen sich Messgenauigkeiten von etwa 0,01 % des Endwertes erreichen, wenn die Integrierzeit T1 durch Frequenzteilung von der Oszillatorfrequenz bestimmt wird. Die mittlere Umsetzungszeit liegt bei etwa 5 ms. AD-Wandler nach diesem Prinzip haben eine große Verbreitung gefunden.

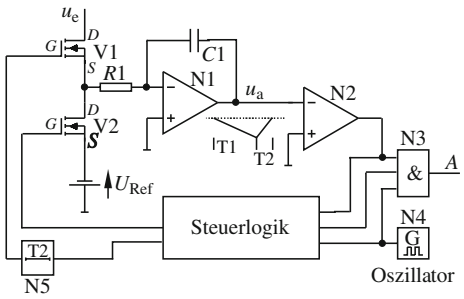


Abb. 44.3 Dual-Slope-Verfahren

44.5 Flash-Wandler

Beim Verfahren mit parallelen Komparatoren nach Abb. 44.4 wird die angelegte Messspannung U_E mit genau festgelegten Referenz- und Teilreferenzspannungen verglichen. Der hier dargestellte 3-Bit-AD-Umsetzer macht 7 Komparatoren und ebenso viele Vergleichsspannungen erforderlich, die über einen Spannungsteiler erzeugt werden. Solange die angelegte Messspannung U_E kleiner ist als die über R8 abfallende Vergleichsspannung, zeigt der Ausgang des Komparators N7 eine logische „0“ an, die in eine logische „1“ übergeht, sobald die angelegte Messspannung U_E größer wird als die über R8 abfallende Vergleichsspannung. Dieser Trend setzt sich mit größeren Messspannungen U_E entsprechend fort.

Ist die angelegte Messspannung U_E größer als die Spannungssumme über den Widerständen R5 bis R8, aber kleiner als die Spannungssumme über den Widerständen R4 bis R8, zeigt der Ausgang des Komparators N4 eine logische „1“ an, während der Ausgang des Komparators N3 eine logische „0“ verzeichnet.

Der X/Y-Dekoder ermittelt zunächst den Komparator, bei dem die Vergleichsspannung erstmalig größer ist als die Messspannung. Das wird durch eine UND-Verknüpfung eines jeden Komparators mit dem nächsten erfolgen. Mit weiteren Codierstufen erhält man hier das 3-Bit-Ausgangssignal an den Ausgängen Q_A bis Q_C. Für einen 8-Bit-AD-Wandler sind 255 Komparatoren mit den entsprechenden Vergleichsspannungen aufzuwenden.

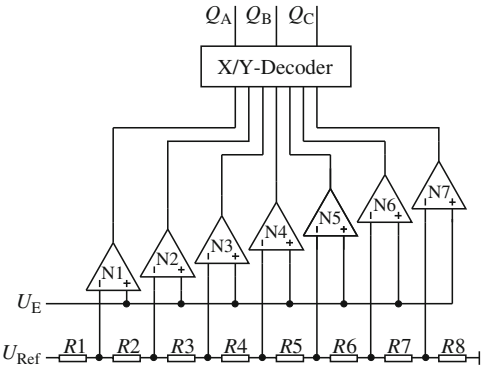


Abb. 44.4 Prinzipschaltung eines Flash-Wandlers

AD-Umsetzer nach diesem Verfahren mit parallelen Komparatoren ermöglichen Umsetzfrequenzen von mehr als 100 MHz. Diese Umsetzer werden als Flash-Wandler bezeichnet. Es gibt sie als IC-Bausteine (8- und 12-Bit-Umsetzer).

44.6 Wandler nach dem Wägeverfahren

Das Wägeverfahren ist ein Vergleichsverfahren, bei dem die digitale Aussage an die gegebene analoge Eingangsgröße durch schrittweise Annäherung (sukzessive Approximation) ermittelt wird.

Die digitale Ausgangsgröße wird über einen Operationsverstärker, der als bewerteter Addierer geschaltet ist, in die analoge Form (Treppenspannung) zurückversetzt und dann über einen Komparator mit der analogen Messgröße verglichen. Dabei wird zunächst das höchste Bit des Speichers auf „1“ gesetzt.

Das Zahlenbeispiel (Tab. 44.1) soll die schrittweise Umsetzung des Wertes 165 bei einem 8-Bit-AD-Wandler verdeutlichen.

Das Umsetzungsverfahren benötigt zwar viele Schritte, arbeitet aber insgesamt wesentlich schneller als ein Wandler nach dem Integrationsverfahren. Die mittlere Umsetzungszeit beträgt für einen 10-Bit-AD-Wandler etwa 10 µs.

Tab. 44.1 Zahlenbeispiel zur schrittweisen Umsetzung

1. Schritt	$128 < 165$	$Q_7 = 1$	128
2. Schritt	$165 - 128 = 37$ $64 > 37$	$Q_6 = 0$	0
3. Schritt	$32 < 37$	$Q_5 = 1$	32
4. Schritt	$37 - 32 = 5$ $16 > 5$	$Q_4 = 0$	0
5. Schritt	$8 > 5$	$Q_3 = 0$	0
6. Schritt	$4 < 5$ $5 - 4 = 1$	$Q_2 = 1$	4
7. Schritt	$2 > 1$	$Q_1 = 0$	0
8. Schritt	$1 = 1$	$Q_0 = 1$	1
			$\sum 165$

44.7 Integrierte Wandler

Analog-Digital-Umsetzer sind als integrierte CMOS-Schaltungen für vielfältige Anwendungen verfügbar. Je nach der vorgesehenen weiteren Verwendung der digitalen Ausgangsgrößen werden diese ICs mit dual-codierten Ausgängen, parallelen BCD-Ausgängen oder *n*-Digit-BCD-Multiplexausgängen unterschieden.

BCD-Ausgänge werden für die Ansteuerung von digitalen Anzeigeeinheiten verwendet, während dual codierte Parallelausgänge für die digitale Datenverarbeitung mit Computern benötigt werden.

Alle Wandler sind durch zwei wesentliche Kenngrößen charakterisiert: die Auflösung und die Umsetzzeit. Die Auflösung sagt aus, wie fein ein Analogwert von dem Wandler in einen Digitalwert umgesetzt wird. Sie wird in Bit angegeben. 4 Bit entsprechen einer Genauigkeit von ca. 7 %, während 8 Bit eine Genauigkeit von 0,4 % bieten und dem Standard entsprechen. Die Umsetzzeit gibt an, wieviel Zeit der Wandler benötigt, um den Wert der anliegenden Spannung in ein Digitalwort umzusetzen. Bei der digitalen Bildverarbeitung werden extrem kurze Zeiten verlangt, während sie bei anderen Gelegenheiten eher von untergeordneter Bedeutung sind. Je nach gewünschter Umsetzzeit werden die verschiedenen Verfahren eingesetzt.

Mit sukzessiver Approximation arbeiten die AD-Wandler der ADC-08xx-Familie. Sie haben eine Auflösung von 8 Bit, eine Umsetzzeit von 100 µs und eine Zugriffszeit von 135 ns bei einer einfachen Versorgungsspannung von 5 V. Die Referenzspannung ist wählbar. Der Wandler ist für den direkten Anschluss an den Mikroprozessor 8080 und seine Verwandten konzipiert worden.

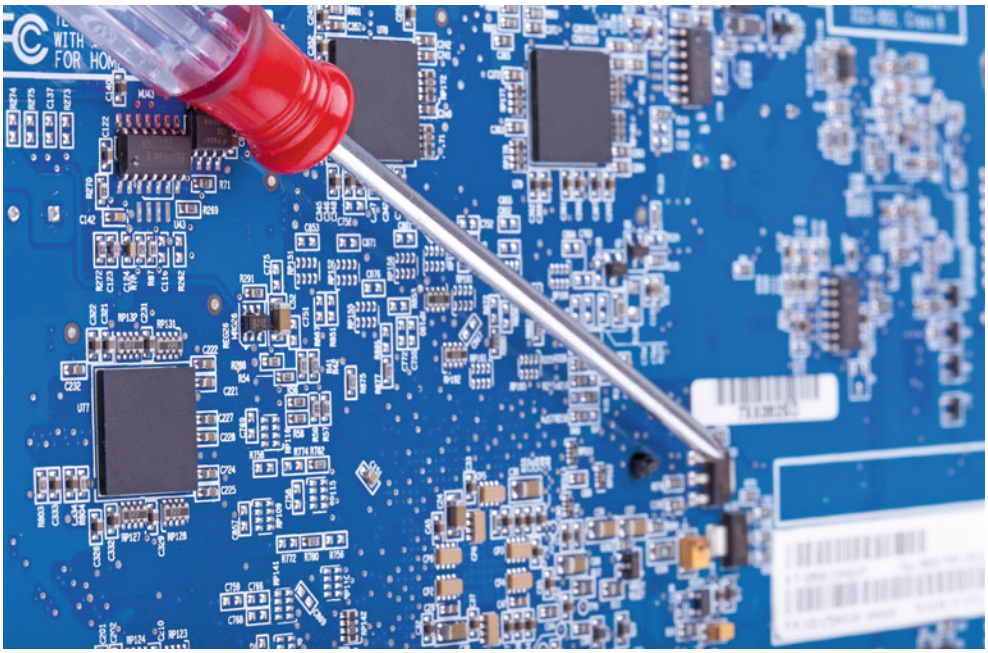
Der von der Fa. Datel hergestellte ADC 847 hat bei einer Auflösung von 8 Bit eine Umsetzzeit von 9 µs.

Der MAX132 der Fa. Maxim ist ein relativ langsamer 18-Bit-AD-Wandler mit seriellem Ausgang, der nach dem Multi-Slope-Verfahren arbeitet. Es können bis zu 100 Wandlungen pro Sekunde erfolgen. Er eignet sich besonders für den Einsatz in Datenerfassungssystemen, Waagen und Schalttafel-Instrumenten.

Digital-Analog-Wandler

45

Peter Döring



45.1 Grundlagen 542
45.2 D/A-Wandler-Varianten 542
45.3 Integrierte Wandler 543

Teil V

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_45

45.1 Grundlagen

Um vorhandene digitale Signale in verwertbare analoge Signale umzuwandeln, benötigt man Digital-Analog-Wandler. Diese Wandler stellen das unverfälschte analoge Signal wieder her und ermöglichen seine analoge Weiterverarbeitung (zum Beispiel Verstärkung).

Voraussetzung ist in den meisten Anwendungsfällen, dass das digitale Signal ein parallel anstehendes Datenwort ist, das eine mehrstellige Binärzahl darstellt.

45.2 D/A-Wandler-Varianten

Eine einfache Schaltung eines 3-Bit-DA-Wandlers mit Operationsverstärker zeigt Abb. 45.1. Es handelt sich hierbei um einen invertierenden OP, bei dem die Eingangswiderstände entsprechend der Wertigkeit der digitalen Stelle mit nachgeschalteter Umkehrstufe gewählt werden müssen.

Das 3-Bit-Eingangssignal Q_A bis Q_C liegt an den Eingängen E_1 bis E_3, so dass $R_1 = R$, $R_2 = R/2$ und $R_3 = R/4$ gewählt werden müssen. Ein weiterer Eingang E4 mit einem Eingangswiderstand $R_4 = R/8$ erweitert diesen Wandler zu einem 4-Bit-DA-Wandler.

Dieser 3-Bit-DA-Wandler wandelt ein 3-Bit-Datenwort in eine analoge Treppenspannung U_{A2} nach Abb. 45.2 am Ausgang um.

Der 4-Bit-DA-Wandler lässt sich für weitere Dekaden nach dem 8-4-2-1-BCD-Code erweitern. Für jede weitere Dekade sind vier Widerstände erforderlich, deren Widerstandswerte um den Faktor 0,1 je Dekade kleiner werden. Diese Art der Erweiterung kann theoretisch beliebig erweitert werden.

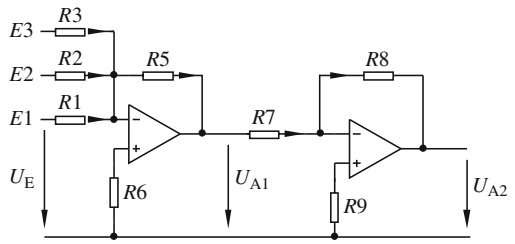


Abb. 45.1 Schaltung eines 3-Bit-DA-Wandlers mit OP

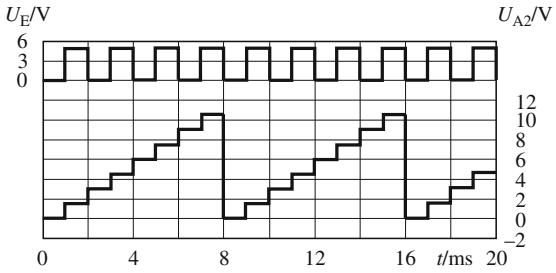


Abb. 45.2 Liniendiagramm zur Schaltung nach Abb. 45.1

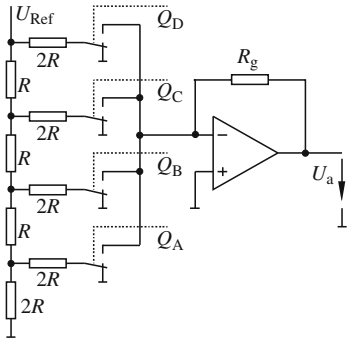


Abb. 45.3 R-$2R$-Netzwerk für einen 4-Bit-Wandler

Die Anforderungen an die Genauigkeit der Widerstände werden bei den Stellen höchster Wertigkeit sehr groß und lassen sich kaum noch erfüllen. Darum benutzt man ein Schaltungsprinzip, bei dem die verwendeten Widerstände in derselben Größenordnung liegen.

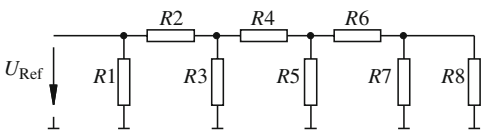
Im Prinzip wird in der Schaltung nach Abb. 45.1 eine Summation der bewerteten Einzelströme durchgeführt. Das lässt sich auch mit einem Kettenleiter oder R-$2R$-Netzwerk für einen 4-Bit-Wandler nach Abb. 45.3 bewerkstelligen.

Entsprechend dem logischen Zustand der digitalen Eingänge Q_A bis Q_D werden die Umschalter entweder auf Masse gelegt oder auf den invertierenden Eingang des OP geschaltet, so dass der bewertete Strom im Addierer summiert wird (oder gegen 0-Potential abfließt) und als Ausgangsspannung U_a erscheint.

Die Umschalter werden durch elektronische Schalter realisiert. Der Gegenkopplungswiderstand R_g lässt sich je nach gewünschter analoger Ausgangsspannung U_a berechnen, da durch ihn der Summenstrom fließt.

Integrierte Schaltkreise der Industrie enthalten zusätzlich oft einen Speicher für das digitale Datenwort. Üblich sind 8-Bit-DA-Wandler.

Die Schaltung nach Abb. 45.4 soll die Grundlage für den Einstieg in eine knappe Netzwerkberechnung bieten. Die Widerstände R7 und R8 liegen zueinander parallel, haben also insgesamt den Widerstand R. Als Reihenschaltung mit R6 ergibt sich wieder der Wert $2R$. Der Widerstand R5 mit $2R$ liegt



$R2 = R4 = R6 = R$

$R1 = R3 = R5 = R7 = R8 = 2R$

Abb. 45.4 Schaltung zur Netzwerkberechnung

nun parallel zum Ersatzwiderstand der bisher berechneten Widerstände R6, R7 und R8.

Das Netzwerk hat einen Gesamtersatzwiderstand von R. Die Referenzspannungsquelle U_{ref} wird also mit R belastet. Der ihr entnommene Strom teilt sich auf R1 und R2 zu gleichen Teilen auf, was sich an allen Knoten wiederholt.

Durch die senkrecht gezeichneten Widerstände des Netzwerkes nach Abb. 45.4 fließen Ströme, deren Werte sich jeweils halbieren (von der Spannungsquelle aus betrachtet), aber in allen Schaltumständen konstant bleiben.

45.3 Integrierte Wandler

Der DAC 0808 ist ein monolithisch integrierter Digital/Analog-Wandler mit einer Einstellzeit von 150 ns für Vollaussteuerung des Ausgangsstroms. Die digitalen Eingänge sind TTL- und CMOS-kompatibel. Bei einer Versorgungsspannung von ± 5 V werden maximal 33 mW aufgenommen. Einen Digital/Analog-Umsetzer für eine Ausgangsspannung von 10 V zeigt die Schaltung nach Abb. 45.5.

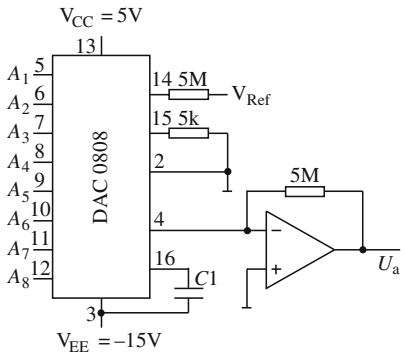


Abb. 45.5 Digital/Analog-Umsetzer mit dem IC/DAC 0808

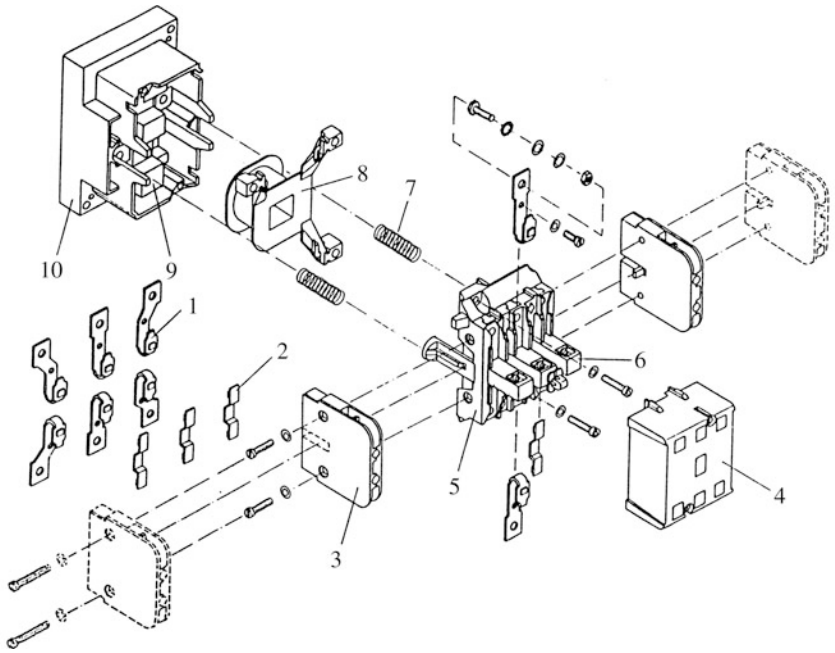
Ausgangsspannung

$$U_a = 10\,\text{V} \cdot \left(\frac{A_1}{2} + \frac{A_2}{4} + \ldots + \frac{A_8}{256} \right) \qquad (45.1)$$

Als Operationsverstärker lässt sich der LM741 verwenden. Für den Kondensator C1 reichen 0,1 µF aus. Als digitale Eingänge stehen die Anschlüsse A_1 bis A_8 zur Verfügung.

Technische Kommunikation/Technisches Zeichnen

VI

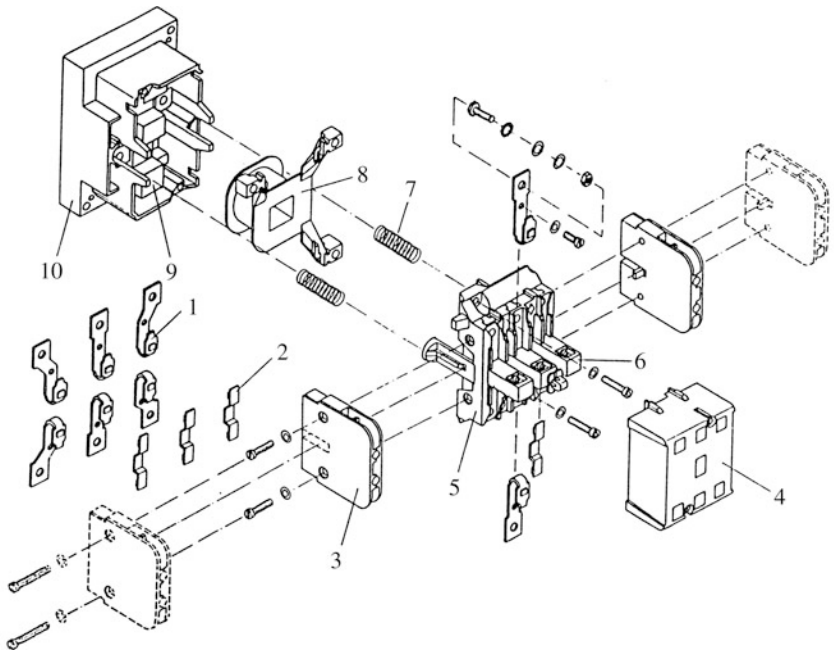


46 Grundlagen der zeichnerischen Darstellung 547
47 Schaltungsunterlagen 561
48 Schaltungssynthese und -analyse 577
49 CAD-Technik 581

Grundlagen der zeichnerischen Darstellung

46

Peter Döring



46.1 Zeichengeräte 548
46.2 Normen für Technische Zeichnungen 548
46.3 Normteile und Konstruktionselemente 549
46.4 Nutzen der Normung 559

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_46

46.1 Zeichengeräte

Eine *Technische Zeichnung* muss nach DIN 6774 Teil 1 in der Weise angefertigt werden, dass sie *übersichtlich, unmissverständlich*, auch *in verkleinertem Maßstab lesbar* bleibt, kostengünstig reproduzierbar und dauerhaft archivierbar ist. Zu dem Zweck benötigt man entsprechendes Papier und angepasstes Zeichengerät.

Für die *Entwurfzeichnung* verwendet man Bleistifte, heute in der Form *nachfüllbarer Feinminenhalter* mit Minen *definierter Härte* und Linienbreite.

Buchstaben kennzeichnen die Härtegrade von Minen:

B = schwarz (weich)
H = hart
HB = hart, schwarz (mittelhart)
F = fest.

Ziffern verweisen auf feinere Abstufungen der Härtegrade.

Zirkel, Lineale und *Schablonen erleichtern und beschleunigen die Zeichenarbeit*. Zur Erstellung der endgültigen Zeichnung werden *Tuschefüller* und Schrift- sowie Schreibschablonen verwendet, die der *Prüfnorm DIN 6775* entsprechen. Sie können für das *normgerechte Zeichnen und Beschriften nach DIN 15* Teil 1 und 2 verwendet werden.

Auf die Geräte und Maschinen, die beim *computergestützen Zeichnen* (CAD = computer aided design) Verwendung finden, soll in Kap. 49 eingegangen werden.

Zeichenpapier ist meist *dicker als normales Schreibpapier* und für die Verwendung von Tuschefüllern geeignet. Häufig wird festes, durchscheinendes *Transparentpapier* verwendet, das für das *Lichtpausverfahren geeignet* ist.

Die *Blattgröße*, sprich das Papier-Format, ist *nach DIN 6771 T6 festgelegt*. Diese Normreihe ist so gestaltet, dass durch fortgesetztes Falten bzw. Schneiden des Blattes in der Mitte jeweils das nächst kleinere Format entsteht. Das Ausgangsformat ist gleich der metrischen Flächeneinheit, d. h., A0 entspricht der Größe 1 m^2.

Alle Zeichenblattgrößen können in der Hoch- oder Querlage verwendet werden. Bevorzugt sind die Formate A4 bis A0 zu verwenden. Bei A4-Format wird die Hochlage bevorzugt.

46.2 Normen für Technische Zeichnungen

Die moderne Produktion ist gekennzeichnet durch eine extreme Arbeitsteilung. Die *Technische Zeichnung* ist hierbei als *Informationsträger über Planungsvorgaben, die Vorbereitung und Ausführung von Anlagen und Geräten* zu betrachten. Es handelt sich um Beschreibungen, die vollständig alle Fakten wiedergeben, aber keine unnötigen Angaben enthalten. Die *Darstellung ist weitestgehend symbolhaft* mit einem Minimum an Textangaben, was die internationale Kooperation vereinfacht.

Tab. 46.1 Auswahl wichtiger Normen für das technische Zeichnen

DIN EN ISO 128	Linien
DIN EN ISO 3098	Normschrift
DIN EN ISO 5455	Blattgröße, Maßstäbe
DIN EN ISO 5457	Faltungen von Technischen Zeichnungen
DIN 6771	Schriftfelder und Stücklisten
DIN 6774	Ausführungsrichtlinien
DIN ISO 5455	Maßstäbe
DIN 6789	Zeichnungssystematik
DIN ISO 286	Allgemeine Toleranzen
DIN ISO 1302	Oberflächenangaben
DIN ISO 5456	Dreidimensionale Projektion
DIN ISO 128	Ansichten, Schnittdarstellung
DIN 406	Arten und Regeln der Maßeintragung
DIN ISO 6410	Darstellung von Gewinde
DIN 461	Graphische Darstellungen
DIN EN 61082	Schaltungsunterlagen, graf. Symbole für Schaltpläne
DIN EN 61082 T. 1	Schaltungsunterlagen, Begriffe, Einteilung
DIN EN 61082 T. 2	Regeln für Stromlaufpläne, Funktions- und Schaltpläne
DIN EN 61082 T. 3	Verbindungspläne und -listen, Geräteverdrahtungspläne
DIN 40719 T. 2	Kennzeichnung von Betriebsmitteln
DIN IEC 60971	Stromrichterbenennungen und -kennzeichen
DIN VDE 0281/0293	Leitungen
DIN EN 60617	Schaltzeichen für Schaltungsunterlagen
DIN EN 60617-6	Elektrische Maschinen
DIN EN 60617-11	Messgeräte, Zähler, Anzeigen, Messgrößenumformer
DIN VDE 0530	Anschlussbezeichnungen und Drehsinn
DIN 40717	Schaltpläne, Installationspläne
DIN 40712	Veränderbarkeit, Einstellbarkeit, Widerstände
DIN 40713	Schaltglieder, Antriebe, Auslöser
DIN 40714	Transformatoren, Drosselspulen, Wandler
DIN 40719 T9	Ausführung von Anschlussplänen
DIN 40719 T10	Ausführung von Anordnungsplänen
DIN 40719 T11	Zeitablaufdiagramme – Schaltfolgediagramme
DIN 66001	Informationsverarbeitung
IEC 617-12	Computertechnik
DIN 46199	Anschlussbezeichnungen

Die *Übersichtlichkeit und Eindeutigkeit von Technischen Zeichnungen* wird durch das *strikte Einhalten von vereinbarten Regeln* erreicht. Diese Regeln werden als Zeichnungsnormen vom Deutschen Institut für Normung (DIN) herausgegeben, wobei diese die Normen und Empfehlungen der Internationalen Normenorganisation (ISO = International Organization for Standardization) zur weltweiten Vereinheitlichung berücksichtigen (Abb. 46.1).

Eine Auswahl wichtiger *Normen für das technische Zeichnen*, hier besonders für die Elektrotechnik, sind in Tab. 46.1 zusammengefasst.

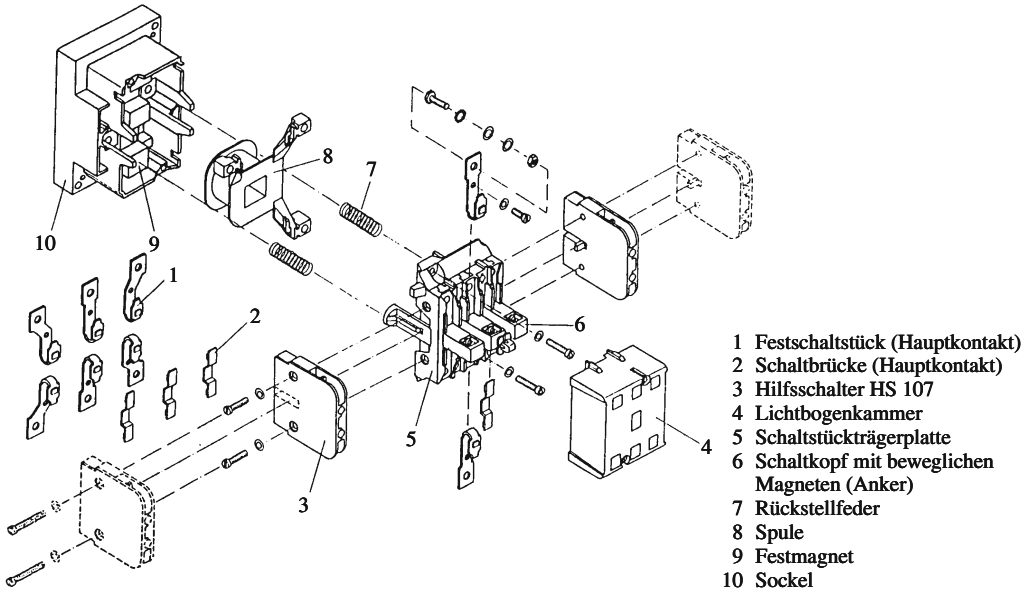


Abb. 46.1 Explosionszeichnung als Serviceplan

46.3 Normteile und Konstruktionselemente

Nach DIN 57100/IEC 364-4-41/VDE 0100 Teil 410 sind Schutzmaßnahmen gegen gefährliche Körperströme zu treffen. Zum Schutz gegen direktes Berühren von aktiven Teilen elektrischer Betriebsmittel können die aktiven Teile als solche isoliert werden oder der Schutz kann durch Abdeckungen oder Umhüllungen, sprich Gehäuse, bewirkt werden. Es handelt sich also um konstruktive Maßnahmen.

Im Zusammenhang mit dem Schutz gegen gefährliche Körperströme werden die elektrischen Betriebsmittel (Geräte und Anlagen) nach VDE 0100 in Schutzklassen, z. B. Schutzklasse II (Schutzisolierung) eingeteilt. Bei der Schutzisolierung werden elektrisch nichtleitende Materialien zur Herstellung der Gehäuse verwendet, so dass aktive Teile räumlich und elektrisch vom Benutzer bzw. Anwender völlig getrennt sind. Gehäuse aus elektrisch leitfähigen Materialien werden mit einem Anschluss für den Schutzleiter PE/PEN versehen.

Die Einteilung der Betriebsmittel in Schutzarten hat eine andere Bedeutung. Nach IEC 529 (1976)/DIN 40050 (1980) werden mechanische Eigenschaften eines Betriebsmittels festgelegt, nämlich die Abdichtung, sprich Kapselung, gegen feste Fremdkörper und gegen Wasser. Die jeweils gültige Schutzart wird durch ein Kurzzeichen angegeben, das aus den Buchstaben IP (engl.: international protection) und zwei Kennziffern nach Tab. 46.2 besteht. Die erste Kennziffer gibt den Berührungs- und Fremdkörperschutz und die zweite Kennziffer den Wasserschutz an.

Aus Betriebsgründen notwendige Gehäuseöffnungen (z. B. Steckverbindungen oder zur Kühlung erforderliche Öffnungen) müssen so beschaffen sein, dass mit dem „IEC-Prüffinger“ nach DIN 57470/VDE 0470 keine aktiven Teile berührt werden können.

Tab. 46.2 Schutzarten nach IEC 529/DIN 40050

	Schutzumfang IP xx	
	Berührungs- und Fremdkörperschutz	Wasserschutz
0	Kein Schutz gegeben	Kein Schutz gegeben
1	Schutz gegen Fremdkörper $d > 50$ mm	Schutz gegen senkrecht fallendes Tropfwasser
2	Schutz gegen Fremdkörper $d > 12$ mm	Schutz gegen schräg fallendes Tropfwasser
3	Schutz gegen Fremdkörper $d > 2{,}5$ mm	Schutz gegen Sprühwasser
4	Schutz gegen Fremdkörper $d > 1$ mm	Schutz gegen Spritzwasser
5	Schutz gegen Staubablagerung	Schutz gegen Strahlwasser
6	Schutz gegen Staubeintritt	Schutz bei Überflutung
7	–	Schutz beim Eintauchen
8	–	Schutz beim Untertauchen

Kunststoff-Leergehäuse für die Aufnahme von Schaltern und Schützen werden nach IP 54, IP 55 oder IP 65 ausgefertigt. Aus Kostengründen ist die Schutzart nicht zu hoch anzusetzen.

Zusätzliche Schutzarten nach VDE 0170/171 betreffen den Schlagwetter- und Explosionsschutz.

Die Leitungseinführung in die Gehäuse erfolgt über Kabeltüllen bzw. Kabelverschraubungen nach DIN 43620 mit normierten Bohrungen mit Pg-Gewinde oder Vorprägungen im Gehäuse.

Für die Errichtung von Verteileranlagen in Schaltanlagen gilt die Norm DIN VDE 0660/IEC 947, EN 60947, für die Errichtung von Installationsverteilern und Zählerplätze gilt DIN VDE 0603.

Tab. 46.3 Schaltglieder, Antriebe und Schaltgeräte nach DIN 40900/IEC 617

Benennung	DIN 40900/IEC 617
Schaltglieder	
Schließer	
Öffner	
Wechsler mit Unterbrechung	
Voreilender Schließer eines Kontaktsatzes	
Nacheilender Öffner eines Kontaktsatzes	
Schließer, schließt verzögert bei Betätigung	
Öffner, schließt verzögert bei Rückfall	
Antriebe elektromechanisch, elektromagnetisch	
Elektromechanischer Antrieb, allgemein, Relaisspule, allgemein	
Antrieb mit besonderen Eigenschaften, allgemein	
Elektromechanischer Antrieb mit Ansprechverzögerung	
Elektromechanischer Antrieb mit Rückfallverzögerung	
Elektromechanischer Antrieb mit Ansprech- und Rückfallverzögerung	
Elektromechanischer Antrieb eines Thermorelais	
Antriebe	
Handantrieb, allgemein	
Betätigung durch Drücken	
Betätigung durch Ziehen	
Betätigung durch Drehen	
Betätigung durch Schlüssel	
Betätigung durch Rolle, Fühler	
Kraftantrieb allgemein	
Schaltschloß mit mechanischer Freigabe	
Betätigung durch Motor	
Notschalter	
Betätigung durch elektromagnetischen Überstromschutz	
Betätigung durch thermischen Überstromschutz	
Betätigung durch elektromagnetischer Antrieb	
Betätigung durch Flüssigkeitspegel	
Steuergeräte	
Druckschalter (nicht rastend)	
Tastschalter mit Schließer und Öffner, handbetätigt durch Drücken	
Tastschalter mit Raststellung und 1 Schließer, handbetätigt durch Drücken	
Tastschalter mit Raststellung und 1 Öffner, handbetätigt durch Schlagen (z.B. Pilzdrucktaster)	
Grenz-/Endschalter	
Schließer	
Öffner	
Näherungsschalter induktiv, Schließerverhalten	
Druckwächter, öffnend	
Schwimmerschalter, schließend	
Schaltgeräte	
Schütz (Schließer)	
3poliges Schütz mit drei elektrothermischen Überstromauslösern	
3poliger Trennschalter	
3poliger Leistungsschalter	
3poliger Schalter mit Schaltschloß mit drei elektrothermischen Überstromauslösern, drei elektromagnetischen Überstromauslösern, Motorschutzschalter	

Die Gehäuse können Schnellbefestigungsschienen (Hutschiene u. ä.) für Kleinschütze, Schmelzsicherungen, Zeitrelais, LS-Automaten und weitere Kleingeräte nach DIN 46277/EN 50022 enthalten. Für Gehäuseabmessungen sollte das 25 mm Rastergrundmaß nach DIN 43660 verwendet werden, also z. B. 150, 175, 187,5, 225, . . . , 750 mm.

Schalter sind Betriebsmittel, die elektrische Geräte oder Anlagen von anderen Anlagen oder Geräten in der Weise trennen, dass sie Strompfade verbinden, unterbrechen oder trennen. Schalter werden nach VDE 0660 unterschieden hinsichtlich ihrer Bauform, ihrem mechanischen Verhalten in den Schaltstellungen, nach der Betätigungsart, nach dem Schaltvermögen, nach dem Verwendungszweck und nach der Art der Lichtbogenlöschung.

Tab. 46.3 gibt die wichtigsten grundlegenden Schaltzeichen nach DIN 40900/IEC 617 für die Darstellung von Schaltern und Schaltgeräten wieder. Nicht alle Schaltereigenschaften sind durch Schaltzeichen darstellbar.

Die Schaltglieder nach Tab. 46.3 können einzeln oder in Gruppen beliebig kombiniert werden. Der Antrieb der Schaltglieder

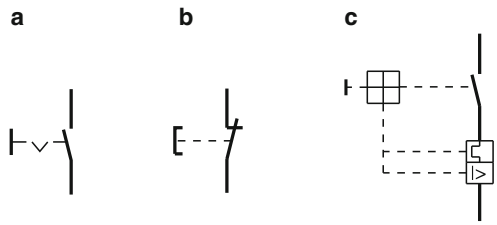


Abb. 46.2 Schaltglieder mit verschiedenen Antrieben. **a** Schließer als Rastschalter mit Handantrieb, **b** Öffner als Tastschalter, Betätigung durch Drücken, **c** einpoliger Schlossschalter mit elektrothermischer und -magnetischer Auslösung

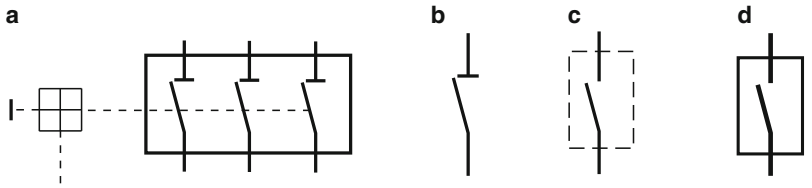


Abb. 46.3 Schaltgeräte (VDE 0660). **a** dreipoliger Leistungstrenner **b** Trennschalter, Lastschalter, Leistungsschalter

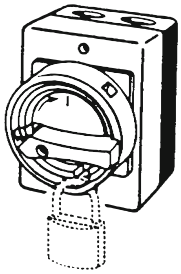


Abb. 46.4 „Abschließbarer" Hauptschalter als Reparaturschalter

kann per Hand, durch Motor, elektromagnetisch oder elektromechanisch erfolgen. Die Antriebssymbole können je nach Funktionsabsicht den Schaltgliedern zugeordnet werden.

Während bei Rastschaltern das Schaltglied in der jeweiligen Schaltstellung bleibt, kehrt das Schaltglied von Tastschaltern bei Aufhebung des Tastdruckes selbsttätig in die Ausgangsstellung zurück. Schlossschalter nach Abb. 46.2 sind Schalter mit Rückstellkraft und mechanischer Sperre, wobei eine Freiauslösung integriert ist. Die Freigabe der Sperre kann per Hand erfolgen, aber auch, wie hier dargestellt, durch elektrothermische und elektromagnetische Überstromauslösung.

Die gestrichelte Wirklinie wird häufig weggelassen, wenn die Funktion eindeutig zugeordnet ist. Mehrpolige Schlossschalter nach Abb. 46.3 werden nach DIN 40900/IEC 617 auch als Schaltgeräte bezeichnet.

Die Einteilung der Schalter nach ihrem Schaltvermögen erfolgt nach VDE 0660. Leerschalter bzw. Trennschalter dienen ausschließlich zum fast stromlosen Schalten von Stromkreisen. Sie sollten nur als NH-Sicherungstrenner vor einem Leistungsschalter verwendet werden.

Leistungsschalter haben ein Einschalt- und Ausschaltvermögen in Höhe der möglichen Kurzschlussströme, während Motorschalter zum Schalten von Motoren geeignet sind und für den Anlaufstrom der Motoren bemessen sind. Leistungsschalter werden vornehmlich an Netzschaltstellen mit Dauereinschaltung oder geringer Schalthäufigkeit verwendet. Sie können nach VDE 0113 auch als Hauptschalter verwendet werden, die für Be- und Verarbeitungsmaschinen vorgeschrieben sind und im Reparaturfall eine Maschine gänzlich vom Netz trennt. Es werden häufig Nockenschalter mit abschließbarem Antrieb nach Abb. 46.4 verwendet. Ein einfaches Einhängebügelschloss bringt hier Sicherheit.

Lastschalter haben ein Schaltvermögen bis etwa zum doppelten Nennstrom.

VDE 0660 beschreibt auch Hilfsstromschalter, die vornehmlich in Steuerstromkreisen als Taster, Steuerkontakte an Schützen und Hilfsschalter verwendet werden. An die Kontakte werden nur geringe Leistungsanforderungen gestellt.

Haupt- oder Leistungsschütze verfügen über Kontakte für große Leistungen und zusätzliche sogenannte Hilfskontakte, während Hilfsschütze ohne Leistungskontakte gebaut werden, da sie ausschließlich zu Steuerungszwecken gebaut werden.

Nach DIN EN 50005 und DIN EN 50011/50012 erfolgt die Anschlusskennzeichnung von Schützen und Hilfsschützen gemäß Abb. 46.5. Die Schützspule hat die Anschlusskennung A1-A2, die Hauptkontakte haben die Einerziffern 1–2/3–4/5–6 während die Steuerkontakte mit Doppelziffern gekennzeichnet sind. Mit der 1. Ziffer werden die Kontakte fortlaufend nummeriert (Ordnungszahl) während die 2. Ziffer einen Öffner, Schließer oder Wechsler (Funktionsziffer) kennzeichnet.

Zeitrelais haben eine Kontaktkennung nach DIN EN 50042 gemäß Abb. 46.6, während die Kontaktkennung für die Wechsler

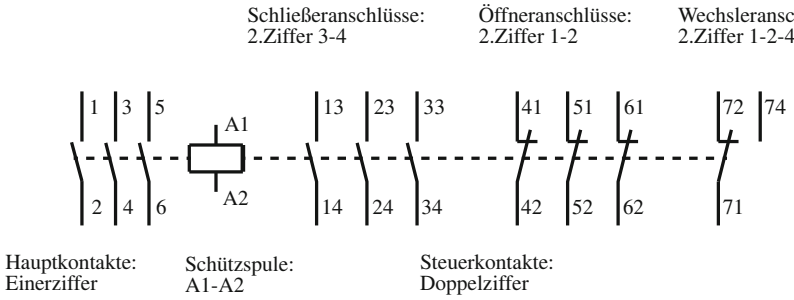


Abb. 46.5 Anschlusskennzeichnung von Schützen und Hilfsschützen nach DIN EN 50005 und DIN EN 50011/50012

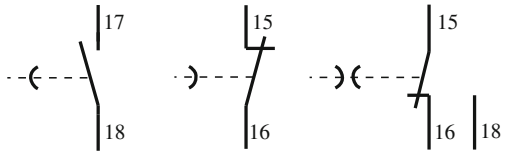


Abb. 46.6 Kennzeichnung von zeitverzögerten Steuerkontakten

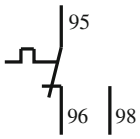


Abb. 46.7 Kennzeichnung von thermischen Überstromauslösern

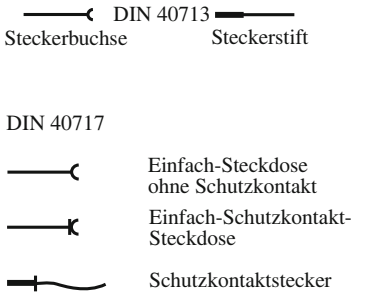


Abb. 46.8 Einfache Steckverbindungen nach DIN 40713 und DIN 40717

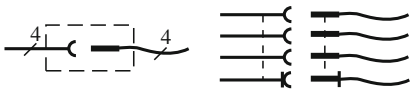


Abb. 46.9 Dreipoliger Steckverbinder mit Schutzkontakt

an thermischen Überstromauslösern (z. B. Motorschutzrelais) nach Abb. 46.7 erfolgt.

Steckvorrichtungen sind zweiteilige Schaltgeräte zum Verbinden von ortsfesten Betriebsmitteln mit beweglichen Geräten oder Leitungen. Sie bestehen aus Stecker und Steckdose bzw. Kupplung innerhalb eines Leitungszuges. Es werden Steckvorrichtungen mit und ohne Schutzleiter verwendet sowie zweipolige und vielpolige Steckvorrichtungen für vielfältige Aufgaben zum Einsatz gebracht.

Steckverbinder (Messer- und Federleisten), Steckverbinderleisten nach DIN 41612 und andere (z. B. HF- und NF-Steckverbinder nach DIN 47284) werden als Steckerbuchse(n) und Steckerstift(e) nach DIN 40713 gemäß Abb. 46.8 gezeichnet.

Damit die elektrischen Geräte verschiedener Hersteller von der Konstruktion her kompatibel sind, wurde mit DIN 41494 eine einheitliche „Bauweise für elektronische Einrichtungen„ erstellt. Diese Norm entspricht der sog. „19-Zoll“-Bauweise. Das hierbei bevorzugte Leiterplattenformat für die Messerleisten ist die sog. „Europakarte“. Normal ist, dass die Federleiste beim Einbau dem Gestell, sprich Gehäuse, zugeordnet ist.

Als zweipolige Steckvorrichtung mit Schutzleiter wird in Deutschland das sog. „Schuko®-System“ mit Steckgeräten gemäß DIN 49440 verwendet. Die zeichnerische Darstellung erfolgt nach DIN 40717 gemäß Abb. 46.8. Dreipolige Steckverbinder mit Schutzkontakt werden nach Abb. 46.9 gezeichnet.

Für die Herstellung von mehrpoligen Industriesteckvorrichtungen, bekannt unter der nicht mehr gültigen Norm CEE 17, gelten VDE 0623 Teil 1 und Teil 20, die den Normen IEC 309-1 und IEC 309-2 (ab 1995 allein gültig) entsprechen. Industriesteckvorrichtungen haben ein rundes Profil gemäß Abb. 46.10b mit einer Unverwechselbarkeitsnut bzw. -nase, die der Sicherstellung der richtigen Einstecklage des Steckers dient. Die Buchsen bzw. Stifte sind auf einem Kreis (Teilkreis) angeordnet. Die

Pos. und Bild	I in A	Stromart	Netzspannung in V	C
a	16	E	220 bis 240	6[h]
b	32		380 bis 415	9[h]
c		D	220/380 bis 240/415	6[h]
d		E	220 bis 240	6[h]
e	63		380 bis 415	9[h]
f	125	D	3× 380 bis 415	6[h]
g			220/38ß bis 240/415	6[h]

C Code für die Lage der Schutzleiterbuchse als Uhrzeigerstellung („Uhrzeit“);
E Einphasen-Wechselstrom;
D Drehstrom

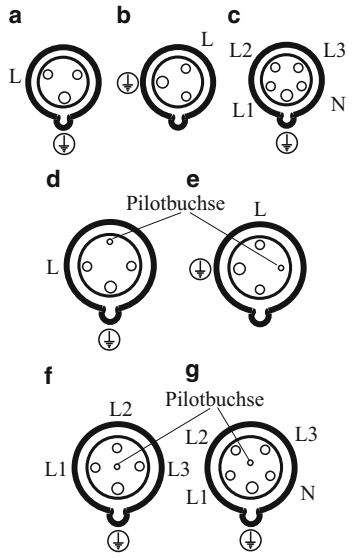


Abb. 46.10 Beispiele von Steckdosen für die genormten Netzspannungen nach DIN IEC 38

Tab. 46.4 Anwendungsklassen und Zuverlässigkeitsangaben nach DIN 40040

Kennbuchstabe	1. Stelle	2. Stelle	3. Stelle Feuchtebeanspruchung					4. Stelle Ausfallrate	5. Stelle Beanspruchungsdauer in h	6. Stelle Mechanische Beanspruchung	7. Stelle Luftdruck	
	ϑ_{min} [°C]	ϑ_{min} [°C]	Jahresmittel	30 Tage im Jahr	60 Tage im Jahr	Übrige Tage	Bemerkungen zu Bauelementen			Beispiele des Bauelementeeinsatzes	Untere Druckgrenze in mbar	Max. Höhe über NN in m
A		400	≤ 100				BE dauernd nass					
B		350										
C		300	≤ 95	100		100	BE betaut					
D		250	≤ 80	100		90	BE betaut	0,1				
E	−65	200	≤ 75	95		85	BE leicht betaut	0,3				
F	−55	180	≤ 75	95		85	BE nicht betaut	1				
G	−40	170	≤ 65		85	75	BE nicht betaut	3				
H	−25	155	≤ 50		75	65	BE nicht betaut	10				
I	−10	140	≤ 50	≤ 50	≤ 50	≤ 50		30				
K	0	125						100				
L	+5	110						300				
M		100						1000				
N		90						3000			840	1000
P		85						10.000				
Q		80						30.000	300.000	An Verbrennungsmotoren angebaut		
R		75	≤ 90	100		95	BE betaut	100.000	100.000	Für Bordelektronik	700	2200
S		70						300.000	30.000	Für Bordelektronik	600	3500
T		65						1.000.000	10.000	In tragbaren Geräten	530	4300
U		60						3.000.000	3000	In Schiffsanlagen & Autoradios	300	8500
V		55						10.000.000	1000	In ortsfesten Stromerzeugern & nicht erschütterungsfreien Anlagen	85	16.000
W		50						30.000.000	300	In erschütterungsfreien Geräten & Anlagen	44	20.000
Y		40									20	26.000
Z												

Schutzkontakte müssen beim Einführen des Steckers früher Kontakt geben und sich später trennen als die den Betriebsstrom führenden Kontakte. Die Lage der Buchsen und Stifte für die genormten Netzspannungen nach DIN IEC 38 stellt Abb. 46.10a dar. Hier wird die Unverwechselbarkeit zusätzlich durch die Lage der Kontakte gewährleistet.

Sogenannte Pilotkontakte nach Abb. 46.10d bis 46.10g sorgen bei Bedarf dafür, dass die Steckverbindung nur im spannungslosen Zustand betätigt werden kann. Diese Kontakte schließen beim Einstecken nach den übrigen Kontakten bzw. öffnen beim Herausziehen vor den übrigen Kontakten, so dass z. B. nach Abb. 46.11 eine Schützschaltung die Leitung spannungslos schalten kann.

Elektronische Bauelemente werden nach ihrer technischen Funktion und Leistung unter Berücksichtigung ihrer individuellen Grenz- und Kennwerte ausgewählt. Zusätzlich werden sie aber auch nach dem Grad ihrer klimatischen und mechanischen Beanspruchung und hinsichtlich ihrer Zuverlässigkeit, sprich ihrer Ausfallrate bestimmt.

Nach DIN 40040 werden die „Anwendungsklassen und Zuverlässigkeitsangaben für Bauelemente der Nachrichtentechnik

Tab. 46.5 Buchstabenkombination und ihre Bedeutung

1. Stelle	Untere Grenztemperatur	Klimatische Anwendungsklasse
2. Stelle	Obere Grenztemperatur	
3. Stelle	Feuchtebeanspruchung	
4. Stelle	Ausfallquotient	Zuverlässigkeitsangabe
5. Stelle	Beanspruchungsdauer	
6. Stelle	Mechan. Beanspruchung	Mechan. Anwendungsklasse
7. Stelle	Luftdruck	
8. Stelle	Sonderbeanspruchung	

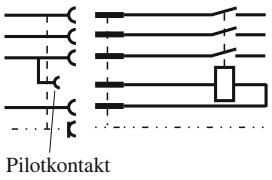


Abb. 46.11 Steckverbinder mit Pilotkontakt

und Elektronik„ bestimmt und gekennzeichnet. Tab. 46.4 bietet eine Übersicht. Zur Kennzeichnung der Anwendungsklassen werden drei bis achtstellige Buchstabenkombinationen verwendet, die entsprechend dem Schlüssel nach Tab. 46.5 zusammengefügt werden:

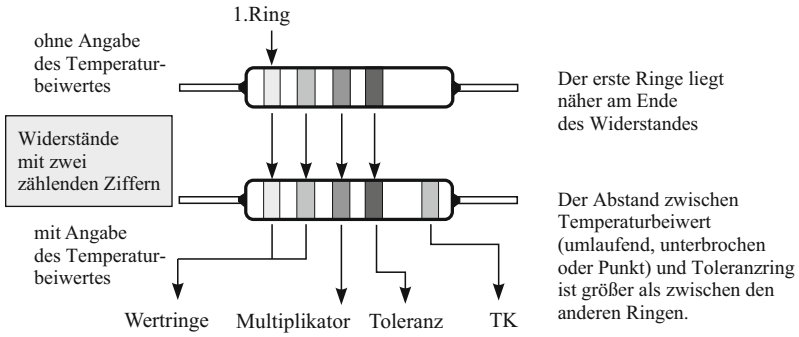
Über Stellen, die den *Kennbuchstaben X* tragen, werden *keine Angaben* gemacht. Der *Kennbuchstabe Z* weist auf *Einzelbestimmungen* hin, die den Herstellerangaben direkt zugeordnet werden müssen.

Widerstände und Kondensatoren werden als einstellbare (im Wert variierbare) Bauteile hergestellt, aber auch als Festwiderstände bzw. Festkondensatoren.Die Kennzeichnung erfolgt durch Farbcode nach DIN 41429 oder durch Zahlen und Buchstaben nach DIN 40825. Bei DIN 41429 kann die Codierung mit zwei oder drei zählenden Ziffern erfolgen. In Abb. 46.12 ist die Kennzeichnung von Widerständen in der ersten Version hinsichtlich der Leserichtung und der Codierung gemäß Farbauswahl dargestellt.

Da bei der Herstellung von Widerständen und Kondensatoren bei großer Serien- und Massenfertigung die ausgewiesenen Nennwerte ohne erhöhte Kosten nicht genau eingehalten werden können, erhalten diese unter Berücksichtigung ihrer Funktion und der erforderlichen Genauigkeit maximal zugelassene Abweichungen, Toleranzen. Von festgelegten Toleranzen ausgehend werden nach DIN 41426 Staffelwerte für sogenannte E-Reihen (Internationale Normreihen) berechnet als geometrische Reihen.

Es sind folgende E-Reihen genormt:

$$\text{E6} = \pm 20\,\%; \quad \text{E12} = \pm 10\,\%; \quad \text{E24} = \pm 5\,\%$$
$$\text{E48} = \pm 2\,\%; \quad \text{E96} = \pm 1\,\%; \quad \text{E192} = \pm 0{,}5\,\%$$



Kennfarbe	Wertziffer	Multiplikator	Toleranz	Temperaturbeiwert a_R $10^{-6}\,K^{-1}$	Bemerkungen
keine	–	–	±20	–	Die Farben gold und silber sind leitend und deshalb nicht immer verwendbar. Ausweichmöglichkeiten: statt gold: für 10^{-1} weiß, für ±5% grün; statt silber: für 10^{-2} grau, für ±10% weiß
silber	–	$\times 10^{-2}\ \Omega$ = 0,01 Ω	±10	–	
gold	–	$\times 10^{-1}\ \Omega$ = 0,1 Ω	± 5	–	
schwarz	0	$\times 10^{0}\ \Omega$ = 1,0 Ω	–	±200	
braun	1	$\times 10^{1}\ \Omega$ = 10 Ω	±1	±100	
rot	2	$\times 10^{2}\ \Omega$ = 100 Ω	±2	± 50	
orange	3	$\times 10^{3}\ \Omega$ = 1 kΩ	–	± 15	
gelb	4	$\times 10^{4}\ \Omega$ = 10 kΩ	–	± 25	
grün	5	$\times 10^{5}\ \Omega$ = 100 kΩ	± 0,5 %	–	
blau	6	$\times 10^{6}\ \Omega$ = 1 MΩ	± 0,25 %	–	
violett	7	$\times 10^{7}\ \Omega$ = 10 MΩ	± 0,1 %	–	
grau	8	$\times 10^{8}\ \Omega$ = 100 MΩ	–	–	
weiß	9	$\times 10^{9}\ \Omega$ = 1000 MΩ	–	–	

Abb. 46.12 Kennzeichnung von Widerständen durch Farbcode (Auszug)

E-Reihen (Auszug)																								Nach DIN 41426
E6	1,0				1,5				2,2				3,3				4,7				6,8			
E12	1,0		1,2		1,5		1,8		2,2		2,7		3,3		3,9		4,7		5,6		6,8		8,2	
E24	1,0	1,1	1,2	1,3	1,5	1,6	1,8	2,0	2,2	2,4	2,7	3,0	3,3	3,6	3,9	4,3	4,7	5,1	5,6	6,2	6,8	7,5	8,2	9,1
E48	1,00 1,05 1,10 1,15		1,21 1,27 1,33 1,40		1,47 1,54 1,62 1,69		1,78 1,87 1,96 2,05		2,15 2,26 2,37 2,49		2,61 2,74 2,87 3,01		3,16 3,32 3,48 3,65		3,83 4,02 4,22 4,42		4,64 4,87 5,11 5,36		5,62 5,90 6,19 6,49		6,81 7,15 7,50 7,87		8,25 8,66 9,09 9,53	
DIN-Reihe																								Nach DIN 323
R5	1,00				1,60				2,50				4,00				6,30							
R10	1,00		1,25		1,60		2,00		2,50		3,15		4,00		5,00		6,30		8,00					
R20	1,00	1,12	1,25	1,40	1,60	1,80	2,00	2,24	2,50	2,80	3,15	3,55	4,00	4,50	5,00	5,60	6,30	7,10	8,00	9,00				
R40	1,00 1,06	1,12 1,18	1,25 1,32	1,40 1,50	1,60 1,70	1,80 1,90	2,00 2,12	2,24 2,36	2,50 2,65	2,80 3,00	3,15 3,35	3,55 3,75	4,00 4,25	4,50 4,75	5,00 5,30	5,60 6,00	6,30 6,70	7,10 7,50	8,00 8,50	9,00 9,50				
Bei jeder Reihe können die üblichen Toleranzen gewählt werden.																								

Abb. 46.13 E-Reihen und DIN-Reihen (Auszüge)

Schicht-Festwiderstände							
Schichtmaterial	Kohle			Kohle-gemisch	Metall	Metall-oxid	Metall-glasur
Anforderungen	allgemein	erhöht	erhöht	allgemein	erhöht	erhöht	erhöht
DIN	44051 (983)	44052 (9.83)	44055 (9.83)	44054 (9.83)	44061 (9.83)	44063 (9.83)	44064 (9.83)
Anwendungsklasse	FKF (FHF)	FKF	FKF	FKF	FKF	FHF FZF	FHF
Widerstandswertereihe	E24	E24	E24	E24	E24	E24	E24
Widerstands-abweichung bei Anlieferung	±2 % ±5 %	±2 % ±5 %	±1 % ±2 %	± 2 % ±10 %	±0,5 % ±1 % ±2 %	±2 % ±5 %	±0,5 % ±1 % ±2 %
Temperatur-koeffizient in 10^{-6}/K (zwischen 20 °C und 70 °C)	–150 bis –1500			±1300	B[1]): 0 ± 100 C: 0 ± 50 D: 0 ± 25 E: 0 ± 15	±250	B[1]): 0 ± 100 C: 0 ± 50
ΔR_{zul} nach 1000 h Dauerprüfung bei P_{70}	≤ ±(5 % · R + 0,1 Ω) bis 1 MΩ +10 % · R –5 % · R über 1 MΩ	≤ ±(2 % · R + 0,05 Ω) bis 1 MΩ +4 % · R –2 % · R über 1 MΩ	≤ ±(1 % · R + 0,05 Ω) bis 1 MΩ +2 % · R –1 % · R über 1 MΩ	≤ ±4 % · R –2 · R +0,1 Ω	≤ +1 % · R –0,5 % · R +0,05 Ω	≤ ±2 % · R +0,1 Ω	≤ ±1 % · R +0,1 Ω
ΔR_{zul} nach 8000 h Dauerprüfung bei P_{70}	≤ (+10 % · R –1 % · R + 0,1 Ω) bis 1 MΩ +20 % · R –5 % · R über 1 MΩ	≤ (+4 % · R –2 % · R + 0,05 Ω) bis 1 MΩ +8 % · R –2 % · R über 1 MΩ	≤ (+2 % · R –1 % · R + 0,05 Ω) bis 1 MΩ +4 % · R –1 % · R über 1 MΩ	≤ (+4 % · R –15 % · R +0,1 Ω)	≤ (+2 % · R –0,5 % · R +0,05 Ω)	≤ (+4 % · R –3 % · R +0,1 Ω)	≤ (+2 % · R +0,1 Ω)
[1]) Diese Buchstaben sind der 2. Kennbuchstabe der Baugrößebeziehung.							

Abb. 46.14 Übersicht über Schicht-Festwiderstände

Außer nach E-Reihen sind Widerstandswerte ebenfalls nach sog. (DIN-)R-Reihen gemäß DIN 323 genormt. Abb. 46.13 stellt Auszüge aus E- und DIN-Reihen dar.

Eine Übersicht über Schicht-Festwiderstände auf der Basis des bisher gesagten liefert Abb. 46.14. Die Bauformen von Widerständen nach Abb. 46.15 sind nur teilweise genormt. Dies gilt auch für Kondensatoren nach Abb. 46.16, wobei hier die Anzahl aufgrund der verwendeten Materialien größer ist.

Halbleiter-Bauelemente wie Dioden, Transistoren und ICs sind ebenfalls nur z. T. genormt. Bei den ICs und den Bausteinen aus dem Computerbau ist die Normung der Bauformen am weitesten fortgeschritten, was die Kompatibilität von Mikrochips erheblich erhöht. Die Übersicht nach Abb. 46.17 zeigt nur eine Auswahl an genormten und nicht genormten Bauformen, wobei zu beachten ist, dass zahlreiche Bauformen das Rastermaß von 2,54 mm einhalten, was die Layouterstellung vereinfacht.

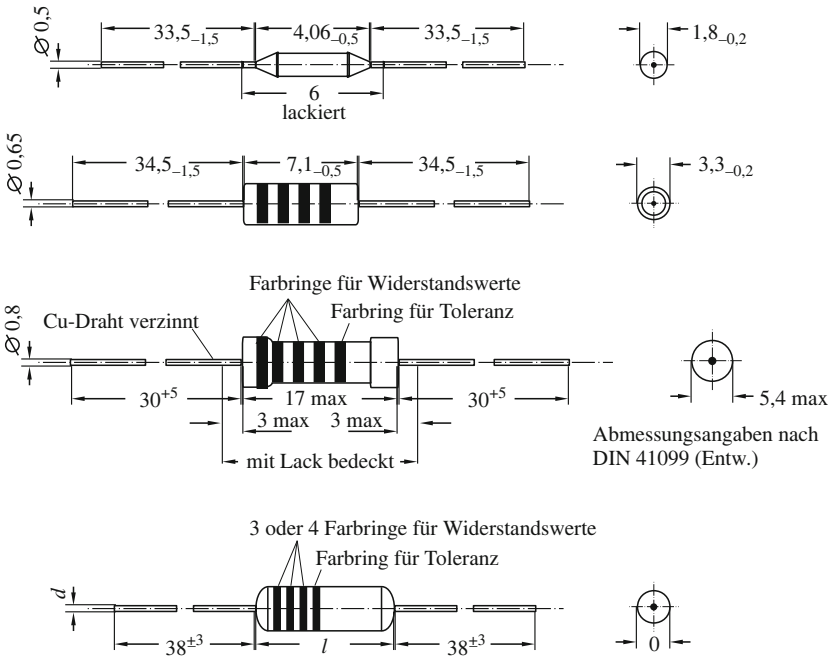


Abb. 46.15 Bauformen von Widerständen (teilweise genormt)

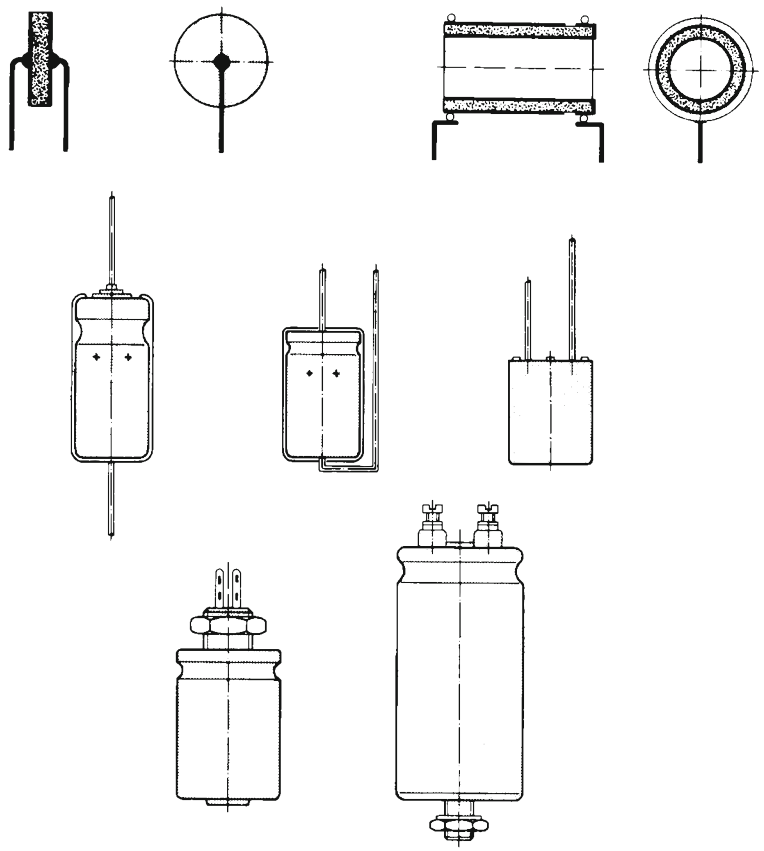


Abb. 46.16 Auswahl an Kondensator-Bauformen (teilweise genormt)

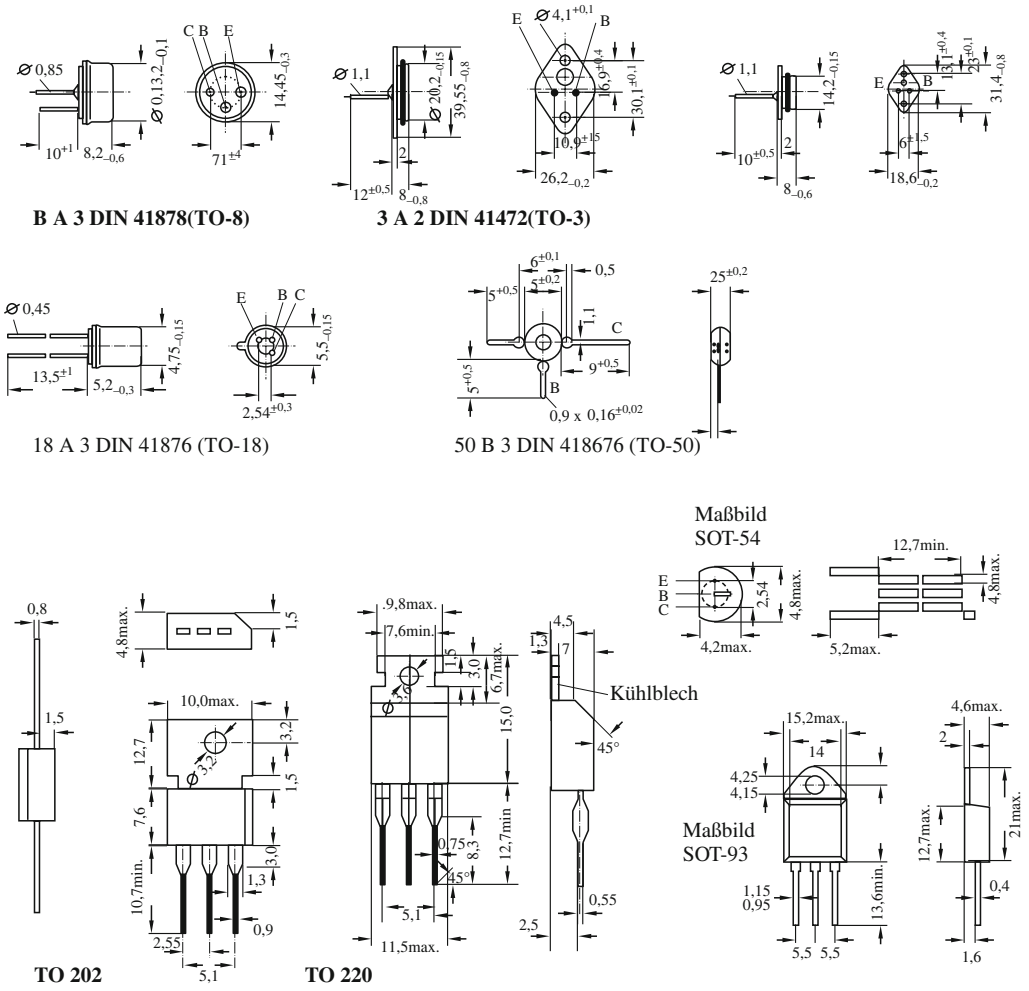


Abb. 46.17 Bauformen von Dioden, Transistoren und ICs (teilweise genormt)

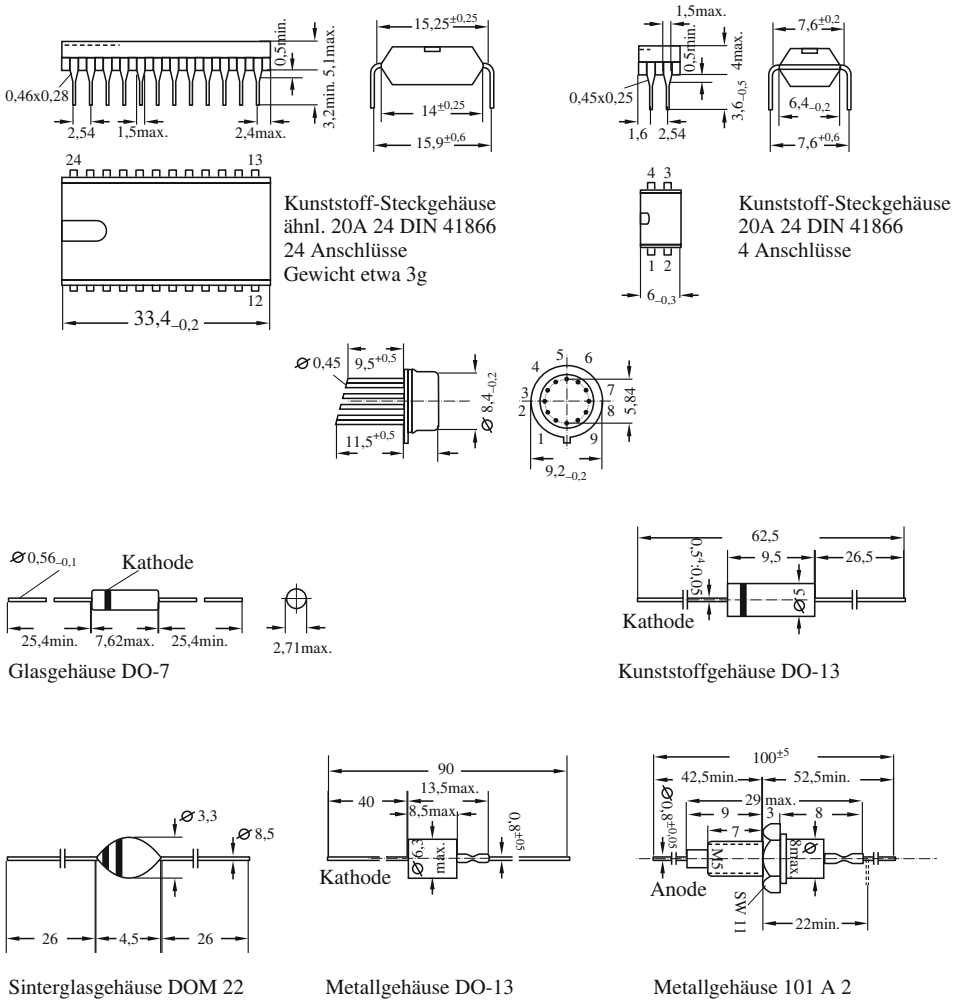


Abb. 46.17 (Fortsetzung)

46.4 Nutzen der Normung

Die *Technische Zeichnung als Informationsträger* dient bei einem Fertigungs- bzw. Erstellungsauftrag als Verständigungsmittel zwischen der Planungs- bzw. Projektierungsstelle und der Werkstatt bzw. der Montagestelle. Durch das durchgängige Verwenden der Zeichnungsnormen wird die *Zeichenarbeit erleichtert* und eine *klare Darstellung des Werkstückes bzw. der Anlage* erreicht.

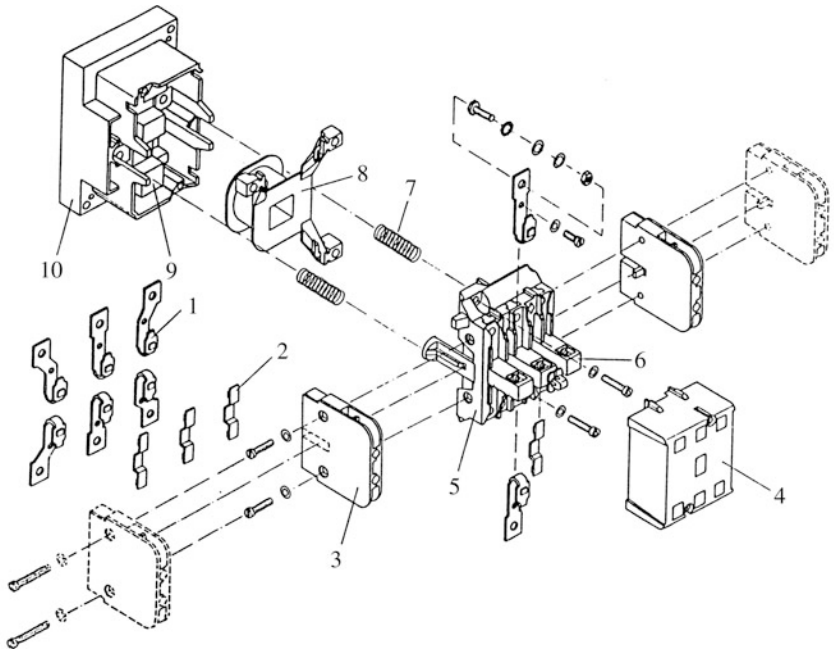
Die Normung stellt durch Vereinheitlichung eine zweckdienliche Ordnung von Erzeugnissen und Verfahren her. Sie legt die Form, Größe und Ausführung gleichartiger Erzeugnisse fest, so dass eine beliebige Austauschbarkeit von Teilen herbeigeführt wird. Dieser Umstand ermöglicht die Serien- und Massenfertigung und die damit verbundene Automatisierung. Normen sind bewährte Lösungen für häufig wiederkehrende Aufgaben.

Technische Fortschritte können immer nur mit zeitlichem Verzug eingearbeitet werden, so dass die Normen eher konservativ denn innovativ sind.

Schaltungsunterlagen

47

Peter Döring



47.1 Schaltzeichen nach DIN 564
47.2 Elektrische Betriebsmittel 566
47.3 Schaltungsunterlagen der Energietechnik 566
47.4 Schaltungsunterlagen der Elektronik 571
47.5 Projektierung 575

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_47

Tab. 47.1 Wichtige VDE-Bestimmungen

VDE 0100	Bestimmungen für das Errichten von Starkstromanlagen bis 1000 V
VDE 0101	Bestimmungen für das Errichten von Starkstromanlagen über 1 kV
VDE 0102	Leitsätze für die Berechnung der Kurzschlussströme
VDE 0105	VDE-Bestimmungen für den Betrieb von Starkstromanlagen
VDE 0107	Bestimmungen für elektrische Anlagen in medizinisch genutzten Räumen
VDE 0108	Bestimmungen für das Errichten und den Betrieb von Starkstromanlagen in Versammlungsstätten, Waren- und Geschäftshäusern, Hochhäusern, Beherbergungsstätten und Krankenhäusern
VDE 0128	Vorschriften für Leuchtröhrenanlagen mit Spannungen von 1000 V und darüber Merkblatt für den Betrieb elektrischer Anlagen in landwirtschaftlichen Betrieben
VDE 0132	Merkblatt für die Bekämpfung von Bränden in elektrischen Anlagen
VDE 0134	Anleitungen zur Ersten Hilfe bei Unfällen (VDE-Druckschrift)
VDE 0141	Bestimmungen und Richtlinien für Erdungen in Wechselstromanlagen für Nennspannungen über 1 kV
VDE 0160	Bestimmungen für die Ausrüstung von Starkstromanlagen mit elektronischen Betriebsmitteln
VDE 0165	Bestimmungen für die Errichtung elektroscher Anlagen in explosionsgefährdeten Betriebsstätten
VDE 0168	Bestimmungen für das Errichten und den Betrieb elektrischer Anlagen in Tagebauen, Steinbrüchen und ähnlichen Betrieben
VDE 0190	Bestimmungen für das Einbeziehen von Rohrleitungen in Schutzmaßnahmen von Starkstromanlagen mit Nennspannungen bis 1000 V
VDE 0193	Richtlinien für den Anschluss und die Anbringung von Elektroden-Durchlauferhitzern
VDE 0210	Bestimmungen für den Bau von Starkstrom-Freileitungen über 1 kV
VDE 0211	Bestimmungen für den Bau von Starkstrom-Freileitungen mit Nennspannungen bis 1000 V
VDE 0228	VDE-Bestimmungen für Maßnahmen bei Beeinflussung von Fernmeldeanlagen durch Starkstromanlagen
VDE 0410	Regeln für elektrische Messgeräte
VDE 0411	VDE-Bestimmungen für elektronische Messgeräte und Regler
VDE 0413	Bestimmungen für Geräte zum Prüfen der Schutzmaßnahmen in elektrischen Anlagen
VDE 0414	Bestimmungen für Messwandler
VDE 0426	Bestimmungen für einpolige Spannungssucher bis 250 V Wechselspannung gegen Erde
VDE 0510	Bestimmungen für Akkumulatoren und Akkumulatoren-Anlagen
VDE 0530	Bestimmungen für umlaufende elektrische Maschinen
VDE 0532	Bestimmungen für Transformatoren und Drosselspulen
VDE 0541	Bestimmungen für Stromquellen zum Lichtbogenschweißen mit Wechselstrom
VDE 0550	Bestimmungen für Kleintransformatoren
VDE 0612	BDE-Bestimmungen für Baustromverteiler für Nennspannungen bis 380 V Wechselspannung und für Ströme bis 630 A
VDE 0620	Vorschriften für Steckvorrichtungen bis 750 V 100 A
VDE 0660	Bestimmungen für Niederspannungsschaltgeräte
VDE 0675	Leitsätze für den Schutz elektrischer Anlagen gegen Überspannungen
VDE 0680	Bestimmungen für Schutzbekleidung, Schutzvorrichtungen und Werkzeuge zum Arbeiten an unter Spannung stehenden Betriebsmitteln
VDE 0701	Bestimmungen für die Instandsetzung, Änderung und Prüfung gebrauchter elektrischer Verbrauchsmittel (Geräte)
VDE 0710	Vorschriften für Leuchten mit Betriebsspannungen unter 1000 V
VDE 0712	Bestimmungen für Entladungslampenzubehör mit Nennspannungen bis 1000 V
VDE 0800	Bestimmungen für Errichtung und Betrieb von Fernmeldeanlagen einschließlich Informationsverarbeitungsanlagen
VDE 0855	Bestimmungen für Antennenanlagen
VDE 0860	Bestimmungen für netzbetriebene Rundfunk- und verwandte elektronische Geräte
VDE 0871	Bestimmungen für die Funk-Entstörung von Hochfrequenzgeräten und -anlagen
VDE 0874	VDE-Leitsätze für Maßnahmen zur Funk-Entstörung
VDE 0875	Bestimmungen für die Funk-Entstörung von Geräten, Maschinen und Anlagen für Netzfrequenzen von 0 bis 10 kHz
VDE 0877	Leitsätze für das Messen von Funkstörungen

Beim Zeichnen von Schaltplänen werden alle zugehörigen Maschinen, Geräte, Schaltteile und Leitungen nicht maßstabsgetreu und der tatsächlichen Ausführung ähnlich gezeichnet, sondern durch *genormte graphische Symbole* dargestellt.

Bei der Erstellung von *Schaltungsunterlagen für die Elektrotechnik* sind zusätzlich zur rein zeichnerischen Darstellung die einschlägigen *elektrotechnischen Vorschriften und Normen*, insbesondere DIN 57100/VDE 0100, zu beachten.

Normen werden erstellt und publiziert von der Internationalen Elektrotechnischen Kommission (IEC) in *IEC-Publikationen* und vom Europäischen Komitee für elektrotechnische Normung (CENELEG) in *Europäischen Normen (EN)* für nahezu alle europäische Länder, z. B. in DIN EN 50005 „Industrielle Niederspannungsgeräte, Anschlussbezeichnungen und Kennzahlen„ oder in sog. Harmonisierungsdokumenten, z. B. HD 21 als DIN 57281/VDE 0281 „VDE-Bestimmungen für Starkstromleitungen“.

Für nationale Normung ist die Deutsche Elektrotechnische Kommission im DIN und *VDE (Verband Deutscher Elektrotechniker)* zuständig, wobei möglichst IEC-Publikationen mit eingearbeitet oder übernommen werden bzw. auch umgekehrt.

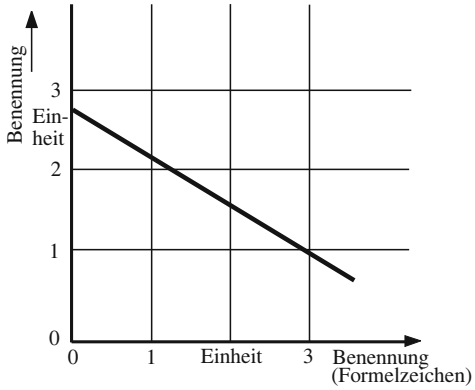


Abb. 47.1 Diagrammdarstellung nach DIN 461 und DIN 40719

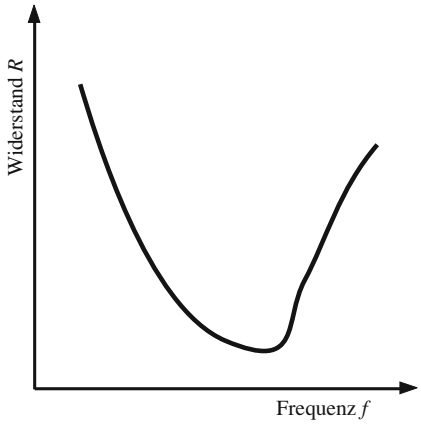


Abb. 47.2 Diagramm im kartesischen Koordinatensystem

Eine Auswahl der wichtigsten VDE-Bestimmungen zeigt Tab. 47.1.

Schaltungsunterlagen sind Schaltpläne, Tabellen, Diagramme und Beschreibungen, die Angaben für das Fertigen, Errichten und die Erhaltung elektrischer Anlagen vermitteln.

Die Norm *DIN 40719* legt die *Einteilung von Schaltungsunterlagen* und die entsprechenden Begriffe, wie sie in der Elektrotechnik üblich sind, fest. Schaltungsunterlagen werden in zwei Gruppen eingeteilt und zwar nach dem Zweck und der Art der Darstellung.

Ein *Schaltplan (engl.: diagram)* zeigt die Art, in der die verschiedenen elektrischen Betriebsmittel zueinander in Beziehung stehen und miteinander verbunden sind.

Ein *Diagramm (engl.: chart)* ist die graphische Darstellung errechneter oder gemessener Werte und sollte nach DIN 461 dargestellt werden. Die Linienbreiten werden nach DIN 15 etwa im Verhältnis Netz zu Achsen zu Kurve wie 1 : 2 : 4 gewählt in der Darstellung nach Abb. 47.1.

Prinzipiell unterscheidet man *Diagramme im kartesischen Koordinatensystem* (Abb. 47.2, 47.3 und 47.4), *Diagramme im Polarkoordinatensystem* nach Abb. 47.5, *Flächendiagramme* nach Abb. 47.6 und *Nomogramme* nach Abb. 47.7.

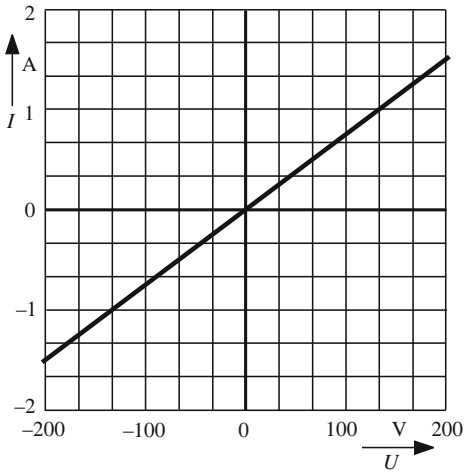


Abb. 47.3 Diagramm mit linearer Einteilung

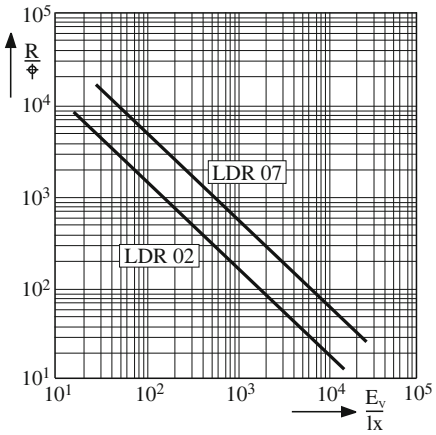


Abb. 47.4 Diagramm mit doppelt-logarithmischer Einteilung

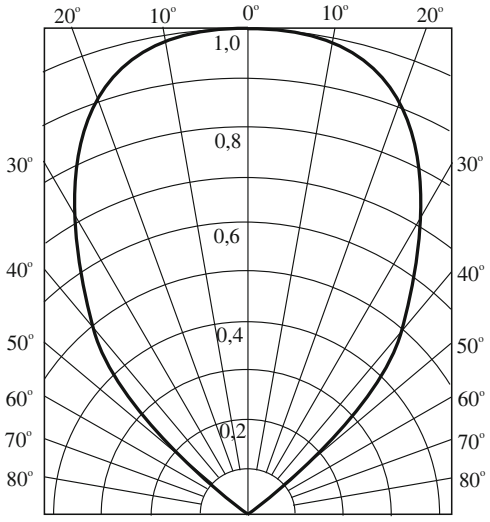


Abb. 47.5 Diagramm im Polarkoordinatensystem

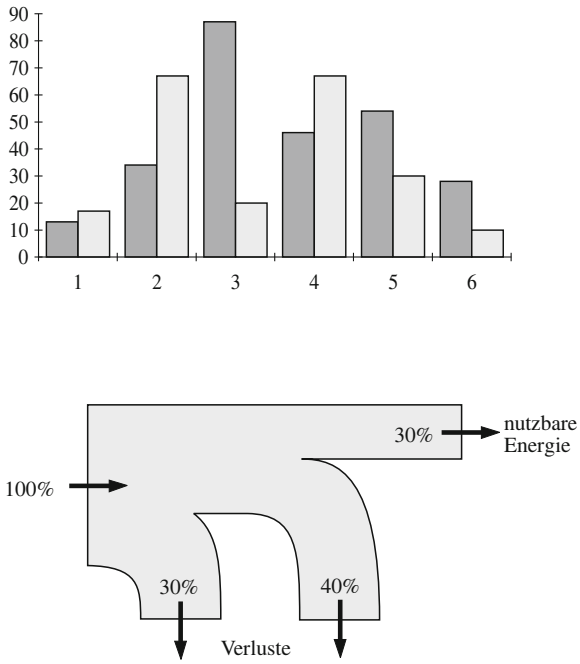


Abb. 47.6 Flächendiagramme



Abb. 47.7 Nomogramm

Das Übersichtsdiagramm (z. B. charakteristische Darstellung) nach Abb. 47.2 wird ohne Achseneinteilung gezeichnet; auf eine nichtlineare Teilung ist hinzuweisen.

Diagramme zur quantitativen Darstellung erhalten an den Achsen eine bezifferte Teilung, wobei jeder negative Wert mit dem entsprechenden Vorzeichen zu versehen ist. Mindestens der erste und der letzte Teilstrich ist mit dem Wert der Einteilung zu versehen. Die *Einteilung kann linear* nach Abb. 47.3, *doppelt-logarithmisch* nach Abb. 47.4 erfolgen oder auch in *linear-logarithmischer* Einteilung. Die *Einheit* darf *nicht in Klammern* gesetzt werden.

Bei mehreren Kurven nach Abb. 47.4 werden an jede Kennlinie der Parameter oder andere Hinweise mit evtl. näherer Erläuterung in der Bildunterschrift verwendet.

47.1 Schaltzeichen nach DIN

In DIN 40700 bis DIN 40717, letztere teilweise ersetzt durch DIN 40900, sind die Schaltzeichen als „Graphische Symbole für Schaltungsunterlagen“ der Elektrotechnik mit ihrer Benennung und Bedeutung festgelegt.

DIN 40900 Teil 1 bis Teil 13, „Graphische Symbole für die Elektrotechnik“, liegt in der harmonisierten Fassung IEC 617

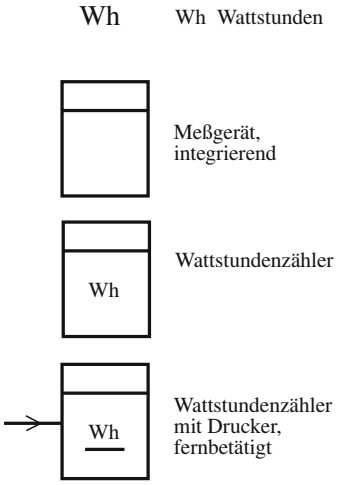


Abb. 47.8 Schaltzeichen aus Symbol- und Grundelementen

Teil 1 bis Teil 13 der Internationalen Elektrotechnischen Kommission (IEC) als Norm vor.

Betriebsmittel einer Schaltung werden durch genormte Schaltzeichen dargestellt. Es muss das Schaltzeichen gewählt werden, das für eine beabsichtigte Aussage gerade ausreichend ist, also ist das Schaltzeichen so einfach wie möglich zu halten. Ein gewähltes Schaltzeichen sollte durchgängig in der Zeichnung verwendet werden.

Schaltzeichen setzen sich aus Symbolelementen und Grundelementen wie in Abb. 47.8 zusammen. Blocksymbole nach Abb. 47.9 sind vereinfachte Darstellungen von Funktionseinhei-

Tab. 47.2 Schaltzeichen nach DIN 40900 (Auszug)

Passive Bauelemente	DIN 40900 Teil 4
Widerstand, allgemein	
Widerstand, veränderbar, allgemein	
Widerstand mit Schleifkontakt	
Widerstand, spannungsabhängig Varistor	
Widerstand, temperaturabhängig Heißleiter	
Kondensator, allgemein	
Kondensator, gepolt, z. B. Elektrolyt-Kondensator	
Kondensator, veränderbar	
Spule, Wicklung, Induktivität	
Spule mit Magnetkern	
Piezoelektrischer Kristall, Schwingquarz	

Halbleiter-Bauelemente	DIN 40900 Teil 5
Halbleiterdiode, allgemein	
Z Diode, Esaki-Diode	
Kapazitätsdiode	
Tunneldiode	
Zweirichtungsdiode, Diac	
Thyristortriode, Kathode gesteuert	
Thyristortriode, bidirektional Triac	
NPN-Transistor	
Unijunction Transistor, Basis N-Typ	

Tab. 47.2 (Fortsetzung)

Halbleiter-Bauelemente	DIN 40900 Teil 5
Sperrschicht-FET mit N-Kanal (JFET)	
Isolierschicht-FET, Anreicherungstyp mit N-Kanal (IGFET, MOS-FET)	
Isolierschicht-FET, Verarmungstyp mit N-Kanal (IGFET, MOS-FET)	
Hall-Generator	

Fotoelektrische Bauelemente	DIN 40900 Teil 5
Fotowiderstand	
Fotodiode	
Fotoelement Fotozelle	
Fototransistor NPN-Typ	
Leuchtdiode	
Optokoppler	

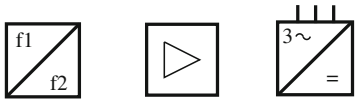


Abb. 47.9 Blocksymbol als Funktionseinheit

ten oder Baueinheiten durch ein einzelnes Schaltzeichen. Wenn für ein konkretes Betriebsmittel kein genormtes Schaltzeichen existiert, so kann durch die Kombination von Grundsymbolen, Symbolelementen, Kennzeichen oder Schaltzeichen ein neues Schaltzeichen entworfen werden.

Die Lage, in der Schaltzeichen in den Normblättern dargestellt sind, ist nicht die einzig gültige. Schaltzeichen dürfen nach Abb. 47.10 gedreht oder gespiegelt werden, wenn ihre Bedeutung dadurch nicht verändert wird. Werden Schaltzeichen verkleinert oder vergrößert, so sollen ihre Proportionen wie in Abb. 47.11 erhalten bleiben.

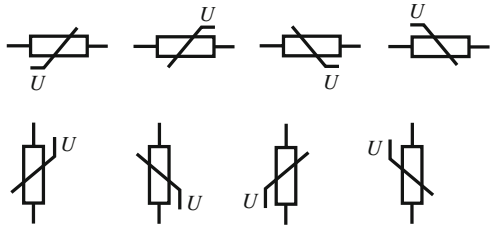


Abb. 47.10 Variable Lage von Schaltzeichen im Schaltplan

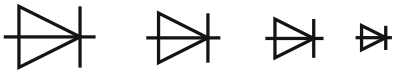


Abb. 47.11 Darstellungsgrößen von Schaltzeichen

47.2 Elektrische Betriebsmittel

Die Kennzeichnung von Betriebsmitteln erfolgt nach IEC 750 (1983)/DIN 40719 Teil 2 (1978) und stellt die Beziehung her zwischen dem Betriebsmittel in der Anlage und den verschiedenen Schaltungsunterlagen, wie z. B. Schaltplänen, Stücklisten, Stromlaufplänen und Anweisungen. Sie erfolgt in Kennzeichnungsblöcken mit Vorzeichen zur sicheren Unterscheidung der Blöcke (Tab. 47.3 und 47.4).

Bei senkrechtem Leitungsverlauf nach Abb. 47.12 steht die Betriebsmittelkennzeichnung links und die Anschlusskennzeichnung rechts; bei waagerechtem Verlauf dagegen steht die Betriebsmittelkennzeichnung unten.

Kann auf die Kennzeichnung der Anlage und des Ortes verzichtet werden, so kann die Kennzeichnung der Betriebsmittel und der Anschlüsse nach Abb. 47.13 erfolgen. In den meisten Fällen reicht der Kennzeichnungsblock 3 aus.

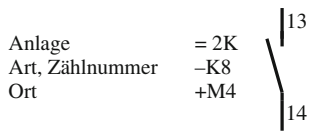


Abb. 47.12 Lage der Betriebsmittelkennzeichnung im Plan

Tab. 47.3 Kennzeichnungsblöcke nach DIN 40719

Kennzeichnungsblock				Vorzeichen	Beispiel	Erklärung zum Beispiel
1	Anlage			=	= B3	Anlage B3
2	Ort			+	+ D4	Stockwerk D, Raum 4
3	Art	Zählnummer	Funktion	–	– K2T	Schütz, Nr. 2, Zeitrelais
4	Anschluss			:	: 12	Anschluss Nr. 12

Nur zur Kennzeichnung erforderliche Blöcke angeben, Vorzeichen kann entfallen, wenn Verwechselung des Blockes ausgeschlossen ist Mindestangabe in Block 3 ist die Zählnummer.

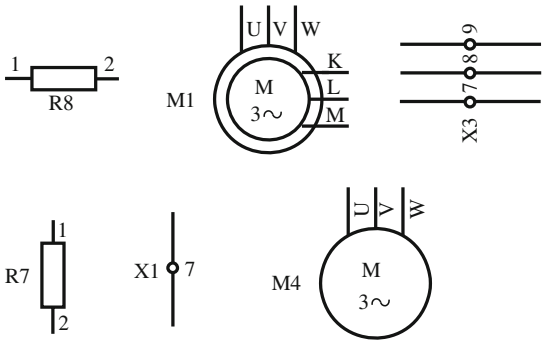


Abb. 47.13 Vereinfachte Betriebsmittelkennzeichnung

Die Kennzeichnung der Funktion in Kennzeichnungsblock 3 kann entfallen, wird sie jedoch vorgenommen, so sind die Kennbuchstaben nach Tab. 47.5 entsprechend DIN 40719 zu verwenden.

Für die Kennzeichnung der Anlage und des Ortes benennt DIN 40719 weitere Kennbuchstaben, die hier aber nicht aufgeführt werden sollen.

47.3 Schaltungsunterlagen der Energietechnik

Bei der *Ausfertigung elektrischer Schalt- und Verdrahtungspläne* genügt oft eine einzige Linienbreite nach DIN 15. Für A4- und A3-Formate ist die Linienbreite 0,5 günstig.

Stromwege werden geradlinig und möglichst kreuzungsfrei dargestellt. Die Verbindungslinien verlaufen parallel zu den Rändern der Zeichnung. Die Anschlussstellen an Betriebsmittel werden nicht besonders dargestellt, sie werden aber gemäß DIN 42400 mit einer Anschlusskennzeichnung versehen nach Abb. 47.14. Zusammengehörende Anschlüsse werden mit 1 und 2 gekennzeichnet. Von der Energieseite her gesehen gilt 1 = Eingang und 2 = Ausgang.

Hat ein Betriebsmittel nach Abb. 47.15 mehrere Anschlüsse, so wird fortlaufend numeriert. Die Anschlüsse von Drehstrom-

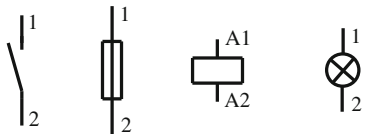


Abb. 47.14 Kennzeichnung der Anschlussstellen von Betriebsmitteln

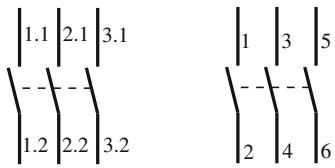


Abb. 47.15 Anschlussnummerierung mehrpoliger Betriebsmittel (allgemein)

Tab. 47.4 Kennbuchstaben für die Art der Betriebsmittel

Kennbuchstabe	Art des Betriebsmittels	Beispiele
A	Baugruppen, Teilbaugruppen	Verstärker, Magnetverstärker
B	Umsetzer von nichtelektrischen auf elektrische Größen und umgekehrt	Messumformer, thermoelektrische Fühler, Thermozellen, Mikrofon, u. ä.
C	Kondensatoren	
D	Verzögerungseinrichtungen, Speichereinrichtungen, binäre Elemente	Plattenspeicher, Magnetbandgeräte, Verzögerungsleitungen
E	Verschiedenes	Beleuchtungseinrichtungen, Heizeinrichtungen; Einrichtungen, die ansonsten hier nicht benannt werden
F	Schutzeinrichtungen	Sicherungen, Schutzrelais, Trennsicherungen, Überspannungsableiter
G	Generatoren, Stromversorgungen	Rotierende Generatoren, Batterie
H	Meldeeinrichtungen	Optische und akustische Meldegeräte
K	Relais, Schütze	Leistungsschütze, Hilfsschütze, Zeitrelais
L	Induktivitäten	Drosselspulen
M	Motoren	
N	Verstärker, Regler	Operationsverstärker
P	Messgeräte, Prüfeinrichtungen	Anzeigende, schreibende und zählende Messeinrichtungen, Uhren
Q	Starkstrom-Schaltgeräte	Leistungsschalter, Schutzschalter, Motorschutzschalter
R	Widerstände	Einstellbare Widerstände, Heißleiter
S	Schalter, Wähler	Taster, Endschalter, Steuerschalter
T	Transformatoren	Spannungswandler, Stromwandler
U	Modulatoren, Umsetzer	Diskriminator, Frequenzwandler, Umformer, Wechselrichter
V	Röhren, Halbleiter	Elektronenröhren, Dioden, Transistoren, Thyristoren
W	Übertragungswege, Hohlleiter, Antennen	Schaltdrähte, Sammelschienen, Dipole
X	Klemmen, Stecker, Steckdosen	Trennstecker und -steckdosen, Prüfstecker, Lötleisten, Klemmenleisten
Y	Elektrisch betätigte mechanische Einrichtungen	Bremsen, Kupplungen, Ventile
Z	Abschluss, Ausgleichseinrichtungen, Filter, Begrenzer, Gabelanschlüsse	Kabelnachbildungen

Tab. 47.5 Kennbuchstaben für die allgemeine Funktion

Kennbuchstabe	Allgemeine Funktion
A	Hilfsfunktion
B	Bewegungsrichtung
C	Zählung
D	Differenzierung
F	Schutz
G	Prüfung
H	Meldung
J	Integration
K	Tastbetrieb
L	Leiterkennzeichnung
M	Hauptfunktion
N	Messung
P	Proportional
O	Zustand (Stop, Start, Begrenzung)
R	Rückstellen, Löschen
S	Speichern, aufzeichnen
T	Zeitmessung, verzögern
V	Geschwindigkeit (beschleunigen, bremsen)
W	Addieren
X	Multiplizieren
Y	Analog
Z	Digital

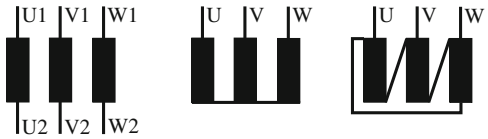


Abb. 47.16 Anschlusskennzeichnung von Drehstrommotoren

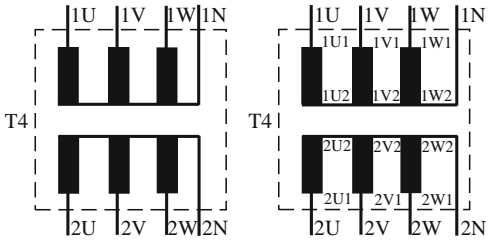


Abb. 47.17 Anschlusskennzeichnung von Drehstromtransformatoren

motoren werden nach Abb. 47.16 gekennzeichnet, während die Anschlüsse von Drehstromtransformatoren gemäß Abb. 47.17a gekennzeichnet werden, wobei unter Umständen auch die Innenanschlüsse nach Abb. 47.17b zu kennzeichnen sind.

Leitungsverbindungen ohne Klemmpunkt nach Abb. 47.18 sind für zu plottende Zeichnungen vorgesehen; sie werden auch mit

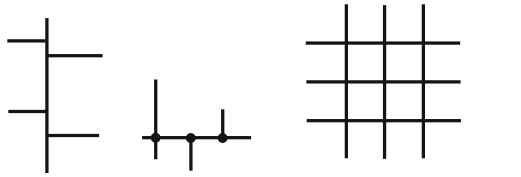


Abb. 47.18 Leitungsverbindungen mit/ohne Klemmpunkt

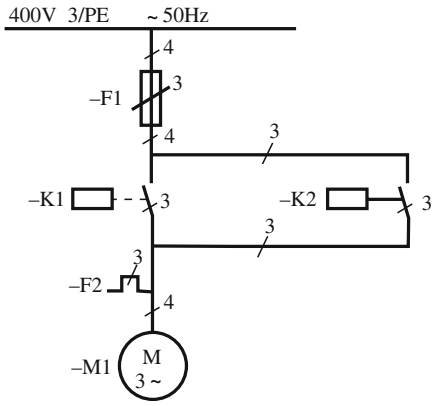


Abb. 47.19 Übersichtsschaltplan einer Wendeschützschaltung

ausgefülltem Punkt dargestellt. Leitungskreuzungen haben keine elektrische Verbindung.

Bei der Art der Darstellung von Schaltungsunterlagen unterscheidet DIN 40719 zwischen ein- oder mehrpoliger Darstellung. Bei der *einpoligen Darstellung (engl.: single-line representation)* werden *zwei oder mehr Leiter durch eine einzige Linie* mit entsprechender symbolhafter Ergänzung *dargestellt*. Mehrere gleiche Geräte oder Bauelemente können durch ein einziges Schaltzeichen dargestellt werden.

Bei *mehrpoliger Darstellung (engl.: multiline representation)* wird jedes Betriebsmittel durch ein Schaltzeichen dargestellt.

Bezüglich der Darstellung der Schaltzeichen kann eine weitere Unterscheidung vorgenommen werden. Bei der *zusammenhängenden Darstellung (engl.: assembled representation)* werden alle Schaltzeichen eines Betriebsmittels (z. B. Relais mit Schaltkontakten) zusammenhängend gezeichnet. Bei der *aufgelösten Darstellung (engl.: detached representation)* werden *Schaltzeichen für elektrische Betriebsmittel (z. B. Relais mit Schaltkontakten) getrennt gezeichnet* und so angeordnet, dass *jeder Stromweg* (Strompfad) *geradlinig verläuft* und somit *gut zu verfolgen* ist.

Ein *Übersichtsschaltplan (engl.: block diagram)* nach DIN 40719 ist die vereinfachte Darstellung einer Schaltung,

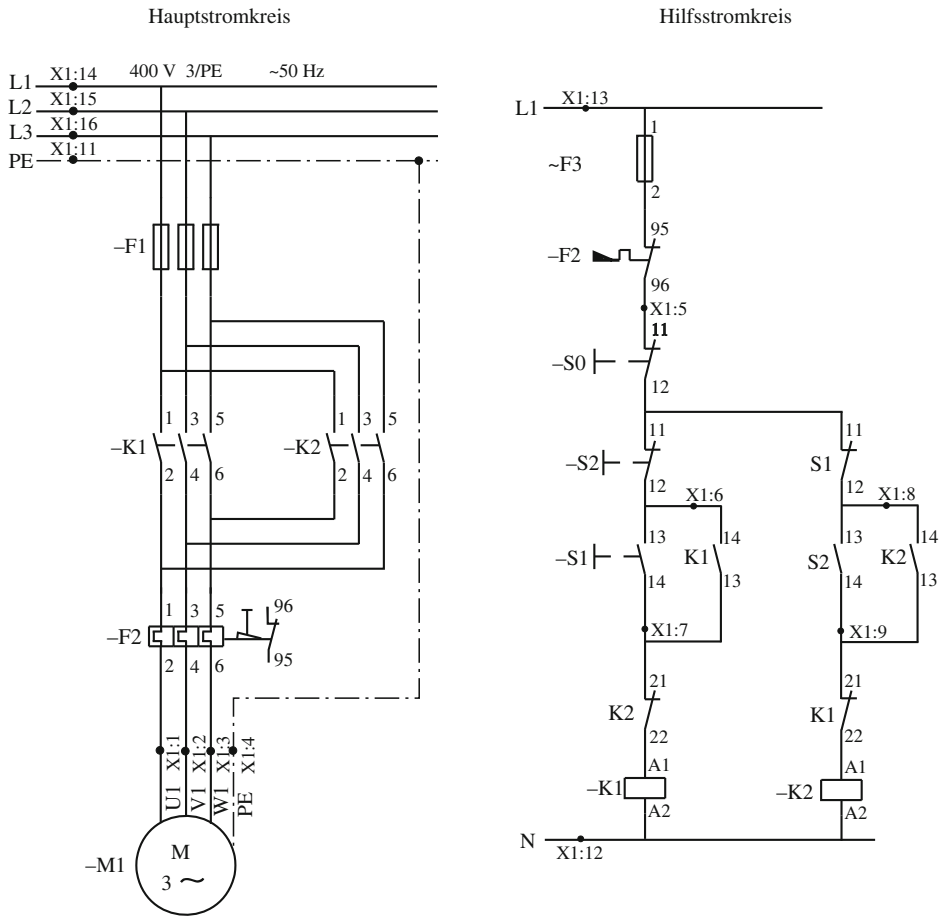


Abb. 47.20 Stromlaufpläne in aufgelöster Darstellung für Haupt- und Hilfsstromkreis

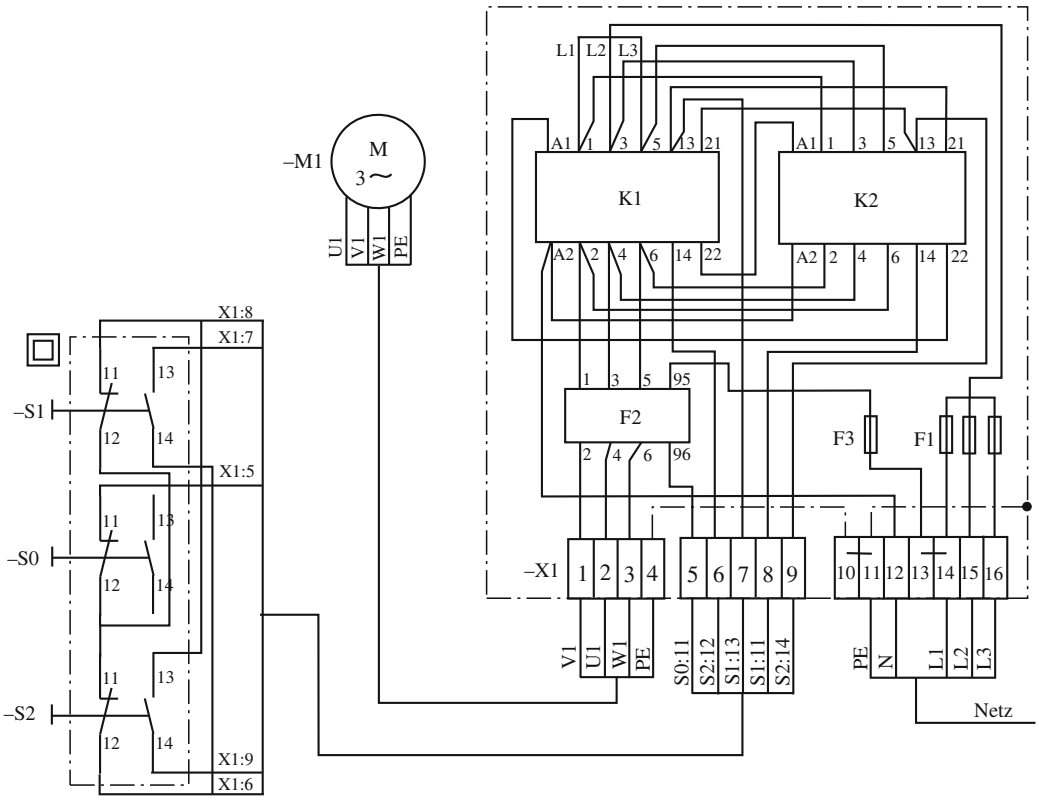


Abb. 47.21 Verdrahtungsplan einer Wendeschützschaltung

wobei nur die wesentlichen Teile zur Gliederung elektrischer Einrichtungen und ihrer Systembeschreibung berücksichtigt werden. Im Beispiel nach Abb. 47.19 ist der einpolige Übersichtsschaltplan einer Wendeschützschaltung ohne Steuerleitungen dargestellt.

Ein *Stromlaufplan* (engl.: *circuit diagram*) ist die ausführliche Darstellung einer Schaltung mit ihren Einzelteilen. Er zeigt die Arbeitsweise einer elektrischen Einrichtung. Er kann in aufgelöster und zusammenhängender Darstellung zur Anwendung kommen. Abb. 47.20 zeigt die *Stromlaufpläne in aufgelöster Darstellung* für den Hauptstromkreis und die Steuerung einer Wendeschützschaltung. Die Kontaktbelegungspläne der Schütze fehlen im Steuerstromkreis, aber die Anschlussbezeichnungen für die Klemmleisten sind integriert. Diese Form der Stromlaufpläne ist die meist verwendete Darstellungsform. Aus ihm wird der Verdrahtungsplan entwickelt.

Verdrahtungspläne (engl.: wiring diagrams) zeigen die inneren und/oder äußeren *Verbindungen zwischen elektrischen Betriebsmitteln*, ohne Aufschluss über die Wirkungsweise zu geben. Verdrahtungspläne können durch entsprechende Tabellen ersetzt werden. Der Verdrahtungsplan nach Abb. 47.21 gibt die Schaltungen nach Abb. 47.19 und II-20 wieder. Die Schütze K1 und K2 werden wie auch das Motorschutzrelais F2 durch ein Rechteck dargestellt.

Die Verbindungsleitungen können einzeln oder zusammengefasst gezeichnet werden. Die Anschlussstellen werden nach Abb. 47.22 mit Leitungsnummern oder mit den Zielbezeichnungen versehen, d. h., am Ende einer Verbindungsleitung wird angegeben, mit welcher Anschlussklemme das andere Leitungsende verbunden ist. Der Verdrahtungsplan dient als Unterlage für die Fertigung und Montage und wird bei umfangreichen Anlagen unterteilt in Geräteverdrahtungspläne, Verbindungspläne und Anschlusspläne.

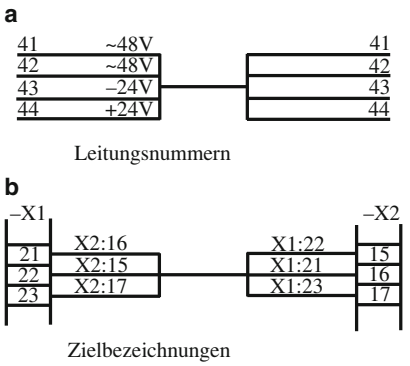


Abb. 47.22 Kennzeichnung von Verbindungsleitungen mit **a** Leitungsnummern oder **b** Zielbezeichnungen

Die *Verbindungen innerhalb eines Gerätes* stellt der Geräteverdrahtungsplan *(engl.: unit wiring diagram)* dar, während der *Verbindungsplan (engl.: interconnection diagram)* die *Verbindungen zwischen den verschiedenen Geräten* einer Anlage darstellt.

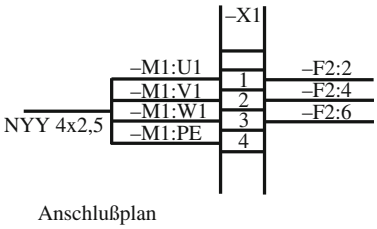


Abb. 47.23 Anschlussplan der Klemmenleiste X1

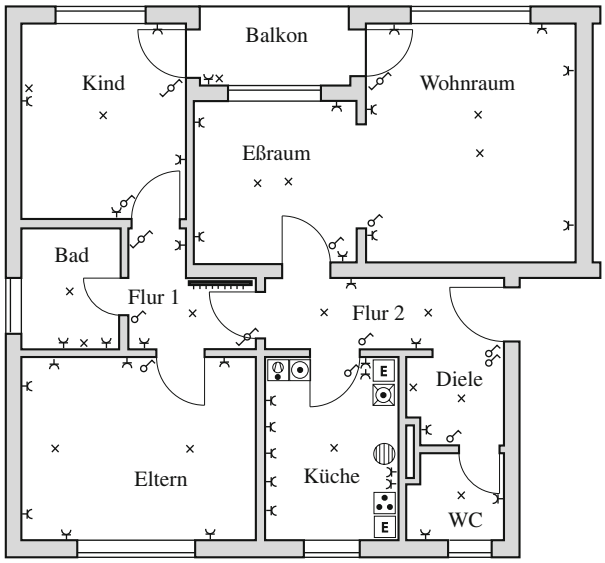


Abb. 47.24 Elektro-Installationsplan einer Wohnung

Ein *Anschlussplan (engl.: terminal diagram)* zeigt die *Anschlusspunkte einer elektrischen Einrichtung* und die daran angeschlossenen inneren und äußeren leitenden Verbindungen. Der Anschlussplan nach Abb. 47.23 zeigt die Belegung der Klemmenleiste X1. Die zum Motor M1 am Planort T4 abgehende Leitung W2 ist vom Typ NYY $4 \times 2{,}5$.

Ein *Ersatzschaltplan* (engl.: *equivalent circuit diagram*) ist eine besondere Ausführung eines erläuternden Schaltplanes für die Analyse und Berechnung von Stromkreiseigenschaften.

Ein *Ablaufdiagramm* oder eine Ablauftabelle *(engl.: sequence chart* or table) zeigt Arbeitsvorgänge in einer festgelegten Reihenfolge, während *Zeitablaufdiagramme (engl.: time sequence chart* or table) zusätzlich die Zeitabstände zwischen aufeinanderfolgenden Arbeitsvorgängen zeigen.

Gerätelisten schlüsseln die für eine Fertigung erforderlichen Konstruktionselemente, Geräte, elektrischen Einrichtungen und Leitungen auf nach Anzahl, Typ und auch Bestellnummern der Hersteller.

Im Installationsplan nach DIN 40719 Teil 5 werden die Betriebsmittel mit Schaltzeichen in einpoliger Darstellung annähernd lagerichtig in den Gebäudegrundriss eingezeichnet. Abb. 47.24 zeigt den Elektro-Installationsplan für eine Wohnung mit mehreren Installationsbereichen nach VDE 0100. DIN 18015 Teil 1 legt sowohl Installationszonen als auch die Leitungsführung und die räumliche Lage der meisten Betriebsmittel fest.

Im Übersichtsplan nach Abb. 47.25 werden die Stromkreise festgelegt und die Nennwerte der Betriebsmittel eingetragen. Nach DIN 57100/VDE 0100 ergibt sich auch die Verwendung entsprechender Leitungen und Schutzmaßnahmen, z. B. FI-Schutzschalter.

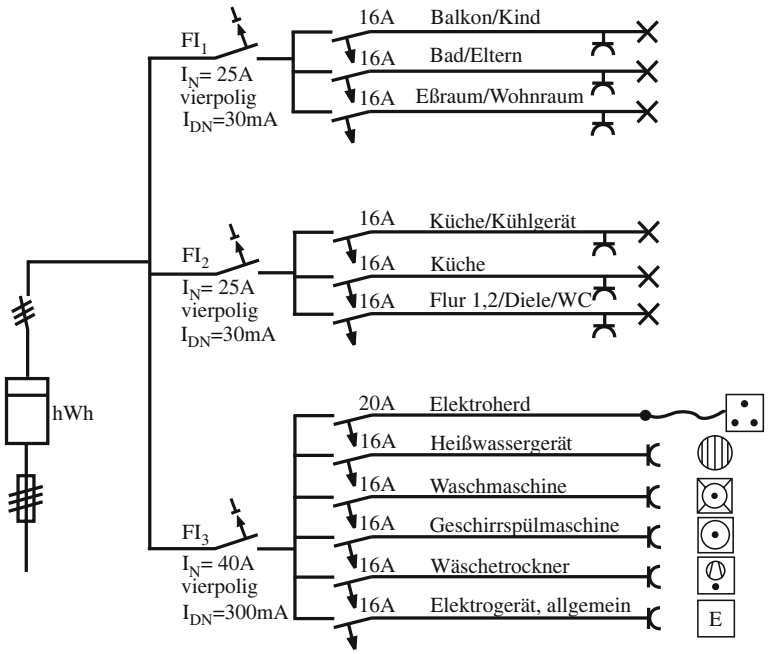


Abb. 47.25 Übersichtsplan zum Installationsplan

47.4 Schaltungsunterlagen der Elektronik

Ersatzschaltpläne, bzw. Ersatzschaltbilder sind nach DIN 40719 besondere Ausführungen eines erläuternden Schaltplanes für die Analyse und Berechnung von Schaltkreisen und realen Bauteileigenschaften im Vergleich zu „idealen“ Bauelementen.

So hat z. B. jeder Kondensator außer der gewünschten Kapazität noch weitere elektrische Eigenschaften, die sich als Komponenten in einem Ersatzschaltbild nach Abb. 47.26 darstellen und analysieren lassen. Die Serieninduktivität L ergibt sich als Folge der Wirkung der Anschlussdrähte bei höheren Frequenzen und der Serienwiderstand R_S als Kontaktwiderstand an den Anschlüssen. Die Kondensator-Kapazität C ist die gewünschte Größe, während der Parallelwiderstand R_{is} als Isolationswiderstand des Dielektrikums zu verstehen ist, der bei bestimmten Kondensatoren für die Selbstentladung des Kondensators verantwortlich ist.

Die Frequenzabhängigkeit eines realen Kondensators lässt sich mit entsprechenden Ortskurve nach Abb. 47.27 darstellen, während sich z. B. die Temperaturabhängigkeit mit einem doppeltlogarithmischen Diagramm ($Z = f(f)$ mit der Temperatur ϑ als Parameter) nach Abb. 47.28 darstellen lässt.

47.4.1 Allgemeines

Die Darstellung logischer Elemente und Verknüpfungen wird in Katalogen und Datenbüchern unterschiedlich gehandhabt. Abb. 47.29 zeigt daher eine Gegenüberstellung der logischen Grundfunktionen, die noch häufig anzutreffen sind. Die Darstellung der Schaltzeichen nach DIN 40900 setzt sich aber immer weiter durch.

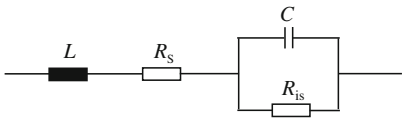


Abb. 47.26 Ersatzschaltbild eines Kondensators (reales Verhalten)

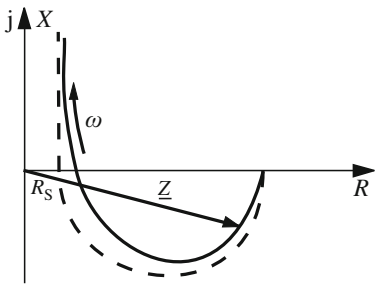


Abb. 47.27 Ortskurve eines Kondensators, d. h. Beschreibung des Frequenzverhaltens

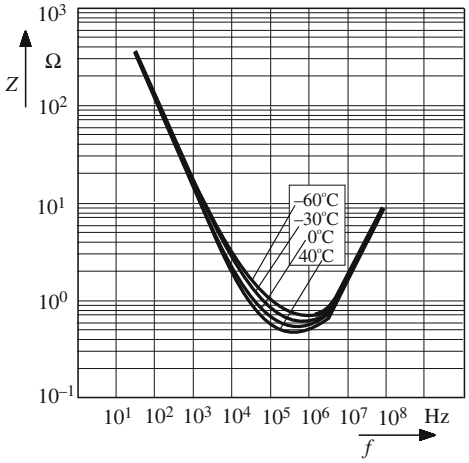


Abb. 47.28 Scheinwiderstand eines Kondensators in Abhängigkeit von der Temperatur

	DIN 40900	DIN alt	USA
NICHT	1		
UND	&		
ODER	≧1		
NAND	& &		
NOR	≧1 ≧1		

Abb. 47.29 Schaltsymbolvergleich

47.4.2 Schaltzeichen nach DIN 40900 Teil 12 Binäre Elemente

Symbolaufbau

Ein Symbol besteht aus einer Kontur, dem Block, oder aus Kombinationen von Konturen und den Funktionskennzeichen. Das Seitenverhältnis der Konturen ist ohne Bedeutung. Die Eingangsleitungen werden nach Abb. 47.30 links und die Ausgangsleitungen rechts angeordnet. Die Funktionskennzeichen nach Abb. 47.31 beschreiben die logische Funktion der Schaltung.

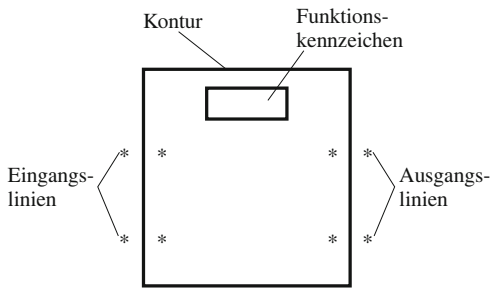
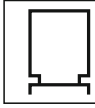


Abb. 47.30 Kontur

Zeichen	Bedeutung
&	UND-Verknüpfung
≥1	ODER-Verknüpfung
=1	Exclusiv-ODER-Verknüpfung
Σ	Summierer
COMP	Komparator
CTR	Zähler
DIV	Teiler
MUX	Multiplexer
P–Q	Subtrahierer
SRG	Schieberegister
X/Y	Codierer oder Codeumsetzer
⎍	Schmitt-Trigger
⎍	Monostabile Kippstufe (allgemein)
1⎍	Monostabile Kippstufe (nicht nachtriggerbar)
G ⎍	Astabile Kippstufe (allgemein)

Abb. 47.31 Funktionskennzeichen

Symbol	Beschreibung
	Element-Kontur
	Steuerblock-Kontur

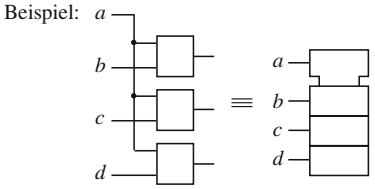


Abb. 47.32 Steuerblock-Kontur

Besitzt eine Schaltung einen oder mehrere Eingänge, die zu mehr als einem Element der Schaltung führen, kann ein Steuerblock nach Abb. 47.32 benutzt werden (siehe auch Anwendungsbeispiele).

Ein- und Ausgänge

Die Symbole der Ein- und Ausgänge nach Abb. 47.33 legen die Beziehungen zwischen den internen und externen Logikzuständen fest. Die Negation am Ein- und Ausgang kann unterschiedlich dargestellt werden. DIN 40900 empfiehlt auf nationaler Ebene den Negationskreis zu verwenden. In Datenbüchern findet man häufig die zweite Darstellungsweise.

Abhängigkeitsnotation

Mit Hilfe der Abhängigkeitsnotation werden Zusammenhänge zwischen Eingängen, Ausgängen oder Ein- und Ausgängen angegeben, ohne interne Elemente und Zwischenverbindungen im einzelnen genau darzustellen. Somit lassen sich kompakte und aussagekräftige Symbole herstellen. Abb. 47.34 gibt eine Übersicht über Arten der Abhängigkeit und zeigt einige

Symbol	Beschreibung
	Negierter Eingang
	Negierter Ausgang
	Negierter Eingang
	Negierter Ausgang
	Eingang mit Hysterese
	Ausgang mit offenem Kollektor
	3-state-Ausgang
	Freigabeeingang
	Schiebeeingang rechts
	Schiebeeingang links
	aktiv bei ansteigender Flanke } Dynamischer Eingang
	aktiv bei abfallender Flanke

Abb. 47.33 Kennzeichnung von Ein- und Ausgängen

Beispiele hierzu. Die Adressen-Abhängigkeit (*A*) kennzeichnet Adresseneingänge eines Speichers. Die Steuer-Abhängigkeit (*C*) kennzeichnet Zeitsteuer- oder Takteingänge. Das Kennzeichen ist häufig bei Flipflops zu sehen.

Anwendungsbeispiele

Bei Anwahl eines Eingangs über die Abhängigkeitsnotation *G* nimmt der interne Logikzustand des Ausgangs den internen Zustand des Eingangs an. Liegt an den drei Eingängen beispielsweise „1“, „1“, „1“ wird der Eingang 7 zum Ausgang geschaltet. Mit dem Eingang EN kann der Multiplexer aktiviert werden (low-aktiv), (Abb. 47.35).

Die gesteuerten Eingänge des Richtungstreibers sind UND-verknüpft. Die freigewählte Zahl (hier 3) wird neben dem Buchstaben *G* geschrieben und an jedem weiteren UND-verknüpften Eingang (EN). 1 und 2 des Freigabeeingangs geben jeweils in Verbindung mit dem Verstärker die Richtung an, wie sie die Zahlenfolge am 3-state-Ausgang zuordnet (Abb. 47.36).

Über die Steuereingänge *M* wird die Betriebsart wie links schieben, rechts schieben oder paralleles Laden dargestellt. Den Einfluss des Takteingangs *C* wird durch die entsprechende Zahl und die Schieberichtung durch zusätzliche Pfeile gekennzeichnet. Über die Eingänge 4,5D werden anstehende Daten mit der ansteigenden Flanke an C5 parallel geladen (Abb. 47.37).

Über den Steuereingang C3 kann der BCD-Zähler die 3D-Eingänge parallel laden. Die Betriebsart: Zählrichtung (vorwärts „+“, rückwärts „−“) erfolgt über den Eingang M2 und mit G1 erfolgt die Zählfreigabe. Mit dem Rücksetzeingang *R* wird der Zähler rückgesetzt. Nach jeder erneuten Zählung liefert der Ausgang PIN 7 einen negativen Impuls (Abb. 47.38).

Bei Speichern können die Speicherstellen durch Adresseingänge (Kennzeichen A) angewählt wer- den. Der dargestellte Speicher besitzt 1024 Speicherstellen (1 k) mit den Adressen 0 bis 1023. Die Daten können am Ausgang ((0)A . . . (7) A, 8 Bit) mit „*H*“ am Steuereingang gelesen und mit „*L*“ über die mit 2D gekennzeichneten Leitungen eingeschrieben werden (Abb. 47.39).

Kennbuchstabe und Abhängigkeitsform	Wirkung bei „1“ am Eingang	Wirkung bei „0“ am Eingang
A Adresse		
C Steuerung	Erlaubt die Funktion	Verhindert die Funktion
EN Freigabe	Erlaubt die Funktion	Verhindert Zugriff der Eingänge ∇ Ausgänge hochohmig
G UND	Erlaubt die Funktion	Ausgang „0“
M Modus	Erlaubt die Funktion bei gewählter Betriebsart	Verhindert die Funktion
N Exclusiv-ODER		
R Rücksetzen		
S Setzen		

Beispiele:

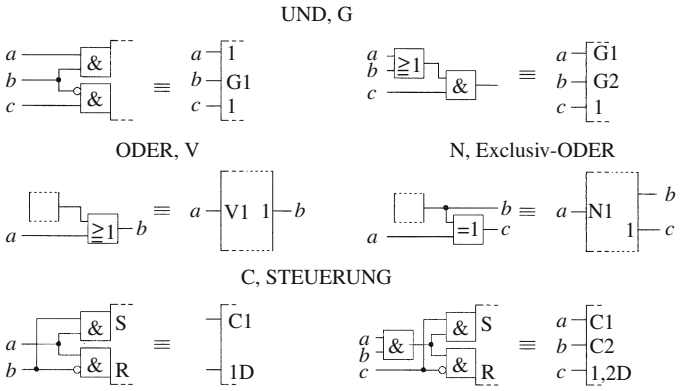


Abb. 47.34 Abhängigkeitsnotation

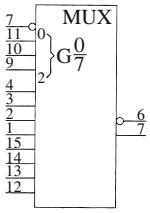


Abb. 47.35 Multiplexer

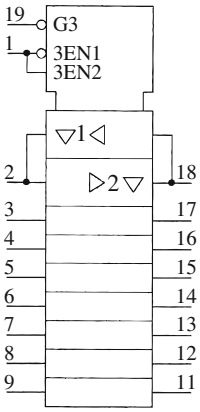
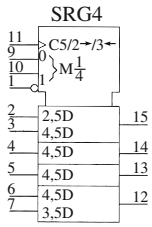


Abb. 47.36 Richtungstreiber



Betriebsart (MODE)

M	
0 0	Halt
0 1	rechts schieben →
1 0	links schieben ←
1 1	paralleles Laden

Abb. 47.37 Schieberegister

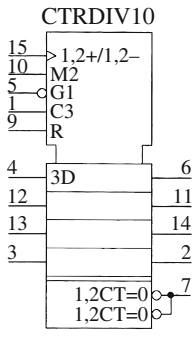


Abb. 47.38 Zählerbaustein

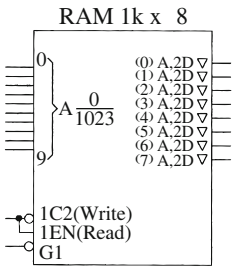


Abb. 47.39 RAM (*r*andom *a*ccess *m*emory)

47.4.3 Entwurf von Schaltungen

47.4.3.1 Verdrahtungsplan mit Universalplatinen

Kleinere elektronische Einzelschaltungen lassen sich mit geringem Aufwand durch Universalplatinen realisieren. Solche Platinen sind Lochrasterplatten aus einem meist 1,5 mm starkem Isoliermaterial (Hartpapier oder Pertinax) mit Bohrungen im 1/10″-Raster (2,54 mm) und 1 mm Durchmesser oder sie bestehen aus einem Isolierstoff mit Längsleiterbahnen oder Lötaugen aus Kupfer (Auflage 35 mm) mit einem Rastermaß 1/10″ und einem gleichmäßigen Rastermaß mit Löchern von ebenfalls 1 mm Durchmesser. Außer Streifen- und Punktrasterplatien sind Printplatten mit zusätzlich verschiedenen Steckerleistenanschlüssen nach DIN 41612 im Angebot der Herstellerfirmen. Die gebräuchliche Europlatine hat die Abmessungen 160 mm× 100 mm. Tab. 47.6 führt gängige Maße von Printplatinen auf.

Ausgangslage zur Schaltungsrealisierung ist in allen Fällen der Schaltplan und die Stückliste aller benötigten Bauelemente. Die Bauelemente sollten steckbar sein und müssen in das 1/10″-Raster passen. Nach Festlegung der Platinengröße wird eine Lageskizze (Entwurf) mit der Lage der Bauelemente und den notwendigen Verbindungen erstellt. Bei der Bauelementepositionierung kann für einen übersichtlichen Aufbau der Eingang links und der Ausgang rechts vorgesehen werden oder alle Anschlüsse werden auf eine Leiterplattenseite gelegt. Die Verbindungen sollten kurze Wege aufweisen und Kreuzungen nach Möglichkeit nicht vorhanden sein.

Bei Verwendung einer Lochrasterplatine werden die Bauelemente lagerichtig auf die Platine gesteckt und auf der Unterseite mit Silberdraht verlötet.

Wird eine Platine mit vorhandenen Leiterbahnen eingesetzt, sollte der Entwurf die notwendigen Unterbrechungen und Brücken aufweisen.

Tab. 47.6 Maße von Printplatinen

Plattenstärke	1, 1,5, 2 mm
Lochdurchmesser	0,8, 1 mm
Rastermaße	2,54, 5,08 mm
Leiterbreite	> 0,8 mm
Leiterabstand	> 0,8 mm
Kupferauflage	35, 70 µm

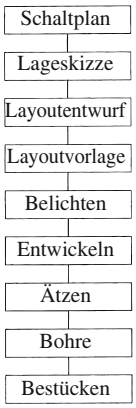


Abb. 47.40 Arbeitsschritte zur Herstellung eines Labormusters

47.4.3.2 Entwurf und Herstellung gedruckter Schaltungen

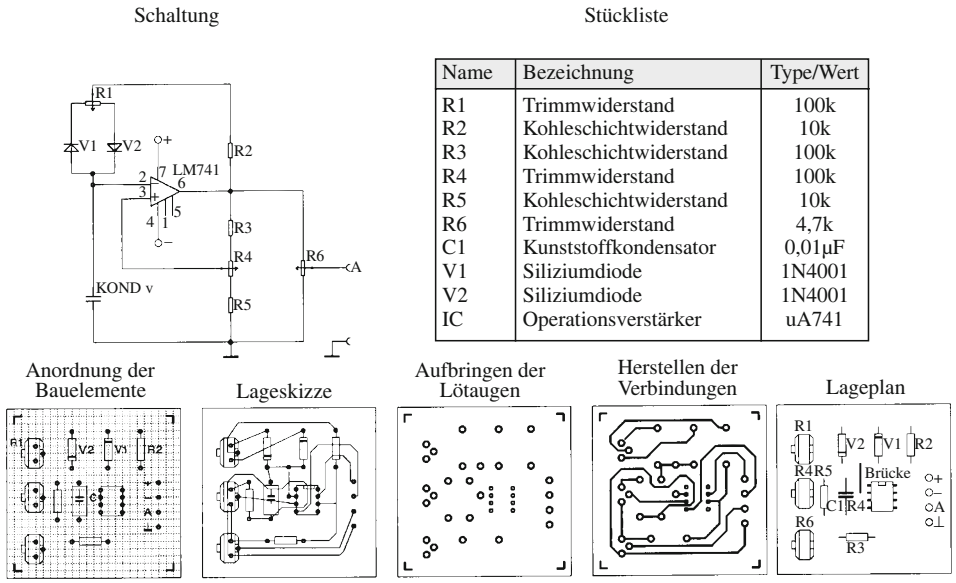
Gedruckte Schaltungen sind heute Bestandteil der meisten elektronischen Geräte. Ihr Vorteil liegt in der computergesteuerten Fertigung und der schnellen Austauschbarkeit im Servicefall.

Zur Herstellung gedruckter Schaltungen verwendet man heute vorwiegend fotobeschichtete Platinen aus Hartpapier oder Epoxyd-Hartgewebe als Trägermaterial. Sie können einseitig oder beidseitig fotobeschichtet sein. Ihre Kupferauflage ist meist 35 mm und die Platinengröße unterschiedlich (Beispiele: 60 mm × 100 mm, 250 mm × 250 mm). Die Bauelemente werden abschließend über Verbindungsleitungen aus Kupfer miteinander verbunden. Die folgende Abb. 47.40 zeigt die notwendigen Schritte bei der Herstellung eines Labormusters bis zur fertigen Platine und Abb. 47.41 wählt als Beispiel die Schaltung eines Rechteckgenerators zur Erläuterung der Arbeitsabläufe.

Die Anordnung der Leiterbahnen auf der Platine nennt man *Layout*. Zur Layouterstellung benötigt man transparentes 1/10″-Rasterspapier, eine Folie, Klebesymbole, Leiterstreifen und Lötaugen. Von der *Lageskizze* werden auf einem ersten Blatt entsprechend der Größe die Lage der Bauelemente und ihre Anschlüsse übertragen. Anschließend wird ein zweites Blatt über das erste gelegt und die Verbindungsleitungen eingezeichnet. Dieses Blatt dient als Vor- lage für alle Klebematerialien. Zunächst werden alle Lötaugen aufgebracht, danach die Leiterbahnen.

Das völlig lichtundurchlässige Layout wird unter die fotobeschichete Platine nach Abb. 47.42 gelegt und von unten ca. 2 Minuten mit ultraviolettem Licht (UV-Licht) bestrahlt, um die Leiterbahnen auf die Fotoschicht zu übertragen. Die Platine wird daraufhin in ein Entwicklerbad gegeben, wo die belichteten Stellen von einer Schutzschicht (Fotolack) entfernt werden. Beim anschließenden Ätzen werden alle Kupferflächen, die nicht zum Layout gehören, entfernt.

Der Ätzvorgang dauert bei Verwenden von Eisen- III-Chlorid und einer Temperatur von 40 °C etwa 10 bis 20 Minuten. Die Platine wird gewässert, um Ätzmittelreste abzuspülen und mit einem Lösungsmittel (Azeton) kann der Abdecklack entfernt werden.



Name	Bezeichnung	Type/Wert
R1	Trimmwiderstand	100k
R2	Kohleschichtwiderstand	10k
R3	Kohleschichtwiderstand	100k
R4	Trimmwiderstand	100k
R5	Kohleschichtwiderstand	10k
R6	Trimmwiderstand	4,7k
C1	Kunststoffkondensator	0,01µF
V1	Siliziumdiode	1N4001
V2	Siliziumdiode	1N4001
IC	Operationsverstärker	uA741

Abb. 47.41 Arbeitsschritte am Beispiel eines Rechteckgenerators

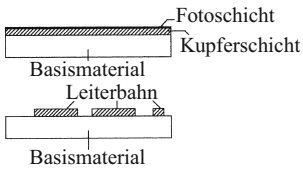


Abb. 47.42 Vor und nach dem Ätzvorgang

Als Ätzmittel wird häufig Eisen-III-Chlorid ($FeCl_3$), Ätznatron (NaOH) oder Ammoniumpersulfat $(NH4)_2S_2O_8$ verwendet. Aus Sicherheitsgründen sollten bei Umgang mit allen Chemikalien Handschuhe und Schutzbrille getragen werden. Des Weiteren ist eine sachgerechte Entsorgung zu gewährleisten. Der Ätzvorgang selbst hängt von der Konzentration des Ätzmittel, der Temperatur und der Ätzzeit ab. Bei zu langer Ätzzeit oder zu hoher Ätzkonzentration kommt es zu Unterätzungen.

Leiterbahnen werden vor Oxidation geschützt, wenn sie mit Zinn, Silber oder Gold überzogen oder mit Lötlack eingesprüht werden.

Die Breite der Leiterbahnen sollte in Abhängigkeit der durchfließenden Ströme gewählt werden, damit keine unzulässigen Erwärmungen (i. d. R. max. 65 °C) auftreten. Ein Leiter von 35 mm Dicke kann bei 1 bis 2 mm Breite mit etwa 1,5 bis 2 A belastet werden. Hierbei tritt eine Temperaturerhöhung von 10 K auf. Um Spannungsüberschläge zu vermeiden dürfen Leiterbahnen nicht beliebig eng zueinander verlegt werden.

Nach der Platinenherstellung werden die Löcher in den Lötaugen mit einer Ständerbohrmaschine mit Hartmetallbohrern gebohrt. Anschließend kann die Bestückung erfolgen. Diese so von Hand erstellten Layouts haben den Nachteil, dass Fehler nie ganz ausgeschlossen werden können.

47.5 Projektierung

Werden technische Geräte oder Schaltungen entwickelt und konstruiert ist eine umfangreiche Planung erforderlich. Zunächst muss das Problem definiert werden mit den jeweiligen Forderungen an die Schaltung und ihren Eigenschaften. Bei komplexen Aufgaben empfiehlt es sich im Team zu arbeiten und nach Lösungsmöglichkeiten mit den auftretenden Vor- und Nachteilen zu suchen. Ist eine Entscheidung getroffen, werden Aufgaben und Prioritäten aufgelistet und Verantwortlichkeiten zugewiesen. Ein festgelegter Zeitplan ist genauso einzuhalten wie die Obergrenze der zugestandenen Geldmittel. Nach Durchführung der Arbeit erfolgt eine Auswertung.

Beispiel

Universal-Eingabeplatine für einen Einplatinencomputer

Definition: Platine mit Stromversorgung, Eingangsbuchsen für den Analog-Digitalteil und Ausgangsbuchsen für den Digital-Analogteil.

Ziel: Die Schaltung soll innerhalb einer Woche erstellt und getestet sein. Die Einzelteile sollen weniger als 25,00 € kosten. Vorhanden ist eine veränderbare Gleichspannungsquelle 2 bis 15 V/2 A.

Die konstruierte Schaltung muss zur Verfügung stellen:

- +5 V stabil für die Computerplatine
- +12 V für externe Versuchsschaltungen
- 2 DA-Ausgänge
- 4 AD-Eingänge ◄

Durchführungsplan:

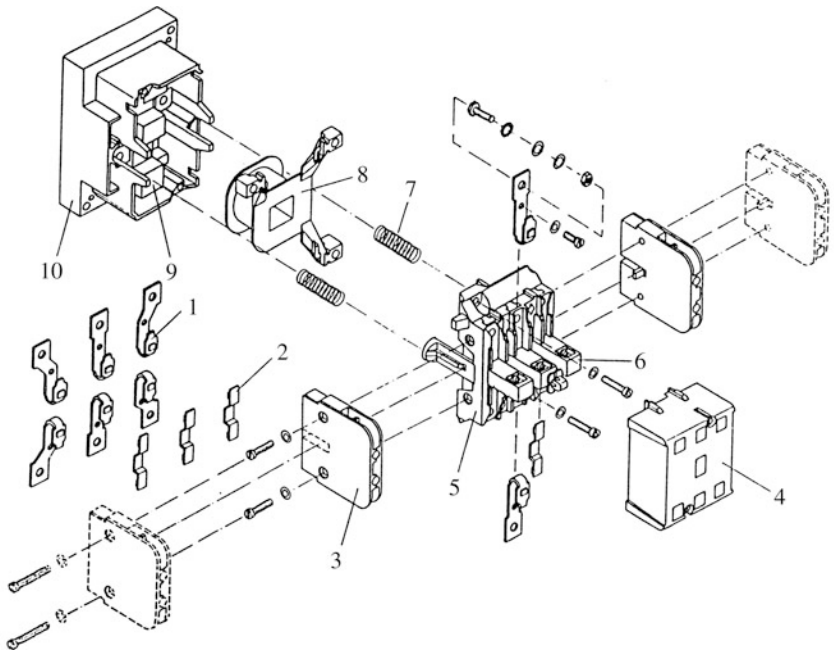
- Entwurf der Schaltung
- Beschaffung der Materialien
- Aufbau des Prototyps auf einer Steckplatine
- Herstellung der Platine
- Erprobung der Platine

Die Schaltung (Abb. 48.3) wird in Abschn. 48.2 beschrieben und in Kap. 49 die Realisierung der Platine.

Schaltungssynthese und -analyse

48

Peter Döring



48.1 Beispiele aus der Elektrotechnik 578
48.2 Beispiele aus der Elektronik 578

Teil VI

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_48

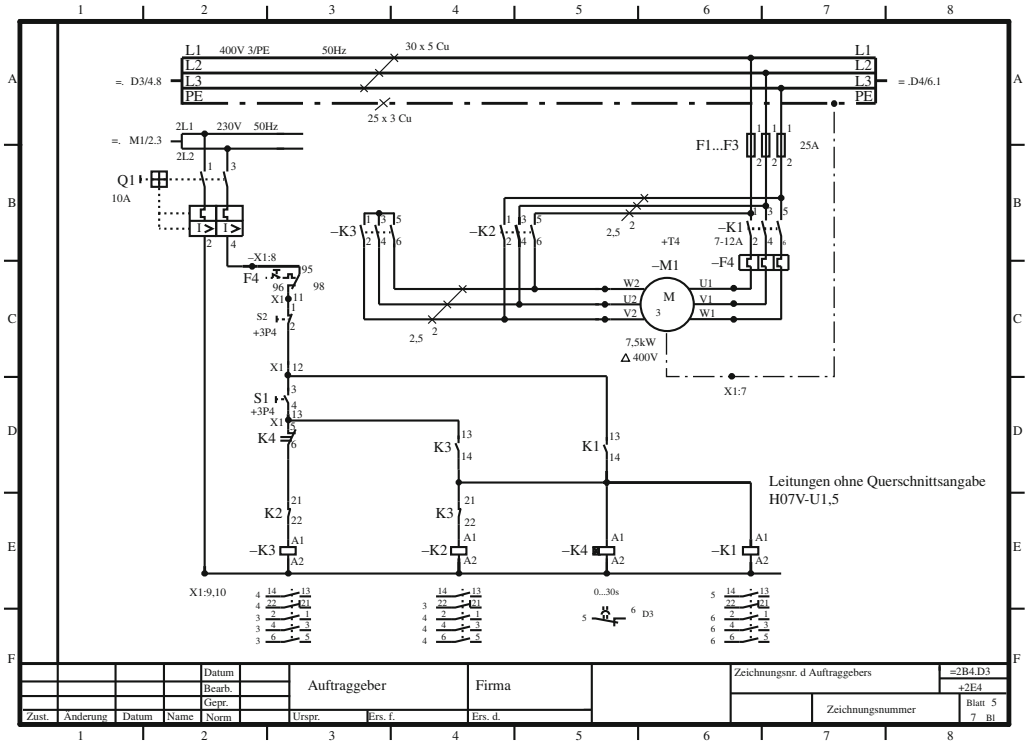


Abb. 48.1 Stern-Dreieck-Schützschaltung

48.1 Beispiele aus der Elektrotechnik

Für die Schaltungssynthese wie auch die -analyse ist die Kenntnis der Schaltzeichen, der Kennzeichnung von Betriebsmitteln und der Anschlusskennzeichnung notwendig. Auch die umfangreichsten und komplexesten Schaltungsaufgaben setzen sich aus der Kombination bzw. Verknüpfung einiger Grundschaltungen zusammen. Weil das Problem einer zu planenden Steuerung in der Regel durch die Darstellung einer Schaltung in einem Stromlaufplan gelöst wird, ist dieser besonders wichtig und darum im Beispiel nach Abb. 48.1 dargestellt.

Im Schriftfeld der Zeichnung werden unten rechts die Blattnummer und die Anzahl der Blätter für eine Anlage angegeben, hier also Blatt 5 von 7 Blättern. Darüber im Schriftfeld angeordnet steht mit einem Gleichheitszeichen als Vorzeichen der Kennzeichnungsblock „Anlage“ (hier =2B4.D3) und mit einem Pluszeichen als Vorzeichen der Kennzeichnungsblock „Einbauort“ (hier Schaltschrank +2E4). Diese Kennzeichnung gilt für alle auf diesem Blatt dargestellten Betriebsmittel und muss auf diesem Blatt folglich nicht mehr angegebenen werden.

Der Motor ist am Aufstellungsort +T4 installiert, während die Tasterkombination im Schaltpult +3P4 integriert ist. Der Hinweis auf ein anderes Blatt wird nach einem Schrägstrich, der Planabschnitt nach einem Punkt geschrieben; die Hauptstromversorgung kommt also von Blatt 4 (= .D3/4..; 2B4-Anlage mit Funktionsgruppe D = Leistungskreis) und geht nach Blatt 6 (= .D4/6..), um dort in Planquadrat 1 (= .D4/... 1) fortgeführt zu werden. Die Steuerspannung kommt von Blatt 2 aus Planquadrat 3.

Für die Wartung und Instandhaltung sind zusätzliche Informationen einzufügen, wie z. B. die Nennstromwerte für Sicherungen und Auslösebereich der Motorschutzschalter bzw. -relais. Die Reihenklemmenanschlüsse für den Motor (X1: 1, 2, 3, 4, 5, 6, 7) und die Tasterkombination (X1: 11, 12, 13) sind ebenfalls angegeben.

48.2 Beispiele aus der Elektronik

Schaltungen in der Elektronik werden so dargestellt, dass sich der Signalfluss von links nach rechts ergibt. Dies bezieht sich auch auf die Nummerierung der Bauelemente.

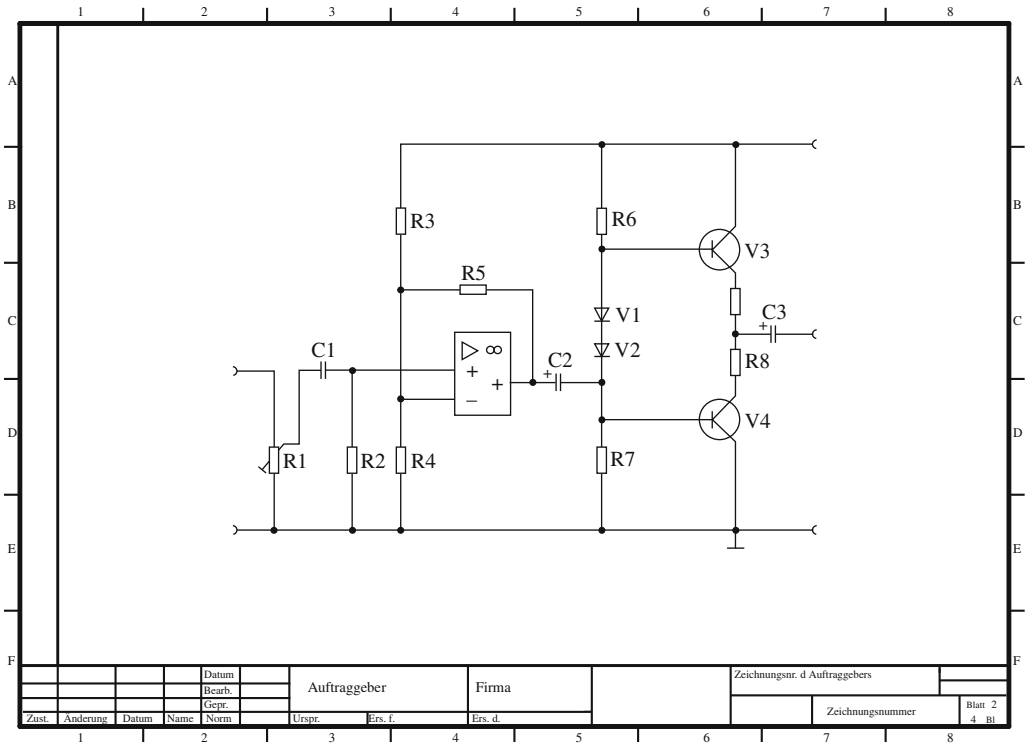


Abb. 48.2 Kopfhörerverstärker

Beispiel 1

Abb. 48.2 zeigt beispielhaft die Schaltung eines Kopfhörerverstärkers.

Die Eingangsspannung wird über R1 eingespeist. Dieser Widerstand dient auch zur Lautstärkeeinstellung. Über den Operationsverstärker wird das Signal verstärkt und gelangt über C2 zur Komplementärendstufe. Diese besteht aus einem npn- und pnp-Transistor der jeweils die negative oder positive Signalhalbwelle über C3 dem Kopfhöreranschluss zuführt. Beide Transistoren arbeiten in Kollektorgrundschaltung. Ihre Basisspannung und damit ihr Ruhestrom wird über zwei in Durchlassrichtung betriebene Dioden V1 und V2 eingestellt. Ohne V1 und V2 würden Übernahmeverzerrungen entstehen.

Die Schaltung ist für eine unsymmetrische Spannung ausgelegt. Sie benötigt daher R3 und R4, um den invertierenden Eingang auf die halbe Betriebsspannung zu legen. ◀

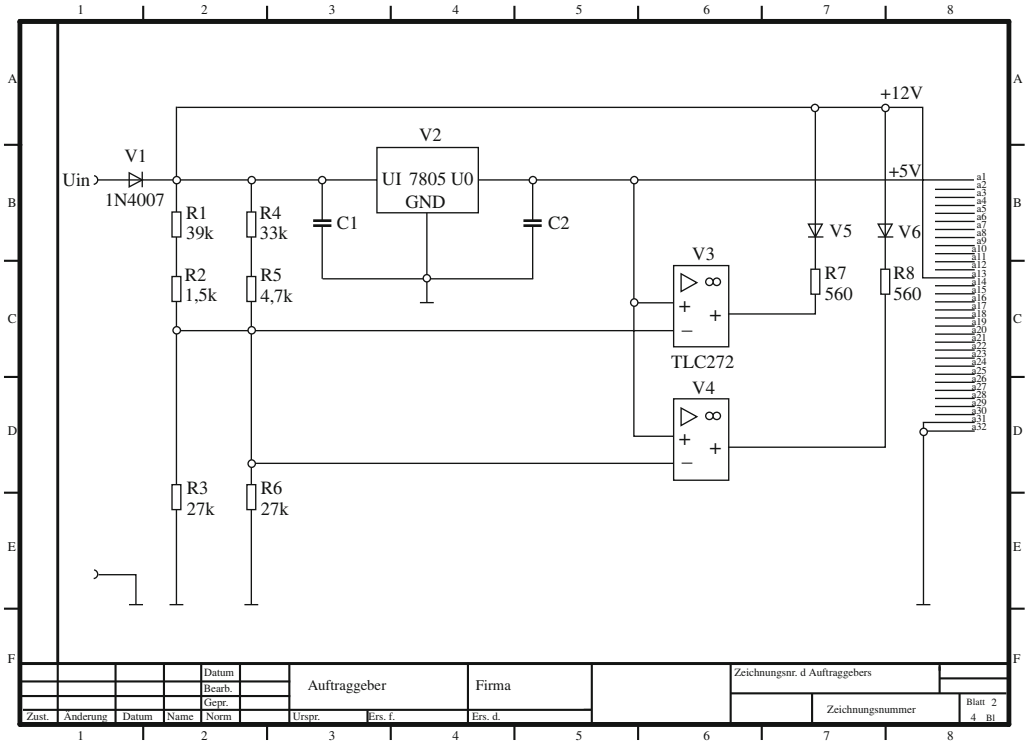


Abb. 48.3 Stromversorgung für einen Einplatinencomputer

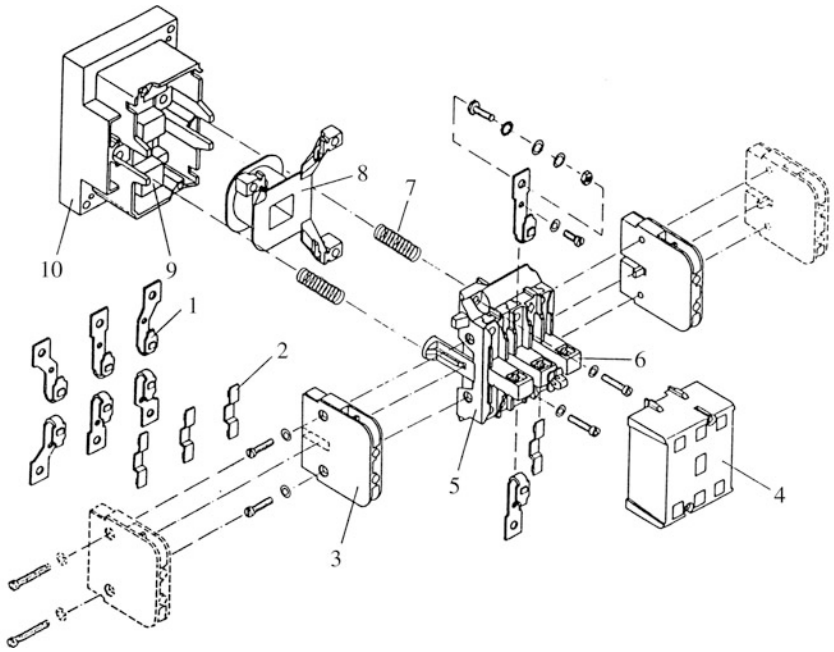
Beispiel 2

Abb. 48.3 zeigt die Stromversorgung für einen Einplatinencomputer. Die Schaltung selbst wird mit einer Gleichspannung betrieben. V1 dient zur Vermeidung einer falschen Polung. Die notwendigen 5 V für den Computer werden mit dem Stabilisierungs-IC 7805 erreicht. Die 12-V-Spannung dient zur Spannungsversorgung zusätzlicher Versuchsschaltungen. Um ohne Messgerät auszukommen wurden zwei Komparatoren eingesetzt, die über Leuchtdioden die gewünschte Spannung anzeigen (LED grün) oder bei zu hoher Spannung warnen (LED rot). ◀

CAD-Technik

49

Peter Döring



49.1 Allgemeines 582
49.2 Hardware und Software 582
49.3 Erstellen von Schaltplänen 582
49.4 Erstellen von Layouts 582
49.5 Anwendungen in der Elektronik 586
49.6 Auswahl von CAD-Systemen 586

Teil VI

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_49

49.1 Allgemeines

Unter CAD (Computer Aided Design) versteht man ganz allgemein computergestütztes konstruieren. Hierbei steht die Erstellung technischer Unterlagen im Vordergrund. Für die Elektrotechnik und Elektronik sind dies einmal Stromlauf- und Funktionspläne, Verdrahtungspläne, Lagepläne oder Layouts zur Herstellung gedruckter Schaltungen und zum anderen Simulationsprogramme, die Entwicklungsarbeiten erheblich verkürzen.

Vorteile beim Einsatz dieser Technik liegen in der schnellen Korrigierbarkeit von Fehlern.

49.2 Hardware und Software

Unter Hardware versteht man alle Geräte die zu einer Computeranlage gehören. Hierbei unterscheidet man Eingabegeräte wie Tastatur, Maus, Grafiktablett mit Lupe und Scanner. Ausgabegeräte sind Bildschirm, Drucker und Plotter.

Die Zeichnungen werden häufig mit Hilfe der Maus durchgeführt. Hierbei werden die Bewegungen der Rollkugel in Signale umgewandelt und von einem Maustreiber (spezielles Programm) ausgewertet. Der Scanner wird eingesetzt, um über eine Optik Punkt für Punkt ganze Bilder einzulesen. Neben dem Drucker wird als Pheripheriegerät zur Ausgabe von Grafiken der Plotter eingesetzt. Mit seinem Schreibstift kann direkt auf Papier oder Folie geschrieben werden. Beim Fotoplotter belichtet ein Lichtstrahl das Papier.

Oft arbeiten CAD-Systeme mit zwei Bildschirmen, wobei der eine als Grafik- und der andere als Textbildschirm genutzt wird. Grundvoraussetzung vieler CAD-Programme sind Rechner ab 486 mit 16 MB-RAM. Beispiele für Zeichenprogramme sind AUTOCAD, AUTOSKETCH und besonders auf die Bedürfnisse in der Elektrotechnik und Elektronik abgestellt ORCAD und WSCAD. Zur Layouterstellung wird häufig EAGLE eingesetzt und im Bereich der Simulationsprogramme sind PSPICE und MICROCAP und für den Ausbildungsbereich ELECTRONIC WORKBENCH zu nennen.

Um CAD-Systeme richtig und effektiv einzusetzen bedarf es einer gründlichen Einarbeitung.

49.3 Erstellen von Schaltplänen

Bei der Erstellung technischer Zeichnungen am Bildschirm können Fehler schnell korrigiert werden. Häufig verwendete Symbole sind in Bibliotheken gesammelt und können dort abgerufen werden. Sie lassen sich beim Laden beliebig platzieren und können gedreht oder gespiegelt werden. Nachträgliches Verschieben von Baugruppen, Bauteilen, Leitungen oder Schriften ermöglichen eine optimale Darstellung. Mit Hilfe von Bauteileeditoren können Bauteilebibliotheken selbst erweitert werden. Eigene Symbole lassen sich somit herstellen und Vergößerungen und Verkleinerungen sind problemlos durchzuführen. Da Zeichnungen reproduziert werden können sind auch nachträgliche Änderungen möglich. Die Zeichnungsausgabe erfolgt auf einem Drucker oder Plotter. CAD-Zeichnungen lassen sich abspeichern und in andere Programme wie beispielsweise Textverarbeitungsprogramme einfügen. Der Einsatz von CAD bedeutet eine erhebliche Zeitersparnis gegenüber Arbeiten am konventionellen Zeichenbrett.

49.4 Erstellen von Layouts

EAGLE ist ein bekanntes und weitverbreitetes CAD-Programm und für den Entwurf von Platinenlayouts optimiert. Die Ausgabe der Layouts kann über Matrixdrucker, Plotter, Fotoplotter, Laserdrucker und Bohrmaschinen erfolgen. Mit dem Layout-Editor lassen sich beliebige Leiterbahnbreiten, Lötaugenformen und -durchmesser herstellen. Er kann für konventionelle und SMD-Technik eingesetzt werden.

Schaltpläne können auch mit anderen CAD-Programmen beispielsweise mit ORCAD gezeichnet werden. Die im Anschluss an die Zeichenarbeit erstellte Netzliste kann in EAGLE-Script-Dateien konvertiert werden. In der Netzliste (Netlist) befinden sich alle verwendeten Bauelemente und ihre Verbindung untereinander.

Beispiel

Entwurf einer gedruckten Schaltung mit dem Programm EAGLE

Mit dem EDIT-Befehl wird der Schaltplan (Schematic) und mit USE-Befehl die gewünschte Bibliothek (Library) geöffnet. Aus dieser kann das entsprechende Bauteil gewählt und in einem zuvor gewählten Rahmen (Frames) platziert werden. Der Schaltplan nach Abb. 49.1 kann mit den Befehlen „Schieben (MOVE), Löschen (DELETE) und Drehen (ROTATE)„ korrigiert und geändert werden. Bauelemente werden mit Namen (NAME) und Wert (VALUE) eindeutig gekennzeichnet. Mit dem NET-Befehl lassen sich die entsprechenden Verbindungen zeichnen und Verbindungspunkte mit einem Lötpunkt (JUNCTION) versehen. ◄

Der Befehl ERC (Electrical Rule Check) ermöglicht eine Überprüfung der Verbindungen auf Kurzschlüsse. Mit dem Speicherbefehl (WRITE) wird die Schaltung abgespeichert.

Der BOARD-Befehl lässt auf dem Bildschirm eine Leerplatine erscheinen mit allen Bauelementen des Schaltplans nach Abb. 49.2. Die dem Bauelement im Schaltplan zugeordnete

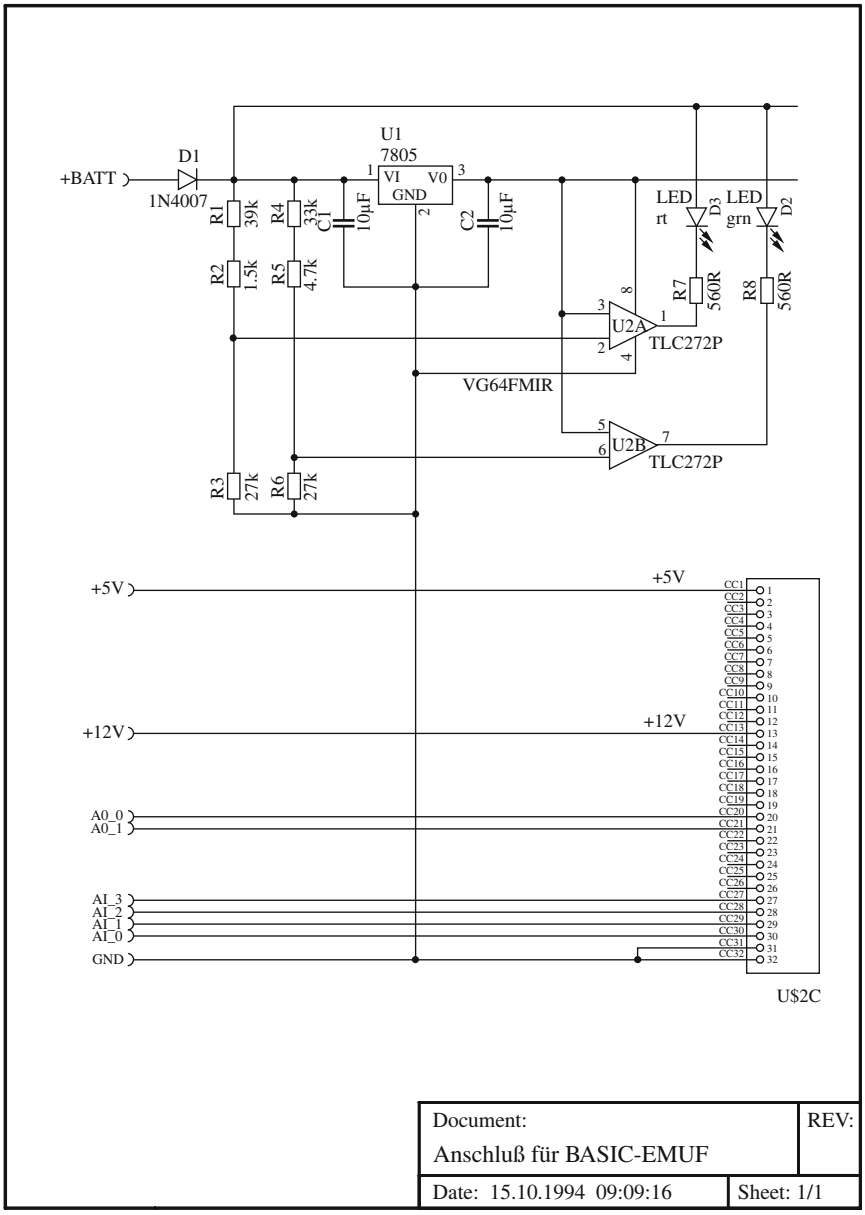


Abb. 49.1 Schaltplan

Bauform erscheint automatisch. Man braucht sich beispielsweise keine Gedanken mehr zu machen, ob das IC TLC272 mit 8 oder 14 PINs (Anschlüsse) versehen ist. Die Bauelemente sind durch Luftlinien (Gummileitungen) miteinander so verbunden, wie der Schaltplan dies vorgegeben hat.

Die Elemente können nach den Wünschen des Entwicklers verschoben und auf die vorgegebene Leerplatine platziert werden (Abb. 49.3).

Mit dem Befehl AUTO erfolgt ein automatisches Routen (Herstellen des Layouts, Leiterplattenentflechtung) oder mit dem Befehl ROUTE ein manuelles. Am BOARD lassen sich anschließend auch noch Korrekturen durchführen. Dies ist in der Praxis not-wendig, nicht weil Verbindungen falsch gelegt oder vergessen worden sind, sondern weil beispielsweise Versorgungsspannungsleitungen breiter ausgelegt werden sollen. Nach Abschluss der Designarbeit erzeugt das Programm die verschiedenen Daten zur Platinenherstellung. Hierzu gehören Layout, Lötmaske, Lageplan und Stücklisten. Abb. 49.4 zeigt den Lagenplan und Abb. 49.5 das entsprechende Layout.

Von Hand erstellte Layouts haben den Nachteil, dass Fehler nie ganz ausgeschlossen werden können. Bei CAD-Systemen treten keine Fehler mehr auf, wenn das Programm richtig bedient wird.

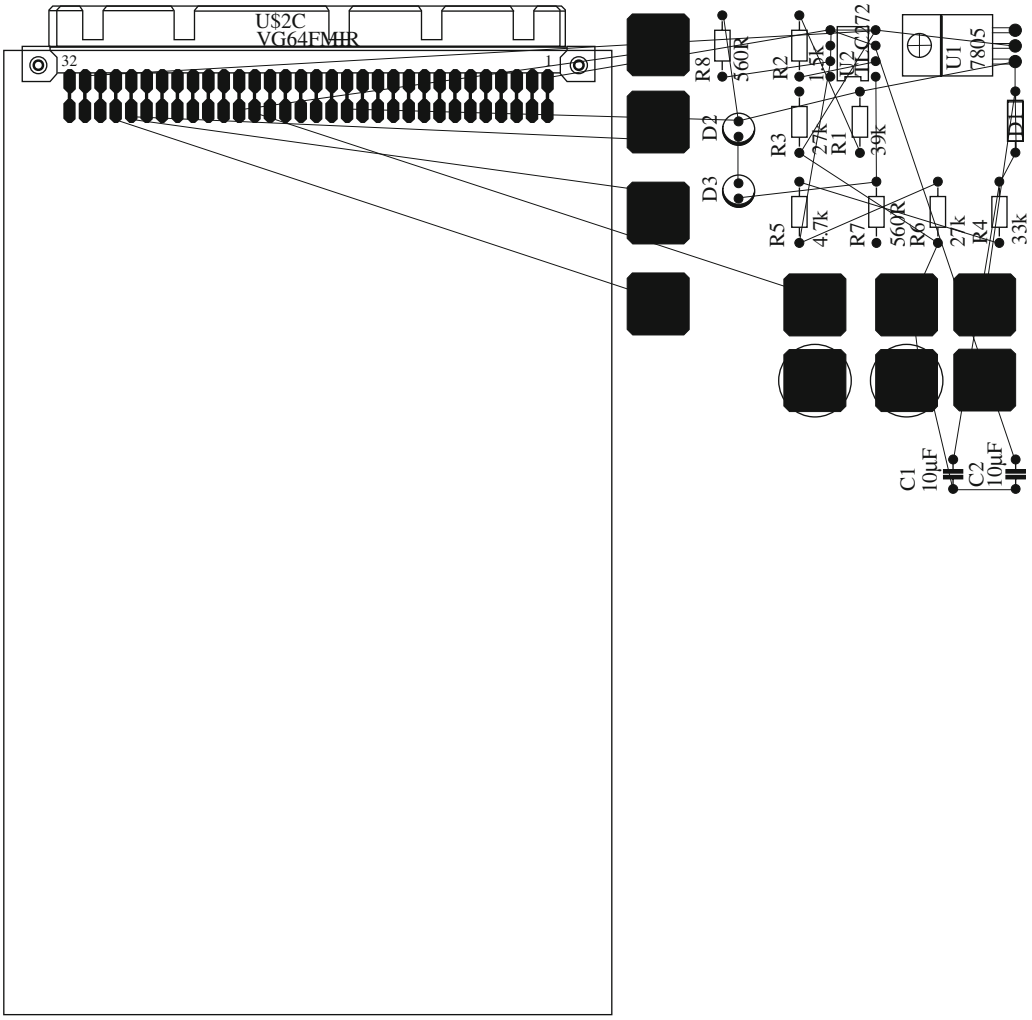


Abb. 49.2 Leerplatine und Bauelemente

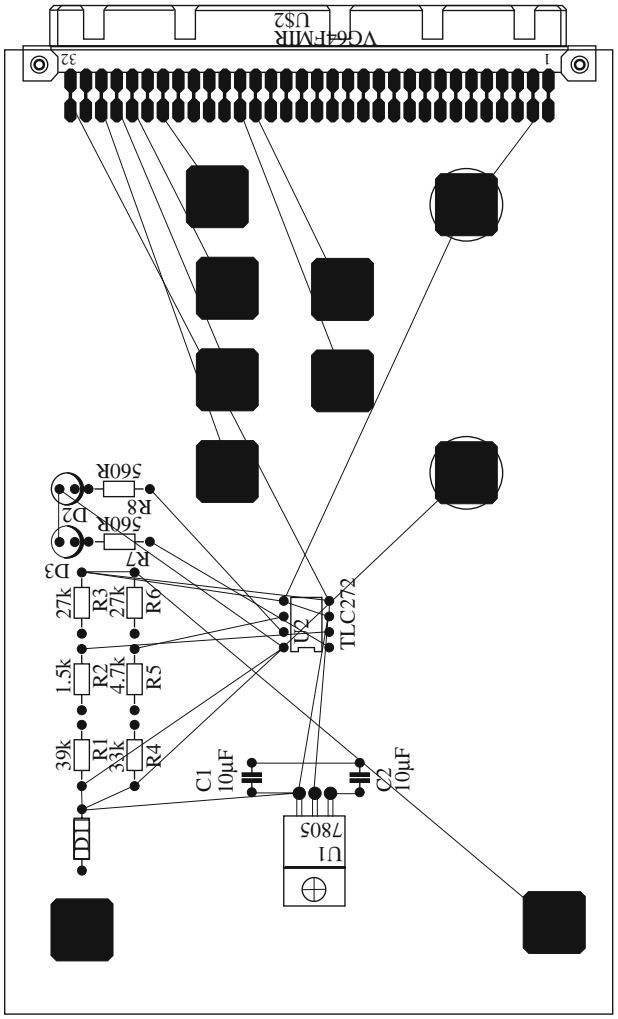


Abb. 49.3 Anordnung der Bauelemente auf der Platine

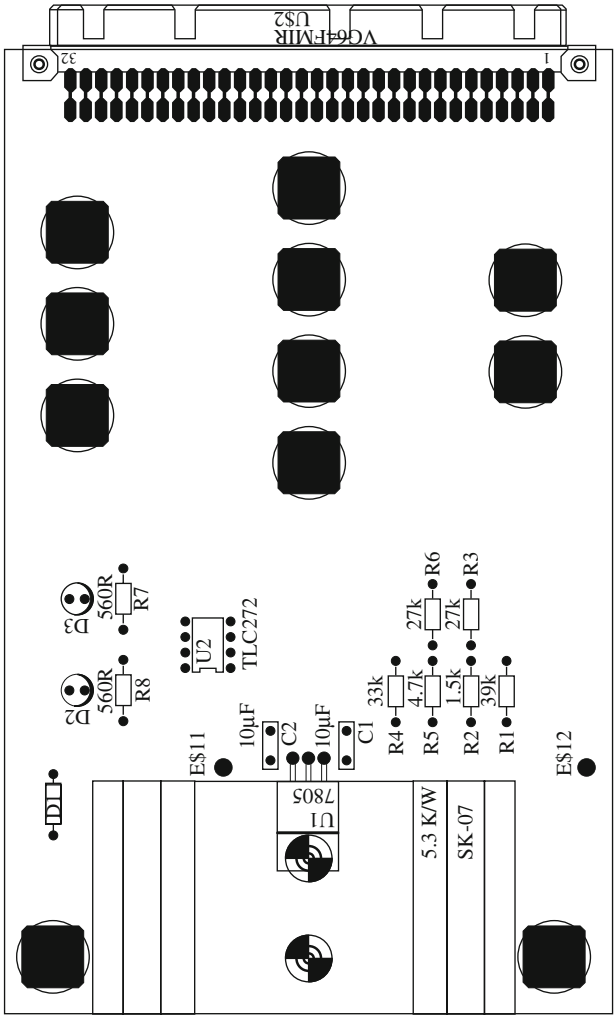


Abb. 49.4 Lageplan

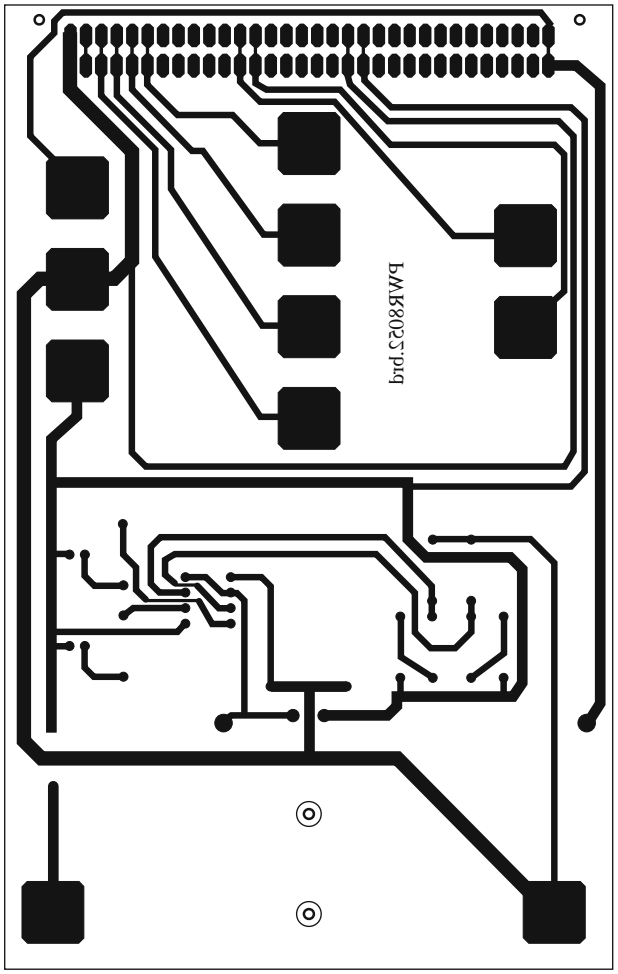


Abb. 49.5 Layout

49.5 Anwendungen in der Elektronik

- Elektronische Schaltungen
- Schaltungsentwurf
- Entwurf gedruckter Schaltungen
- Layout
- Leiterplattenentflechtung
- Stücklistenerstellung
- Bestückungsplan

49.6 Auswahl von CAD-Systemen

Entwurf gedruckter Schaltungen

- Richtlinien zum Entwurf
- Herstellung gedruckter Schaltungen
- Stromlaufplan
- Layouterstellung (Freihand, CAD)
- Bestückungsliste
- Stückliste
- Maske für Lötstoplack
- Bearbeitungszeichnung

Datentechnik

VII



50 Digitaltechnik 589
51 Integrierte Schaltkreise der Digitaltechnik 633
52 Mikrocomputertechnik 661
53 Computertechnik 729
54 Programmiertechnik 741
55 Datenkommunikation 747

Digitaltechnik

50

Teil VII

Heribert Gierens



50.1 Grundbegiffe der Digitaltechnik 590
50.2 Logische Grundschaltungen 590
50.3 Schaltalgebra 593
50.4 Zahlensysteme in der Digital- und Datenverarbeitung 602
50.5 Codes 606
50.6 Digitale Grundschaltungen 611

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_50

50.1 Grundbegiffe der Digitaltechnik

Als Teilgebiet der Elektronik, Automatisierungs-, Nachrichten- und Datenverarbeitungstechnik kann mit Hilfe der Digitaltechnik eine große Zahl technischer Probleme gelöst und das Verständnis komplexer Systeme gefördert werden. Im Gegensatz zu analogen Schaltungen, wo Werte über einen Bereich kontinuierlich ausgegeben werden, können Ein- und Ausgangssignale digitaler Schaltungen nur zwei mögliche stabile Zustände annehmen. Man spricht daher von einem *binären System*, das aufgrund der Zweiwertigkeit logische Entscheidungen ermöglicht. Die Schaltungen selber nennt man logische Schaltungen. Hierbei sind die binären Zustände zwei Spannungswertbereichen (Pegeln) zugeordnet. Der niedrigere Spannungsbereich (näher an minus unendlich) wird mit LOW und der positivere Spannungsbereich (näher an plus unendlich) wird mit HIGH bezeichnet. Die genauen Werte sind von der schaltungstechnischen Realisierung abhängig. Soll technologieunabhängig gearbeitet werden, also ohne auf die besonderen Bedingungen zu achten, mit denen die Schaltung später vielleicht realisiert werden soll, werden die binären Zustände mit den Zeichen „0“ und „1“ beschrieben. Je nach gewähltem Bezugspunkt wird zwischen positiver und negativer Logik unterschieden. Dabei entspricht bei positiver Logik die „1“ dem H- und die „0“ dem L-Pegel; bei negativer Logik dagegen die „0“ dem H- und die „1“ dem L-Pegel (Abb. 50.1). Die Zeichen „0“ und „1“ heißen *Bits*. Das *Bit* (Binary Digit = Binärzeichen) ist die kleinste Informationseinheit. Vier Bits ergeben eine Tetrade oder Nibble, und 2 Tetraden ergeben ein Byte. Die nachfolgend größeren Informationseinheiten sind mit Potenzen zur Basis 2 festgelegt (Abb. 50.2). Ein Byte nennt man in 8-Bit-Systemen auch *„Wort“*. Allgemein kann aber ein „Wort“ auch aus mehreren Bytes zusammengesetzt sein. Die höchste und niederwertigste Stelle eines Wortes wird besonders gekennzeichnet. Man spricht vom MSB (Most Significant Bit) und vom LSB (Least Significant Bit).

Die Grundelemente digitaler Schaltungen sind logische Grundbausteine. Sie besitzen voneinander unabhängige Eingänge und einen Ausgang, dessen Ausgangsvariable jeweils von Eingangsvariablen abhängt. Bausteine mit solchen Grundverknüpfungen nennt man *Gatter*. Jede Verknüpfung kann durch ein Schaltzeichen, eine Wertetabelle (Wahrheitstabelle, Funktionstabelle), eine Gleichung oder einen Impulszeitplan (Funktionszeitplan) dargestellt werden. Die grafische Darstellung mit Schaltzeichen wird verwendet, um größere Zusammenhänge aufzuzeigen. Die Wertetabelle drückt in übersichtlicher Weise aus, welche Ausgangszustände sich bei allen möglichen Eingangszuständen ergeben. Der Impulszeitplan stellt den Zusammenhang zwischen Eingangs- und Ausgangssignalen dar und erlaubt damit den Zugang zu den Messtechniken.

Gleichungen sind Hilfsmittel, um mit dem Einsatz der Schaltalgebra vereinfachte Schaltungen zu finden. Die Aussage jeder dieser genannten Darstellungsformen ist dabei dieselbe. Alle logischen Zusammenhänge in der Digitaltechnik lassen sich durch die drei Grundverknüpfungen beziehungsweise Grundfunktionen NICHT (Negation), UND (Konjunktion) und ODER

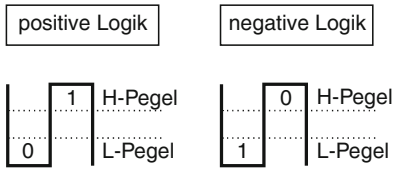


Abb. 50.1 Pegelzuordnung

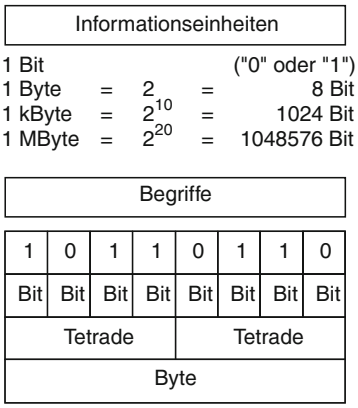


Abb. 50.2 Informationseinheiten und Begriffe

(Disjunktion) beschreiben. Weitere abgeleitete Verknüpfungen vereinfachen die Realisierung digitaler Schaltungen. Es handelt sich dabei um die Funktionen: NAND, NOR, Exclusiv ODER (Antivalenz) und Exclusiv NOR (Äquivalenz).

Als Übersicht zeigt Abb. 50.3 die Schaltzeichen aller Grundverknüpfungen mit einer Funktionsbeschreibung in Form einer Wertetabelle, eines Impulszeitplans und einer Funktionsgleichung. Besitzt eine Verknüpfung nur einen Eingang, sind zwei Eingangskombinationen, bei zwei Eingängen vier verschiedene Eingangskombinationen möglich. Bei n Eingängen gibt es 2^n verschiedene Eingangskombinationen.

50.2 Logische Grundschaltungen

50.2.1 Grundverknüpfungen

Einfache Schaltungen aus Leuchtdiode und Schalter, bei der die Eingangsgröße der Schalter mit den Zuordnungen: Schalter offen = „0“ und Schalter geschlossen = „1“, und die Ausgangsgröße die Leuchtdiode mit den Zuordnungen: LED leuchtet nicht = „0“ und LED leuchtet = „1“ ist, verdeutlichen die jeweilige Funktion (Abb. 50.4).

50.2.1.1 NICHT-Verknüpfung

Die NICHT-Verknüpfung (NOT) hat nur eine Eingangsvariable. Da die Ausgangsvariable immer den entgegengesetzten Wert der Eingangsvariablen annimmt, spricht man auch von einem *Inverter*. Ist der Schalter geöffnet („0“), kann Strom durch die

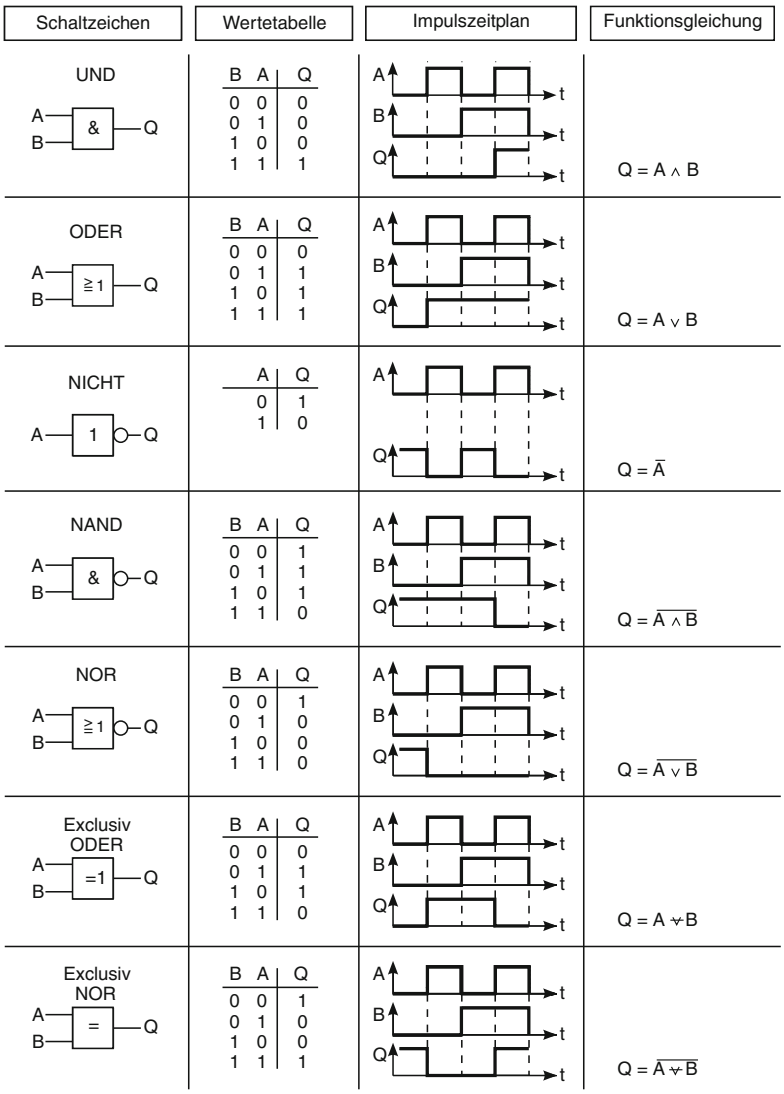


Abb. 50.3 Logische Verknüpfungen

Leuchtdiode fließen und sie leuchtet („1“). Bei geschlossenem Schalter („1“) ist die Diode kurzgeschlossen. Damit fließt durch die Leuchtdiode kein Strom und sie leuchtet nicht („0“).

50.2.1.2 UND-Verknüpfung

Eine UND-Verknüpfung (AND) besitzt mindestens zwei Eingangsvariable. Der Ausgang wird nur „1“, wenn alle Eingänge „1“ sind, wenn also A UND B UND *n* gleichzeitig 1 ist. Die Schaltung zeigt, dass die Diode nur leuchten kann („1“ ergibt), wenn der eine UND der andere Schalter geschlossen ist (also „1“ aufweist).

50.2.1.3 ODER-Verknüpfung

Eine ODER-Verknüpfung (OR) erhält am Ausgang dann eine „1“, sobald eine Eingangsvariable „1“ ist, wenn also ein ODER zwei ODER *n* Eingänge „1“ sind. In der Schaltung leuchtet die Diode („1“), sobald einer der Schalter geschlossen („1“) ist.

50.2.2 Realisierungsmöglichkeiten logischer Verknüpfungen

Logische Verknüpfungen lassen sich durch Mechanik (Schalter, Schütze), Elektronik oder Software realisieren. In der Elektronik werden Schalter durch Dioden und Transistoren ersetzt. Abb. 50.5 zeigt Grundverknüpfungen solcher Bauelemente. Man spricht auch von Dioden-Widerstandslogik oder Transistor-Diodenlogik. Ihr Einsatz ist so heute nicht mehr üblich. Die Schaltungen dienen lediglich zum Grundverständnis logischer Schaltungen.

Der Eingang ist jeweils mit A beziehungsweise B und der Ausgang mit Q gekennzeichnet. Bei der UND-Schaltung stellt sich beispielsweise nur die Spannung +5 V („1“) am Ausgang ein, wenn A *und* B an +5 V („1“) liegen, weil dann beide Dioden gesperrt sind. In den drei möglichen anderen Fällen ergibt sich eine Spannung von ca. 0,7 V („0“) am Ausgang.

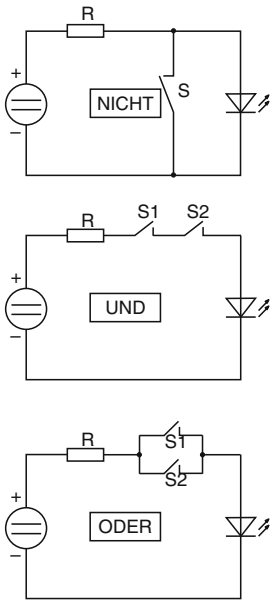


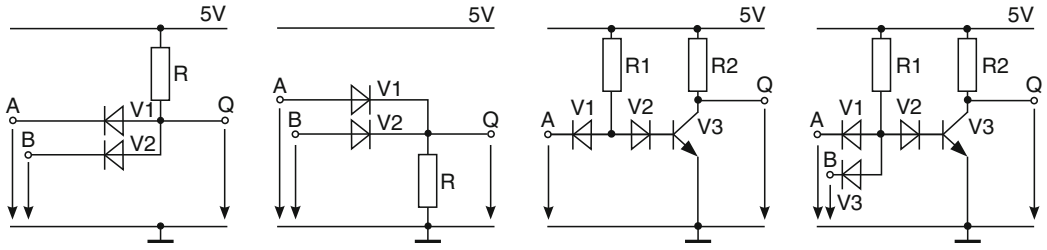
Abb. 50.4 Grundverknüpfungen

Bei der ODER-Schaltung reicht es, wenn A *oder* B auf +5 V liegt, um am Ausgang ungefähr 5 V („1“) zu erzeugen.

Liegt bei der NICHT-Schaltung A auf Masse („0“), ist der Transistor gesperrt, und die Ausgangsspannung nimmt den Betriebsspannungswert +5 V („1“) an. Der Transistor schaltet nur durch, wenn der Eingang A mit +5 V („1“) verbunden wird. Die Ausgangsspannung beträgt in diesem Fall etwa 0,2 V („0“).

Heute setzt man *integrierte Standardbausteine* ein. Die angebotenen Logikfamilien unterscheiden sich durch unterschiedliche Pegel, durch Verlustleistungen und Laufzeiten. Die Bausteine selbst besitzen oft mehrere Gatter (Abb. 50.6).

Ihre PIN- oder Anschlussbelegung können aus den entsprechenden Datenbüchern entnommen werden. Bei Betrachtung des IC's von oben (On Top) ist die Zählrichtung immer von links oben nach rechts oben. Bei der Darstellung von Funktionsplänen werden die Betriebsspannungsanschlüsse nicht mit gezeichnet. Als Beispiel dient ein Baustein mit vier NAND-Gattern. Weitere Beispiele sind in den Abschn. 42.1 „Integrierte Schaltungen“ und 51.8.3 „Programmierbare Logikbausteine“ zu finden.



UND

UA	UB	UQ
0V	0V	0,7V
0V	5V	0,7V
5V	0V	0,7V
5V	5V	5,0V

ODER

UA	UB	UQ
0V	0V	0,7V
0V	5V	5,0V
5V	0V	5,0V
5V	5V	5,0V

NICHT

UA	UQ
0V	5,0V
5V	0,2V

NAND

UA	UB	UQ
0V	0V	5,0V
0V	5V	5,0V
5V	0V	5,0V
5V	5V	0,7V

Abb. 50.5 Dioden- und Transistorlogik

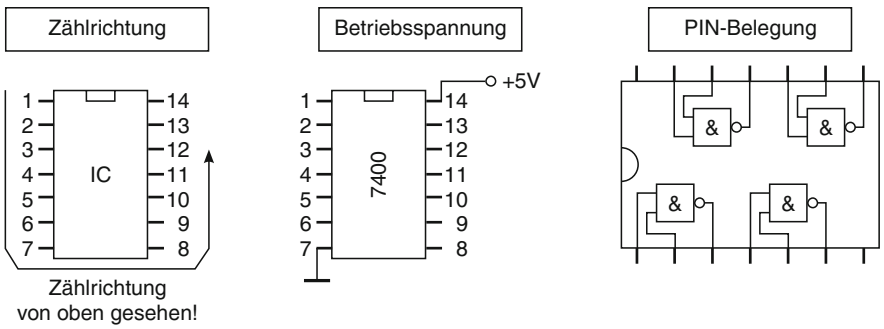


Abb. 50.6 Standardbaustein mit NAND-Gattern

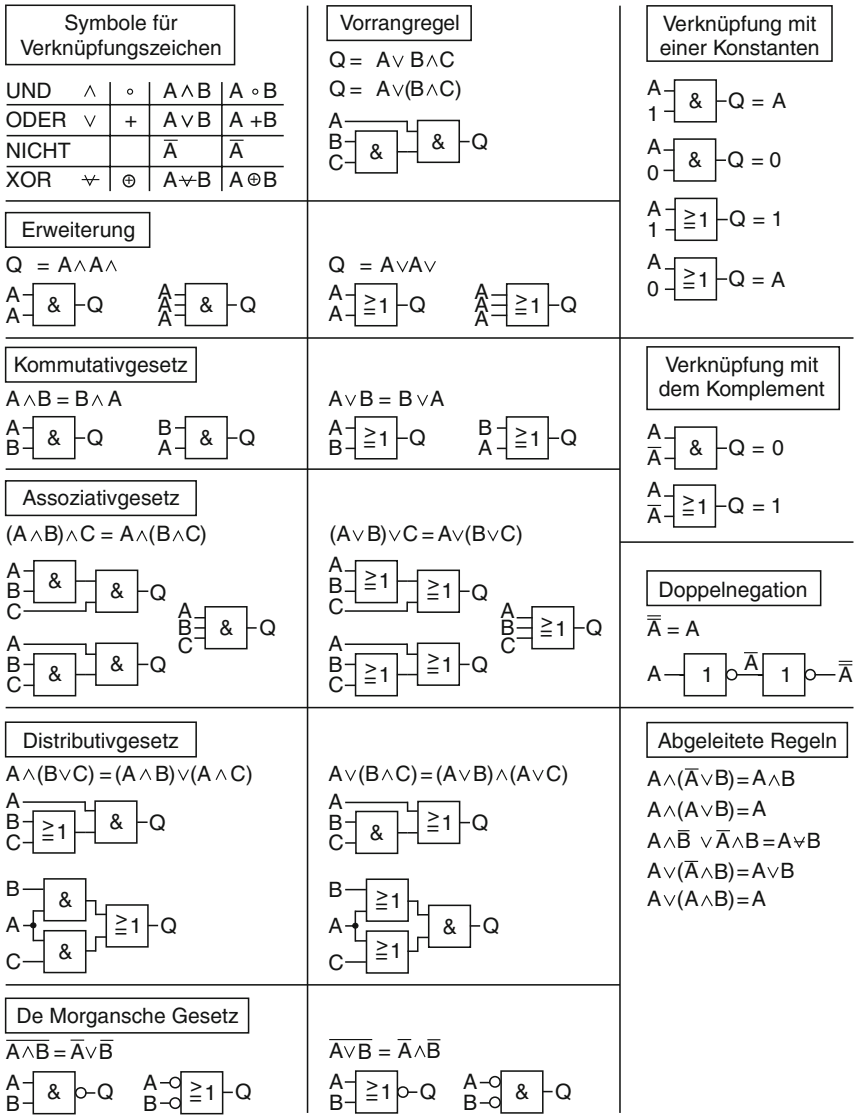


Abb. 50.7 Regeln und Gesetze der Schaltalgebra

50.3 Schaltalgebra

50.3.1 Allgemeines

Die Schaltalgebra ist ein Hilfsmittel zur mathematischen Beschreibung logischer Funktionen. Mit ihren Regeln und Gesetzen können Gleichungen umgeformt, vereinfacht und so Schaltungen optimal realisiert werden (Synthese von Schaltungen). Für die Verknüpfungszeichen von Variablen existieren unterschiedliche Darstellungen. Die beiden in der Digitaltechnik häufig angewandten Formen sind in der Übersicht (Abb. 50.7) mit aufgeführt.

In der Schaltalgebra muss eine bestimmte Rangfolge der Operationen eingehalten werden. Sie besagt, dass NICHT-Verknüpfungen vor UND- und ODER-Verknüpfungen durchzuführen sind (Vorrangregel). Das *Kommutativgesetz* wird auch Vertauschungsgesetz genannt. Danach können die einzelnen Eingangsvariablen vertauscht werden. So darf beispielsweise nach Abb. 50.4 bei der UND-Verknüpfung zuerst S1 und dann S2 oder umgekehrt geschaltet werden. Die Reihenfolge der Schalter hat keinen Einfluss auf die Funktion der Gesamtschaltung. Dasselbe gilt auch für die Parallelschaltung der Schalter bei der ODER-Verknüpfung.

Mit dem *Assoziativgesetz* werden Variable einer UND- oder einer ODER-Verknüpfung zusammengefasst. Weil das Assoziativgesetz Variable durch Klammern verbindet, heißt das Gesetz auch *Verbindungsgesetz*. Klammern können hierbei gesetzt oder fortgelassen werden. Wird für eine Schaltung eine festgelegte Anzahl von Eingängen benötigt, lässt sich die Aufgabe mit diesem Gesetz lösen.

Das *Distributiv-* oder *Verteilungsgesetz* ist aus der Algebra bekannt. Hiermit kann eine gemeinsame Variable ausgeklammert werden.

Wertetabelle

C	B	A	Q
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

Minterme

$\overline{A} \wedge \overline{B} \wedge \overline{C}$ $\quad \overline{A} \wedge \overline{B} \wedge \overline{C}$

$\overline{A} \wedge \overline{B} \wedge C$ $\quad \overline{A} \wedge \overline{B} \wedge C$

$\overline{A} \wedge B \wedge \overline{C}$

$A \wedge B \wedge C$

$A \wedge \overline{B} \wedge \overline{C}$ $\quad A \wedge \overline{B} \wedge \overline{C}$

$A \wedge B \wedge C$

$A \wedge B \wedge \overline{C}$ $\quad A \wedge B \wedge \overline{C}$

$A \wedge B \wedge C$

Maxterme

$\overline{A} \vee \overline{B} \vee \overline{C}$

$\overline{A} \vee \overline{B} \vee C$

$\overline{A} \vee B \vee \overline{C}$ $\quad \overline{A} \vee B \vee \overline{C}$

$\overline{A} \vee B \vee C$ $\quad \overline{A} \vee B \vee C$

$A \vee \overline{B} \vee \overline{C}$

$A \vee \overline{B} \vee C$ $\quad A \vee \overline{B} \vee C$

$A \vee B \vee \overline{C}$

$A \vee B \vee C$ $\quad A \vee B \vee \overline{C}$

Gleichungen

Konjunktive Normalform

$$Q = \overline{A} \wedge \overline{B} \wedge \overline{C} \vee \overline{A} \wedge \overline{B} \wedge C \vee A \wedge \overline{B} \wedge \overline{C} \vee A \wedge B \wedge \overline{C} \quad \text{(I-1)}$$

$$\overline{Q} = \overline{A} \wedge B \wedge \overline{C} \vee \overline{A} \wedge B \wedge C \vee A \wedge \overline{B} \wedge C \vee A \wedge B \wedge C \quad \text{(I-2)}$$

$$\overline{\overline{Q}} = Q \quad \text{(I-3)}$$

Disjunktive Normalform

$$Q = (\overline{A} \vee B \vee \overline{C}) \wedge (\overline{A} \vee B \vee C) \wedge (A \vee \overline{B} \vee C) \wedge (A \vee B \vee C) \quad \text{(I-4)}$$

Abb. 50.8 Disjunktive und konjunktive Normalform

Das *De Morgan'sche Gesetz* hilft bei der Umsetzung der Grundfunktionen in NOR- und NAND-Technik. Eine NAND-Verknüpfung der Variablen A und B kann durch Negierung dieser Variablen in eine ODER-Verknüpfung umgewandelt werden, und bei einer NOR-Verknüpfung ist durch Negieren der Variablen ein Ersatz durch UND möglich (siehe auch Abb. 50.11). Da alle Grundfunktionen durch NAND oder NOR nachgebildet werden können (Abb. 50.12), lässt sich die Lagerhaltung durch den Einsatz dieser Techniken verkleinern.

Im Abb. 50.7 sind die Verknüpfungen mit einer Konstanten und wichtige abgeleitete Regeln zusammengefasst. Sind beispielsweise Doppelnegierungen bei einer Variablen vorhanden, so heben sich diese praktisch auf. Eine Erweiterung mit einer einmal vorhandenen Variablen darf immer durchgeführt werden, weil hiermit das Resultat nicht beeinflusst wird.

Vom *Absorptionsgesetz* spricht man, wenn eine der Variablen keine Auswirkung auf das Ergebnis hat. Bei der *Analyse* logischer Schaltungen können mit Hilfe der Schaltalgebra Schaltfunktionen in die entsprechende Wertetabelle umgesetzt werden. Erfolgt die Programmierung in der SPS mit nicht grafikfähigen Programmiergeräten, sind ebenfalls Kenntnisse der Schaltalgebra zur richtigen Programmierung notwendig.

50.3.2 Normalform einer binären Funktion

50.3.2.1 Disjunktive Normalform

Geht man von der Wertetabelle in Abb. 50.8 aus, kann für jede Zeile dieser Tabelle ein Funktionsterm geschrieben werden. Für Zeile 1 zum Beispiel: /A /B /C (sprich: A nicht UND B nicht UND C nicht). Die so durch UND verknüpften Funktionsterme nennt man *Minterme*. Von Interesse ist, wann die Funktion „1" ist. Sie soll nach der Wertetabelle den Wert „1" erhalten in der Zeile 0, ODER in Zeile 1, ODER in Zeile 4, ODER in Zeile 6. Damit lässt sich (I-1) für Q aufstellen. Man spricht von der *disjunktiven* (ODER-) *Normalform* (DNF). Sind in einer Wertetabelle die Anzahl der „0"-Funktionen für Q geringer, ist es günstiger, die Gleichung für /Q aufzustellen und anschließend das Ergebnis zu invertieren.

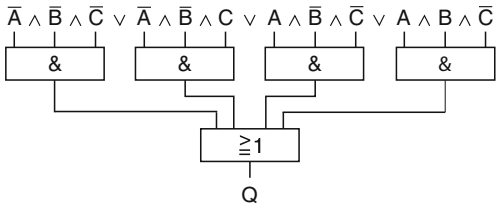


Abb. 50.9 Umsetzung einer Gleichung in eine Schaltung

50.3.2.2 Konjunktive Normalform

Die *konjunktive* (UND-) *Normalform* (KNF) erhält man, wenn die *Maxterme*, also alle ODER-Verknüpfungen, die Q = „0" ergeben, UND-verknüpft werden (I-4). Jede Schaltung lässt sich durch eine Gleichung in disjunktiver oder konjunktiver Normalform beschreiben.

50.3.2.3 Umwandeln der Gleichung in Schaltzeichen

Gleichungen lassen sich in die genormte Darstellung umwandeln, indem die entsprechenden Schaltzeichen der Verknüpfungen gezeichnet werden, wie das beispielhaft für (I-1) im Abb. 50.9 gezeigt wird.

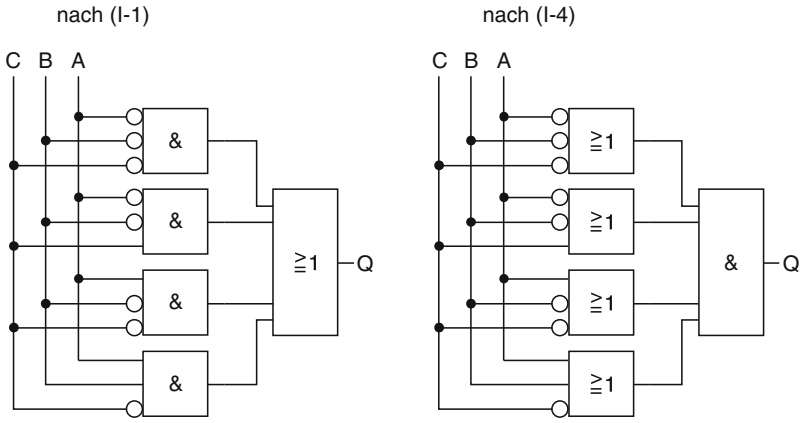


Abb. 50.10 Schaltungen aus (I-1) und (I-4)

$$Q = \overline{A} \wedge \overline{B} \wedge \overline{C} \vee \overline{A} \wedge \overline{B} \wedge C \vee A \wedge \overline{B} \wedge \overline{C} \vee A \wedge B \wedge \overline{C}$$

$$Q = \overline{A}\,\overline{B}\,\overline{C} \vee \overline{A}\,\overline{B}\,C \vee A\,\overline{B}\,\overline{C} \vee A\,B\,\overline{C}$$

$$Q = \overline{A}\,\overline{B} \quad (\overline{C} \vee C) \quad \vee A\,\overline{C} \quad (\overline{B} \vee B)$$

$$Q = \overline{A}\,\overline{B} \quad \wedge \quad 1 \qquad\quad A\,\overline{C} \quad \wedge \quad 1$$

$$Q = \overline{A}\,\overline{B} \qquad\qquad \vee A\,\overline{C}$$

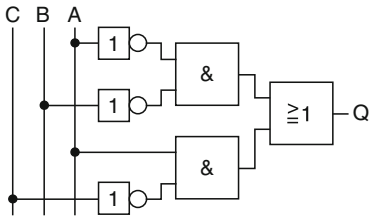


Abb. 50.11 Schaltungsminimierung

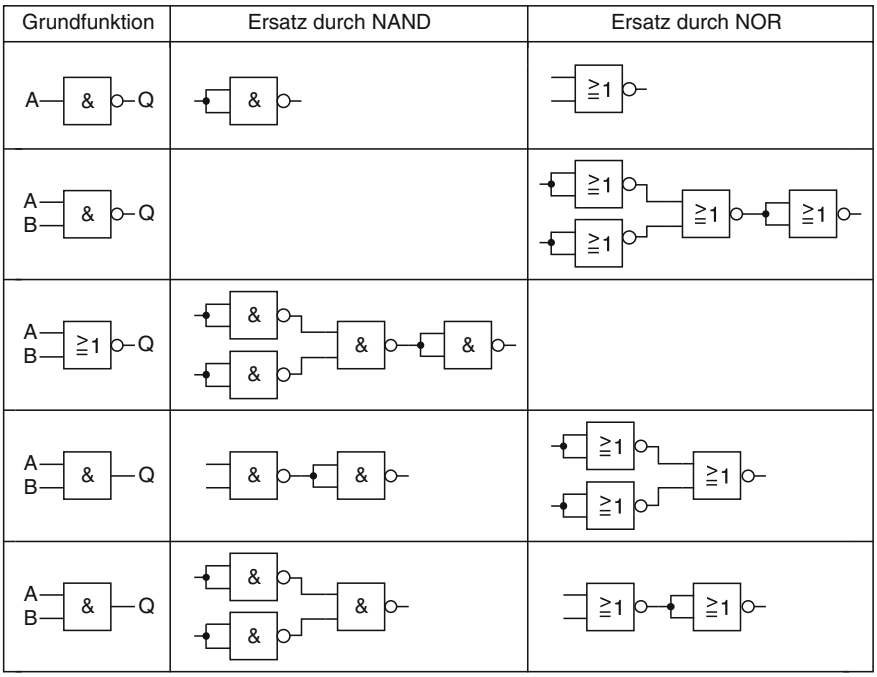


Abb. 50.12 Ersatz der Grundfunktionen durch NAND- und NOR-Technik

Oft sind keine negierten Eingänge wie /A, /B, /C vorhanden. Dann heißen die Eingänge A, B und C, und die notwendige Invertierung wird durch ein NICHT-Gatter realisiert. Abb. 50.10 verdeutlicht dies an der Umsetzung von (I-1) und (I-4), wobei die Eingangsinverter durch die Invertierungszeichen dargestellt sind.

50.3.2.4 Schaltungsminimierung mit Hilfe der Schaltalgebra

Mit Hilfe der Schaltalgebra lassen sich die Gleichungen vereinfachen und in eine digitale Schaltung umsetzen. Abb. 50.11 zeigt die Vereinfachung von (I-1).

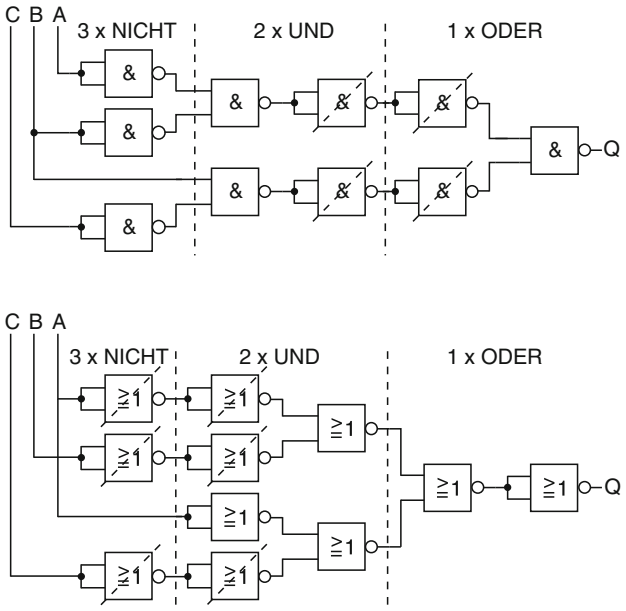


Abb. 50.13 Umsetzung einer Schaltung in NAND- und NOR-Technik

Mit dem Distributivgesetz erhält man eine ODER-Verknüpfung mit dem Komplement der Variablen. Da dies „1“ ergibt, kann die entsprechende Variable aufgrund der UND-Verknüpfung mit dieser Konstante entfallen. Das Ergebnis bildet die Gleichung (I-5). Zur besseren Übersicht werden bei den Gleichungen die UND-Zeichen fortgelassen.

50.3.2.5 Umsetzung in NAND- oder NOR-Technik

Jede Grundfunktion lässt sich nach Abb. 50.12 durch NAND- oder NOR-Funktionen ersetzen.

Besteht nun der Wunsch, die Schaltung nach (I-4) nur aus NAND- oder NOR-Bausteinen aufzubauen, werden die Bausteine entsprechend eingesetzt, und man erhält eine funktionsfähige Schaltung. Treten hierbei Doppelnegierungen auf, dürfen Gatter entfallen, wie dies im Abb. 50.13 am Beispiel der Umsetzung in NAND- und NOR-Technik gezeigt wird.

Die Gesamtschaltung benötigt dann eine geringere Anzahl an Gattern.

Das umständliche Verfahren, andere Grundfunktionen so durch NAND zu ersetzen, kann umgangen werden, wenn die Ausgangsgleichung sofort nach de Morgan umgewandelt wird. Hierbei wird bei Änderung des Verknüpfungszeichens (aus UND wird ODER und umgekehrt) eine Negierung über beide Variable aufgetrennt.

Abb. 50.14 weist dies in Form einer Tabelle nach. Die eigentliche Umwandlung kann ganz formal erfolgen (Abb. 50.15). Gleichung (I-5) wird zweimal negiert. Dadurch ändert sich die Gleichung nicht. Nun wird, unter Ersatz des ODER-Zeichens durch das UND-Zeichen, die untere Invertierung überall dort, wo sich das ODER-Zeichen befindet, unterbrochen. Jede Invertierung wird dann zu einem NAND erweitert.

Der Vorteil in diesem Verfahren liegt darin, dass man in (I-6) schon die Schaltung „sehen“ kann.

Tritt die Forderung nach NAND mit nur zwei Eingängen auf, werden die entsprechenden Doppelnegierungen in eine Glei-

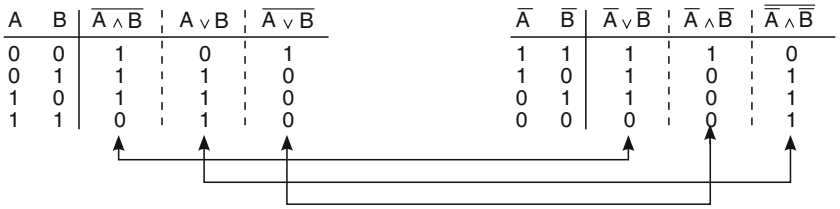


Abb. 50.14 Umwandlung nach de Morgan

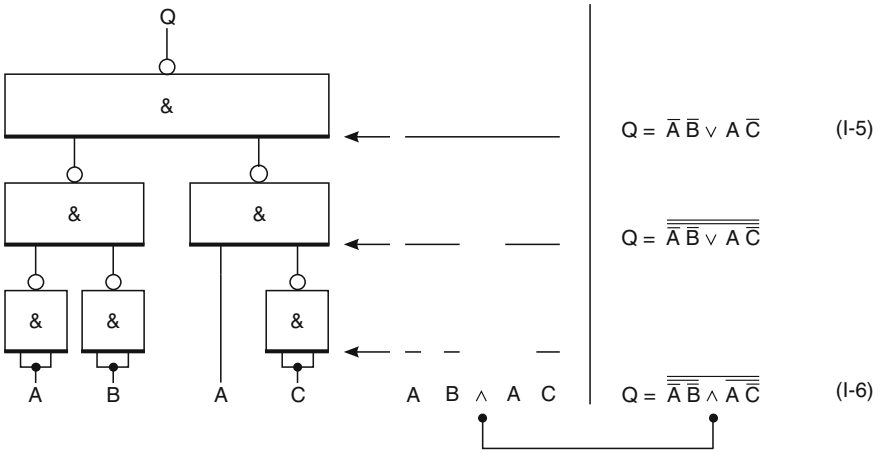


Abb. 50.15 Umsetzung nach NAND

$$\overline{A}\,\overline{B}\,C\,D \vee \overline{A}\,B\,\overline{C}\,D \vee A\,\overline{B}\,\overline{C}\,D \qquad \text{(I-7)}$$

$$\overline{\overline{A}\,\overline{B}\,C\,D} \vee \overline{\overline{A}\,B\,\overline{C}\,D} \vee \overline{A\,\overline{B}\,\overline{C}\,D} \qquad \text{(I-8)}$$

$$\overline{\overline{A}\,\overline{B}\,C\,D} \vee \overline{\overline{A}\,B\,\overline{C}\,D} \vee \overline{A\,\overline{B}\,\overline{C}\,D} \qquad \text{(I-9)}$$

Abb. 50.16 NAND-Umwandlung für zwei Eingänge

chung eingefügt. Die Umsetzung von (I-7) in NAND-Technik zur NAND-Form (I-8) und der Ergebnisbildung (I-9) werden im Abb. 50.16 verdeutlicht.

Eine aus dieser Gleichung erstellte Schaltung ist recht aufwendig. Sie kann rationeller hergestellt werden, wenn vorhandene Verknüpfungen mehrmals genutzt werden, wie das in Abb. 50.17 aus der Entwicklung von (I-7) dargestellt ist. Man spricht dann von einer *Vermaschung*.

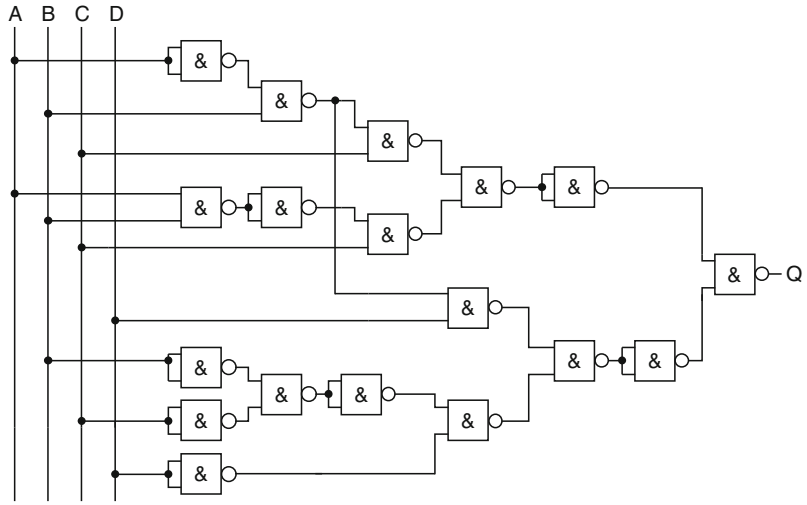


Abb. 50.17 Vermaschung

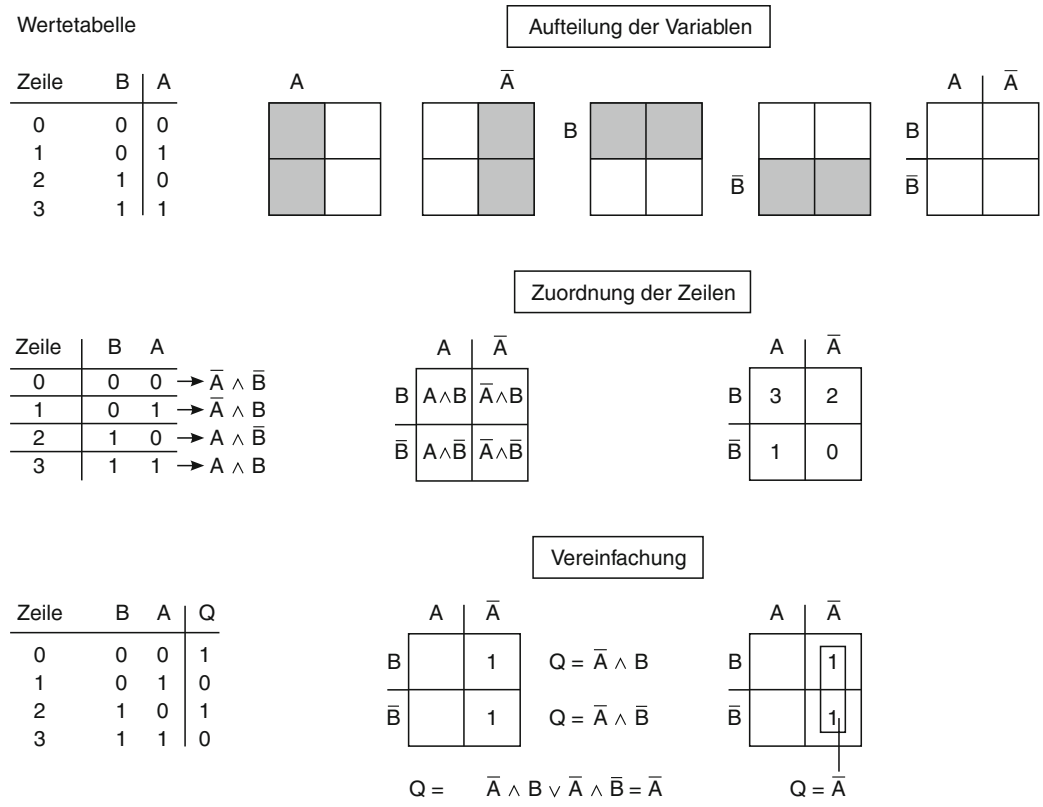


Abb. 50.18 Entwicklung der KV-Tabelle

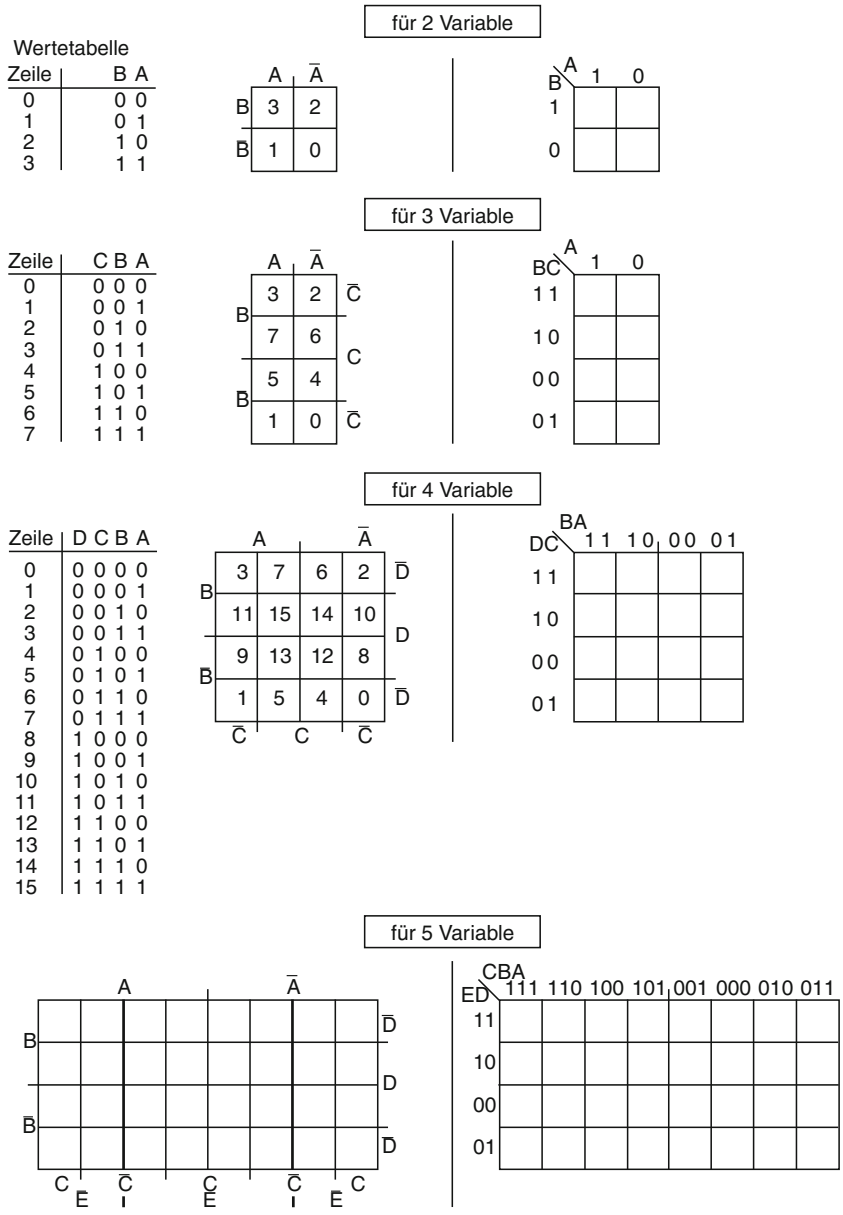


Abb. 50.19 KV-Diagramme für 2 bis 5 Variable

50.3.2.6 KV-Tabelle

Das von *Karnaugh* und *Veitch* entwickelte Verfahren ist praktisch eine besondere Art, eine Wertetabelle zu schreiben. Jedem Kästchen des Diagramms ist eine Zeile der Wertetabelle zugeordnet. Damit ergibt sich die Anzahl der Felder aus der Zahl der Eingangskombinationen. Die Anordnung erfolgt so, dass sich jeweils von Feld zu Feld nur eine Variable ändern darf (Abb. 50.18).

Im Bild sind die entsprechenden Felder schraffiert und die Zuordnung der Zeilen zu den Feldern gekennzeichnet. Vergleicht man zwei benachbarte Felder und verknüpfen die dafür stehenden konjunktiven (UND-) Verknüpfungen disjunktiv (ODER), wie in der Gleichung dargestellt, dann lässt sich erkennen, dass nach einer Vereinfachung nur eine Variable das Ergebnis bildet. Damit sind Vereinfachungen durch Schleifenbildung möglich. Abb. 50.19 zeigt verschiedene Darstellungen von KV-Tafeln für 2 bis 5 Variable.

Verknüpft, d. h. zu Schleifen zusammengefasst, werden dürfen immer nur $2, 4, 8, \ldots$ also 2^n Variable in senkrecht oder waagerecht nebeneinander liegenden benachbarten Feldern. Die Randfelder gelten ebenfalls als benachbart.

Abb. 50.20 zeigt hierzu Beispiele mit den sich hieraus ergebenden Gleichungen. Alle Schleifen sind durch UND verbunden. Für die Ergebnisbildung muss jede dieser Schleifen durch ODER verknüpft werden. Eine teilweise gegenseitige Überdeckung der Schleifen ist hierbei zulässig. Es sollten möglichst

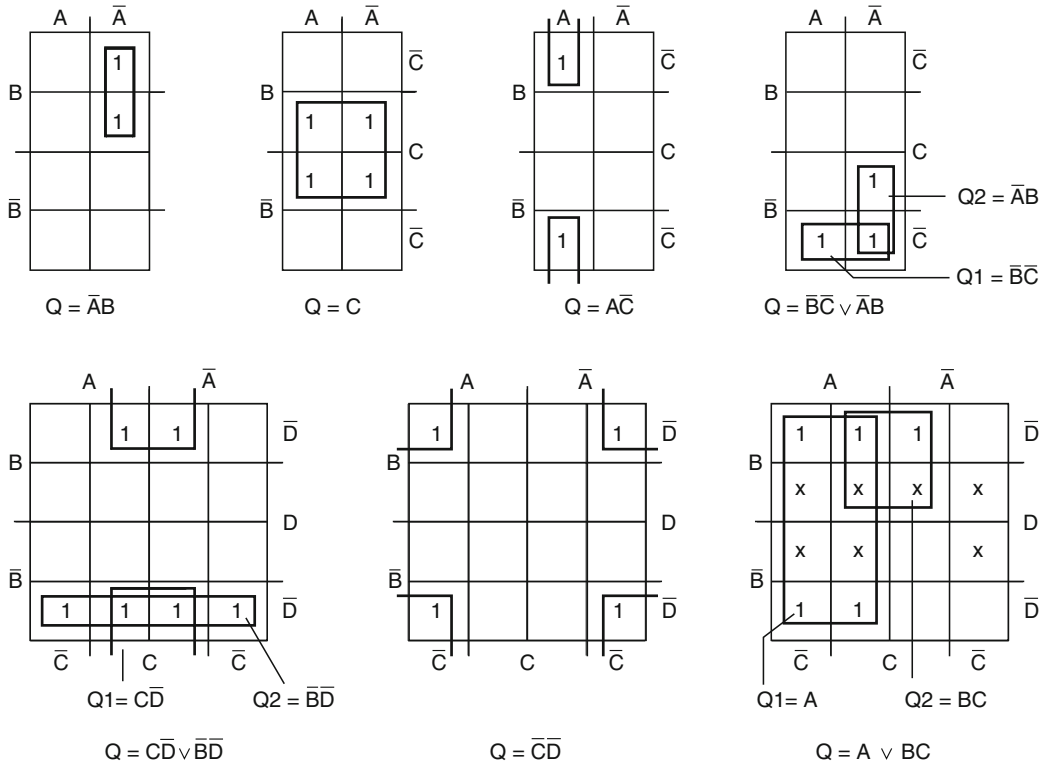


Abb. 50.20 Beispiele für Schleifenbildung

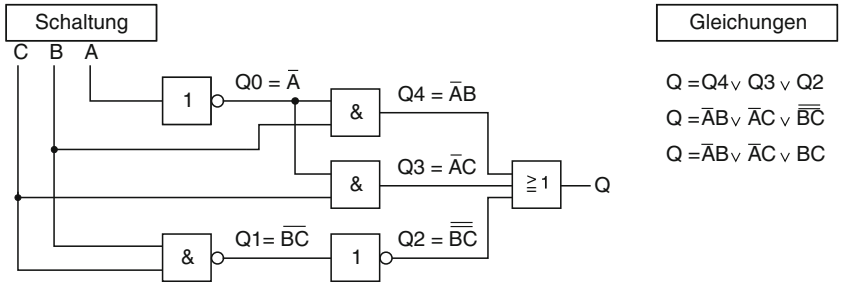


Tabelle:

C B A	Q0 = Ā	Q1= BC̄	Q2 = B̄C̄	Q3 = AC̄	Q4 = ĀB	Q = Q2 ∨ Q3 ∨ Q4
0 0 0	1	1	0	0	0	0
0 0 1	0	1	0	0	0	0
0 1 0	1	1	0	0	1	1
0 1 1	0	1	0	0	0	0
1 0 0	1	1	0	1	0	1
1 0 1	0	1	0	0	0	0
1 1 0	1	0	1	1	1	1
1 1 1	0	0	1	0	0	1

Nachweis:

C B A	Q
0 0 0	0
0 0 1	0
0 1 0	1
0 1 1	0
1 0 0	1
1 0 1	0
1 1 0	1
1 1 1	1

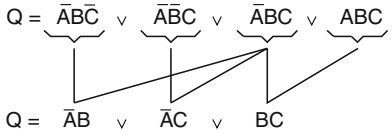


Abb. 50.21 Analyse einer Schaltung

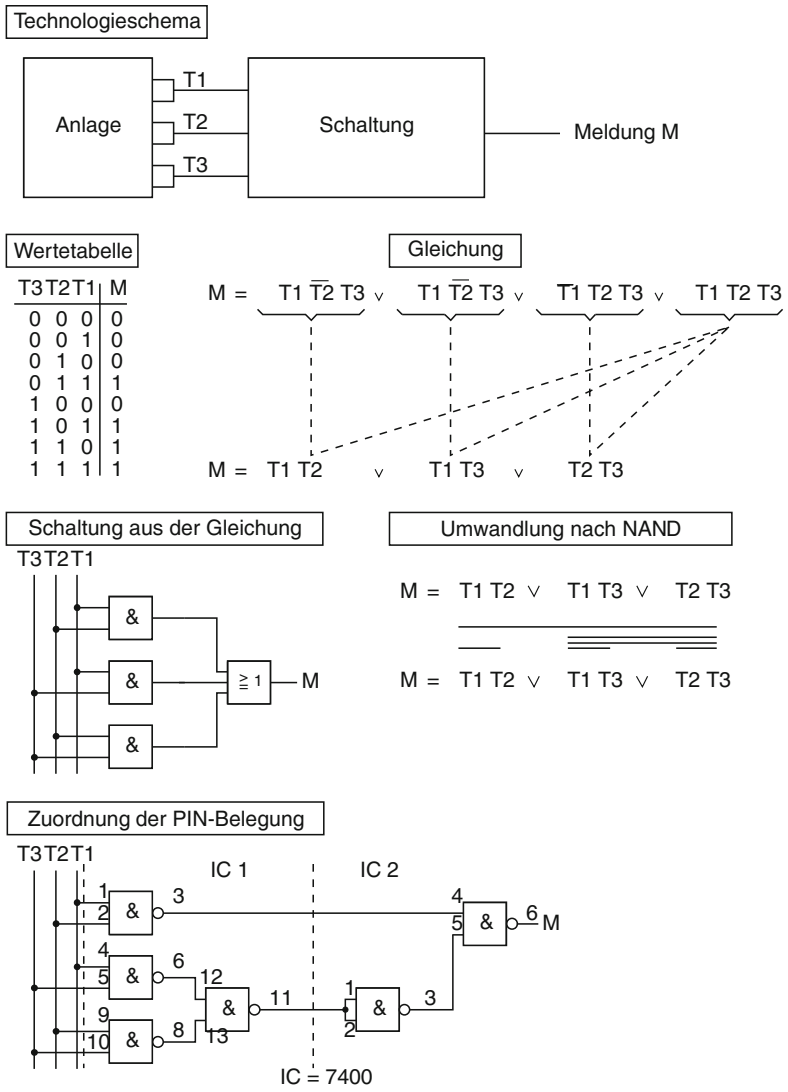


Abb. 50.22 Synthese einer Schaltung

viele Variable in die Schleifenbildung einbezogen werden, um so den Gesamtausdruck der Gleichung zu verkleinern. Damit reduzieren sich die benötigten Bauelemente einer Schaltung unter Umständen erheblich. Können bestimmte Eingangskombinationen nicht auftreten, dann darf dies in der KV-Tafel gekennzeichnet werden (z. B. durch ein x). Für eine optimale Lösung werden diese Felder mit in die Schleifenbildung einbezogen.

50.3.2.7 Analyse logischer Schaltungen

Soll eine Übersicht über alle möglichen Ausgangszustände in Abhängigkeit der jeweiligen Eingänge erstellt werden, ist die Wertetabelle eine geeignete Darstellungsform. Wird sie aus einer Schaltung ermittelt, spricht man von der *Analyse* einer digitalen Schaltung.

Abb. 50.21 zeigt beispielhaft einen Lösungsweg auf. An die Ausgänge der Gatter werden die entsprechenden Verknüpfungsgleichungen (hier Q0 bis Q4) geschrieben und mit ihren jeweiligen Zuständen „1“ beziehungsweise „0“ in eine Tabelle eingeordnet. Der Ausgang $Q = Q2 \vee Q3 \vee Q4$ beinhaltet dann die eigentliche Lösung. Es kann eine Überprüfung der gefundenen Gleichung vorgenommen werden, wenn die Lösungsgleichung mit den Regeln der Schaltalgebra vereinfacht wird.

50.3.2.8 Synthese logischer Schaltungen

Soll eine Schaltung *entworfen* werden, spricht man von der Synthese einer Schaltung. Das meist in verbaler Form vorliegende Problem wird skizziert, und Variable und Ausgänge werden entsprechend zugeordnet. Über eine Wertetabelle werden bei allen möglichen Eingangskombinationen die gewünschten oder geforderten Ausgangskombinationen beschrieben. Anschließend erfolgt die Umsetzung der Wertetabelle in eine Gleichung, die dann gegebenenfalls vereinfacht wird. Über diese Gleichung wird die Schaltung gezeichnet. Anschließend wählt man die entsprechenden Standardbausteine zur Realisierung der Schaltung aus.

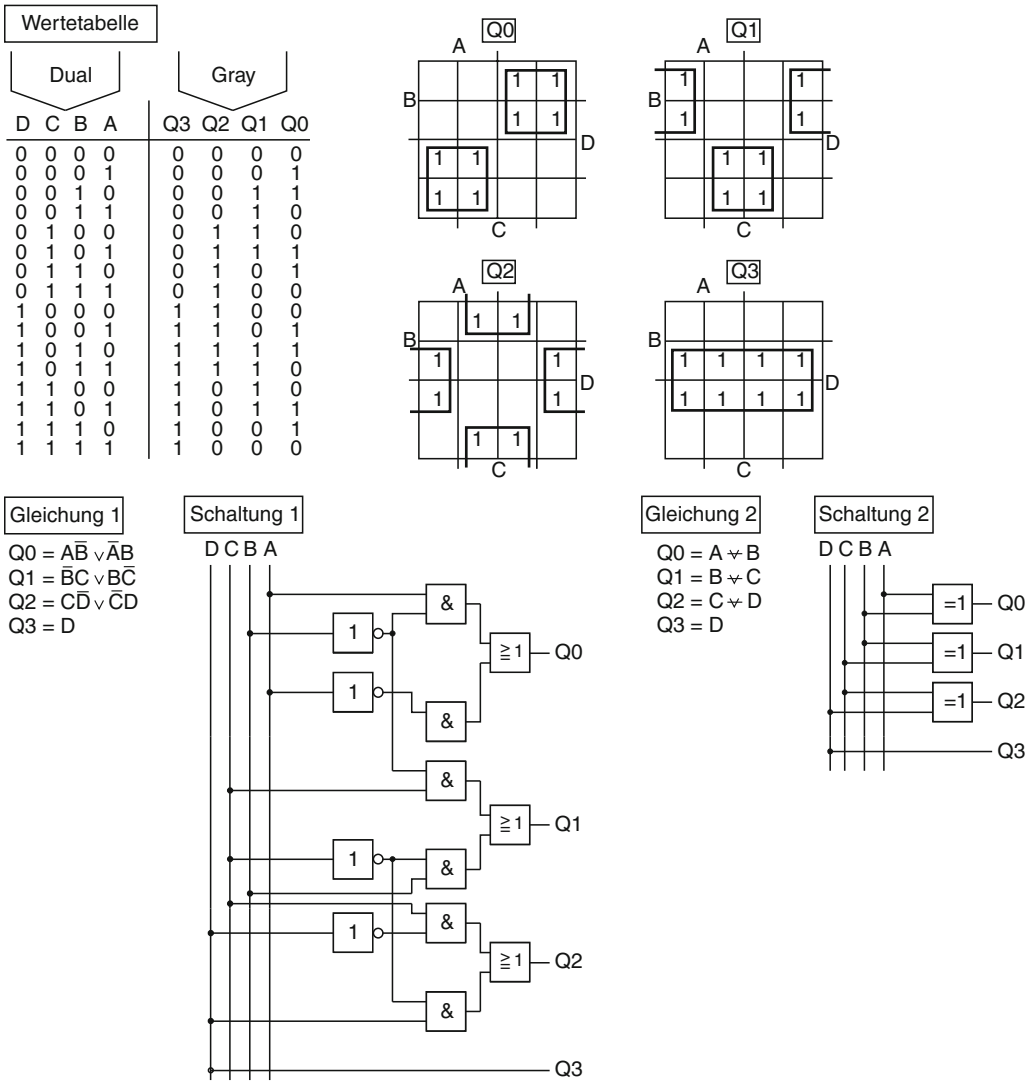


Abb. 50.23 Synthese einer Schaltung

Beispiel 1

Eine Anlage soll von drei Temperatursensoren überwacht werden. Überschreitet die Temperatur einen vorgegebenen Wert, soll eine Meldung durch eine Anzeige erfolgen. Dies ist immer dann der Fall, wenn mindestens zwei der Sensoren „1“ melden. Die Schaltung soll mit dem Standardbaustein 7400 realisiert werden. Abb. 50.22 zeigt das Technologieschema, die Wertetabelle und eine mögliche Lösung. ◄

Beispiel 2

Konzeption eines Dual-Gray-Codierers. Ziel der Schaltung ist es, eine anliegende 4-Bit-Dualzahl in den Graycode umzuwandeln. Die Schaltung besitzt jeweils 4 Ein- und Ausgänge. Zur Lösung der Aufgabe empfiehlt es sich, zunächst wieder die Wertetabelle aufzustellen (Abb. 50.23). ◄

50.4 Zahlensysteme in der Digital- und Datenverarbeitung

Zahlen in der Digitaltechnik werden in *binärer* Form dargestellt. In der Datenverarbeitung sind das Dualsystem und das Hexadezimalsystem von Bedeutung. Jedes Zahlensystem besteht aus einer Anzahl von Ziffern zu einer Basis. Der Exponent bestimmt den Stellenwert, also die Wertigkeit der Ziffer. Die Ziffern selbst sind die Faktoren der Stellenwerte. Beim Dualsystem liegt die Basis 2 mit den Zeichen „0“ und „1“ und beim Hexadezimalsystem die Basis 16 mit den Zeichen „0“ bis „F“ zugrunde. Das Dezimalsystem mit der Basis 10 besitzt einen Ziffernvorrat von „0“ bis „9“, also 10 Zeichen oder Ziffern. Abb. 50.24 zeigt die drei Zahlensysteme im Vergleich.

50.4.1 Dualsystem Synthese

50.4.1.1 Bildung der Dualzahlen und Umwandlung in Dezimalzahlen

Im Dualsystem (Zweiersystem) werden nur die Ziffern „0“ und „1“ verwendet. Die Dualzahlen werden vom Prinzip her wie die Dezimalzahlen gebildet. Der Wert der Dualzahl ist die Summe der Produkte von Ziffer und Stellenwert. An zwei Beispielen wird dies durch Umwandlung einer Dual- in eine Dezimalzahl verdeutlicht (Abb. 50.25).

50.4.1.2 Umwandlung dezimal nach dual

Die Dualzahl aus der Dezimalzahl erhält man, indem die Dezimalzahl jeweils durch zwei dividiert wird. Treten Restbeträge

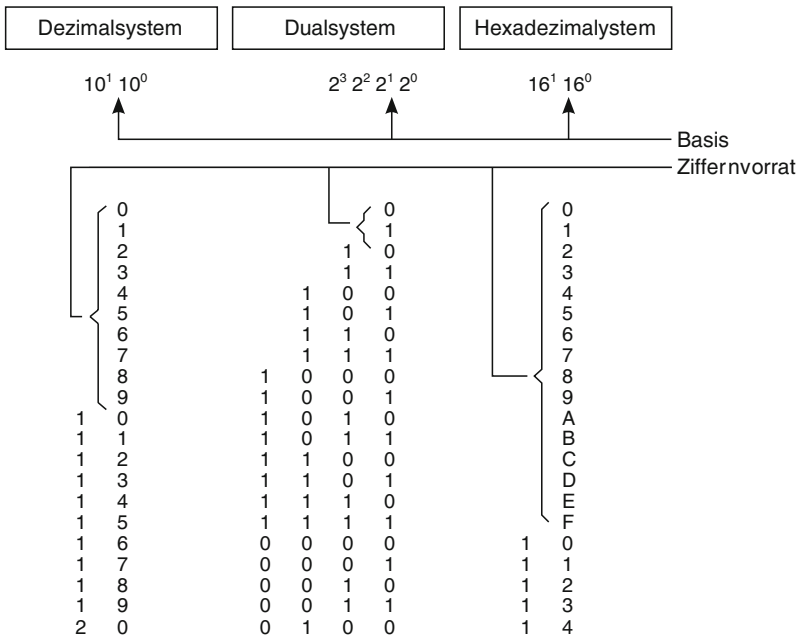


Abb. 50.24 Zahlensysteme

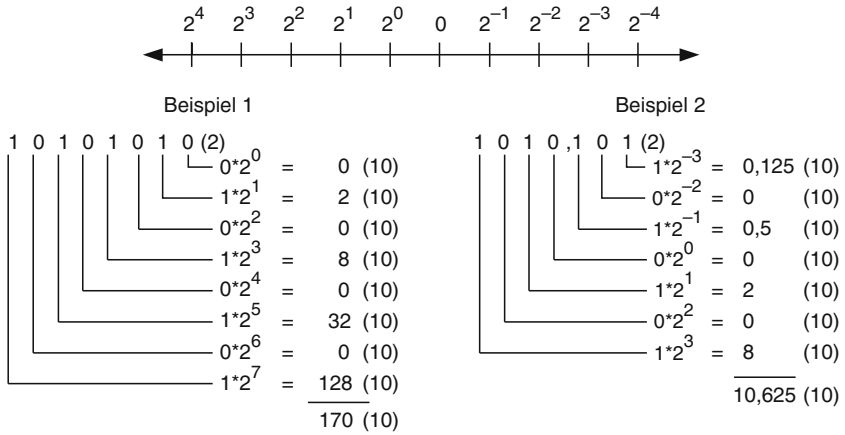


Abb. 50.25 Umwandlung von Dual- in Dezimalzahlen

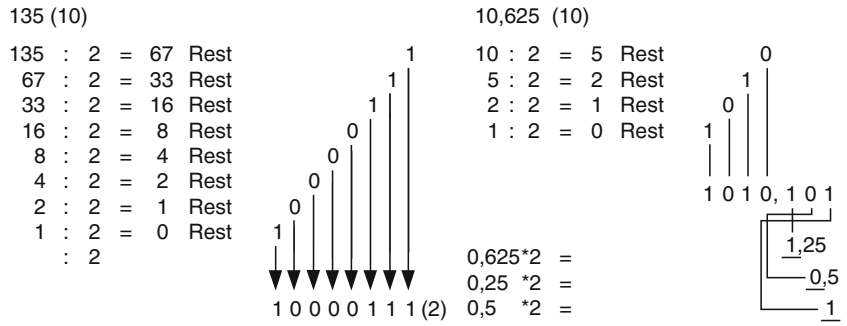


Abb. 50.26 Umwandlung von Dezimal- in Dualzahlen

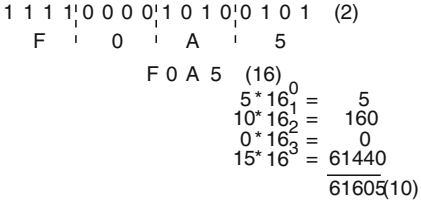


Abb. 50.27 Bildung und Umwandlung von Hexadezimalzahlen

auf, wird dies durch „1“, sonst durch „0“ gekennzeichnet. Die Folge der sich ergebenden 0- und 1-Werte entspricht dann der Dualzahl. Man sucht also die in der Dezimalzahl enthaltenen Zweierpotenzen. Zahlen hinter dem Komma werden mit 2 multipliziert und das Vorkommaergebnis jeweils hinter das Komma der Dualzahl übertragen (Abb. 50.26).

50.4.2 Hexadezimalsystem

Für die Darstellung der dezimalen Zahlen 10 bis 15 werden im Hexadezimalsystem die zusätzlichen Zeichen A, B, C, D, E und F verwendet. Bei großen Zahlenwerten benötigen Dualzahlen sehr viele Ziffern. Sie werden unübersichtlich lang und damit unhandlich. Man verwendet daher Hexadezimalzahlen, die sich auf einfache Weise aus Dualzahlen bilden lassen. Jeder Tetrade wird eine Hexadezimalzahl zugeordnet. Die Umwandlung einer Hexadezimalzahl in eine Dezimalzahl und umgekehrt erfolgt vom Prinzip her wie die Umwandlung beim Dualsystem. Abb. 50.27 führt ein Beispiel auf.

50.4.3 Rechnen mit Dualzahlen

Abb. 50.28 zeigt die Grundregeln für zweistellige Dualzahlen und Beispiele für mehrstellige Dualzahlen, wo die jeweilige Dezimalzahl im Vergleich mit angegeben ist. Bei der Addition entsteht bei der Summenbildung $1 + 1$ ein Übertrag, der zur nächsthöheren Stelle addiert werden muss. Bei der Subtraktion $0 - 1$ muss eine Zahl von der nächsthöheren Stelle geholt (geborgt) werden. Die Multiplikation erfolgt wie bei den Dezimalzahlen durch Teilproduktbildung mit anschließender Addition. Bei der Division wird festgestellt, wie oft der Dividend im Divisor enthalten ist. Dies geschieht wie bei dem Dezimalsystem durch Subtraktion.

50.4.4 Zahlen in Rechenanlagen

50.4.4.1 Darstellung von Zahlen

In Rechenanlagen unterscheidet man die Darstellung als *Festpunkt-* oder *Gleitpunktzahl*. Bei den Festpunktzahlen erfolgt die Kennzeichnung einer positiven Zahl mit dem MSB = „0“ (MSB = Most Significant Bit = Höchstwertigstes Bit), und die negative Zahl besitzt „1“ als MSB (Abb. 50.29). Da die negative Zahl im *Zweierkomplement* dargestellt ist, setzt dies voraus, dass eine bestimmte Wortbreite n festliegt. Damit ist auch der Zahlenbereich bei nur positiven Zahlen (Z) mit $Z_{min} = 0$ und $Z_{max} = 2^n$ und bei positiver und negativer Zahlendarstellung mit $Z_{min} = -2^{(n-1)}$ und $Z_{max} = 2^{(n-1)} - 1$ begrenzt. Das Beispiel führt bei unterschiedlichen Wortbreiten die kleinste und größte darstellbare Zahl auf und zeigt in Form einer Tabelle und am Zahlenkreis die Zahlen -8 bis $+7$ bei einer Wortbreite von $n = 4$.

Bei der Gleitpunktzahlendarstellung ist das MSB wie bei den Festpunktzahlen dem Vorzeichen vorbehalten, es folgen Binärstellen für den Exponent und für die Mantisse. Der darstellbare Zahlenbereich hängt vom Exponenten ab. Häufig werden Doppelworte, wie z. B. zwei 16-Bit-Worte benutzt. Abb. 50.30 beschreibt das Format dualer Gleitpunktzahlen auch anhand eines Beispiels.

50.4.4.2 Einer- und Zweierkomplement

Das *Zweierkomplement* ZK einer Zahl Z ist deren Ergänzung zur nächsthöheren Wertigkeit B^n, wobei n die Stellenanzahl und B Basis des Zahlensystems ist. Dagegen ist das *Einerkomplement* EK die Ergänzung der Zahl Z zur nächsthöheren Wertigkeit $B^n - 1$. Dies sollen die in Abb. 50.31 aufgeführten Beispiele in Form einer Tabelle an Dezimal- und an Dualzahlen verdeutlichen.

Bei den Dualzahlen erhält man das Einerkomplement der Zahl Z durch Invertierung aller Stellen der Zahl und das Zweierkomplement einer Zahl Z durch die Addition des Einerkomplements mit „1“.

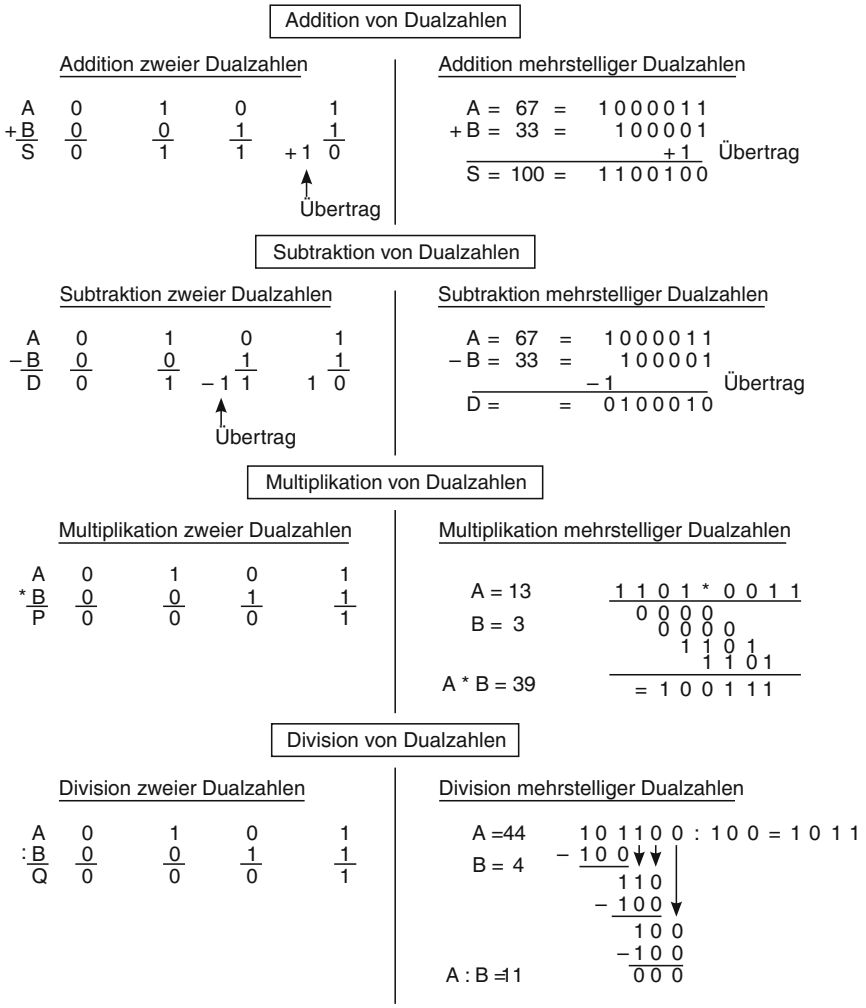


Abb. 50.28 Rechnen mit Dualzahlen

Beispiel

Dualzahl:	01010101			
EK:	10101010	*EK*	10101010	
			+1	
		ZK	10101011	◀

50.4.4.3 Subtraktion mit Hilfe des Komplements

Mit Hilfe des Zweierkomplements lassen sich Subtraktionen durchführen. Der Rechner führt diese Subtraktion praktisch auf eine Addition zurück. Er bildet zunächst das Zweierkomplement der zu subtrahierenden Zahl durch Invertieren der Zahl (EK) und führt anschließend eine Addition mit „1“ durch. Werden Festpunktzahlen mit beliebigen Vorzeichen addiert, muss der Rechner das Ergebnis auf seine Korrektheit überprüfen. Das Ergebnis ist falsch, wenn die Operanden gleiches Vorzeichen besitzen und das Ergebnis ein davon verschiedenes Vorzeichen.

Beispiele

$$\begin{aligned} A &= 10 = 00001010 \\ B &= -2 = 11111110 \\ \hline A + (-B) &= 00001000 = 8 \end{aligned}$$

$$\begin{aligned} A &= 2 = 00000010 \\ B &= -10 = 11110110 \\ \hline A + (-B) &= 11111000 = -8 \end{aligned}$$

Vorgehensweise des Rechners:

$$\begin{aligned} B = 00000010 \rightarrow \text{EK} &= 11111101 \\ + &\quad 1 \\ \hline &= 11111110 = \text{ZK}(B) \end{aligned}$$

$$\begin{aligned} A &= 00001010 \\ \text{ZK}(B) &= 11111110 \\ \hline A + (-B) &= 00001000 \end{aligned}$$

◀

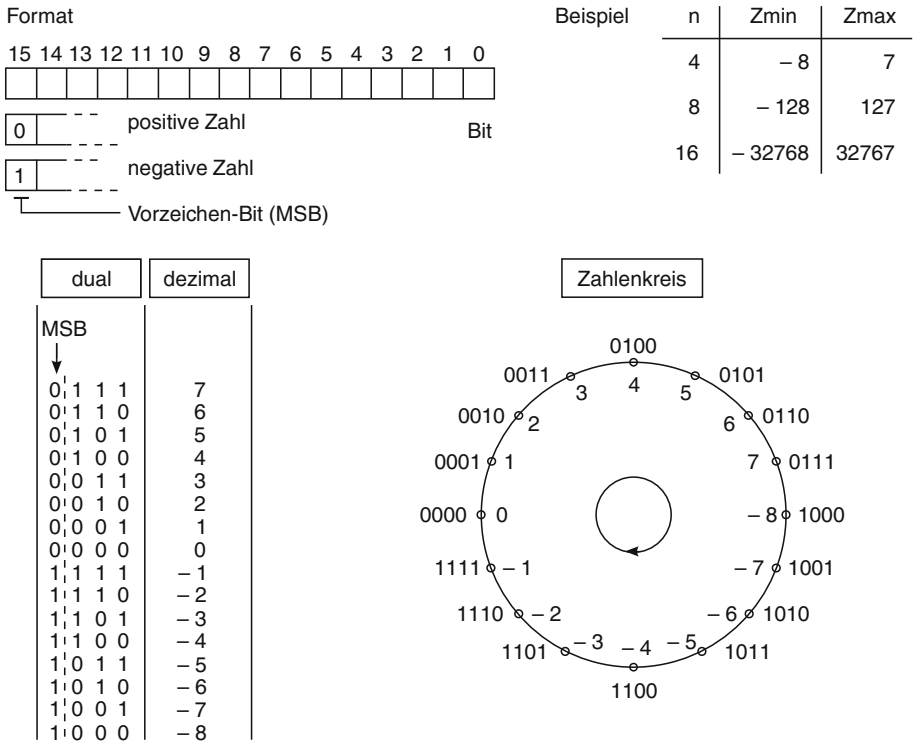


Abb. 50.29 Festpunktzahlen

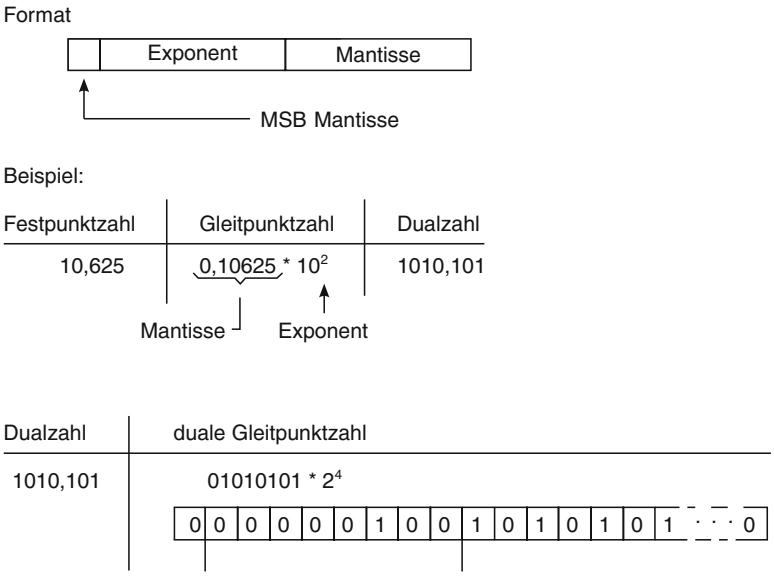


Abb. 50.30 Gleitpunktzahlen

		Z +ZK = B^n					Z * EK = B^n – 1		
Z	ZK	Σ	B	n	B^n	Z	EK	Σ	B^n – 1
22	78	100	10	2	100	22	77	99	99
635	365	1000	10	3	1000	635	364	999	999
101	011	1000	2	3	1000	101	101	111	111
1010	0110	10000	2	4	10000	1010	1010	1111	1111

Abb. 50.31 Bildung des Komplements

50.5 Codes

50.5.1 Allgemeines

Nach DIN 43300 sind Codes eindeutige Zuordnungen von Zeichen eines Zeichenvorrates zu den Zeichen des anderen Zeichenvorrates. *Codierung* ist die Zuordnung eines Zeichens aus dem Zeichenvorrat X zu einem Zeichen des Zeichenvorrats Y, und *Decodierung* ist die Rückführung des Zeichens aus dem Zeichenvorrat Y zum Zeichen des Zeichenvorrats X. Ist die codierte Information eine Zahl, spricht man von einem *numerischen Code*. Sind zusätzlich Buchstaben und Zeichen codiert, ist es ein *alphanumerischer Code*. Kommt in einem Code die Wertigkeit der Stellen zum Ausdruck, ist es ein *bewerteter* (z. B. BCD) oder *gewichteter* Code. Ist die Anordnung im Code ohne Wertigkeit der einzelnen Stelle, spricht man von einem *unbewertetem* oder *ungewichtetem* Code (z. B. 3-Excess-Code). Ändert sich das vorherige Zeichen zum nachfolgenden immer nur um 1 Bit, ist es ein *einschrittiger* (z. B. Gray-Code), ansonsten ein *mehrschrittiger* Code. Die Zuordnungsvorschrift wird in Codetabellen festgelegt. Abb. 50.32 zeigt eine Übersicht gängiger 4-Bit-Codes (tetradische Codes).

50.5.2 Binär-Code

Als Binärcode bezeichnet man Wörter aus Binärzeichen. Die Darstellung erfolgt wie bei den Dualzahlen.

50.5.3 BCD-Code

Im BCD-Code (BCD = Binary Coded Decimal) wird jede Stelle einer Dezimalzahl durch ein binäres 4-Bit-Muster (Tetrade) dargestellt. Da man mit vier Stellen sechzehn Kombinationen bilden kann, werden sechs dieser Bit-Kombinationen, die nicht benötigt werden, *Pseudotetraden* (Pseudo griech. scheinbar) genannt. Allgemein werden diese überschüssigen Elemente als *Redundanz* (siehe auch 1.5.6) bezeichnet. Die einzelnen BCD-Codes unterscheiden sich durch die Lage der Pseudotetraden. Beim BCD-Dual-Code sind dies beispielsweise die Codewörter 1010 bis 1111 und beim Aikencode die Codewörter von 0101 bis 1010. Durch Reihenbildung der jeweiligen Codewörter können beliebig lange Dezimalzahlen codiert werden.

50.5.3.1 BCD-Code

Dieser Code wird allgemein BCD-Code oder auch 8-4-2-1-Code genannt. Da die Stellenbewertung im Dualsystem angegeben ist, ist dieser Code ein *bewerteter Code*.

Beispiel

Wertigkeit: 8 4 2 1

Dualzahl: $1\,0\,0\,1 = 1 \cdot 8 + 0 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$

$= 8 + 1 = 9$

Treten bei der Addition Pseudotetraden auf, ist eine Korrektur mit +6 dual nötig. ◄

Dual–Code					BCD–Code					3–Excess–Code				Aiken–Code					Gray–Code				
Dezimalzahl	8421				Dezimalzahl	8421				Dezimalzahl				Dezimalzahl	2421				Dezimalzahl				
	8	4	2	1		8	4	2	1						2	4	2	1					
0	0	0	0	0	0	0	0	0	0	Pseudo–Tetraden				0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0	1					1	0	0	0	1	1	0	0	0	1
2	0	0	1	0	2	0	0	1	0					2	0	0	1	0	2	0	0	1	1
3	0	0	1	1	3	0	0	1	1	0	0	0	1	1	3	0	0	1	1	3	0	0	1	0
4	0	1	0	0	4	0	1	0	0	1	0	1	0	0	4	0	1	0	0	4	0	1	1	0
5	0	1	0	1	5	0	1	0	1	2	0	1	0	1	Pseudo–Tetraden					5	0	1	1	1
6	0	1	1	0	6	0	1	1	0	3	0	1	1	0						6	0	1	0	1
7	0	1	1	1	7	0	1	1	1	4	0	1	1	1						7	0	1	0	0
8	1	0	0	0	8	1	0	0	0	5	1	0	0	0						8	1	1	0	0
9	1	0	0	1	9	1	0	0	1	6	1	0	0	1						9	1	1	0	1
10	1	0	1	0	Pseudo–Tetraden					7	1	0	1	0	5	1	0	1	1	10	1	1	1	1
11	1	0	1	1						8	1	0	1	1	6	1	1	0	0	11	1	1	1	0
12	1	1	0	0						9	1	1	0	0	7	1	1	0	1	12	1	0	1	0
13	1	1	0	1						Pseudo–Tetraden					8	1	1	1	0	13	1	0	1	1
14	1	1	1	0											9	1	1	1	1	14	1	0	0	1
15	1	1	1	1																15	1	0	0	0

Abb. 50.32 Codes

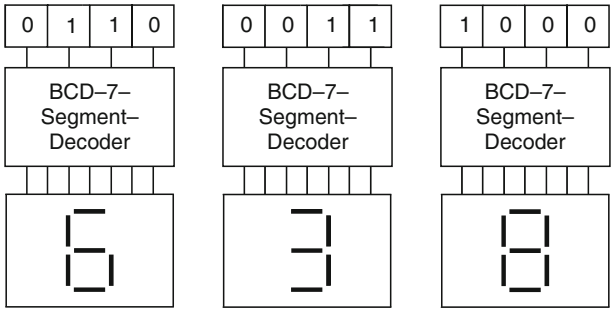


Abb. 50.33 Ansteuerung mit BCD-Code

Beispiel

$$\begin{aligned} A = 15 &= 0001\ 0101 \\ B = 17 &= 0001\ 0111 \\ &\overline{0010\ 1100} \quad \textit{Pseudotetrade} \\ &\ \ 0000\ 0110 \quad \textit{Korrektur} \\ A + B = 32 &= 0011\ 0010 \end{aligned}$$

$$\begin{aligned} A = 11 &= 0001\ 0001 \\ B = 12 &= 0001\ 0010 \\ A + B = 23 &= 0010\ 0011 \end{aligned}$$

Abb. 50.33 zeigt eine häufige Anwendung des BCD-Codes, wie dieser Code auch allgemein genannt wird. Er steuert dort über einen Decoder 7-Segmentanzeigen (Abb. 50.34). Handelsübliche Anzeigen gibt es mit einer gemeinsamen Anode oder einer gemeinsamen Katode. Bei einem Segment mit einer gemeinsamen Anode leuchtet beispielsweise die „0“ auf, wenn die Segmente a, b, c, d, e und f Spannung erhalten („1“ sind). ◀

50.5.3.2 3-Excess-Code

Der 3-Excess-Code ist ein *nicht bewerteter*, *symmetrischer* Code. Seine „Spiegelachse“ liegt zwischen 4_{10} und 5_{10}. Beim Rechnen sind Korrekturen notwendig.

Beispiel

$$\begin{aligned} A = 15 &= 0100\ 1000 \\ B = 17 &= 0100\ 1010 \\ &\overline{1000\ 0010} \\ &\ \ 1101\ 0011 \quad \textit{Korrektur} \\ A + B = 32 &= 0110\ 0101 \end{aligned}$$

Bei diesem Code können gerade und ungerade Zahlen durch die letzte Stelle leicht unterschieden werden und eine Komplementbildung ist durch Vertauschen von „0“ und „1“ ebenfalls einfach auszuführen. ◀

50.5.3.3 Aiken-Code

Der Aiken-Code ist ein *symmetrischer* Code mit der Bewertung 2421.

Beispiel

BCD-Code

Wertigkeit: 2 4 2 1

Dualzahl: $1\ 1\ 0\ 1 = 1 \cdot 2 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$
$= 2 + 4 + 1 = 7$

Ungerade und gerade Zahlen sind wie beim 3-Excess-Code durch die letzte Stelle leicht erkennbar. Dieser Code benötigt beim Rechnen keine Korrekturen. ◀

50.5.4 Gray-Code

Der Gray-Code lässt sich für jede Wortbreite realisieren. Er ist ein *unbewerteter* oder *ungewichteter* Code und unterscheidet sich von dem vorherigen und nachfolgenden Bitmuster jeweils nur um ein Bit. Man spricht auch von einem *einschrittigen* Code. Er wird oft zur Umwandlung analoger Größen in eine digitale

BCD–Code				7–Segment–Code						
2^3	2^2	2^1	2^0	g	f	e	d	c	b	a
0	0	0	0	0	1	1	1	1	1	1
0	0	0	1	0	0	0	0	1	1	0
0	0	1	0	0	0	1	1	0	1	1
0	0	1	1	1	0	0	1	1	1	1
0	1	0	0	1	1	0	0	1	1	0
0	1	0	1	1	1	0	1	1	0	1
0	1	1	0	1	1	1	1	1	0	1
0	1	1	1	0	0	0	0	1	1	1
1	0	0	0	1	1	1	1	1	1	1
1	0	0	1	1	1	0	1	1	1	1

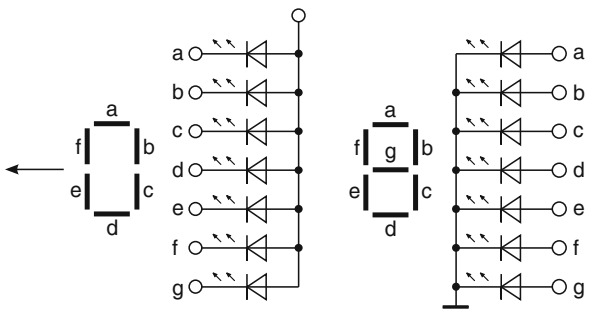


Abb. 50.34 7-Segmentanzeigen und Ansteuerung

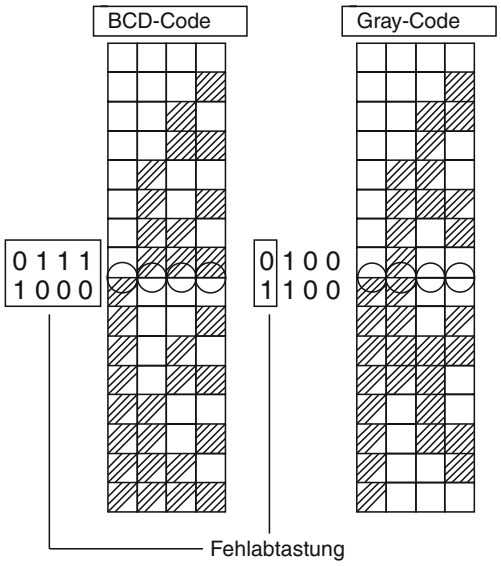


Abb. 50.35 Abtastung und Fehljustierung

Information umgesetzt, wie beispielsweise die Bestimmung einer Position auf einer Wegstrecke. Abb. 50.35 veranschaulicht dies an einem Codelineal, das von Sensoren abgetastet wird. Mögliche Fehljustierungen beschränken bei diesem Code den Fehler auf ein Bit. Bei Einsatz des Dual-Codes würden sich beim Übergang von der Zahl 3 nach 4 drei Stellen verändern und bei Fehljustierung der Sensoren einen erheblichen Fehler erzeugen.

50.5.5 Codierung alphanumerischer Zeichen

Der ASCII-Code (ASCII = American Standard Code of Information Interchange = Amerikanischer Standardcode zum Datenaustausch) (Abb. 50.36, Tab. 50.2) ist genormt nach DIN 6603.

Bit 7 6 5	0_0	0_1	0_{10}	1_0	0_{11}	1_{01}	1_{10}	1_{11}
0 0 0 0	NUL	DLE		0	@	P		p
0 0 0 1	SOH	DC1		1	A	Q	a	q
0 0 1 0	STX	DC2	"	2	B	R	b	r
0 0 1 1	ETX	DC3	#	3	C	S	c	s
0 1 0 0	EOT	DC4	$	4	D	T	d	t
0 1 0 1	ENQ	NAK	%	5	E	U	e	u
0 1 1 0	ACK	SYN	&	6	F	V	f	v
0 1 1 1	BEL	ETB		7	G	W	g	w
1 0 0 0	BS	CAN	(	8	H	X	h	x
1 0 0 1	HT	EM	)	9	I	Y	i	y
1 0 1 0	LF	SUB	*	:	J	Z	j	z
1 0 1 1	VT	ESC	+	;	K		k	
1 1 0 0	FF	FS		<	L		l	
1 1 0 1	CR	GS		=	M		m	
1 1 1 0	SO	RS		>	N		n	
1 1 1 1	SI	US	/	?	O	–	o	DEL

ASCII-CODE

Abb. 50.36 ASCII-Code

Tab. 50.1 Zahlen und Buchstaben im ASCII-Code

0	30	A	41	a	61
1	31	B	42	b	62
2	32	C	43	c	63
3	33	D	44	d	64
4	34	E	45	e	65
5	35	F	46	f	66
6	36	G	47	g	67
7	37	H	48	h	68
8	38	I	49	i	69
9	39	J	4A	j	6A
		K	4B	k	6B
		L	4C	l	6C
		M	4D	m	6D
		N	4E	n	6E
		O	4F	o	6F
		P	50	p	70
		Q	51	q	71
		R	52	r	72
		S	53	s	73
		T	54	t	74
		U	55	u	75
		V	56	v	76
		W	57	w	77
		X	58	x	78
		Y	59	y	79
		Z	5A	z	7A

Er ist ein 7-Bit-Code, mit dem man $2^7 = 128$ Zeichen codieren kann. Außer den Ziffern 0 bis 9 und den Buchstaben a bis z beinhaltet dieser Code Sonder- und Steuerzeichen. Abb. 50.36 beschreibt die Möglichkeit der Zeichenzuordnung zu den jeweiligen Binärzeichen. Benötigt man beispielsweise den Code für A, dann lässt sich der über die Bitfolge Bit7, Bit6, Bit5 mit 100 und Bit4, Bit3, Bit2 und Bit1 mit 0001 finden (A = 1000001). Die häufig benötigten Zahlen und Buchstaben sind in Hex-Darstellung unter der Tab. 50.1 zusammengefasst. Bei Übertragungen wird ein 8. Bit (Parity-Bit = Paritäts-Bit) zur Fehlererkennung mit übertragen.

Beispiel

für die Codierung:

R	O	M
1010010	1001111	1001101

◀

50.5.6 Fehlererkennung und Redundanz

Um einen Fehler bei der Übertragung digitaler Informationen zu erkennen, müssen außer den eigentlichen Daten weitere Zeichen übertragen werden. Ein Maß für diese zusätzlichen Zeichen ist die *Redundanz* (Weitschweifigkeit).

Tab. 50.2 Bedeutung der Abkürzungen im ASCII-Code

NUL	NULL	Null
SOH	START OF HEADING	Beginn der Kopfzeile
STX	START OF TEXT	Textbeginn
ETX	END OF TEXT	Textende
EOT	END OF TRANSMISSION	Übertragungsende
ENQ	ENQUIRE	Anfrage für Übertragung
ACK	ACKNOWLEDGE	Antwort bejahend
BEL	BELL	Klingelzeichen
BS	BACK SPACE	Rückschritt
HT	HORIZONTAL TABULATION	Tabellierung (horizontal)
VT	VERTICAL TABULATION	Tabellierung (vertikal)
LF	LINE FEED	Zeilenvorsprung
FF	FORM FEED	Seitenvorschub
CR	CARRIAGE RETURN	Wagenrücklauf
SO	SHIFT OUT	Dauerumschaltung
SI	SHIFT IN	Rückschaltung
DLE	DATA LINK ESCAPE	Datenumschaltung
DC1...4	DEVICE CONTROL	Steuersignale
NAK	NEGATIVE ACKNOWLEDGE	Antwort verneinend
SYN	SYNCHRON	Synchronisation
ETB	END OF TRANSMISSION BLOCK	Übertragungsblockende
EM	END OF MEDIUM	Aufzeichnungsende
ESC	ESCAPE	Umschaltung
FS	FILE SEPARATOR	Trennzeichen
GS	GROUP SEPARATORS	Gruppentrennzeichen
RS	RECORD SEPARATOR	Untergruppenkennzeichen
US	UNIT SEPARATOR	Teilgruppentrennzeichen
DEL	DELETE	Löschen

Beispiel

Wird ein 4-Bit-Wort übertragen, ergeben sich $2^4 = 16$ Kombinationsmöglichkeiten. Daraus folgt, dass mit n Bits $2^n = Z$ Kombinationen möglich sind. Um beispielsweise die Ziffern 0 bis 9 darzustellen, sind nur 10 Zeichen notwendig. Die gesuchte Anzahl n, die zur Darstellung dieser Zeichen notwendig ist, lässt sich über den Logarithmus zur Basis 2 (Log. dualis) finden (50.1)).

$$\mathrm{ld}\, Z = \frac{\ln Z}{\ln 2} = \frac{\ln 10}{\ln 2} = 3{,}3219 \qquad (50.1)$$

Dies bedeutet, dass 3,32 Bit eigentlich nur notwendig sind. Da zur Darstellung dieser Zeichen aber 4 Bit benutzt werden, ist die Redundanz R nach (50.2) 0,68 Bit.

$$R = n - \mathrm{ld}\, Z = 4\,\mathrm{Bit} - 3{,}32 = 0{,}68\,\mathrm{Bit} \qquad (50.2)$$

Beim ASCII-Code beträgt demnach die Redundanz mit $n = 8$: $R = 8\,\mathrm{Bit} - 7\,\mathrm{Bit} = 1\,\mathrm{Bit}$. Der Dual- und Gray-Code enthält keine Redundanz, da hier alle Codewörter benutzt werden. ◀

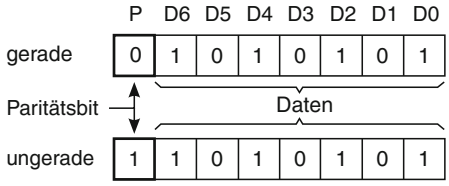


Abb. 50.37 Paritätsbit

50.5.6.1 Einfache Prüfung auf Parität

Wird ein Bit der eigentlichen Information zur Fehlererkennung zugefügt, spricht man von einem Paritätsbit. Es kann das Wort auf eine gerade Anzahl von „1“sen (even parity check) oder ungerade Anzahl von „1“sen (odd parity check) ergänzen (Abb. 50.37).

Erzeugt werden kann das Paritätsbit (gerade Parität) mit Hilfe von EXOR-Verknüpfungen. Abb. 50.38 zeigt eine mögliche Schaltung zu Erzeugung gerader Parität und die Überprüfung eines Datenwortes. Soll eine ungerade Parität erzeugt werden, ist ein zusätzlicher Inverter nötig. Eine nicht korrekte Datenübertragung wird durch Paritätsprüfung im Empfänger erkannt. Die genaue Ortung des Fehlers ist allerdings so nicht möglich.

50.5.6.2 Kreuzsicherungsprüfung

Eine weitere Möglichkeit, Fehler zu erkennen und eine Korrektur vorzunehmen, ist die Kreuzsicherungs- oder *Kreuzpari-*

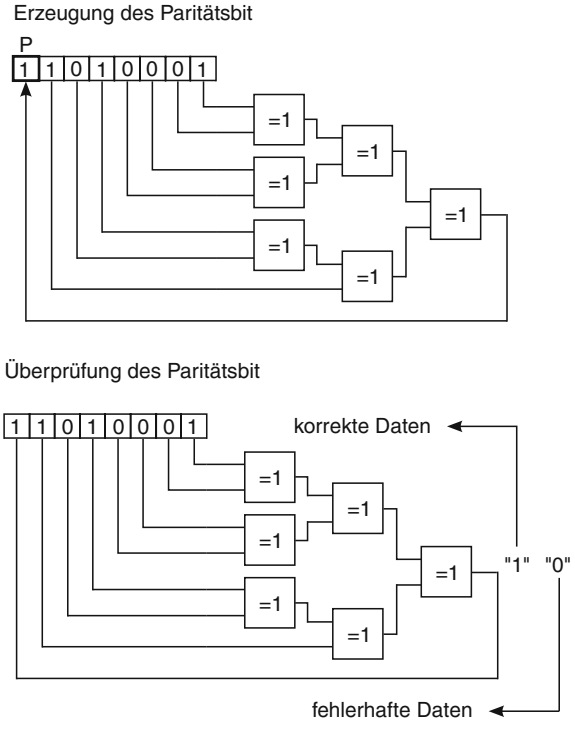


Abb. 50.38 Erzeugung und Überprüfung des Paritätsbits

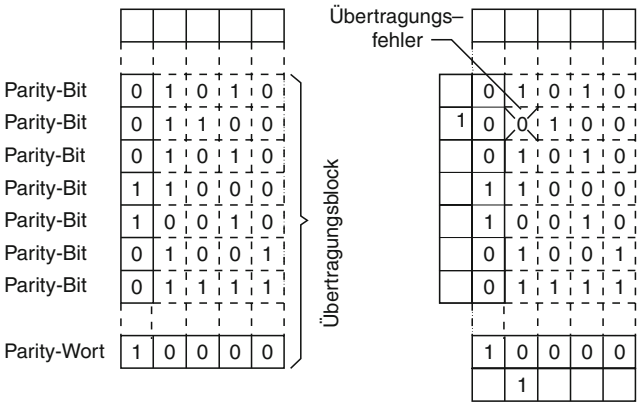


Abb. 50.39 Kreuzparitätsprüfung

tätsprüfung. Hierbei wird ein Datenblock mit Paritätsbits für die Reihen und mit einem Paritätswort für die Spalten des Übertragungsblockes versehen. Im Abb. 50.39 wurde der aus sieben 4-Bit-Worten bestehende Datenblock mit zusätzlichen Paritätsbits übertragen. Im Empfänger erfolgt anschließend eine Prüfung auf Parität. Fehlerhafte Daten werden erkannt, da die Ortung durch den Kreuzungspunkt (im Bild gekennzeichnet) erfolgt. Durch Invertierung kann eine Korrektur erfolgen. Mehrfachfehler können auf diese Weise allerdings nicht korrigiert werden.

50.5.6.3 Hamming-Code

Mit Hilfe des Hamming-Codes ist es möglich, fehlerhafte Bits genau zu bestimmen und danach zu korrigieren. Hierzu werden mehrere Prüfbits übertragen, wobei die Anzahl der Prüfbits

Tab. 50.3 Daten- und Prüfbits

Wortbreite	8	16	32	64
Datenbits	4	11	26	57
Prüfbits	4	5	6	7

von der Informationswortbreite abhängt. Den Zusammenhang zwischen Informationswortbreite und Anzahl der Redundanzbits nennt man auch *Hamming-Distanz*. In Tab. 50.3 sind für übliche Wortbreiten die Anzahl der Datenbits und Prüfbits angegeben.

Beispiel

zur Bildung des Übertragungswortes (Abb. 50.40): P1 wird aus D3, D2 und D0 gebildet. Für die gerade Parität steht „0“ und für ungerade Parität „1“. Die Quersumme der „1“en aus a, b und c ergibt mit dem Prüfbit immer eine gerade Parität. Für eine 4-Bit-Information werden 3 Prüfbits benötigt. Damit wird das Datenwort 7 Bit breit. Nach dem Beispiel werden die Daten mit D2, D4, D5 und D6 ohne Änderung übertragen. Die entsprechenden Prüfbits P0(D0), P1(D1) und P2(D3) werden jeweils in Abhängigkeit der durch Pfeile gekennzeichneten Daten gebildet: a = P0, D0, D2 und D3; b = P1, D0, D1 und D3; c = P2, D0, D1 und D2 und nach Zusammensetzung des zu übertragenden Datenwortes: a = D0, D6, D4 und D2; b = D1, D6, D5 und D2; c = D3, D6, D5 und D4.

Die Prüfung der Daten erfolgt so, dass ein Prüfbit mit mindestens zwei Datenbits in Verbindung gebracht wird.

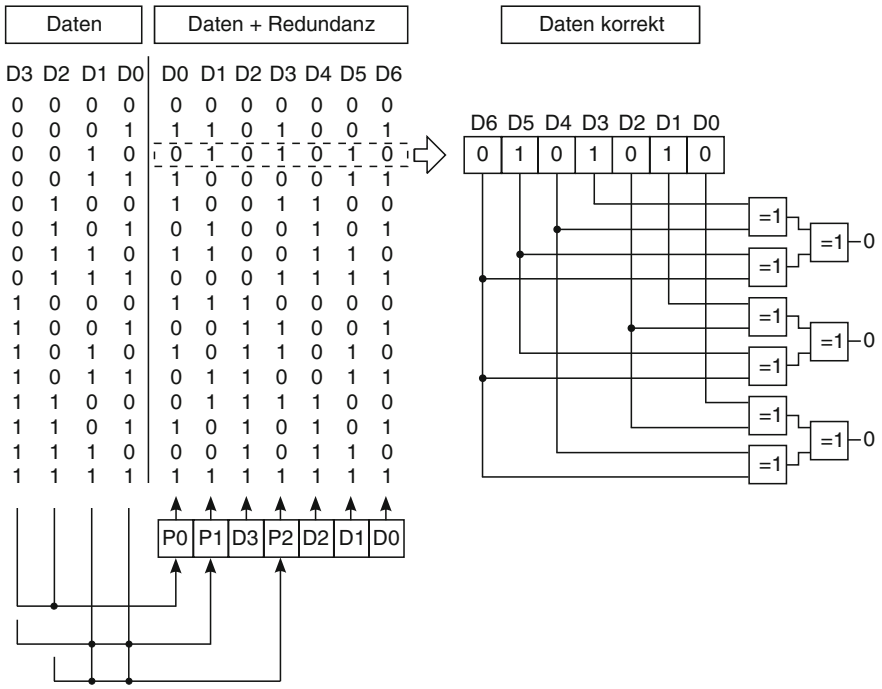


Abb. 50.40 Bildung des Übertragungswortes beim Hamming-Code

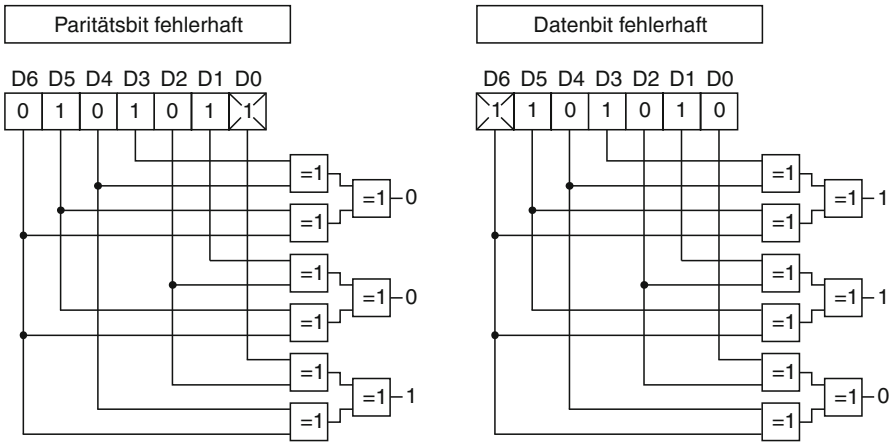


Abb. 50.41 Fehlerermittlung

Je nach 3-Bit-Ausgangswort ist das fehlerhafte Bit, Datenbit oder Paritätsbit, erkannt und kann durch Invertierung korrigiert werden. Die Prüfung nach der Datenübertragung zeigt an, dass alle Daten korrekt übertragen wurden, weil das Prüfausgangswort $0\ 0\ 0_2$ ergibt. Bei fehlerhafter Datenübertragung wird dies durch das Prüfausgangswort in der Weise angezeigt, dass es mindestens zwei „1“sen beinhaltet. Wurde das Paritätsbit selber fehlerhaft übertragen, wird dies nur durch eine „1“ mitgeteilt. Da die Bitkombinationen den entsprechenden Daten- beziehungsweise Paritätsbits zugeordnet sind, kann der Fehler erkannt und behoben werden. Abb. 50.41 veranschaulicht mögliche Fehlerkombinationen. Im ersten Fall ist D0 (Paritätsbit) fehlerhaft und im zweiten Fall das Datenbit D6. ◄

50.6 Digitale Grundschaltungen

50.6.1 Allgemeines

Digitale Schaltungen können nach Schaltnetzen und Schaltwerken unterschieden werden. Schaltnetze sind Schaltungen, deren Ausgänge eindeutig von den Eingangsvariablen bestimmt werden (kombinatorische Schaltungen). Schaltwerke sind Schaltnetze, die ein zusätzliches Speicherverhalten aufweisen können (sequentielle Schaltungen). Die meisten Grundschaltungen brauchen nicht aus einzelnen Gattern aufgebaut werden, sondern werden von den Herstellern als komplexe digitale Schaltung für universelle Einsätze angeboten oder können durch zusätzliche äußere Beschaltung den individuellen Wünschen angepasst werden. In diesem Zusammenhang spielen programmierbare Digitalbausteine wie PAL's und Gal's eine immer größere Rolle. Sie werden in der Automatisierungs-, Mess- und Datentechnik eingesetzt oder sind Teilbausteine größerer hochintegrierter Schaltungen.

50.6.2 Schaltnetze

Logische (kombinatorische) Schaltungen, deren Ausgänge eindeutig von den Eingangsvariablen bestimmt werden, bezeichnet man als Schaltnetze. In diesen Schaltungen befinden sich keine Speicherelemente.

50.6.2.1 Rechennetze

Eine der wesentlichen Rechenschaltungen ist der *Addierer*, weil die Recheneinheit des Prozessors viele Rechenoperationen auf Additionen zurückführt.

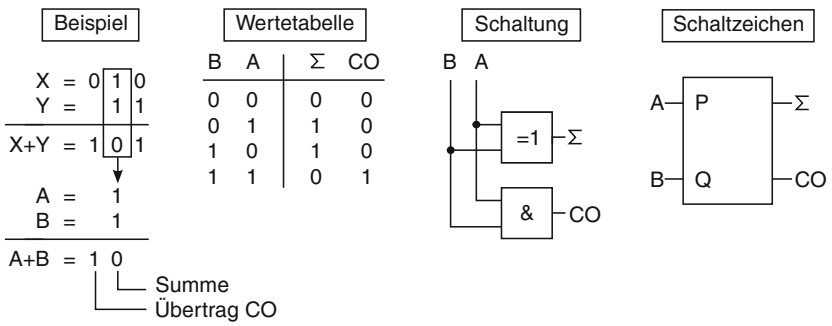


Abb. 50.42 Addition einer Binärstelle und Halbaddierer

50.6.2.1.1 Halbaddierer

Halbaddierer addieren zwei binäre Werte und zeigen die Summe und den Übertrag an (Abb. 50.42).

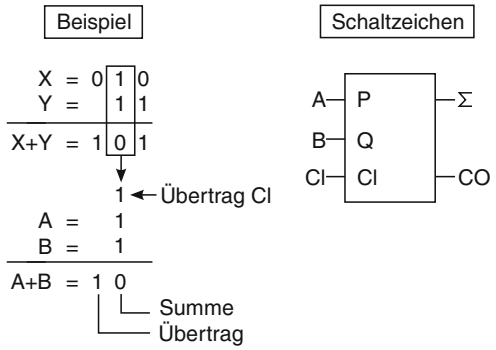


Abb. 50.43 Volladdierer

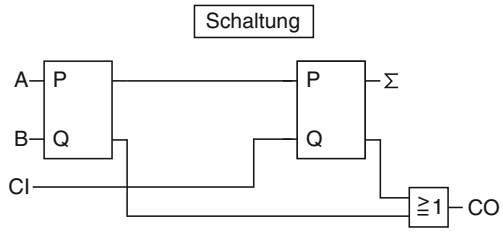


Abb. 50.44 Volladdierer aus zwei Halbaddierern

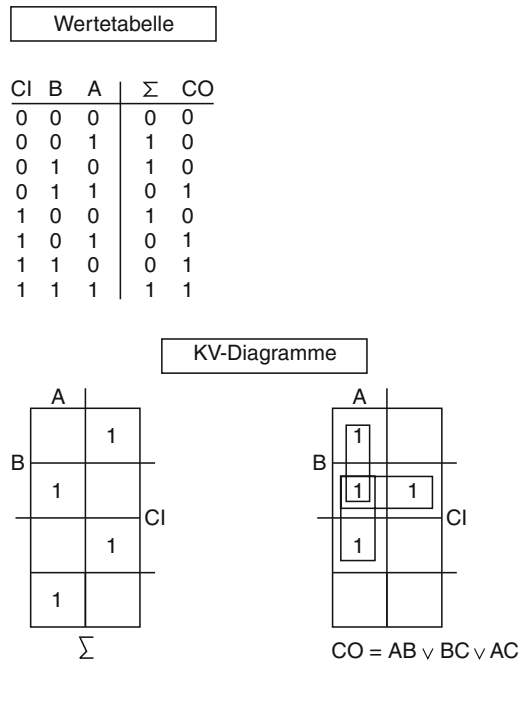


Abb. 50.45 Volladdierer aus Grundgattern

50.6.2.1.2 Volladdierer

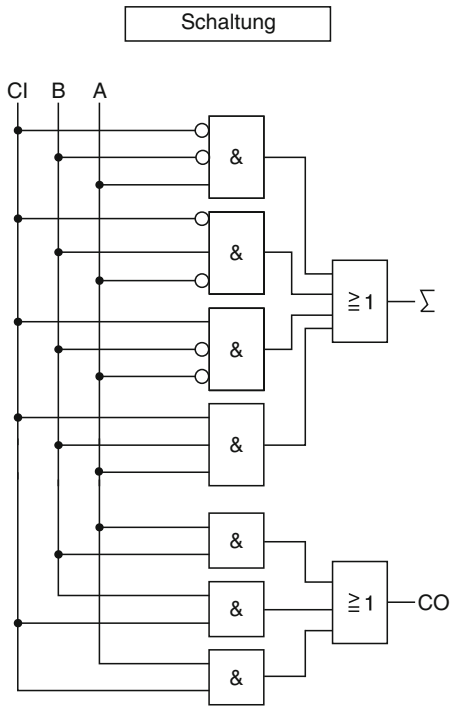
Sollen *n*-stellige Zahlen addiert werden, wird für jede Binärstelle ein Addierwerk benötigt. Ein Halbaddierer ist mit seinen zwei Eingängen hierzu nicht geeignet, weil eventuelle Überträge der niederwertigeren Stelle auch verarbeitet werden müssen. Schaltungen, die den Eingangsübertrag mit verarbeiten können, nennt man Volladdierer. Abb. 50.43 zeigt dies beispielhaft.

Ein Volladdierer lässt sich aus Halbaddierern aufbauen, wenn die Summe $A + B$ eines Halbaddierers einem zweiten Halbaddierer zugeführt und dort mit dem Übertrag der niederwertigen Stelle addiert wird. Der Übertragsausgang muss über ein ODER-Gatter gesondert herausgeführt werden (Abb. 50.44). Der Eingang für den Eingangsübertrag wird mit $CI = Carry\ In$ und der Übertragsausgang mit $CO = Carry\ Out$ abgekürzt.

Die Realisierung eines Volladdierers kann auch über die Wertetabelle vorgenommen werden, wie Abb. 50.45 zeigt. Die Gleichung für den Übertrag lässt sich vereinfachen.

50.6.2.1.3 Serieller *n*-Bit-Addierer

Führt man dem Addierer nacheinander taktweise die Zahlen A und B zu und verarbeitet eventuell auftretende Überträge, die beim nachfolgenden Takt addiert werden müssen, dann ergibt sich aus den Einzelsummen die Gesamtsumme nach n-Takten. Das Beispiel Abb. 50.46 zeigt das Prinzip für zwei 4-Bit-Zahlen. Für die Ein- und Ausgabe werden hier Schieberegister benötigt, die taktsynchron gesteuert werden müssen. Im dargestellten Beispiel steht die Summe nach dem vierten Takt in einem Speicher zur Verfügung.



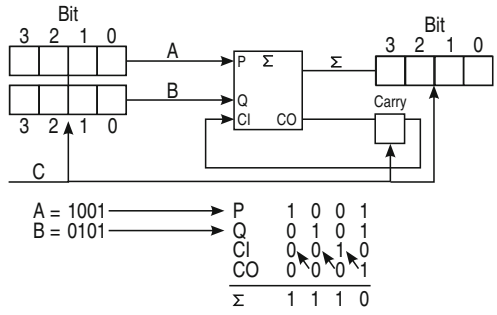


Abb. 50.46 Serielle Addition

50.6.2.1.4 Paralleler *n*-Bit-Addierer

Da der Aufbau eines parallelen *n*-Bit-Addierers über die Wertetabelle zu aufwendig ist, wird er aus Volladdierern hergestellt. Für jede zu addierende Bitstelle ist hierbei ein Volladdierer vorzusehen. Da bei der niedrigsten Stelle kein Eingangsübertrag anfällt, würde hier ein Halbaddierer ausreichen. In der Praxis baut man Addierer nur aus Volladdierern auf und legt den Eingangsübertragseingang der niederwertigsten Stelle auf „0“ (Masse). Das Prinzip eines 4-Bit-Volladdierers und ein Standardbaustein sind im Abb. 50.47 dargestellt.

Die Rechenzeit der Schaltung ist gegenüber der eines einzelnen Volladdierers größer, da die Summe bei diesem erst gebildet werden kann, wenn die Überträge den nachfolgenden Stufen zugeführt wurden. Ergibt sich beispielsweise bei der niederwertigsten Stufe ein Übertrag, wirkt sich dies auf alle folgenden Stufen aus. Die benötigte Zeit wird auch als *Mindestadditionszeit* (Add Time) bezeichnet. Kürzere Rechenzeiten erhält man mit speziellen Rechenbausteinen (Carry Look Ahead). Hier wird der Übertrag durch Verknüpfungsschaltungen „vorausschauend“ direkt aus den Eingangswerten ermittelt. In Abb. 50.48 ist noch einmal die Wertetabelle und die Gleichung für den Übertrag angegeben. Setzt man die Gleichung für beliebige Stellen um, so ergibt sich die Gleichung für ci + 1. Diese Gleichung teilt man auf in die Variable gi (Carry Generate), die angibt, ob ein Übertrag auftreten wird und pi (Carry Propagate), die in Verbindung mit c0 den Übertrag weiterleiten muss.

Das Bild zeigt die Entwicklung einer Stelle und die Gleichungen für einen 4-Bit-Addierer. Mit steigender Stellenzahl erhöht sich der Schaltungsaufwand erheblich, aber die Zeit bleibt auf zwei Gatterlaufzeiten begrenzt. Als Standardbaustein bietet sich Look ahead Carry Generator 74182 an.

50.6.2.1.5 Subtrahierer

Subtrahierer können mit Hilfe von Volladdierern realisiert werden, indem der Eingangsübertragsanschluss CI auf „1“ (+*U*) gelegt wird und der Subtrahend invertiert wird. Diese Vorgehensweise entspricht der Addition mit dem *Zweierkomplement* (Abb. 50.49).

Ändert man die Schaltung nach Abb. 50.50, erhält man ein universelles Additions-Subtraktionsnetz. Die Eingänge S0 bis S4 sind Steuereingänge, mit denen entsprechende Operationen durchgeführt werden können. S0 und S1 dienen zum Durchschalten der Zahlen A und B, und mit S2 und S3 kann jeweils eine Invertierung vorgenommen werden, um das Einer-Komplement zu bilden. Mit S4 wird nach Wunsch zusätzlich „1“

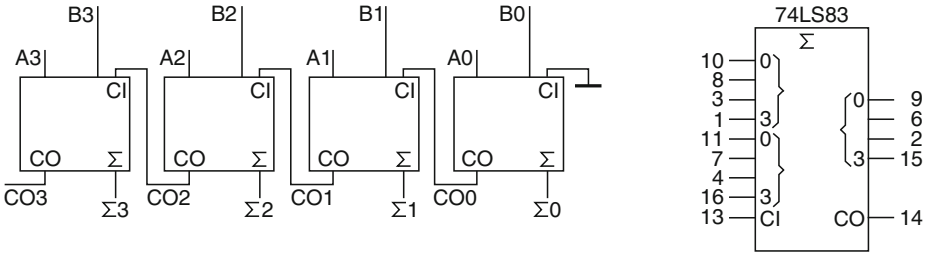


Abb. 50.47 Paralleler 4-Bit-Volladdierer und Standardbaustein

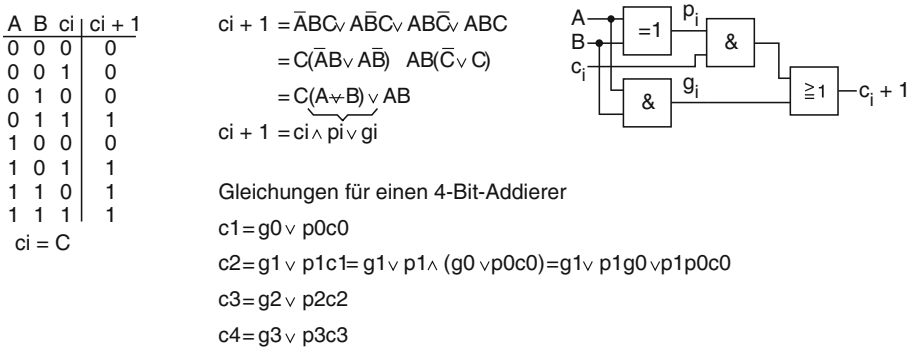


Abb. 50.48 Carry-Look-Prinzip

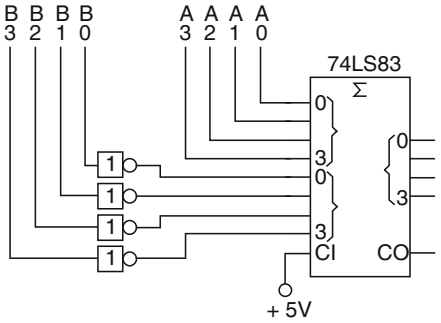


Abb. 50.49 4-Bit-Subtrahierer

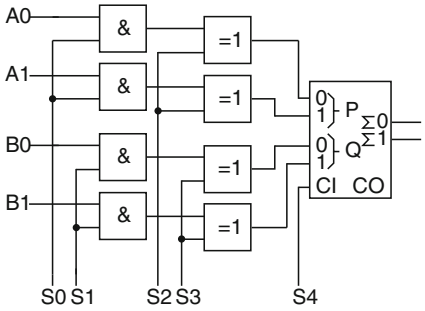


Abb. 50.50 Einfaches Rechenwerk

addiert, um in Verbindung mit S2 beziehungsweise S3 das Zweier-Komplement zu erhalten. Soll die Operation $A - B$ durchgeführt werden, muss S0 und S1 auf „1“, S2 auf „1“, S3 auf „0“ und S4 auf „1“ liegen.

50.6.2.1.6 Addierer für BCD-Dualzahlen

Vom Prinzip lassen sich zwei 4-Bit-Volladdierer zur Addition zweier BCD-Dualzahlen verwenden (Abb. 50.51). Hierbei muss nur berücksichtigt werden, dass eine Summe $> 9_{10}$ einen Übertrag für die nächste Decade anzeigen und eine Ergebniskorrektur erfolgen muss. Die Korrekturschaltung addiert mit Hilfe des 2. Addierers zu jeder Summe des 1. Addierers bei Auftreten einer Pseudotetrade oder eines Übertrages 6_2.

50.6.2.2 Komparatoren

50.6.2.2.1 Einfacher Komparator

Einfache Komparatoren (Abb. 50.52) prüfen zwei Zahlen A und B auf Gleichheit. Nur wenn beide Eingangsvariablen „0“ oder beide Eingangsvariablen „1“ sind, wird der Ausgang Q = „1“ (Wertetabelle). Diese Aufgabe erfüllen die dargestellten Schaltungen.

Für n Bit bietet sich eine einfache Erweiterung durch eine Äquivalenzschaltung für jede Bitstelle und anschließender UND-Verknüpfung an, wie das Beispiel für den 4-Bit-Komparator zeigt. Als Standardbaustein bietet sich der 8-Bit-Komparator 74HC688 an.

50.6.2.2.2 Komparator mit Größer- und Kleiner-Vergleich

Weitergehende Komparatoren vergleichen zusätzlich nach den Gesichtspunkten $A < B$ und $A > B$. Abb. 50.53 zeigt die entsprechende Wertetabelle für den 1-Bit-Vergleich, die Schaltung und das Schaltzeichen.

Ein Komparator für n Bitstellen wird durch Zusammenschalten mehrerer Universalkomparatoren erstellt. Abb. 50.54a–d zeigt die Entwicklung dieses Komparators auf.

Außer den Eingängen für die Vergleichbits besitzt der Universalkomparator drei zusätzliche Eingänge, um bei einem Vergleich mehrerer Binärstellen die einzelnen Stufen miteinander zu verbinden. Der Eingang $A = B =$ „1“ hat die Aufgabe, die Schaltung so zu aktivieren, dass ein Bitvergleich möglich wird. Bei $A > B$ wird mit dem Ausgang $A = B$ die nächstniederwertige Stelle gesperrt und die Meldung $A > B$ über die nachfolgende ODER-Verknüpfung sofort dem Ausgang gemeldet. Weil der Vergleich bei der Stelle mit der höchsten Wertigkeit beginnt, wird die dafür zuständige Stufe nach Abb. 50.55 beschaltet. Die Eingänge $A > B$ und $A < B$ liegen auf „0“ (Masse) und der Eingang $A = B$ auf „1“ ($+U$).

50.6.2.3 Codewandler und Decoder

50.6.2.3.1 Codewandler

Ein Codewandler (Codierer) ordnet einen Code einem anderen Code zu. Er setzt einen Quellcode in einen Zielcode um, er ist praktisch ein „Übersetzer“.

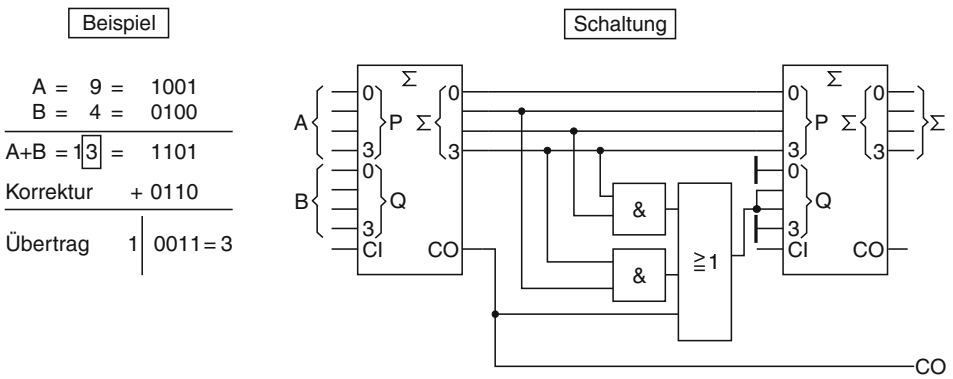


Abb. 50.51 Addierer für BCD-Zahlen

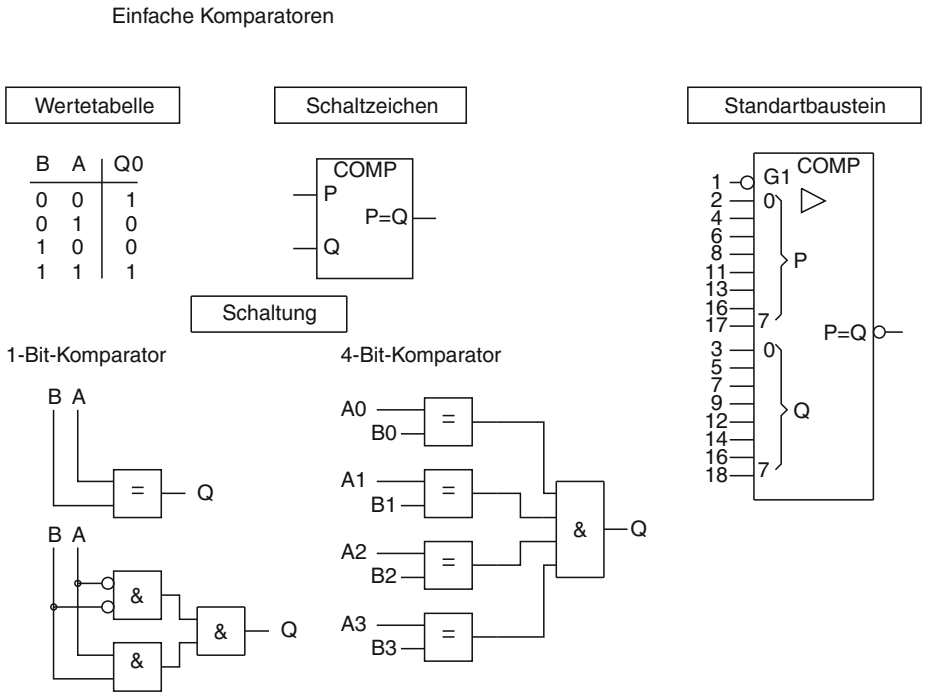


Abb. 50.52 Einfache Komparatoren

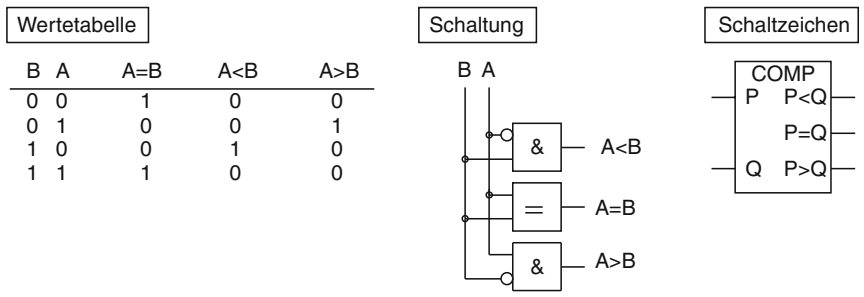


Abb. 50.53 1-Bit-Komparator

Beispiel

Um einen Gray/BCD-Codierer zu entwerfen, wird Quell- und Zielcode in einer Wertetabelle gegenübergestellt. Für jede Stelle muss dann ein KV-Diagramm benutzt und die Gleichung ermittelt werden. Anschließend ist aus der Gleichung die Schaltung zu zeichnen (Abb. 50.56). ◄

50.6.2.3.2 1-aus-*n*-Decoder

Wie aus der Wertetabelle (Abb. 50.57) hervorgeht, erhält jeweils ein zugewiesener Ausgang „1“ in Abhängigkeit einer dualen Eingangsinformation. Q1 schaltet beispielsweise bei der Eingangskombination A = „1“ und B = „0“ auf „1“. Hierbei kann die Schaltung um einen zusätzlichen Freigabeeingang erweitert werden. Ein Steuersignal kann dann den Baustein aktivieren. Ist der Freigabeeingang F = „0“, sind auch alle Ausgänge „0“. Mit Decodern lassen sich z. B. verschiedene Speicherbausteine oder Baugruppen anwählen.

50.6.2.4 Multiplexer und Demultiplexer

Multiplexer und Demultiplexer sind „elektronische Schalter“, die Eingangsinformationen verschiedener Datenleitungen auf einen Ausgang (Multiplexer) oder eine Eingangsinformation auf verschiedene Ausgangsdatenleitungen (Demultiplexer) schalten. Welche Leitungen durchgeschaltet werden, bestimmt die Steuerung des „Schalters“ (Abb. 50.58).

Aufgabe dieser Bausteine kann es sein, die parallel anliegenden Daten eines Sendebausteins über eine Leitung zu übertragen und im Empfängerbaustein die Daten wieder parallel anliegen zu lassen. Dies geschieht mittels *Zeitmultiplexverfahren*, wobei die entsprechenden Datenleitungen zeitlich nacheinander und synchron durchgeschaltet werden.

50.6.2.4.1 Multiplexer

Multiplexer verteilen die Daten D verschiedener Eingänge, die durch Adressen angewählt werden, auf einen Ausgang Q (Abb. 50.59). Soll beispielsweise das Datenbit von D0 zum Ausgang Q gelangen, muss A0 mit „0“ und A1 mit „0“ angewählt

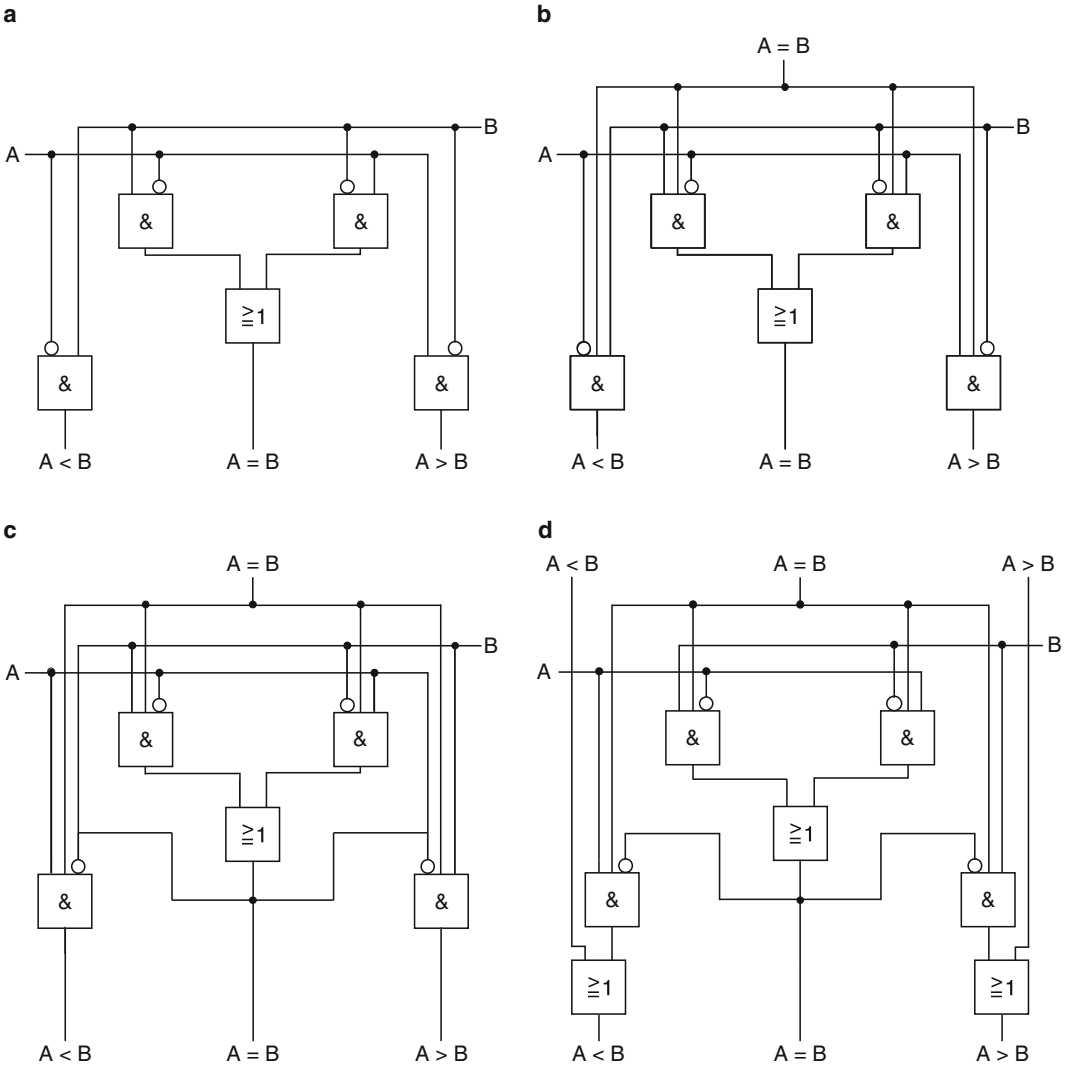


Abb. 50.54 Universalkomparator

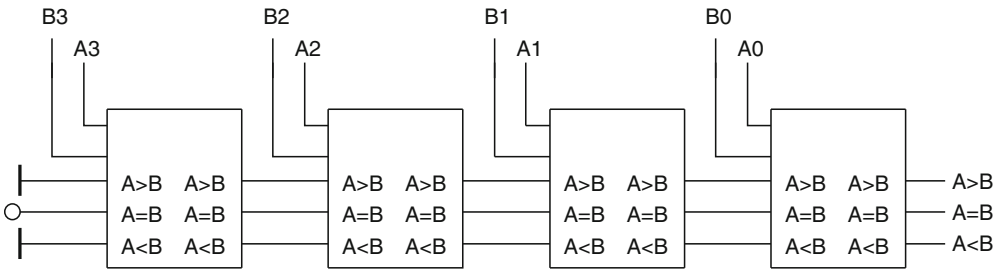


Abb. 50.55 Komparator für 4 Bit

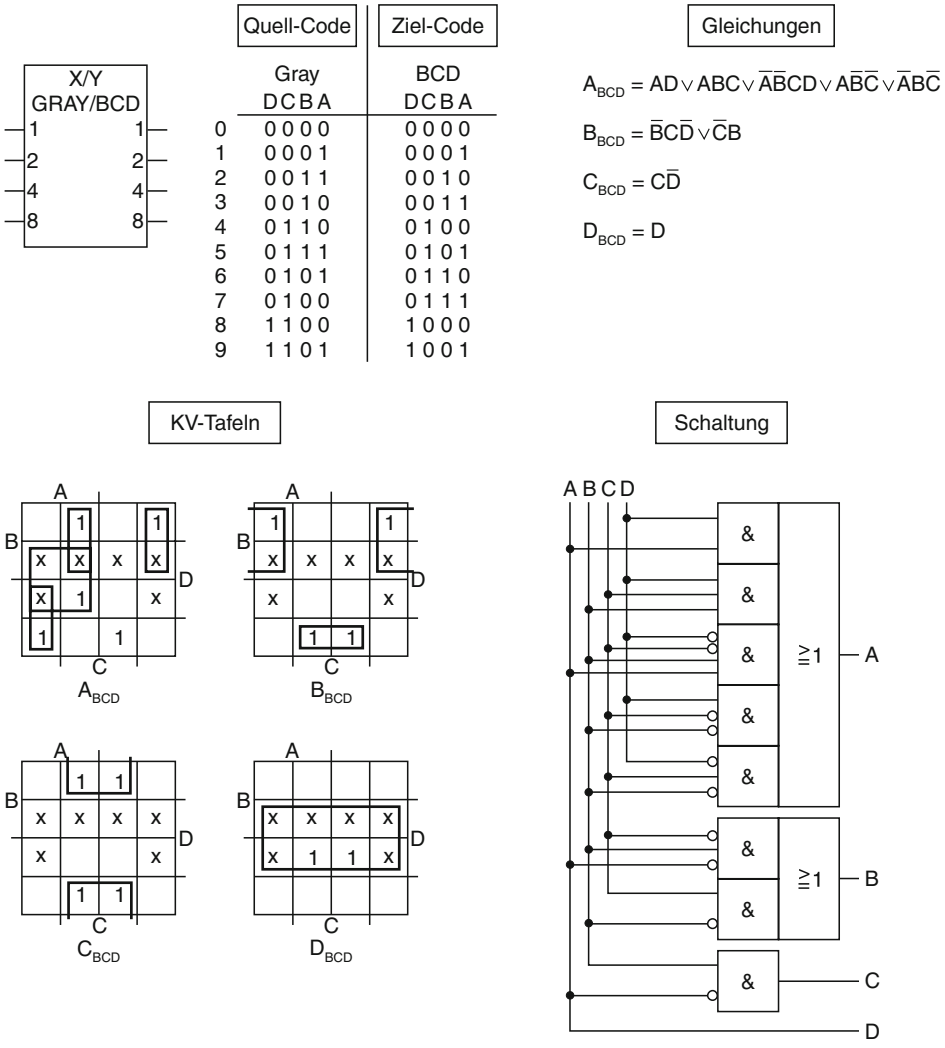


Abb. 50.56 Gray-BCD-Codewandler

werden. Dies wird auch aus der Gleichung deutlich. Die Anzahl der notwendigen Adressleitungen ist von der Anzahl der gewünschten Dateneingänge abhängig.

Nach Abb. 50.60 kann die Schaltung zusätzlich mit einem Freigabeeingang versehen werden. Bei FE = „1“ gelangt das angewählte Datenbit zum Ausgang Q. Standardbausteine besitzen häufig einen 3-state-Ausgang.

Ein Anwendungsbeispiel für Multiplexer zeigt das Prinzip des gemultiplexten Adress-Datenbusses (Abb. 50.61) auf. Einige Prozessoren und Controller besitzen, um Anschlussleitungen einzusparen, so einen Bus. Hierbei werden Adressen und Daten zeitlich nacheinander auf das Bussystem geschaltet. Die Steuerung nimmt die CPU vor.

50.6.2.4.2 Demultiplexer

Mit Hilfe von Demultiplexern kann eine Eingangsinformation auf verschiedene Ausgänge verteilt werden (Abb. 50.62). Die erforderlichen Auswahlleitungen (Adressen) sind abhängig von der Anzahl der Ausgangs-Datenleitungen.

50.6.3 Schaltwerke

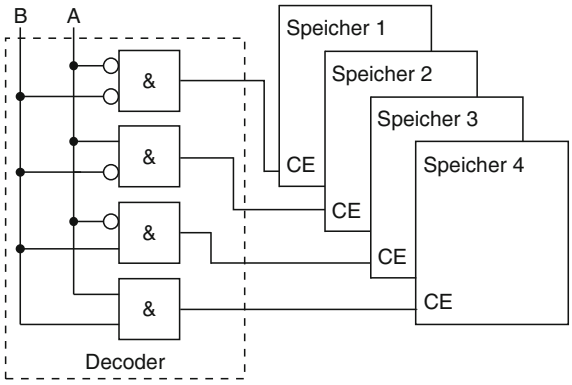
Schaltwerke (Sequentielle Logik) sind Schaltnetze mit einem Speicherverhalten. Die wichtigste Grundschaltung ist das *Flipflop* (bistabile Kippstufe).

50.6.3.1 Speicherbausteine

50.6.3.1.1 Allgemeines Flipflop

Unter Flipflops versteht man Schaltungen, deren zwei Ausgänge immer *komplementäre stabile Zustände* aufweisen. Das Ausgangsverhalten selbst wird über zwei Eingänge beeinflusst. Damit ist die Speicherung einer binären Information möglich.

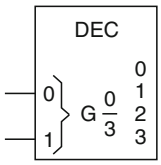
B	A	Funktion
0	0	Speicher 1 angewählt
0	1	Speicher 2 angewählt
1	0	Speicher 3 angewählt
1	1	Speicher 4 angewählt



Wertetabelle

B	A	Q0	Q1	Q2	Q3
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

Schaltzeichen



Schaltungen

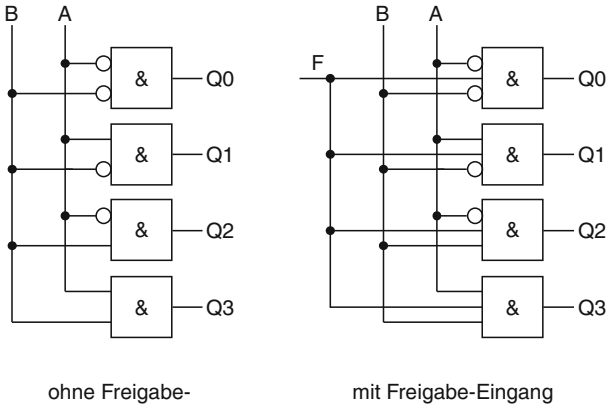


Abb. 50.57 1-aus-4-Decoder und Anwendungsbeispiel

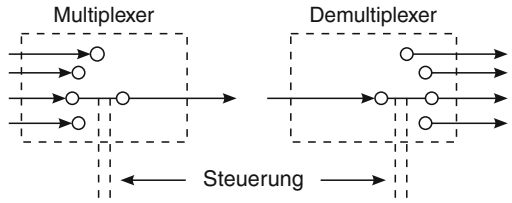


Abb. 50.58 Prinzip

Bei den, durch einen Takt gesteuerten Flipflops, wird nach *taktzustandsgesteuerten* und *taktflankengesteuerten* Flipflops unterschieden. Flipflops lassen sich mit NAND- und NOR-Gattern realisieren.

Flipflop aus NAND-Gattern

Ein NAND erzeugt am Ausgang immer eine „1“, wenn mindestens ein Eingang mit „0“ belegt ist. Damit ergeben sich die Ausgangszustände Abb. 50.63a–c.

Die Wertetabelle fasst alle Schaltzustände zusammen. Q1 = „1“ und Q2 = „1“ entspricht nicht der Funktion eines Flipflops. Diese Eingangskombination, die diesen Ausgangszustand hervorruft, ist daher logisch „verboten“, sie ist irregulär. Sind die Eingangspegel komplementär (ein Eingang „0“ und der andere „1“ oder umgekehrt), schaltet das Flipflop. Bei „1“ an beiden Eingängen bleibt der vorherige Zustand erhalten; man spricht dann vom Speicherzustand. Oft werden Schalter oder Taster als Eingabe von digitalen Systemen eingesetzt. Hierbei kann es zu Prellen kommen. Der Kontakt öffnet sich kurzzeitig wieder, und die konkrete Schaltung erkennt mehrere Impulse. Durch Pull-up-Widerstände an den Eingängen ist die Eingangskombination A = „0“ und B = „0“ nicht möglich. Sobald der Taster gedrückt ist, wird der Zustand gespeichert. Abb. 50.64 zeigt das Prinzip des Prellens und Abhilfe durch ein Flipflop. ◄

50.6.3.1.2 RS-Flipflop

Beim RS-Flipflop (R = Reset = Rücksetzen, S = Set = Setzen) folgt der Ausgang Q immer dem Eingang S. Die Eingangszustände S = R = „1“ sind „verboten“, und die Speicherung

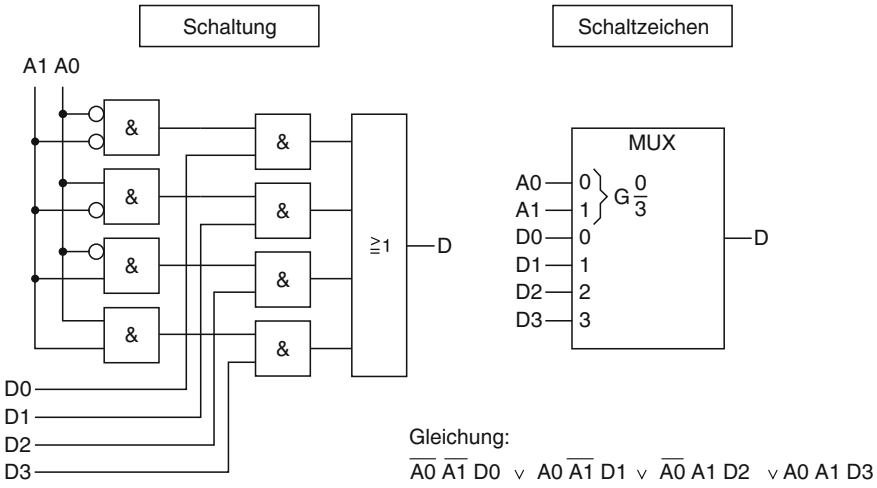


Abb. 50.59 Multiplexer

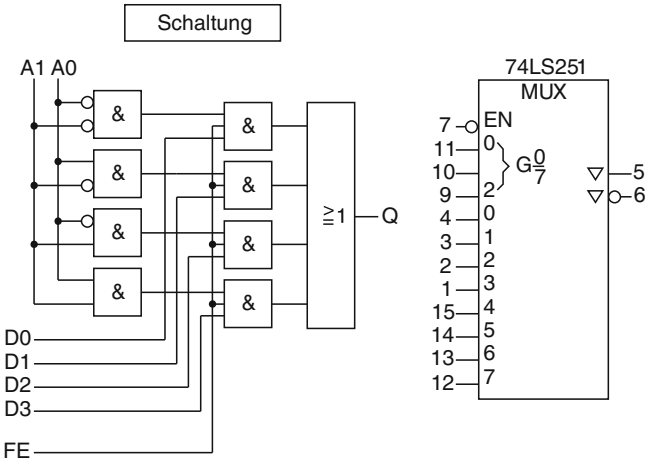


Abb. 50.60 Multiplexer mit Freigabeeingang und Standardbaustein

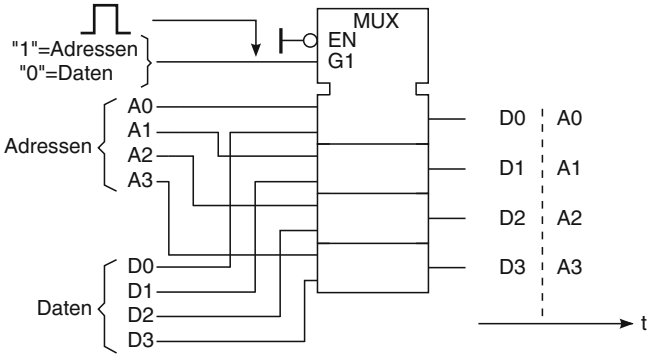


Abb. 50.61 Gemultiplexter Adress-Datenbus

erfolgt bei S = R = „0“. RS-FF’s lassen sich aus NAND- oder NOR-Gattern aufbauen (Abb. 50.65).

50.6.3.1.3 Flipflops mit dominierenden Eingängen

Bei Flipflops mit dominierenden Eingängen (Abb. 50.66) können aufgrund der Eingangsbeschaltung irreguläre Zustände an den Ausgängen nicht auftreten. Das *R-Flipflop* ist ein RS-FF mit dominierendem R-Eingang, weil ein Rücksetzen immer bei R = „1“ erfolgt, unabhängig vom Zustand an S. Das *S-Flipflop* ist ein RS-FF mit dominierendem Setzeingang. Unabhängig vom Zustand am R-Eingang wird das Flipflop bei S = „1“ gesetzt. Das *E-Flipflop* speichert bei gleichen Eingangszuständen R = S immer.

50.6.3.1.4 D-Flipflop

Mit dem D-Flipflop (Data Latch) lässt sich eine binäre Information speichern, solange C = „1“ ist. Irreguläre Zustände treten nicht auf. Abb. 50.67 zeigt ein D-Flipflop aus NAND-Gattern mit der zugehörenden Wertetabelle und dem Schaltzeichen.

Da das Flipflop das Eingangssignal verzögert, spricht man auch von einem *Delay-Flipflop* (siehe Impulsdiagramm Abb. 50.68). Soll der Zustand vom Dateneingang in Abhängigkeit einer Flanke übernommen werden, spricht man vom *einflankengesteuerten* D-Flipflop. Der Eingangswert kann mit der ansteigenden oder abfallenden Flanke des Clock- beziehungsweise Taktsignals übernommen werden. Von einem taktzustandsgesteuertem Flipflop spricht man, wenn die Daten nur bei C = „1“ oder C = „0“ übernommen werden. D-Flipflops werden für kurzzeitige Speicheroperationen eingesetzt. Hierbei werden die an den D-Eingängen anliegenden Informationen mit dem Takt übernommen. Der Takt bestimmt damit den Zeitpunkt der Datenübernahme (Abb. 50.69).

50.6.3.1.5 JK-Flipflop

Erweitert man das RS-FF nach Abb. 50.70 durch zwei UND-Gatter und entsprechender Rückkopplung, werden unzulässige Eingangskombinationen verhindert. Die entgegengesetzten Zustände an den Eingängen J und K werden beim nächsten Takt

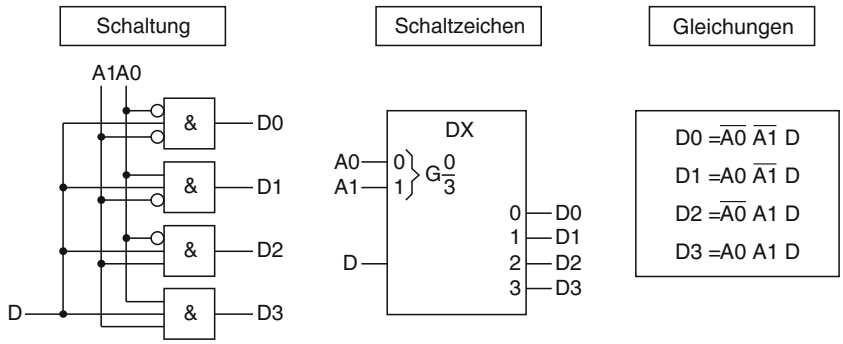
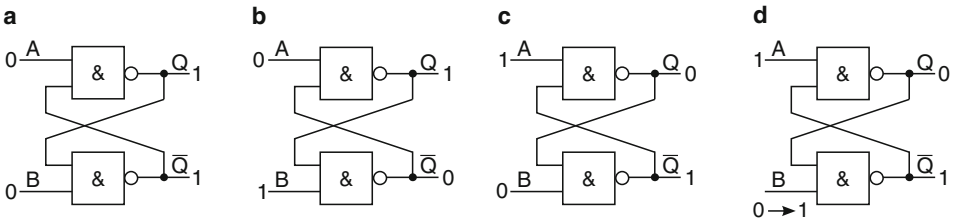


Abb. 50.62 Demultiplexer



Wertetabelle

A	B	Q	$\overline{Q}$	Bemerkungen
0	0	1	1	irregulär
0	1	1	0	
1	0	0	1	
1	1	0	1	Speicherung

Abb. 50.63 Funktion eines Flipflops mit NAND-Gattern

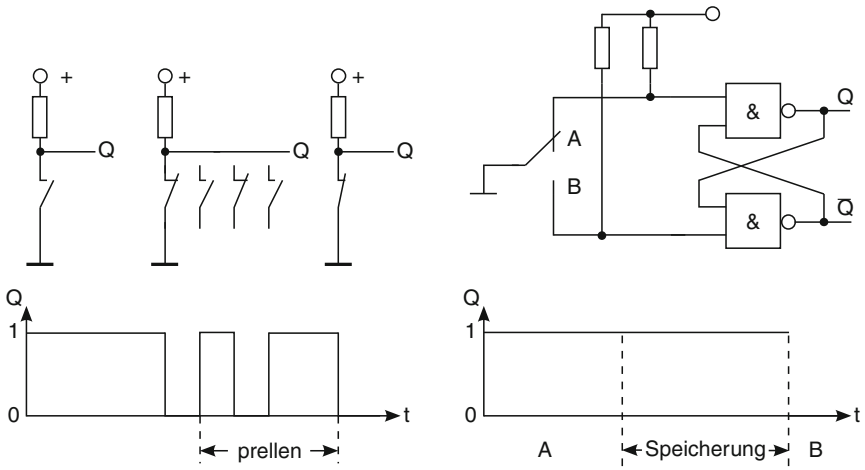
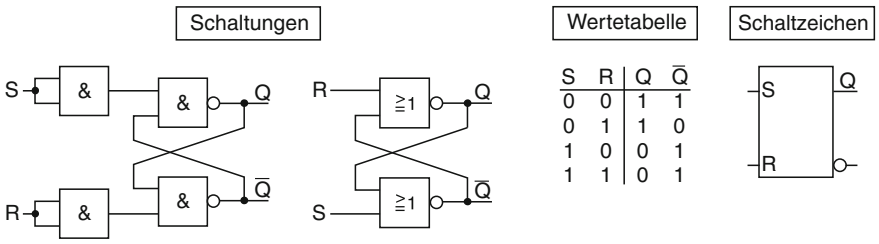


Abb. 50.64 Prinzip des Prellens und entprellter Schalter



S	R	Q	$\overline{Q}$
0	0	1	1
0	1	1	0
1	0	0	1
1	1	0	1

Abb. 50.65 RS-Flipflop aus NAND- und NOR-Gattern mit Wertetabelle und Schaltzeichen

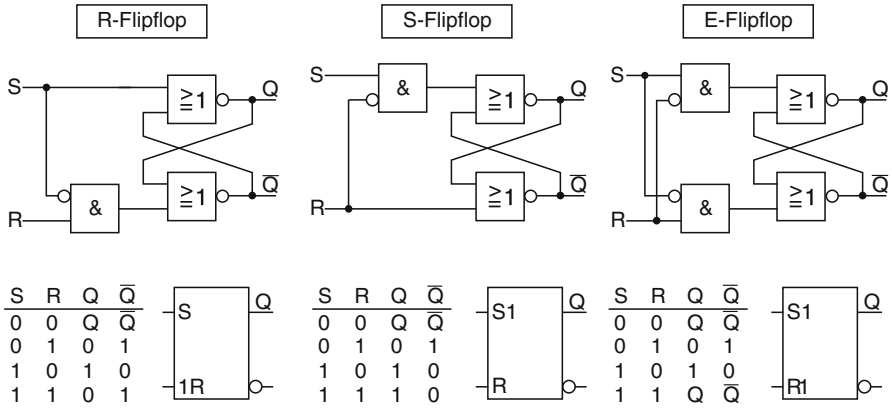


Abb. 50.66 Flipflop mit dominierenden Eingängen

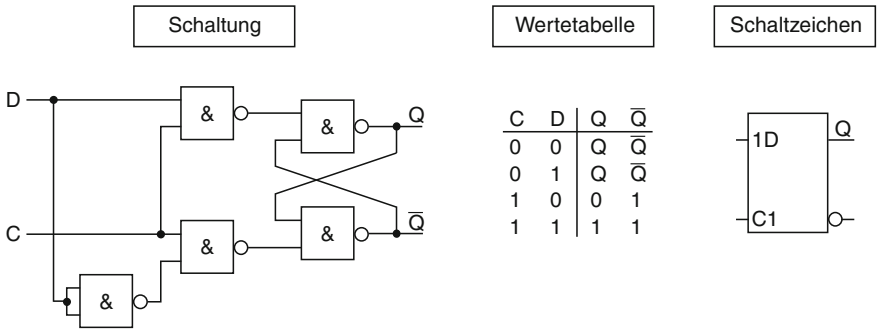


Abb. 50.67 D-Flipflop

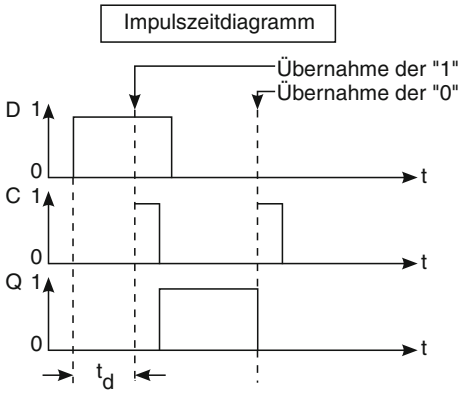


Abb. 50.68 Impulsdiagramm

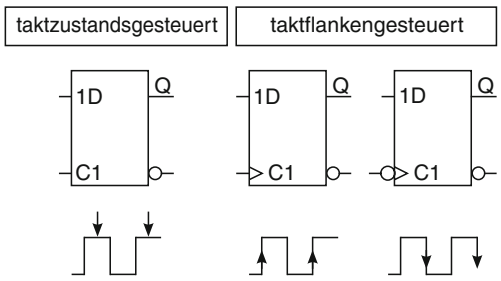


Abb. 50.69 Taktgesteuerte Flipflops

übernommen. Ist J = K = „0“, ändert sich der Ausgang nicht, und bei J = K = „1“ schaltet das Flipflop mit jedem Takt um. Hierbei wirkt das Flipflop dann wie ein T-Flipflop.

50.6.3.1.6 Master-Slave-JK-Flipflop

Das Master-Slave-JK-Flipflop (Abb. 50.71) ist ein *zweiflankengesteuertes* Flipflop, weil es bei ansteigender oder abfallender Flanke die Eingangsinformation aufnimmt und speichert und mit der zur Eingangsübernahme komplementären Flanke die Information zum Ausgang weiter schaltet. Es besteht immer aus zwei JK-Flipflops, dem *Master* (Herr) und dem *Slave* (Diener). Das Zeichen an den Ausgängen deutet einen *retardierenden* Ausgang an. Dies bedeutet nach DIN 40700, dass der Zustandswechsel des Flipflops an den Ausgängen erst dann wirksam wird, wenn die Eingangsvariable zu ihrem ursprünglichen Wert zurückgekehrt ist.

Abb. 50.72 beschreibt die Funktion und zeigt die Auswirkung der Eingangsbeschaltung auf den Ausgang Q. Das Master-Slave-Flipflop übernimmt bei der ansteigenden Flanke die Eingangsinformation. Während dieser Zeit ist der Slave gesperrt und gibt bei abfallender Flanke die Information – hier sperrt das Master-FF – an den Slave weiter.

Betrachtet man die Auswirkung des Clock-Signals in einer zeitlichen Reihenfolge, so ergeben sich die dargestellten Zustände im Master-Slave-JK-FF. Jeweils beim Überschreiten der Schwelle 1 und 2 laufen die dargestellten Reaktionen ab. Mit

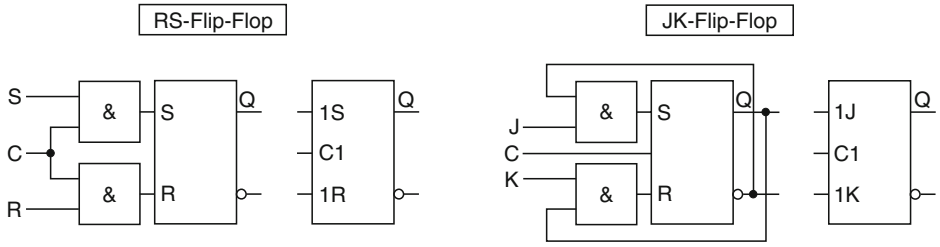


Abb. 50.70 JK-Flipflop

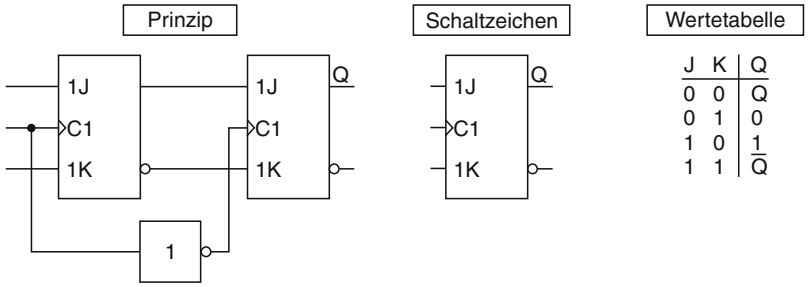


Abb. 50.71 Master-Slave-Prinzip

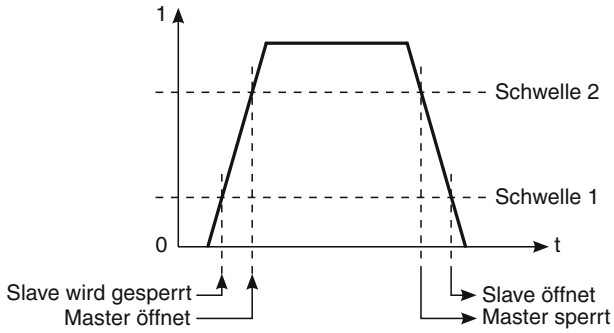


Abb. 50.72 Zeitverhalten des Master-Slave-JK-Flipflop

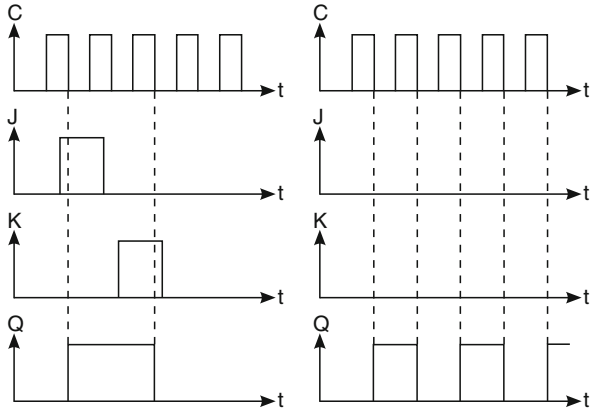


Abb. 50.73 Impulsdiagramm Master-Slave-JK-Flipflop

der ansteigenden Flanke wird der Slave gesperrt, und der Master kann die Eingangsinformation übernehmen, um sie mit der abfallenden Flanke zum Slave zu bringen. Während dieser Zeit ist der Master gesperrt. Abb. 50.73 verdeutlicht diesen Zusammenhang mit dem Impulszeitdiagramm, und Abb. 50.74 zeigt die Anschlussbelegung von drei gängigen Standardbausteinen.

50.6.3.1.7 T-Flipflop

Das T-Flipflop (Trigger-Flipflop, Toggle-Flipflop) ist *flankengesteuert* und besitzt nur den Takteingang. Es schaltet wie das JK-Flipflop bei der Eingangskombination J = „1“ und K = „1“ jeweils mit der Taktflanke die Ausgangszustände um (Abb. 50.75). T-Flipflops werden in Zähler- und Teilerschaltungen eingesetzt.

50.6.3.2 Grundschaltungen aus Speicherbausteinen

50.6.3.2.1 Register

Register sind aus Speicherbausteinen zusammengesetzte Einheiten. Man unterscheidet hierbei *allgemeine Register*, die zur kurzzeitigen Speicherung kleiner Datenmengen wie 8-, 16- oder 32-Bit-Datenworte dienen, und *Schieberegister*, die die gespeicherte Information jeweils mit einem Takt weitergeben. Mit dem Taktsignal (C = Clock) werden die Daten in das Register eingegeben, und mit OE (Output Enable) können die gespeicherten Daten freigegeben werden.

50.6.3.2.2 Schieberegister

Schieberegister bestehen aus hintereinander geschalteten D-Flipflops oder JK-Flipflops mit untereinander verbundenen Takteingängen. Sie geben die gespeicherte Information jeweils mit einem Takt von Flipflop zu Flipflop weiter. Abb. 50.76 zeigt eine mögliche Schaltung und beschreibt die Funktion mit dem Impulszeitdiagramm.

Zur Verdeutlichung sind die Pegel-Ausgänge Q0 bis Q3 gesondert herausgeführt. Das Datenbit liegt am Eingang D an und

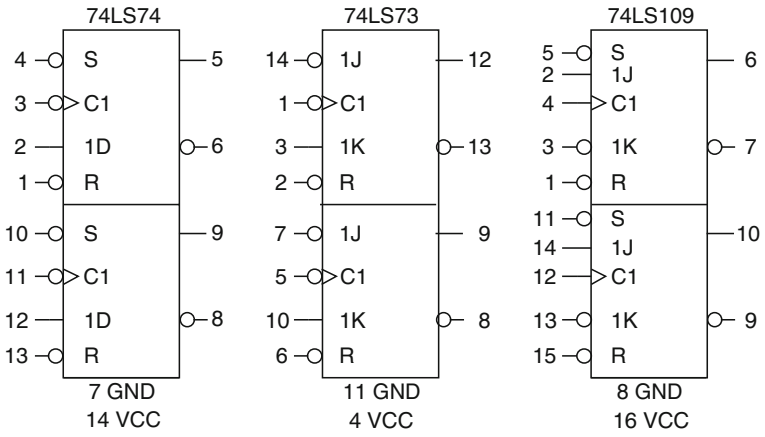


Abb. 50.74 Standardbausteine Master-Slave-JK-Flipflop

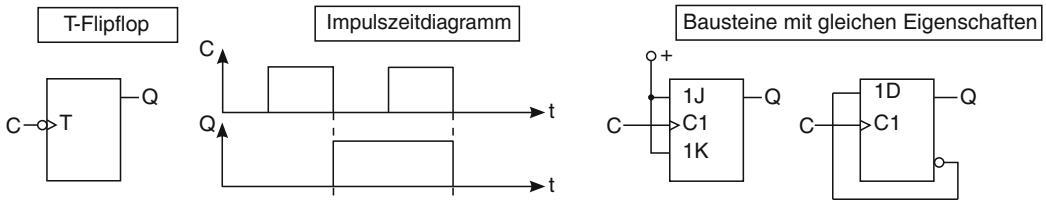


Abb. 50.75 T-Flipflop mit Impulsdiagramm und funktionsgleiche Bausteine

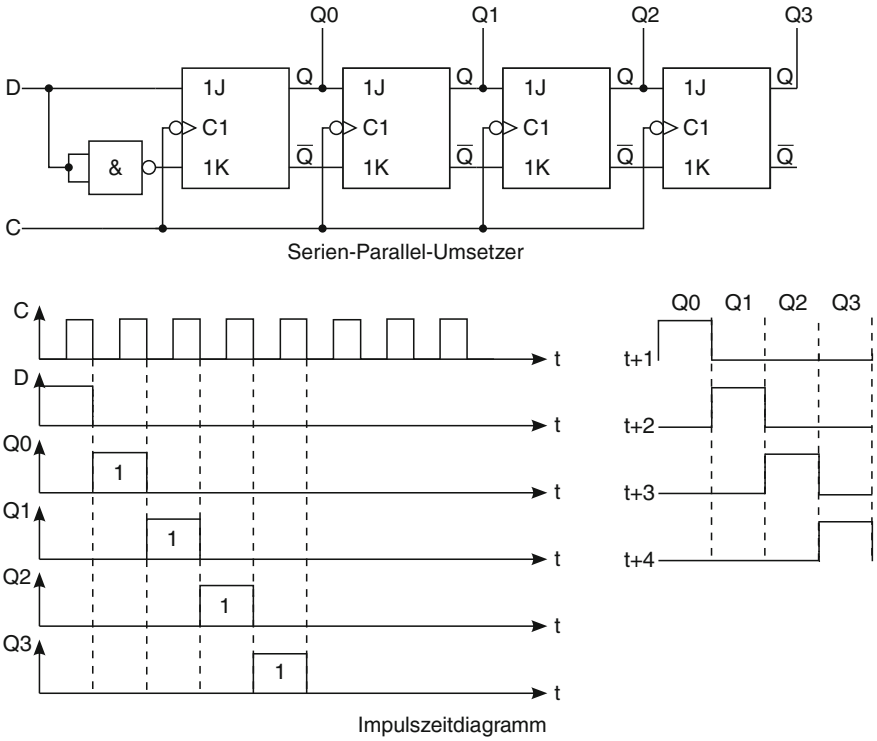


Abb. 50.76 Schieberegister

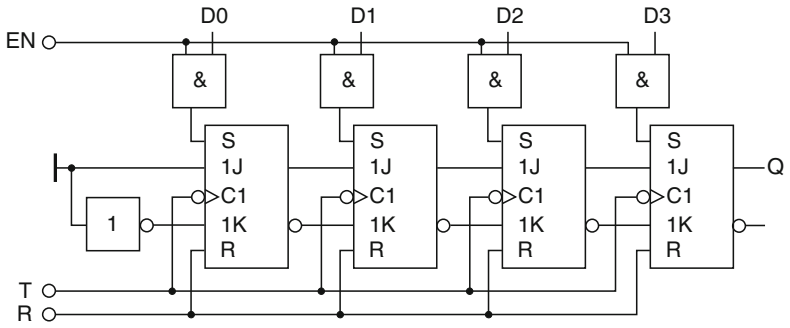


Abb. 50.77 Parallel-Serien-Register

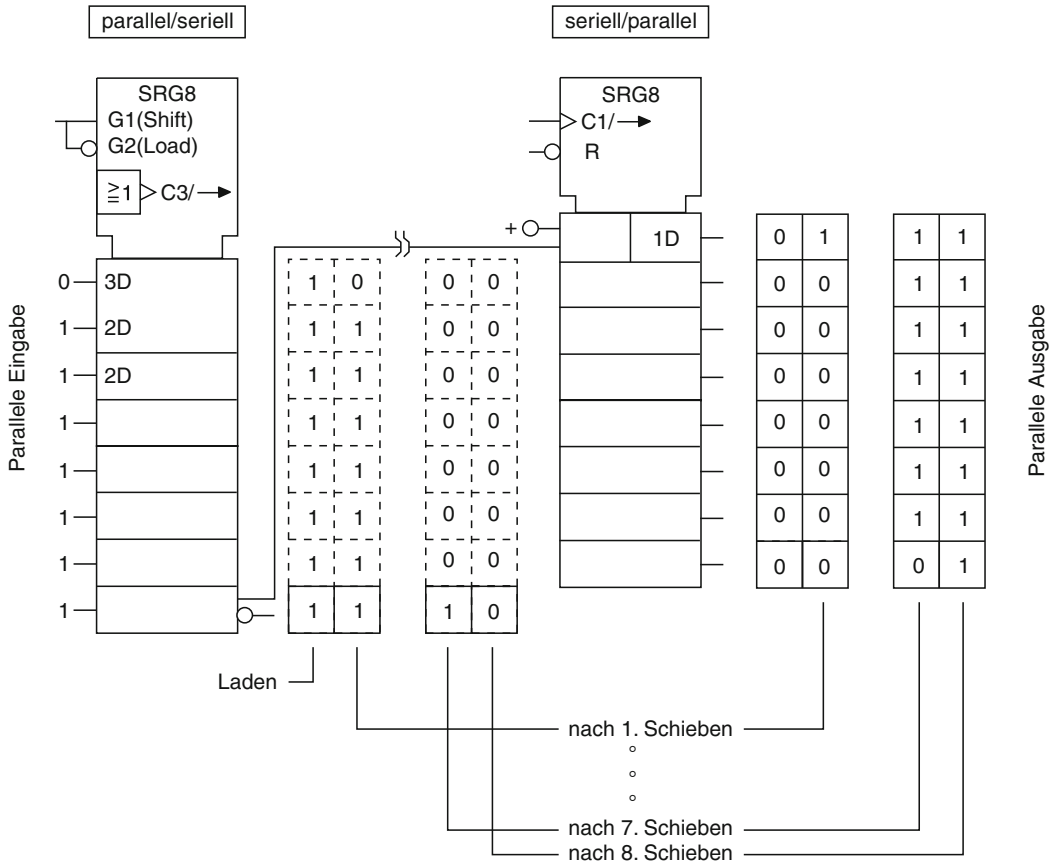


Abb. 50.78 Schieberegisteranwendung

wird mit dem 1.Takt übernommen. Damit wird Q0 „1“. Mit dem nächsten Takt wird „0“ von D übernommen, und die „1“ von Q0 gelangt nach Q2. Nach dem 5. Takt ist das Schieberegister „leer“. Man unterscheidet Schieberegister nach ihrer Länge und ihrer Betriebsart. Je nach Datenein- und ausgabe (seriell, parallel) und Schieberichtung (links, rechts) unterscheiden sich die Betriebsarten der Schiebere- gister. Ist ein Ausgang mit dem Eingang verbunden, spricht man vom *Ringschieberegister*. Im Abb. 50.77 ist ein Schieberegister für die parallele Eingabe dargestellt. S und R sind zusätzliche Vorbereitungseingänge. Die Ansteuerung verhindert das gleichzeitige Anstehen von Setz- und Rücksetzsignal und lässt die parallelen Daten nur mit dem Freigabeeingang zu. Register mit seriellem Dateneingang und parallelen Datenausgängen bezeichnet man auch als *Serien-Parallel-Umsetzer* und Register mit parallelen Eingängen und seriellem Datenausgang als *Parallel-Serienumsetzer*.

Im Abb. 50.78 sind zwei Standardbausteine für eine serielle Übertragung zusammengeschaltet. Die anliegenden parallelen Daten werden geladen und mit jedem Takt über *eine* Leitung dem Baustein 74164 zugeführt. Dieser Baustein stellt nach 8 Takten das eingegebene Datenwort wieder zur Verfügung. In der Mikroprozessortechnik sind solche parallel/seriell- und seriell/parallel-Wandler Bestandteile von Interfaceschaltungen.

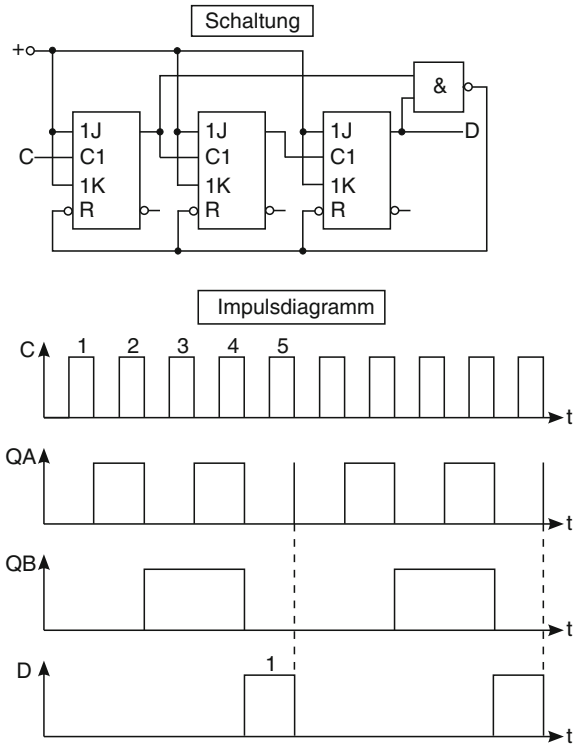


Abb. 50.79 Frequenzteiler 5 : 1

50.6.3.2.3 Frequenzteiler

Frequenzteiler erzeugen nach z Impulsen an ihrem Eingang *einen* Ausgangsimpuls und teilen somit die Eingangsfrequenz durch z. Sie bestehen aus Flipflops, die hintereinander geschaltet sind und deren Ausgang Q jeweils das nachfolgende Flipflop ansteuert. Jedes Flipflop teilt im Verhältnis 2 : 1. Mit vier Flipflops lässt sich ein Teilerverhältnis von $2^4 = 16$ realisieren. Zähler und Teiler bestehen aus gleicher Hardware mit unterschiedlicher Signalauswertung. Der Teiler besitzt immer einen Eingang und einen seriellen Ausgang, während der Zähler seine Information als paralleles Wort ausgibt. Durch entsprechende Verknüpfungen lassen sich ungeradzahlige Teilerverhältnisse realisieren. Daher wird auch nach *geradzahligen* und *ungeradzahligen Teilern* unterschieden. Abb. 50.79 gibt die Schaltung eines 5 zu 1-Teilers wieder.

Nach dem 5. Impuls erfolgt eine Rücksetzung über das NAND-Gatter, um mit Rückstellung aller Flipflops neu zu beginnen. Will man das Impulspausenverhältnis gleich machen, dann muss am Ausgang D ein zusätzliches Flipflop eingesetzt werden. Ein Anwendungsbeispiel ist in Abb. 50.80 dargestellt. Aus einem Rechteckgenerator gewinnt man durch zwei Teilerbausteine unterschiedliche Frequenzen.

50.6.3.2.4 Zähler

Zähler gehören zu den Grundbausteinen der Digitaltechnik. Sie werden z. B. zum Zählen von Stückzahlen, Zählen von Umdrehungen eines Motors oder Zählen einer Zeiteinheit eingesetzt. Die Impulse zur Ansteuerung des Zählers liefert eine Zeitbasis oder unterschiedliche Sensoren. Die Eigenschaft des Zählers, zu addieren oder zu subtrahieren (immer 1) und zu speichern, lässt sich mit Flipflops erreichen. Die gezählten Impulse erscheinen an einem Parallelausgang. Die Zählerkapazität ist von der Anzahl n der Flipflops abhängig. Dabei gilt als maximale Zählerkapazität $K_{max} = 2^n$. Ein Zähler aus 4 Flipflops hat somit $K_{max} = 16$. Er erreicht nach 16 Zählimpulsen wieder Null. Man spricht von einem „Modulo-16- Zähler“. Ist nach dem n-Zählimpuls der Zustand Null erreicht, bezeichnet man den Zähler allgemein als „Modulo-n-Zähler“. Zähler lassen sich nach der *Zählrichtung* als Vorwärts- oder Rückwärtszähler, nach der *Zahlendarstellung*, z. B. als BCD- oder Dualzähler und nach *Art der Taktansteuerung* als Asynchron- oder Synchronzähler einteilen. Abb. 50.81 zeigt einen universellen Standardbaustein.

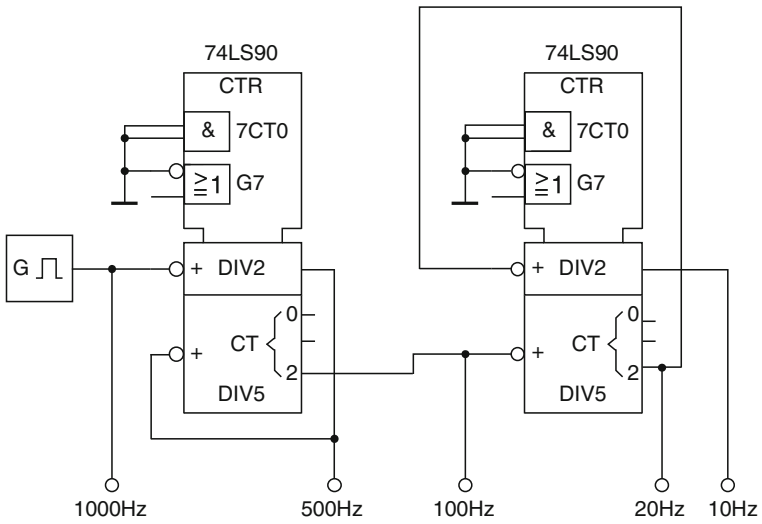


Abb. 50.80 Zählerbausteine als Frequenzteiler

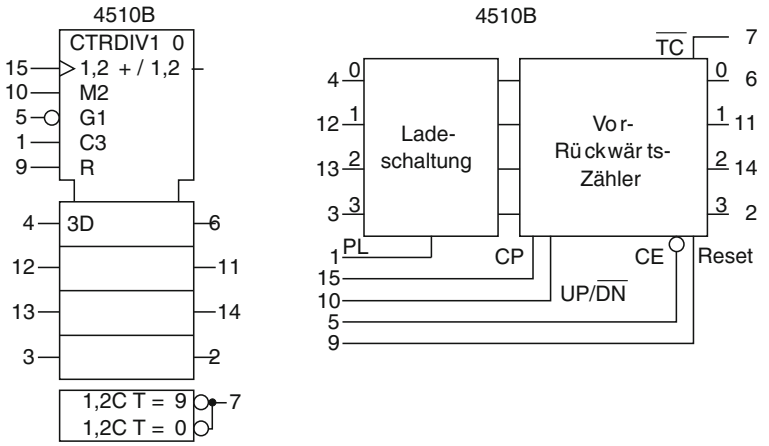


Abb. 50.81 Universalzähler

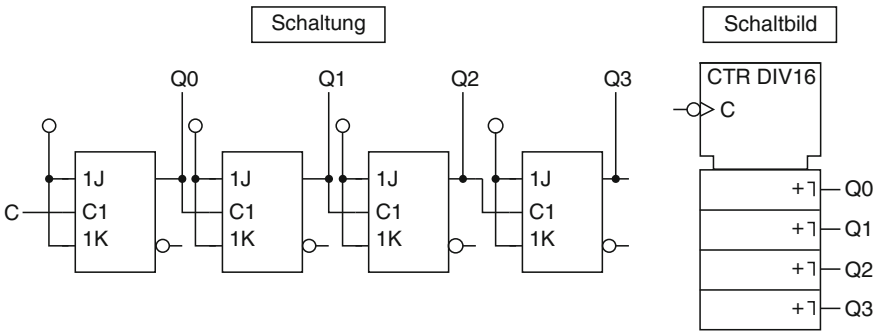


Abb. 50.82 Asynchronzähler

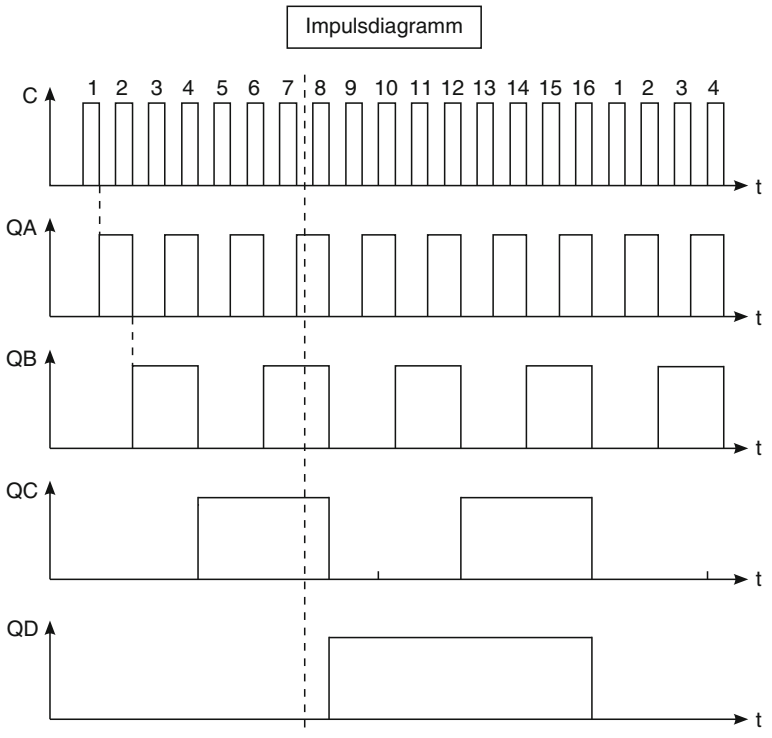


Abb. 50.83 Impulsdiagramm eines Asynchronzählers

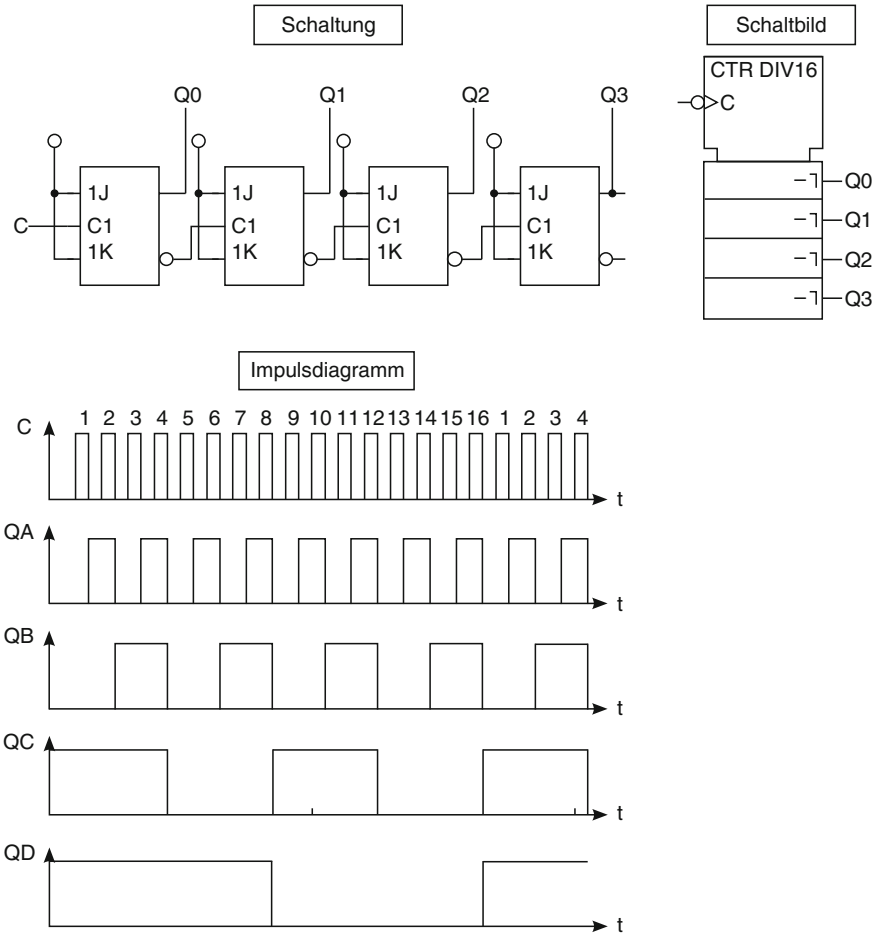


Abb. 50.84 Rückwärtszähler und Impulsdiagramm

50.6.3.2.5 Asynchroner Zähler

Bei einem Asynchronzähler wird der Takt nur dem ersten Flipflop zugeführt. Die nachfolgenden Stufen werden jeweils von der Vorstufe gesteuert. Im Abb. 50.82 wurden zum Aufbau eines *Vorwärtszählers* JK- Flipflops eingesetzt. Die Arbeitsweise der Flipflops entspricht der von T-Flipflops, da J = K = „1“ anliegt.

Geht man davon aus, dass vor Einsetzen des ersten Taktes an den Ausgängen Q0 bis Q3 „0“ anliegt (Abb. 50.83), wird mit der abfallenden Flanke des 1. Taktes Q0 auf „1“ und mit der abfallenden Flanke des nachfolgenden Taktes Q0 auf „0“ gesetzt. Da der Ausgang Q0 als Takt für das nachfolgende Flipflop wirkt, schaltet Q1 mit der abfallenden Flanke von Q0 auf „1“. Nach dem 7. Impuls liegt am parallelen Ausgang der Schaltung die Binärfolge 0111 unter Berücksichtigung der Wertigkeiten Q0 mit 2^0 = „1“, Q1 mit 2^1 = „1“, Q2 mit 2^2 = „1“ und Q3 mit 2^3 = „0“ an. Nach dem 16. Impuls sind alle Ausgänge „0“, und der Vorgang wiederholt sich. Damit handelt es sich um einen Modulo-16-Zähler.

Der *Rückwärtszähler* unterscheidet sich vom Vorwärtszähler schaltungstechnisch durch die Steuerung der einzelnen Flipflops, wie dies in Abb. 50.84 dargestellt ist. Im Gegensatz zum Vorwärtszähler erhält das jeweils nachfolgende Flipflop seinen Zustand von /Q. Damit kehrt sich die Zählrichtung entsprechend um.

Abb. 50.85 zeigt eine mögliche Umschaltung für die Zählrichtung.

Die asynchrone Schaltung hat den Nachteil, dass falsche Ausgangszustände bei höheren Frequenzen auftreten, weil hier Verzögerungszeiten eine Rolle spielen. Beispielsweise können beim Übergang von 3 nach 4 noch die Zustände 2 und 0 (Abb. 50.86) oder beim Übergang von 7 nach 8 die Zustände 6, 4 und 0 auftreten.

50.6.3.2.6 Asynchroner BCD-Vorwärtszähler

Der asynchrone Dualzähler (Abb. 50.87) kann als BCD-Zähler arbeiten, wenn nach der 9 die Rückstellung auf 0 erfolgt.

Dies erreicht man durch eine Auswertschaltung. Der Zählerstand 10 wird zum Rückstellen benutzt (Abb. 50.88). Da die RESET-Schaltung (hiermit werden alle Flipflop-Ausgänge auf „0“ gesetzt) Low-aktiv ist, genügt ein NAND mit Anschluss

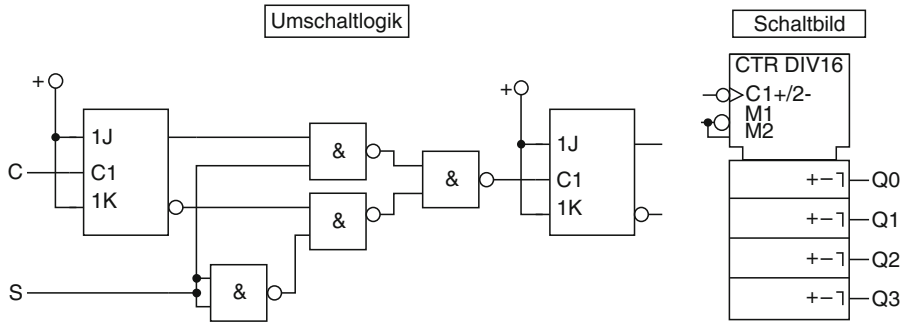


Abb. 50.85 Schaltung zur Zählerumkehr

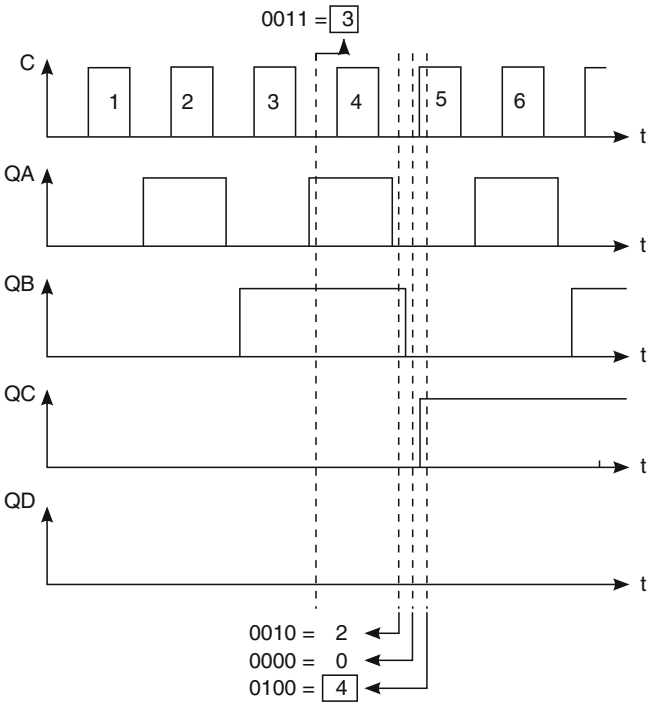


Abb. 50.86 Fehler durch Signallaufzeiten

an die Ausgangswertigkeiten 2^3 = Q3 und 2^1 = Q1. Um ein korrektes Rücksetzen zu erreichen, ist in der Praxis eine Verzögerung notwendig. Hier wird das durch die Laufzeit zweier Gatter realisiert.

50.6.3.2.7 Synchroner Dual-Vorwärts-1-Zähler

Bei den synchronen Zählern werden die Takteingänge aller FFs *gleichzeitig* (synchron) angesteuert, und alle Ausgangszustände ändern sich ebenfalls gleichzeitig. Sie sind daher gegenüber Asynchronzählern schneller. Bei der Realisierung mit JK-Flipflops werden die Vorbereitungseingänge entsprechend geschaltet. Hilfreich ist es hierzu, die Zustände der Flipflops vor und nach dem Taktimpuls zu betrachten und jeweils das Flipflop an der Stelle zu kennzeichnen, wo eine Umschaltung erfolgen muss. Mit dieser Überlegung kann dann durch Verknüpfung entschieden werden, ob J = K = „1“, das heißt das Flipflop durch den Eingangstakt schalten soll. Abb. 50.89 zeigt eine Tabelle mit der entsprechenden Kennzeichnung durch Pfeile.

Da das FF1 hiernach mit jedem Takt umschalten muss, wird J = K auf „1“ gelegt. Das FF2 muss jedesmal schalten, wenn auch Q0 schaltet. Alle Gleichungen sind nachfolgend aufgeführt. Das schaltungstechnische Umsetzen der nachfolgenden Gleichungen führt zur Schaltung Abb. 50.90.

Gleichungen:

für FF1: J = K = 1
für FF2: J = K = Q0
für FF3: J = K = Q0 Q1
für FF4: J = K = Q0 Q1 Q2

50.6.3.2.8 Zähler für mehrere Decaden

Sollen mehrere BCD-Zähler zusammengeschaltet werden, kann das asynchron oder synchron vorgenommen werden. In der Praxis bevorzugt man die synchrone Zusammenschaltung, weil hier keine Zwischenzustände auftreten können. Im Beispiel

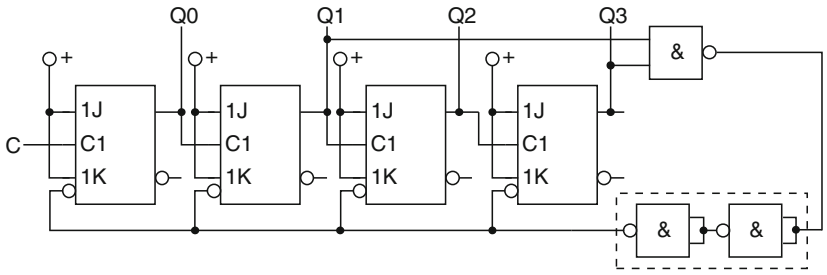


Abb. 50.87 Asynchroner BCD-Vorwärtszähler

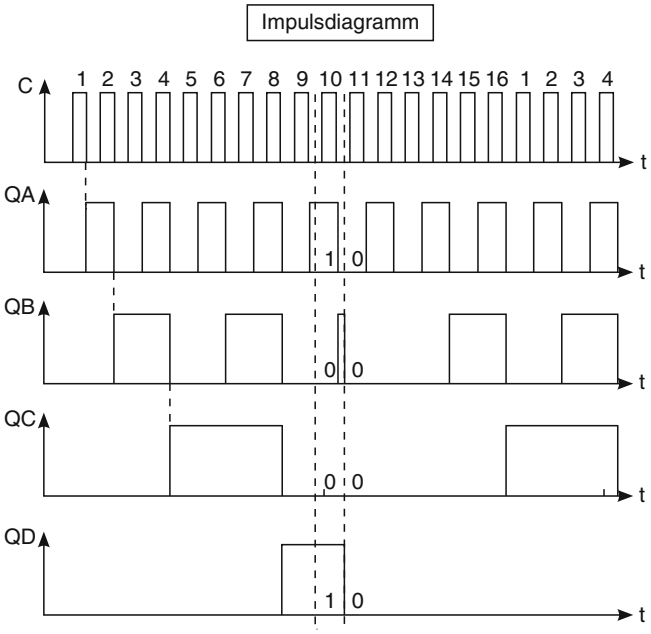


Abb. 50.88 Impulsdiagramm des BCD-Zählers

Q3	Q3+1	Q2	Q2+1	Q1	Q1+1	Q0	Q0+1
0	0	0	0	0	0	0	1
0	0	0	0	0	1	1	0
0	0	0	0	1	1	0	1
0	0	0	1	1	0	1	0
0	0	1	1	0	0	0	1
0	0	1	1	0	1	1	0
0	0	1	1	1	1	0	1
0	0	1	0	1	0	1	0
1	1	0	0	0	0	0	1
1	1	0	0	0	1	1	0
1	1	0	0	1	1	0	1
1	1	0	1	1	0	1	0
1	1	1	1	0	0	0	1
1	1	1	1	0	1	1	0
1	1	1	1	1	1	0	1
1	1	1	0	1	0	1	0

Abb. 50.89 Tabelle zur Realisierung des Synchronzählers

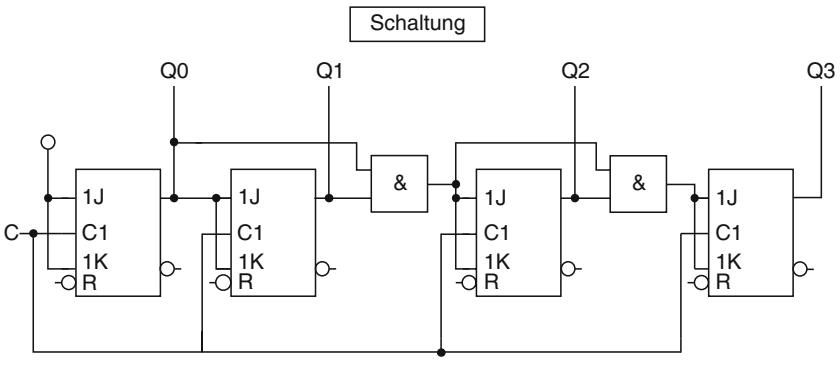


Abb. 50.90 Synchronzähler

Abb. 50.91 ist einmal die asynchrone und die synchrone Zusammenschaltung dargestellt. Bei der synchronen Zusammenschaltung werden beide Zählerbausteine gleichzeitig vom Takt angesteuert. Mit G1 = „0“ ist der Zähler jeweils gesperrt.

50.6.4 Sonderschaltungen

50.6.4.1 Monoflops

Monostabile Kippstufen oder Monoflops besitzen eine stabile Lage und können durch einen Steuerimpuls für eine bestimmte Zeit ti in eine instabile Lage versetzt werden und an einem Ausgang einen Impuls liefern. Hierbei kann je nach Schaltungsdimensionierung eine Impulsverlängerung oder Impulskürzung erfolgen (Abb. 50.92). Der Ausgangsimpuls ist nur von den Daten des Monoflops und nicht von der Dauer des Steuerimpulses abhängig.

In Abb. 50.93 sind zwei Monoflops als Beispiel mit unterschiedlichen Grundgattern aufgeführt. Das NAND-Monoflop verlässt seine stabile Lage bei einer kurzen Ansteuerung mit einem L-Pegel. Dieser erzeugt einen H-Impuls am Ausgang G1 und – weil C zunächst als Kurzschluss wirkt (der Kondensator ist nicht geladen) – damit einen L-Pegel am Ausgang G2. Nach Öffnen des Tasters wird dieser Zustand solange beibehalten, bis der Kondensator C sich aufgeladen und am Eingang von G2 L-Pegel anliegt. Damit erhält der Ausgang G2 H-Pegel und der Ausgang G1 wieder L-Pegel. Die stabile Lage ist wieder erreicht. Die Zeitdauer des L-Pegels am Ausgang G2 entspricht der Zeit ti und hängt von der Zeitkonstanten $R \cdot C$ ab. Die stabile Lage des NOR-Monoflops wird durch einen kurzen H-Impuls unterbrochen. Da der Kondensator C zuvor nicht geladen war (der Ausgang G1 führte H-Pegel und der Eingang G2 über R1 ebenfalls), ergibt der H-Impuls am Eingang ein L am Ausgang G1 und damit ebenfalls am Eingang G2. Der H-Impuls im Rückkopplungszweig hält diesen Zustand solange aufrecht, bis der Kondensator C über R1 aufgeladen die Schaltspannung erreicht hat und am Ausgang wieder einen L-Pegel erzeugt. Die Ansteuerung von Monoflops erfolgt über Differenzierglieder. Die Abb. 50.93 zeigt zwei Beispiele, die je nach gewünschter Steuerflanke gewählt werden können.

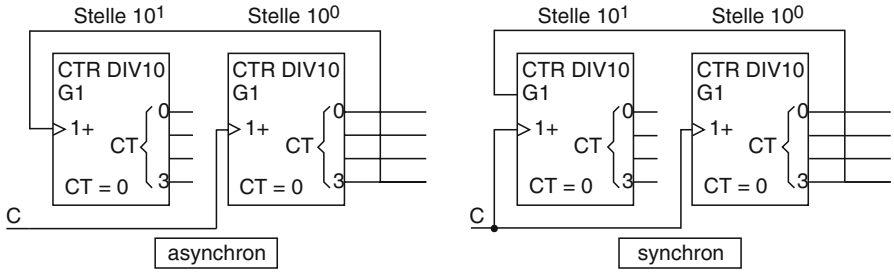


Abb. 50.91 Asynchroner und synchroner Zähler für zwei Stellen

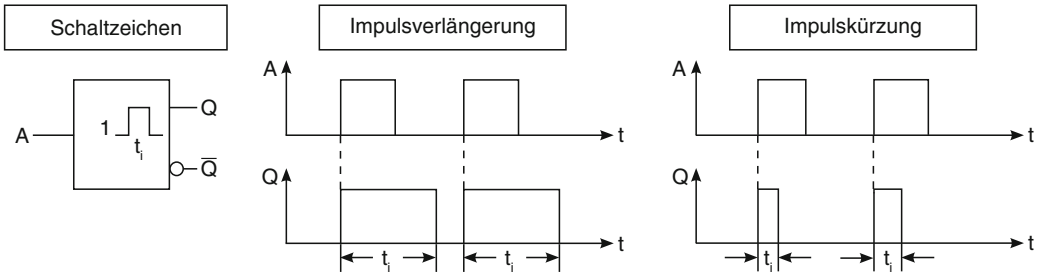


Abb. 50.92 Monoflop und Betriebsarten

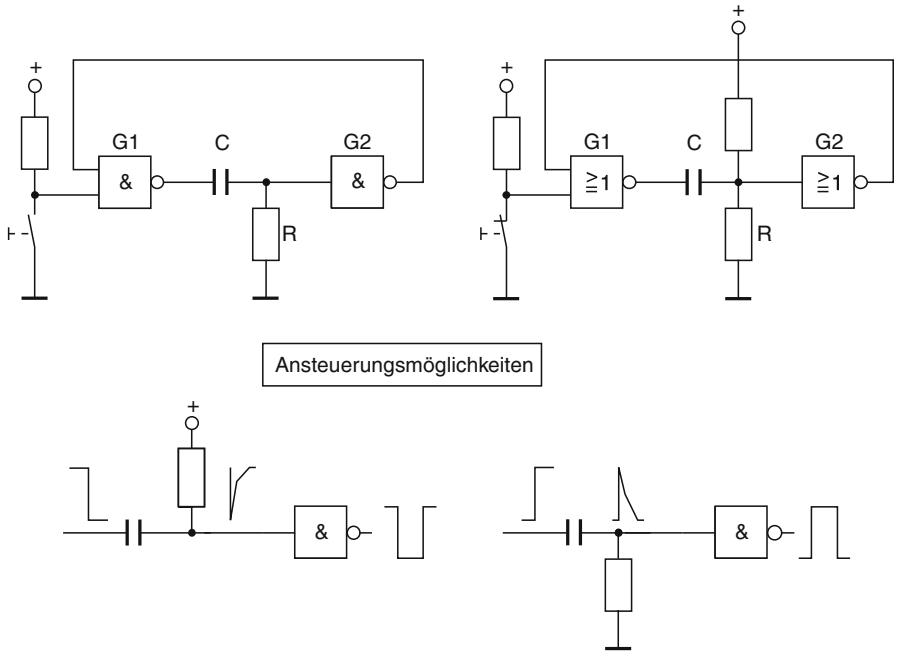


Abb. 50.93 Monoflops aus Grundgattern und Ansteuermöglichkeiten

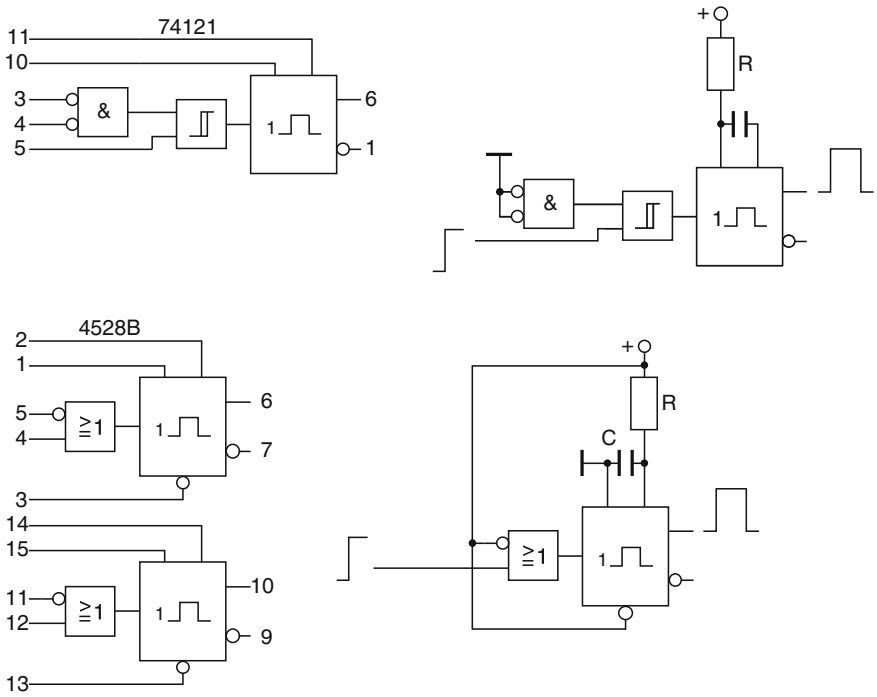


Abb. 50.94 Integrierte Monoflops

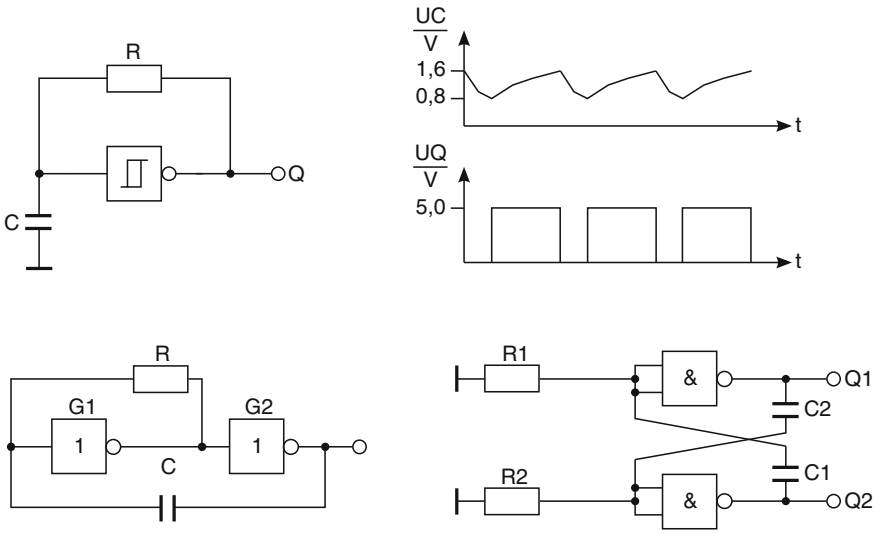


Abb. 50.95 Generatorschaltungen

In der Praxis setzt man integrierte Bausteine ein (Abb. 50.94). Die entsprechende Zeit ti wird durch externe Beschaltung mit einem Widerstand R und einem Kondensator C festgelegt, wobei die Werte den Datenbüchern der Hersteller entnommen werden können. Durch die vorhandenen Eingänge sind unterschiedliche Triggermöglichkeiten vorhanden. Man unterscheidet *nichtretriggerbare* und *retriggerbare* Monoflops. Bei der nichtretriggerbaren Stufe ergeben sich bei erneuter Ansteuerung durch einen Impuls während der aktiven Zeit ti keine Ausgangsänderungen. Der Baustein 4528B besitzt zwei retriggerbare und rücksetzbare Monoflops.

50.6.4.2 Astabile Kippstufen

Astabile Kippstufen sind Rechteckgeneratoren und Bestandteil vieler digitaler Schaltungen. Sie dienen als Zeitbasis digitaler Messgeräte, als Taktoszillator für ein Controller- oder Prozessorsystem oder werden als Baudratengenerator eingesetzt. Sie können mit Hilfe von Gatterschaltungen aufgebaut werden. Abb. 50.95 zeigt eine Schaltung mit einem Schmitt-Trigger, dem zugehörenden Impulsdiagramm und zwei weitere mögliche Schaltungen. Liegt am Ausgang L-Pegel, kann sich der Kondensator über R entladen. Bei Unterschreiten der Einschaltschwelle

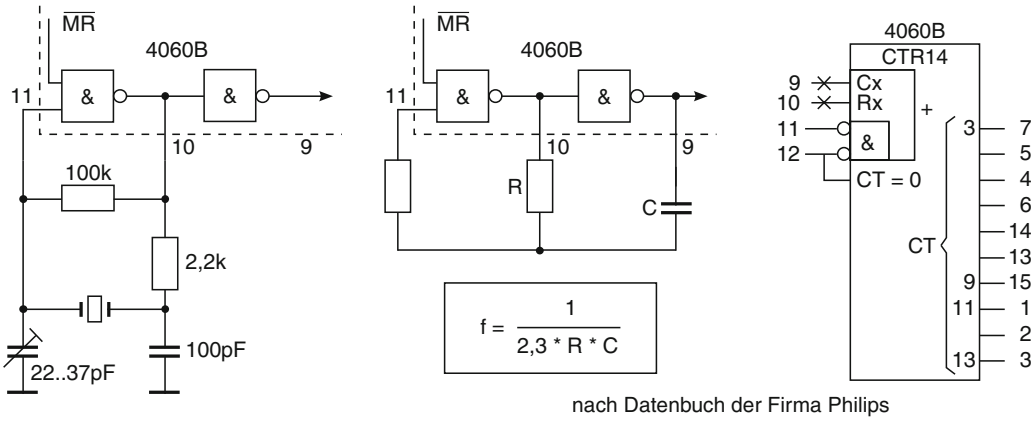


Abb. 50.96 Standardbaustein mit Taktgenerator

am Eingang des Schmitt-Trigger nimmt der Ausgang Q H-Pegel an, und C kann über R wieder aufgeladen werden, bis die Ausschaltschwelle erreicht ist, um an Q wieder einen L-Pegel zu erzeugen. Die Frequenz der Ausgangsspannung ist von der Zeitkonstante tau = R · C abhängig.

Zum Einsatz kommen auch hier integrierte Schaltungen. Als Beispiel dient der Baustein 4060B. Er besteht aus einem Generator und einem sich anschließenden Teiler. Zu seiner Funktion brauchen nur wenige externe Bauelemente zugefügt werden (Abb. 50.96). Seine Oszillatorfrequenz lässt sich bis 214 teilen. Dies hängt vom gewählten Ausgang ab.

Integrierte Schaltkreise der Digitaltechnik

51

Heribert Gierens



51.1 Allgemeines 634
51.2 Umgang mit integrierten Schaltungen 634
51.3 Daten und Begriffe der Logikschaltungen 634
51.4 TTL-Familie 637
51.5 Emittergekoppelte Logik 641
51.6 Integrierte MOS-Schaltungen 642
51.7 Interfaceschaltungen 646
51.8 Anwendungsspezifische integrierte Schaltungen 647
51.9 Gehäuse 659

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_51

51.1 Allgemeines

Integrierte Schaltkreise (Integrated Circuit, IC) bestehen aus einer großen Anzahl von Transistorfunktionen, die auf einem Halbleiterkristall (Chip) integriert sind. Der Chip ist in einem Kunststoff- oder Keramikgehäuse eingebaut. Über Anschlüsse (Pin's) erfolgt die Verbindung zur Schaltung. Mit Hilfe moderner Herstellungsverfahren wird die Integrationsdichte dieser Schaltungen immer größer und die entsprechenden Chips kleiner, leistungsfähiger und kostengünstiger. IC's lassen sich nach Kriterien wie Integrationsdichte, Funktion und Technologie einordnen. Bei der Integrationsdichte wird beispielsweise nach SSI (Small Scale Integration), MSI (Medium Scale Integration), LSI (Large Scale Integration) und VLSI (Very Large Scale Integration) unterschieden. Die Grenzen zwischen den Gruppen sind fließend und nicht einheitlich festgelegt.

Digitalschaltungen (Logik- und Speicherschaltungen) werden in *bipolarer* oder *unipolarer Technik* hergestellt. Bei den bipolaren Logikschaltungen wird nach TTL (Transistor-Transistor-Logic) und ECL (Emitter Coupled Logic) unterschieden, bei den unipolaren Logikschaltungen nach MOS (Metal Oxide Silicon) und CMOS-Typen (Complementary MOS). Sind beide Technologien auf einem Chip realisiert, spricht man von *BiCMOS*-Technik.

Außer den zuvor genannten Standardschaltungen bieten die Hersteller anwendungsspezifische Schaltungen (ASIC = Application Specific IC) an, die sich nach vollkundenspezifischen (full custom) und halbkundenspezifischen (semi custom) Schaltungen unterscheiden. Hierunter versteht man vor allem Gate Arrays, Cell Arrays und PLD's (PLD = Programmable Logic Device). Unter einem Array versteht man allgemein eine Matrix, die verknüpft werden kann.

Abb. 51.1 zeigt eine Übersicht digitaler integrierter Schaltkreise. Speicher und Prozessoren sind in gesonderten Kapiteln aufgeführt.

51.2 Umgang mit integrierten Schaltungen

IC's sind elektrostatisch gefährdete Bauelemente (EGB). Sie müssen daher geschützt werden. Der Hersteller versieht diese Bauteile mit einem besonderen Kennzeichen (Hand als Piktogramm). Bei MOS-Schaltungen ist eine Schutzschaltung häufig integriert. Trotzdem kann es zur Zerstörung der Bauteile bei Berührung der Bausteinanschlüsse durch elektrostatische Entladungen kommen, wenn nicht besondere Regeln beim Umgang mit diesen Bauelementen beachtet werden. Hierzu zählen Schutzmaßnahmen am Arbeitsplatz wie hochohmig leitende Fußbodenbeläge und ein Handgelenkband mit Erdung über einen Ableitwiderstand. Die Werkzeuge müssen ebenfalls aus einem antistatischen Material bestehen bzw. leitend sein. Alle Geräte für Reparaturzwecke sind zu erden. Dies gilt beispielsweise auch für einen Lötkolben, dessen Erdung über einen Widerstand durchgeführt wird. Bei hoher Arbeitsfrequenz der Bauelemente gelangen beim Umschalten Stromspitzen auf die Betriebsspannungsleitung. Bei der Schaltungsentwicklung muss daher gesorgt werden, dass alle Spannungsversorgungen direkt am Versorgungsanschluss des IC's durch einen Kondensator abgeblockt sind. Um die Einkopplung von Fremdsignalen zu vermeiden, ist es weiterhin notwendig, dass keine langen Leiterbahnen bei der Entwicklung eines Layouts verlegt werden.

51.3 Daten und Begriffe der Logikschaltungen

51.3.1 Grenz- und Kenndaten

Grenzdaten sind absolute Grenzwerte. Bei Überschreitung dieser Werte kann die integrierte Schaltung zerstört werden. Kenndaten sind Mittelwerte, die durch Angabe eines garantierten Streubereiches (worst case) ergänzt wurden.

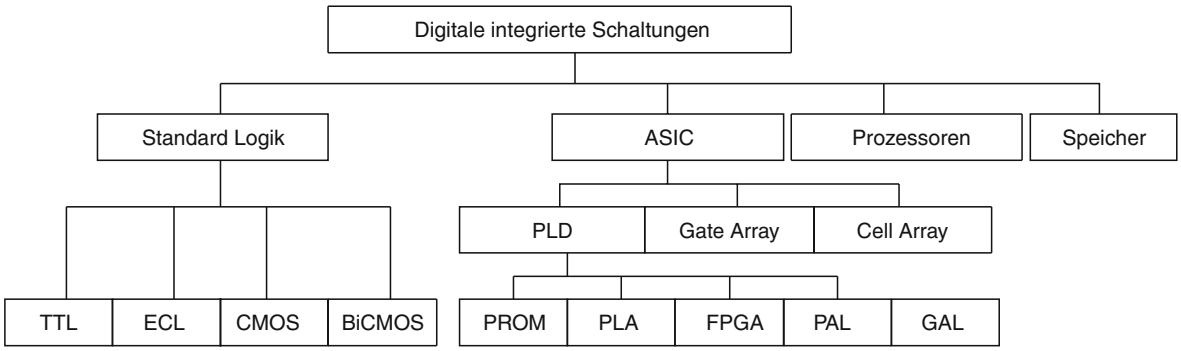


Abb. 51.1 Übersicht: Integrierte digitale Schaltungen

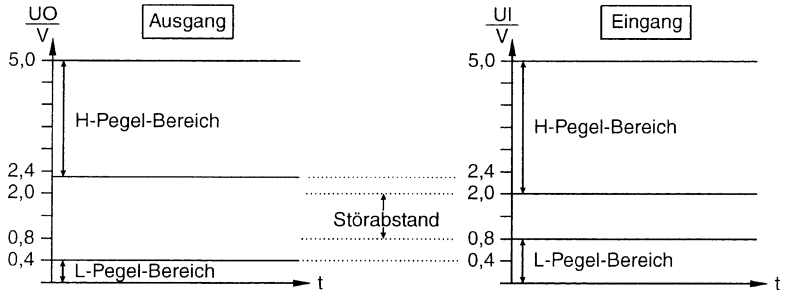


Abb. 51.2 Pegelwerte und Störabstand

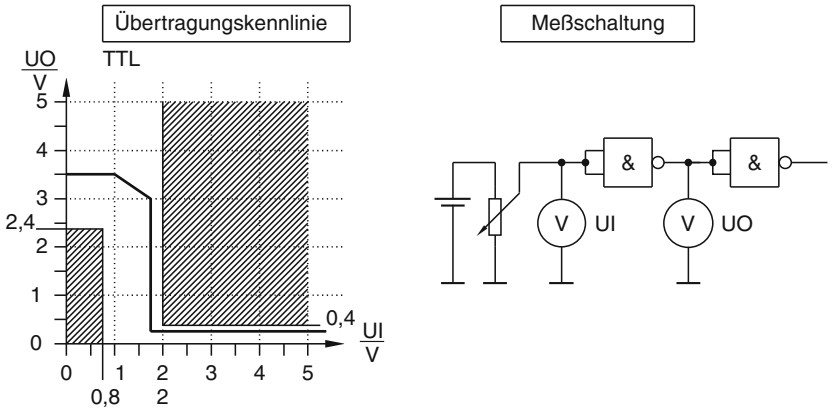


Abb. 51.3 Übertragungskennlinie und Messschaltung

51.3.2 Pegel

H- und L-Pegel sind bestimmten Spannungsbereichen zugeordnet, die eingehalten werden müssen, um die logische Funktion zu gewährleisten. Für Ein- und Ausgang einer Schaltung gelten unterschiedliche Pegelwerte. Abb. 51.2 veranschaulicht dies am Beispiel eines Standard-TTL-Bausteins. Die Schaltung erkennt „L“ von 0 ... 0,8 V (U_{IL}) und „H“ von 2 ... 5 V (U_{IH}). Liefern muss die Schaltung für „L“ 0 ... 0,4 V (U_{OL}) und für „H“ 2,4 ... 5 V (U_{QH}) (siehe auch Tabelle II-1). Der Bereich zwischen „L“ und „H“ ist nicht definiert. Man spricht hier auch von dem „verbotenen Bereich“.

In Datenbüchern wird häufig die Übergangskennlinie, also die Abhängigkeit der Spannung U_O (U Output) von U_I (U Input), angegeben (Abb. 51.3). Die Schraffierung kennzeichnet hier den verbotenen Bereich. Mit der angegebenen Messschaltung kann diese Kennlinie statisch aufgenommen werden.

51.3.3 Störsicherheit

Unter *statischer* Störsicherheit versteht man die Sicherheit gegen eingekoppelte Störspannungen, deren Einwirkung zeitlich größer als die Gatterlaufzeit ist. Man spricht hier auch vom *Störabstand*. Störspannungen bis zur Größe des Störspannungsabstandes haben keinen Einfluss auf die logische Funktion der Schaltung. Bei Standard-TTL beträgt der statische Störspannungsabstand 1,2 V (siehe Abb. 51.2). Unter *dynamischer* Störsicherheit versteht man die Sicherheit gegen Störspannungen mit kürzeren Zeiten als die Gatterlaufzeit. Die Werte des dynamischen Störabstandes sind von der Dauer des Störimpulses, der Temperatur und der jeweiligen Schaltung abhängig.

51.3.4 Lasteinheit

Um die Pegelwerte einhalten zu können, darf das Gatter nur vorgegebenen Belastungen ausgesetzt sein. Jeder Eingang eines Gatters belastet das vorherige Gatter. Bei Standard-TTL fließt bei L ein maximaler Strom von 1,6 mA aus dem Gatter ($I_{IL} = -1{,}6$ mA). Diese Stromgröße wird mit Fan IN = 1 (Eingangslastfaktor FI) bezeichnet. Bei H fließt ein max. Strom von 40 µA in das Gatter ($I_{IH} = 40$ µA). Die Stromrichtung zwischen zwei Gattern bei unterschiedlichen Pegeln ist in Abb. 51.4 deutlich hervorgehoben. Häufig liegen die gemessenen Stromwerte weit unter den maximal zulässigen Werten. Da $I_{QL\,max}$ 16 mA beträgt, dürfen unter worst-case-Bedingungen höchstens 10 Gattereingänge einen Ausgang belasten. Die Ausgangsbelastung (Ausgangslastfaktor) nennt man *Fan Out*. Hiermit wird ausgesagt, mit wieviel Lasteinheiten ein Ausgang belastet wer-

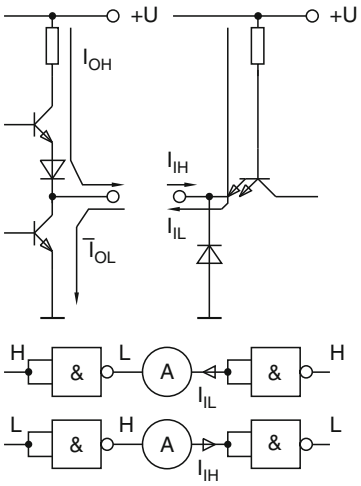


Abb. 51.4 Stromrichtung zwischen den Gattern und Kennzeichnung der Ströme

den darf. Für Standard-TTL gilt ein Fan out von 10 und für CMOS ein Fan out von 50. Sollen mehr als 10 Eingänge an einen Standard-TTL-Ausgang angeschlossen werden, müssen Treiberstufen eingesetzt werden.

51.3.5 Temperaturbereich

Bei der Angabe des Temperaturbereichs wird zwischen *Lagerungstemperatur* und zulässiger *Umgebungstemperatur im Betrieb* unterschieden. So liegt beispielsweise die Lagerungstemperatur bei IC's mit Kunststoffgehäusen zwischen 55 und 125 °C, die zulässige Umgebungstemperatur im Standardbereich bei der 74LS-Serie von 0 bis 70 °C und bei den MOS-Typen der 4000B-Serie von −40 bis 85 °C.

51.3.6 Gatterlaufzeit

Ändert man den Eingangspegel einer Gatterschaltung, z. B. eines Inverter, von L nach H, so dauert es eine gewisse Zeit, bis der Ausgang auf die Eingangsänderung reagiert. Das Zeitverhalten gibt die Gatterlaufzeit (tpd = propagation delay time) wieder, wobei die Angabe sich auf die Differenz der Ein- und Ausgangsflanke in 50 %iger Höhe bezieht. Bei Aufbau von Schaltungen mit *Mehrebenen* können durch unterschiedliche Laufzeiten fehlerhafte und ungewünschte Verhaltensweisen wie kurzzeitige Impulse (Glitches) am Ausgang auftreten. Im Abb. 51.5 durchlaufen die Signale A und B drei Gatter, während C zwei und D

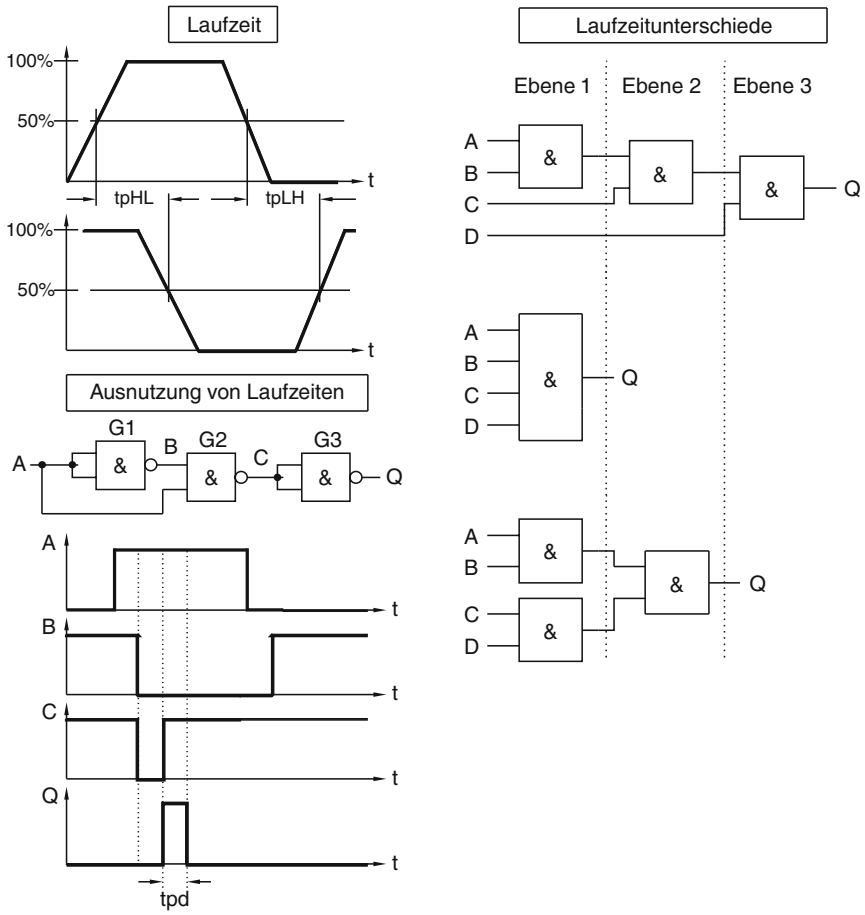


Abb. 51.5 Gatterlaufzeiten

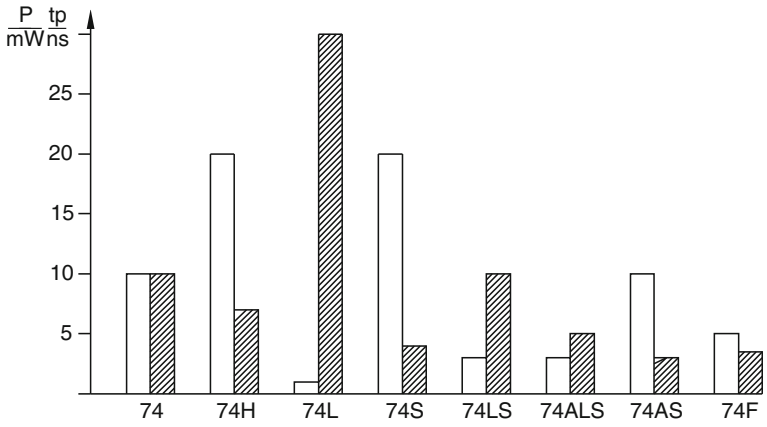


Abb. 51.6 Verlustleistung und Gatterlaufzeit (Prinzipdarstellung, bausteinabhängig)

nur ein Gatter durchläuft. Durch Umorganisation der Schaltung kann erreicht werden, dass die Eingangssignale gleiche Zeiten durchlaufen. Oft wird auch statt der Signallaufzeit die maximale Schaltfrequenz angegeben.

Für kurze Impulse kann die Gatterlaufzeit, wie dargestellt ist, ausgenutzt werden. Das Zeitliniendiagramm beschreibt die Pegel an den einzelnen Gattern. Der Ausgang B ist zunächst „H", da der Eingang A L-Pegel führt. Nimmt A H-Pegel an, schaltet B nach der Verzögerungszeit auf L. C schaltet nach L und der nachfolgende Inverter G3 gibt den Impuls mit der Impulszeit einer Verzögerungszeit heraus. Wünscht man Impulszeiten mehrerer Gatterlaufzeiten, sind zusätzlich jeweils zwei Inverter zwischen G1 und G2 einzusetzen.

51.3.7 Verlustleistung

Bei digitalen Schaltungen wird häufig die Verlustleistung pro Gatter angegeben. Hierbei handelt es sich um einen Mittelwert, der davon ausgeht, dass der Gatterausgang abwechselnd L- und H-Pegel führt. Bei den einzelnen Familien gibt es erhebliche Unterschiede, wie auch dem Abb. 51.6 entnommen werden kann. Vergleicht man ferner die Stromaufnahme eines Standard-TTL-Bausteins, sind auch deutliche Unterschiede der Stromwerte zu erkennen, wenn alle Ausgänge einmal L- und einmal H-Pegel führen. Die größte Verlustleistung tritt im dynamischen Betrieb bei hohen Frequenzen auf. Der Schaltungsentwickler sollte beim Einsatz digitaler Schaltungen darauf achten, dass die Verlustleistung der eingesetzten Bausteine möglichst gering ist. Hierdurch können die Netzteile klein und kompakt dimensioniert werden, und die geringere Wärmeentwicklung der Bauelemente erspart aufwendige Kühlmaßnahmen. Ein Beispiel hierfür ist die 3,3 V-Technologie. In diesem Zusammenhang ist auch der Begriff *Speed-Power-Product* zu nennen, der das Produkt aus Geschwindigkeit und Leistungsaufnahme eines Gatters in Verbindung mit Energiebilanzen angibt.

51.4 TTL-Familie

51.4.1 Eigenschaften und Kenndaten

Abb. 51.6 zeigt die Mitglieder-TTL-Familie im Vergleich von Gatterlaufzeit und Verlustleistung. In Tab. 51.1 sind ihre Kenndaten aufgelistet.

Die Bausteine der einzelnen Untergruppen sind *PIN-kompatibel* (gleiche Funktion und Anschlussbelegung) und besitzen als gemeinsame Kenndaten die Betriebsspannung, Pegelwerte und den Temperaturbereich. Unterschiede bestehen vor allem in der Gatterlaufzeit und der Verlustleistung. Bei der Low-Power- und High-Speed-Untergruppe sind die internen Widerstandswerte gegenüber Standard verändert, um die Leistungsaufnahme geringer zu machen (Low) und die Geschwindigkeit zu erhöhen (High). Die beiden Eigenschaften verlaufen allerdings gegensätzlich.

Weitverbreitet sind die Low-Power-Schottky-Typen mit ihrer gegenüber Standard-TTL geringeren Leistungsaufnahme und

Tab. 51.1 Kennwerte ausgewählter TTL-Serien

Kennwerte	74	74S	74LS	74ALS	74F
Betriebsspannung (V) ±5 %	5	5	5	5	5
UIL (V) <	0,8	0,8	0,8	0,8	0,8
UIH (V) >	2	2	2	2	2
UOL (V) <	0,4	0,5	0,5	0,5	0,5
UOH (V) >	2,4	2,7	2,7	U-2V	2,7
IOL (mA) max	16	8	20	30	24
tp (ns)	10	9	3	5	3
P (mW)	10	2	20	1,2	4

nach Unterlagen der Firma Philips
Abkürzungen: UIL Eingangsspannung für L-Pegel; UIH Eingangsspannung für H-Pegel; UOL Ausgangsspannung für L-Pegel; UOH Ausgangsspannung für H-Pegel; IOL Ausgangsstrom bei L-Pegel

etwa gleicher Geschwindigkeit. Abb. 51.7 zeigt einige LS-Bausteine aus dem umfangreichen Herstellerangebot mit ihren Anschlussbelegungen. Advanced-Schottky und Advanced-Low-Power-Schottky sind verbesserte Nachfolgetypen sowie die FAST-Serie, die gegenüber der LS-Serie ihre Gatterdurchlaufzeit erhöhen konnte.

51.4.2 Standard-TTL

Bei den Standardbausteinen der TTL-Familien haben sich folgende Funktionseinheiten herausgebildet:

- Grundgatter wie z. B. NICHT (Inverter), UND, ODER, NOR und XOR
- Flipflops wie z. B. RS-Flipflop, D-Flipflop und JK-Flipflop
- Multiplexer und Datenselektoren
- Demultiplexer und Decoder
- Rechenschaltungen wie z. B. Addierer
- Komparatoren
- Zähler und Frequenzteiler
- Schieberegister
- Treiber
- Sonderschaltungen wie z. B. Schmitt-Trigger und Zeitglieder

Schaltungen der Standardserie (Bezeichnung 74xx) bestehen aus *bipolaren Transistoren*, die als Schalter arbeiten. Sie sind so dimensioniert, dass sie beim Durchschalten *übersteuert* werden und so voll in den Sättigungsbereich gelangen (gesättigte Logik). Die TTL-Familie ist die Weiterentwicklung der DTL (Dioden-Transistor-Logik) und RTL (Widerstands-Transistor-Logik). Drei typische Vertreter der Standardserie dienen als Beispiel.

Beispiel 1

Baustein IC 7400

Abb. 51.8 zeigt die Innenschaltung des 14poligen IC's und den Aufbau eines NAND-Gatters. Hier erkennt man die Merkmale der weitverbreiteten Logik mit einem *Multiemitter-Transistor* im Eingang (V1) und einer Gegentaktendstufe (Totem-Pol-Stufe) am Ausgang (V3, V4). Von der Funktion her lässt sich diese Schaltung als Weiterentwicklung der RTL (Widerstands-Transistor-Logik) und der DTL (Dioden-Transistor-Logik) ansehen.

Wenn jetzt ein oder beide Eingänge der Schaltung auf Masse liegen (L-Pegel), kann der Transistor V2 nicht durchschalten. Damit ist auch V4 gesperrt, und V3 schaltet durch und erzeugt einen H-Pegel am Ausgang. Liegen beide Eingänge auf +5 V (H-Pegel), schaltet V2 durch, und am Ausgang entsteht ein L-Pegel, da auch V3 durchschaltet. Transistor V4 ist dann gesperrt.

An dieser Schaltung erkennt man auch die Auswirkung offener *Eingänge* auf den Ausgang der Schaltung. Werden beide Eingänge offen gelassen, fließt über die Basis-Kollektor-Diode des Multiemitter-Transistors Strom in den Transistor V2 und schaltet diesen und auch V4 durch. Offene Eingänge wirken wie anliegende H-Pegel. Man legt daher bei NAND- und UND-Gattern die nichtbenötigten Eingänge auf +5 V und bei ODER- und NOR-Gattern auf 0 V.

Müssen Ausgänge verschiedener Gatter logisch miteinander verknüpft werden, ist dies mit *Totem-pole*-Schaltungen nicht möglich. Gatter mit solchen Gegentaktendstufen dürfen an ihren Ausgängen nicht parallel geschaltet werden, da ein Kurzschluss entstehen kann, wenn ein Gatterausgang „L" und der andere „H" liefert. Abhilfe schaffen Gatter mit offenem Kollektor (Open Collector, o.C., siehe Beispiel 2). ◄

Beispiel 2

Baustein IC 7401

Im Abb. 51.9 sind die Innenschaltung des Baustein 7401 und der Aufbau eines Gatters dargestellt. Ein Lastwiderstand RL muss extern eingebaut werden, da der Ausgang ohne Lastwiderstand keinen Pegel ausgeben kann.

Stufen mit offenem Kollektor (o.C.) lassen am RL Betriebsspannungen von 15 oder 30 V zu. TTL-Treiber können damit Kleinverbraucher direkt ansteuern. Gatter mit o.C. dürfen mit ihren Ausgängen parallel geschaltet werden. Der Ausgang erreicht nur dann H-Pegel, wenn alle Eingänge auf H-Pegel liegen; ansonsten liegt L-Pegel am Ausgang an. Man spricht von einer *verdrahteten* UND-Schaltung (Wired AND). Wird der Ausgang mit einem nachfolgenden Inverter verbunden, ergibt sich eine verdrahtete ODER-Verknüpfung (Wired OR). Abb. 51.10 zeigt die beiden genannten Schaltungen und verdeutlicht die Funktion anhand der Wertetabelle und der Gleichungen.

Der Wert des Kollektorwiderstandes lässt sich für eigene Bedürfnisse berechnen oder man nutzt für Standardanwendungen die Tabellen der Hersteller (Tab. 51.2). Das Berechnungsbeispiel bezieht sich auf die Schaltung Abb. 51.11.

$$
\begin{aligned}
\mathrm{RL} &> \frac{\mathrm{U_{max}} - \mathrm{UOL}}{\mathrm{IOL_{max}} - N \cdot \mathrm{IIL}} \\
\mathrm{RL} &< \frac{\mathrm{U_{min}} - \mathrm{UOH}}{n \cdot \mathrm{IOH} + N \cdot \mathrm{IIH}} \\
\mathrm{RL} &> \frac{5{,}25\,\mathrm{V} - 0{,}4\,\mathrm{V}}{16\,\mathrm{mA} - 4 \cdot 1{,}6\,\mathrm{mA}} \\
\mathrm{RL} &> 505\,\Omega \\
\mathrm{RL} &< \frac{4{,}75\,\mathrm{V} - 2{,}4\,\mathrm{V}}{2 \cdot 250\,\mu\mathrm{A} + 4 \cdot 40\,\mu\mathrm{A}} \\
\mathrm{RL} &< 3560\,\Omega
\end{aligned}
\qquad (51.1)
$$

◄

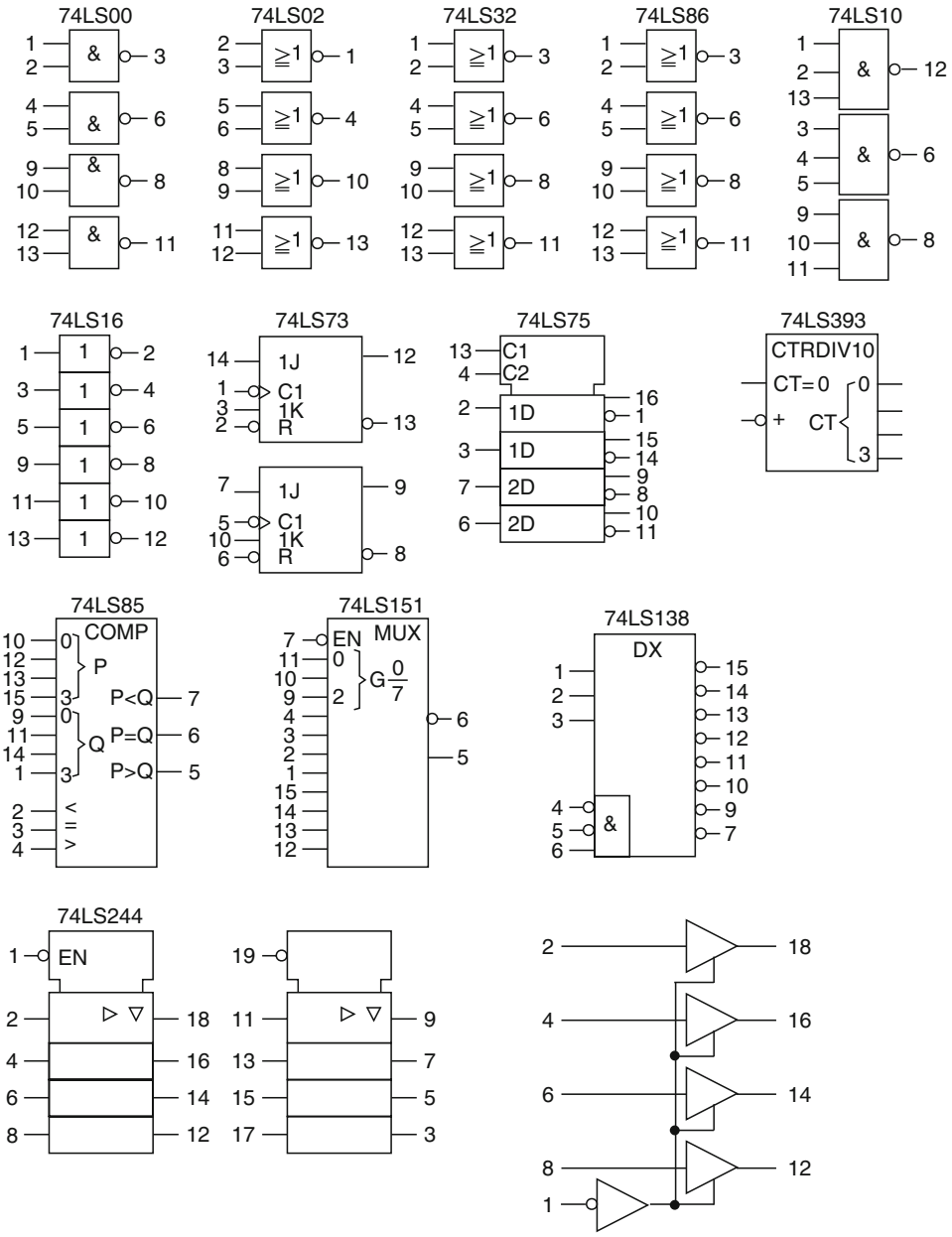


Abb. 51.7 Digitalbausteine mit Anschlussbelegung

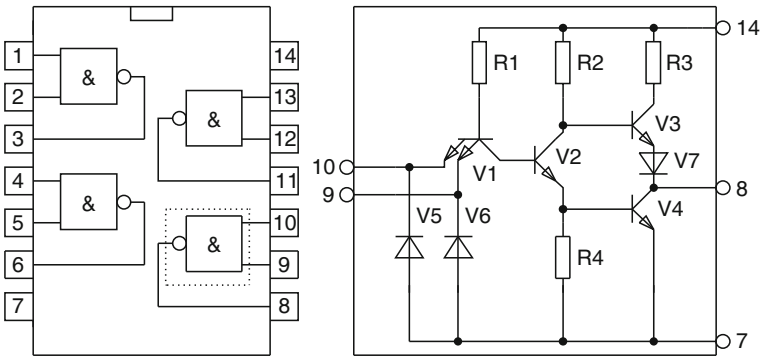


Abb. 51.8 Baustein 7400

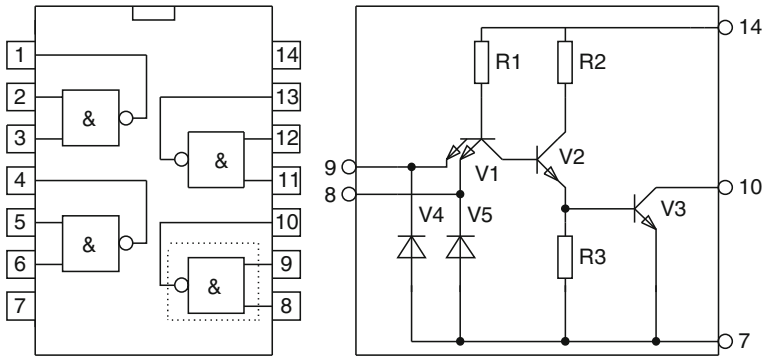


Abb. 51.9 Baustein 7401

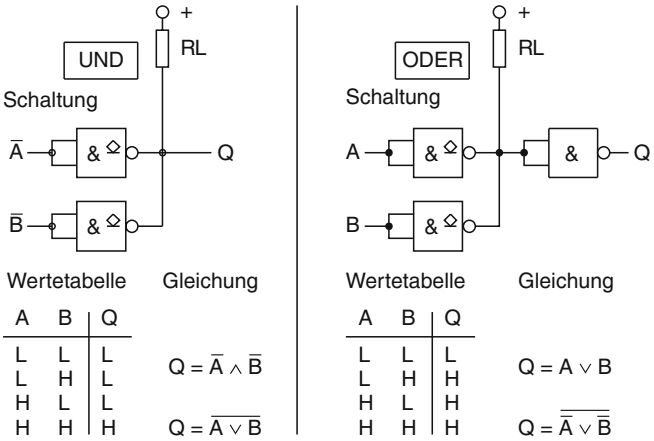


Abb. 51.10 Verdrahtete UND- und ODER-Schaltung

Tab. 51.2 Ermittlung des Kollektorwiderstandes

	RL_{max}							RL_{min}
	N							
n	1	2	3	4	5	6	7	
1	8100	4350	2970	2260	1820	1520	1310	337
2	7120	4050	2830	2170	1760	1480	1280	379
3	6350	3790	2700	2100	1710	1450	1250	433
4	5730	3560	2580	2020	1660	1410	1230	505
5	5220	3350	2470	1950	1620	1380	1200	607
6	4790	3170	2370	1890	1570	1350	1180	758
7	4430	3010	2280	1830	1530	1320	1150	1110
8	4120	2860	2190	1780	–	–	–	1515
9	3850	–	–	–	–	–	–	3030

nach Datenbuch der Firma ITT

Beispiel 3

Baustein IC 7402

Liegt an beiden Eingängen der Schaltung Abb. 51.12 ein L-Pegel, wird V3 und V4 gesperrt und damit auch V6. V5 schaltet durch und erzeugt einen H-Pegel am Ausgang. Bei Anliegen eines H-Pegels an nur einem der Eingänge entsteht am Ausgang ein L-Pegel. Die Schaltung erfüllt damit die NOR-Funktion. ◄

51.4.3 Schaltungen mit 3-state

Tristate-Schaltungen besitzen außer den Pegelausgaben L und H noch eine dritte Möglichkeit. Sie sind in der Lage, mit einem Steuereingang (EN Enable = Freigabe) einen Ausgang hochohmig zu schalten. Damit können diese Schaltungen in parallelen Bussystemen eingesetzt werden. In der Hauptsache handelt es sich bei Einsatz dieser Schaltungen um Leistungstreiber, Richtungstreiber, Zwischenspeicher und Speicherschaltungen. In der Prinzipdarstellung (Abb. 51.13) schaltet ein H-Pegel am Eingang EN die Transistoren V6 und V7 durch und sperrt damit die beiden Endstufentransistoren V3 und V4. Der Ausgang Q besitzt damit keinen definierten H- oder L-Pegel mehr. Er ist praktisch abgeschaltet; man spricht auch davon, dass der Ausgang hochohmig ist.

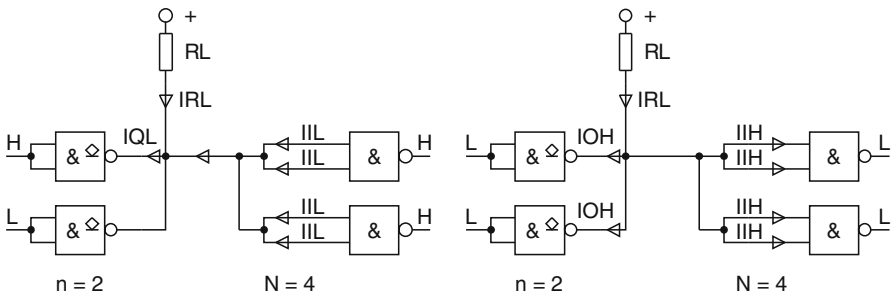


Abb. 51.11 Stromrichtung und -größe bei unterschiedlichen Ansteuerungen

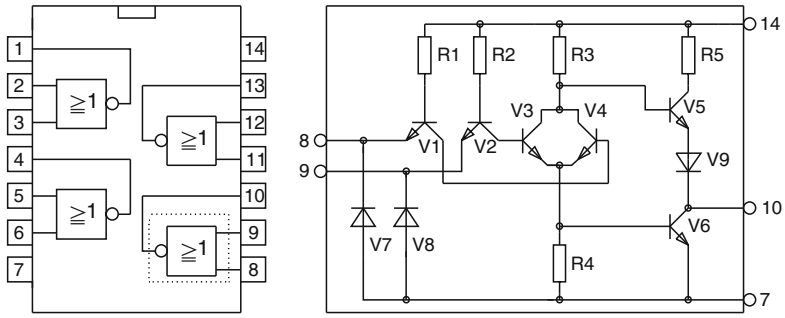


Abb. 51.12 Baustein IC 7402

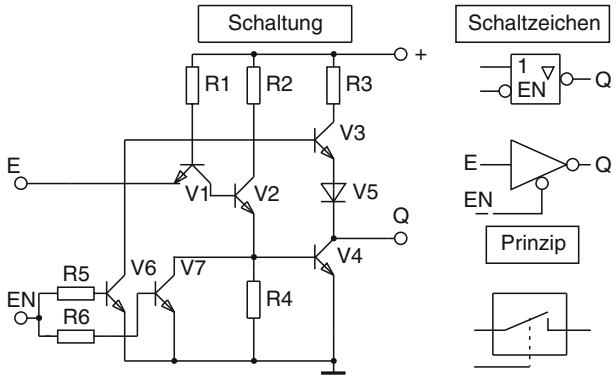


Abb. 51.13 3-state-Gatter

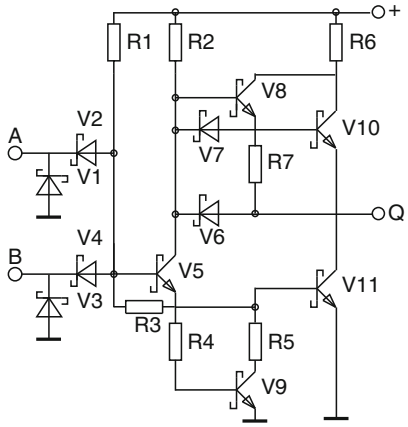


Abb. 51.14 Low-Power-Schottky-Gatter

51.4.4 Schottky-TTL und Low-Power-Schottky

Schottky-Dioden besitzen einen Metall-Halbleiterübergang und Schwellspannungen, die zwischen denen von Germanium- und Silizium-Dioden bei ca. 0,4 V liegen. Ein Schottky-TTL-Gatter besitzt wesentlich kürzere Schaltzeiten als ein Standardgatter; es hat aber den Nachteil, dass die Leistungsaufnahme sehr hoch (20 mW) ist. Weitverbreitet sind heute die Bausteine der LS-Serie (Low-Power-Schottky) wegen der geringeren Leistungsaufnahme von etwa 2 mW pro Gatter und des höheren Fan out gegenüber Standard Typen. Die Schaltzeiten dieser Serie liegen bei 9 ns. Abb. 51.14 zeigt ein Gatter der Schottky-TTL-Serie.

51.5 Emittergekoppelte Logik

Kennzeichen der emittergekoppelten Logik (Emitter Coupled Logic) ist der Differenzverstärker im Eingang und die Auskopplung über den Emitter (Abb. 51.15). Aufgrund des technologischen Aufbaus können ECL-Schaltungen auch zwei Ausgänge führen, deren Pegel invers zueinander sind. Man spricht dann von einem AND/NAND-Gatter oder OR/NOR-Gatter. Bausteine der ECL-Familie sind bedeutend schneller als andere Bausteinfamilien. Sie erreicht Schaltzeiten von kleiner 1 ns. Der Grund ist die Konstantstromquelle des Differenzverstärkers, der die Transistoren nicht in die Sättigung kommen lässt (ungesättigte Logik). Ein weiterer Vorteil ist der relativ hohe Ausgangslastfaktor von 25. Nachteil dieser Technik sind hohe Verlustleistungen bis 40 mW. Auffallend sind die drei Betriebsspannungen UEE, UCC1 und UCC2. UCC1 dient als

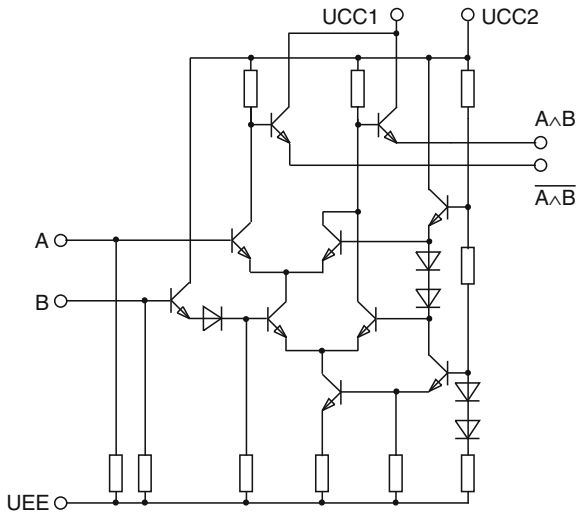


Abb. 51.15 ECL-Gatter mit zwei Ausgängen

Tab. 51.3 Kenndaten

	10K-Reihe	100k-Reihe
Betriebsspannung (V)	−5,2	−4,5
Low-Pegel (V) <	−0,81	−0,88
High-Pegel (V) >	−1,85	−1,81
Ausgangsstrom (mA)	−50	−55
tp (ns)	2	0,80
P (mW)	25	40

Hilfsspannung für Abschlusswiderstände. Häufig werden aber die beiden Anschlüsse UCC1 und UCC2 gemeinsam auf Masse zusammengelegt. Tab. 51.3 fasst die Kenndaten dieser Baugruppe zusammen.

51.6 Integrierte MOS-Schaltungen

Der Vorteil gegenüber Schaltungen in bipolarer Technik ist der geringere Leistungsbedarf, der allerdings mit zunehmender Frequenz steigt. MOS-Schaltungen (MOS = Metal Oxide Semiconductor) sind sehr hochohmig und nehmen praktisch nur im dynamischen Betrieb Leistung auf. Weiter lassen sich hohe Packungsdichten erreichen, und es werden weniger Einzelschritte bei der Herstellung benötigt. Der Nachteil liegt in der Verarbeitungsgeschwindigkeit durch stets vorhandene Streukapazitäten. MOS-Schaltungen sind somit viel langsamer als Standard-TTL-Schaltungen. Auch hier haben sich unterschiedliche Familien ausgebildet.

51.6.1 NMOS- und PMOS-Technik

Kennzeichen dieser Schaltungstechnik ist, dass sie entweder nur mit NMOS- oder nur mit PMOS-Typen aufgebaut sind. Das Einsatzgebiet der MOS-Technik liegt bei hochintegrierten Schaltungen wie z. B. der Speichertechnologie. Abb. 51.16 zeigt das Prinzip eines P- und NMOS-Schalters als Inverter mit der jeweiligen Wertetabelle. Der n-Kanal-Type leitet bei Ansteuerung mit einem H-Pegel und der p-Kanal-Type bei Ansteuerung mit einem L-Pegel.

51.6.2 CMOS-Inverter

Die CMOS-(Complementary MOS-)Logikfamilie besteht aus N- und P-Kanal-MOS-Transistoren. Abb. 51.17 verdeutlicht die Funktion am Beispiel eines Inverters. Je nach Ansteuerung ist immer einer der beiden Transistoren leitend. Bei Ansteuerung mit „L“ leitet der p-Kanal V1 und bei Ansteuerung mit „H“ der n-Kanal V2. CMOS-Gatter nehmen praktisch nur im dynamischen Betrieb Leistung auf. Die statische Verlustleistung ist nahezu Null, weil stets einer der beiden Transistoren gesperrt ist und damit kein Strom mehr durch die Schaltung fließt.

Die Betriebsspannung kann zwischen 2 und 15 V liegen.

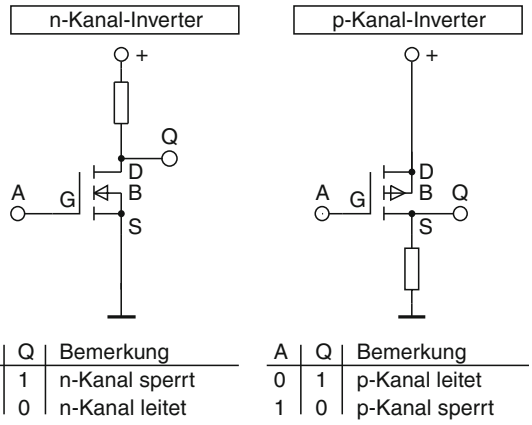


Abb. 51.16 n-Kanal- und p-Kanal-Schalter

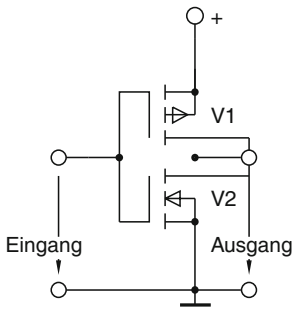


Abb. 51.17 CMOS-Inverter

Gegenüber MOS-Schaltkreisen ist die Leistungsaufnahme bei verbesserten Schaltzeiten und größerem Störspannungsabstand wesentlich kleiner. Gegenüber bipolarer Technik ist die höhere Integrationsfähigkeit hervorzuheben. Die Pegeländerungen von H nach L und umgekehrt liegen genau bei halber Betriebsspannung. Aufgrund des sehr hohen Eingangswiderstandes werden CMOS-Schaltungen gegen statische Aufladungen mit Schutzschaltungen versehen (Dioden im Ein- und Ausgang). Da die FET's direkt zwischen Betriebsspannung und Masse liegen, wirkt der durchgeschaltete Transistor immer als Lastwiderstand. CMOS-Schaltungen besitzen einen großen Spannungsversorgungsbereich (2 bis 15 V) und damit je nach Betriebsspannung einen großen Störabstand. Aufgrund der niedrigen Leistungsaufnahme sind CMOS-Schaltungen auch für den Batteriebetrieb gut einsetzbar.

51.6.2.1 4000-Serie

Bei den CMOS-Typen der 4000-Familie unterscheidet man die ungepufferte Serie A (ältere Serie) und die gepufferte Serie B. Die Ausgänge der B-Serie sind mit Puffern, die aus Doppelinverter bestehen, versehen. Abb. 51.18 zeigt ein NOR-Gatter beider Typen. Sobald ein Eingang H-Pegel besitzt, wird V1 oder V2 gesperrt und ergibt für den Ausgang Q einen L-Pegel. Nur wenn beide Eingänge mit einem L-Pegel angesteuert werden, leitet V1 und V2, während V3 und V4 gesperrt sind. In diesem Fall führt der Ausgang H-Pegel.

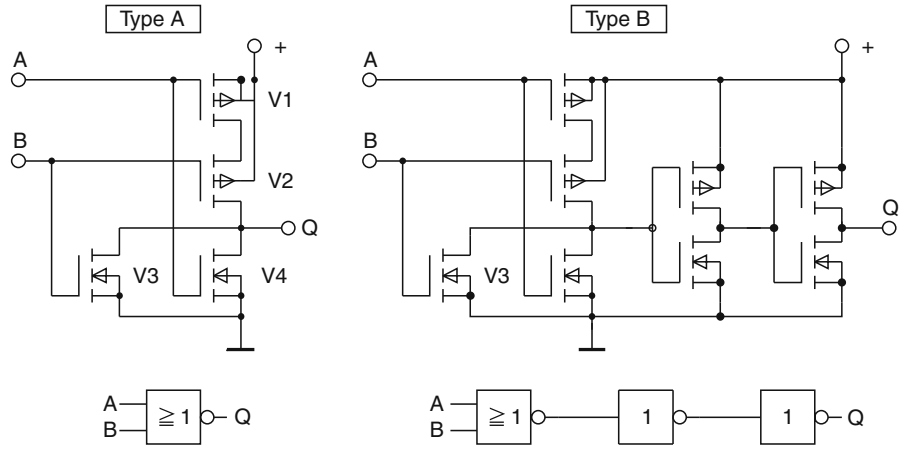


Abb. 51.18 CMOS-NOR

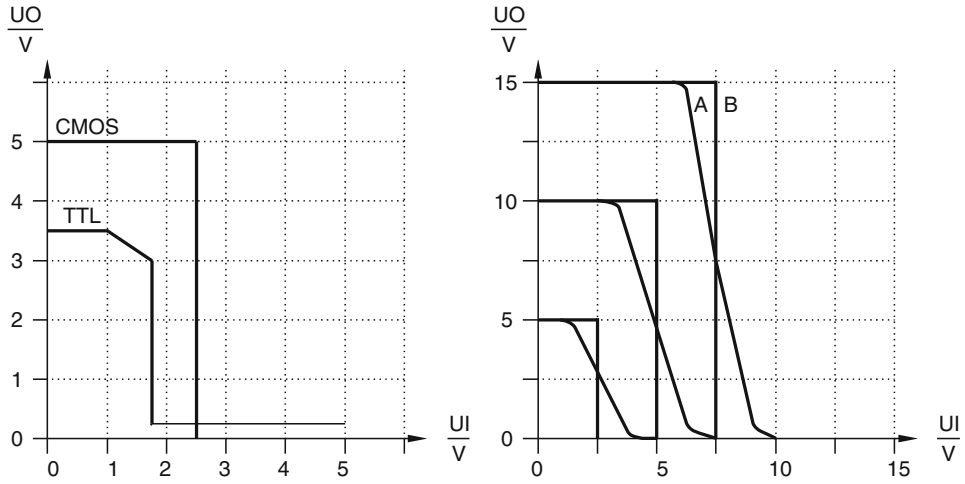


Abb. 51.19 Übertragungskennlinie

Abb. 51.19 vergleicht die Übertragungskennlinien eines Inverters beider Serien bei unterschiedlichen Betriebsspannungen. Das steile Schaltverhalten der B-Serie ist deutlich zu erkennen.

Die Leistungsaufnahme bei unterschiedlichen Frequenzen beschreibt Abb. 51.20. Erst ab etwa 1 MHz erreichen diese Bauelemente die Verlustleistung von LS-Typen.

Tab. 51.4 gibt die Kenndaten der CMOS-Familie wieder. Bei positiver Logik liegt die Betriebsspannung zwischen +3 und +15 V. Die Spannung für den H-Pegel muss mindestens 70 % und für den L-Pegel mindestens 30 % der Betriebsspannung betragen. Offene Eingänge sind nicht zulässig, und unbenutzte Eingänge müssen auf Masse (0) oder Betriebsspannung (1) liegen. CMOS-Gatter dürfen mit ihren Ausgängen nicht parallel geschaltet werden, weil sich je nach Ansteuerung ein undefinierter Ausgangszustand ergibt.

Die Ein- und Ausgänge dieser Schaltungen sind mit Schutzdioden gegen statische Überspannungen gesichert (Abb. 51.21).

Sollen mehrere Bausteine trotzdem eine gemeinsame Signalleitung bekommen, müssen, wie bei den TTL-Bausteinen, 3-

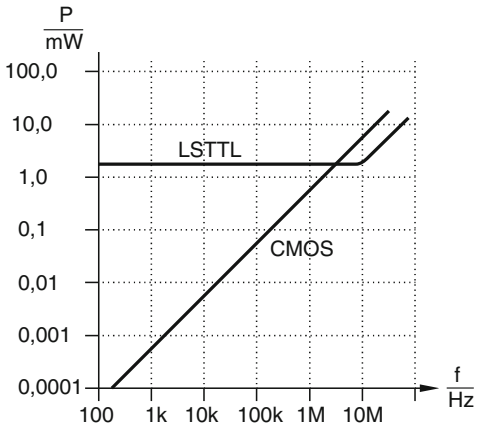


Abb. 51.20 Leistungsaufnahme

state-Schaltungen benutzt werden. Nach Abb. 51.22 wird die Schaltung bei EN = L in den hochohmigen Zustand versetzt. Ist EN = H, wird der anliegende Eingangspegel (L oder H) zum Ausgang Q übertragen.

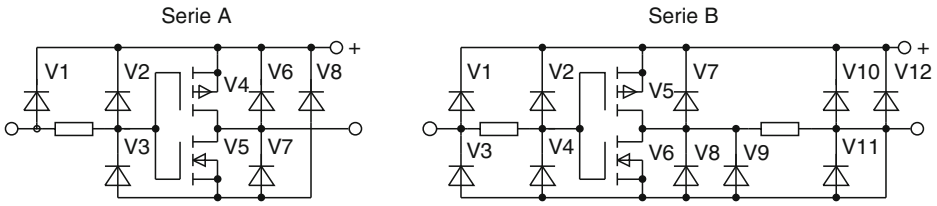


Abb. 51.21 Schaltung mit Schutzdioden

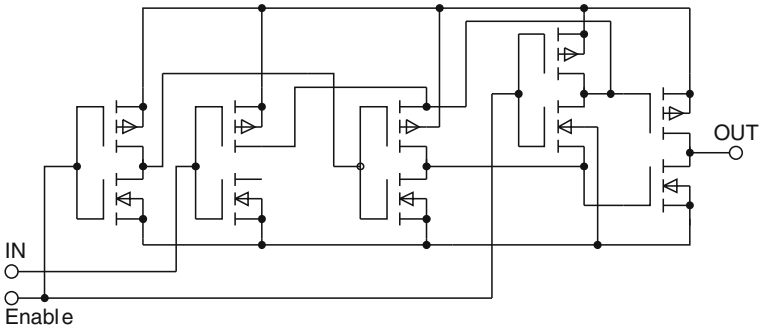


Abb. 51.22 Schaltung mit 3-state

Tab. 51.4 Daten der CMOS-Familie

	4000B	LSTTL	HC	HCT
Betriebsspannung	3–15 V	5 V ± 5 %	2–6 V	5 V ± 10 %
Verlustleistung (statisch)	1 µW	2 mW	2,5 nW	2,5 nW
Verlustleistung (100 kHz)	100 µW	2 mW	75 nW	75 nW
Gatterlaufzeit	94 ns	10 ns	8 ns	8 ns
Max. Taktfrequenz	4 MHz	33 MHz	55 MHz	55 MHz
Ausgangstreiberstrom	0,5 mA	8 mA	4 mA	4 mA

51.6.2.2 CMOS-Schalter CMOS-Bausteine

Der bilaterale CMOS-Schalter besteht aus einem p- und n-Kanal-MOS. Durch einen L-Pegel am Steueranschluss E0 werden beide Transistoren mit den Anschlüssen Y0 und Z0 leitend. Durch einen anliegenden H-Pegel sperren beide Transistoren. Der Schalter kann für analoge und digitale Signale eingesetzt werden. Eingang und Ausgang können beliebig zugeordnet werden. Eine Signalverstärkung erfolgt allerdings nicht. Abb. 51.23 zeigt als Beispiel den Baustein IC 4016 mit seiner Anschlussbelegung und den Aufbau eines der vier Schalter.

CMOS-Schaltungen müssen aufgrund der extrem hohen Eingangswiderstände (> 100 MΩ) gegen statische Aufladungen geschützt werden. Sie besitzen daher im Eingang Schutzschaltungen, die je nach Typ und Hersteller unterschiedlich sind. Wie schon bei der TTL-Familie, bieten auch hier die Hersteller unterschiedliche Bausteine an. Abb. 51.24 zeigt eine kleine Auswahl gebräuchlicher IC's.

51.6.2.3 High-Speed-CMOS

Die HC/HCT-Logikreihe (High-Speed-CMOS/HC-TTL kompatibel) verbindet die Vorteile der bipolaren Technik mit ihrer hohen Schaltgeschwindigkeit und die geringere Leistungsaufnahme der Standard-CMOS-Serie (Tab. 51.2). Dadurch sind

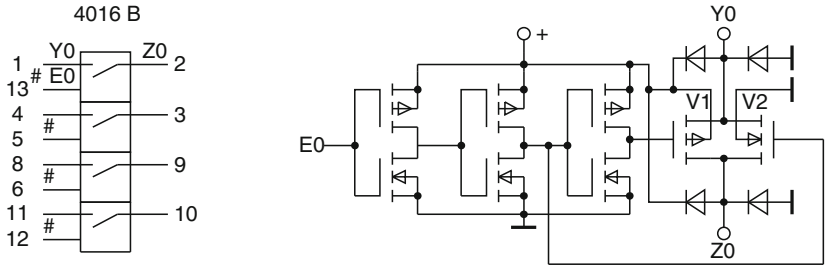


Abb. 51.23 CMOS-Schalter

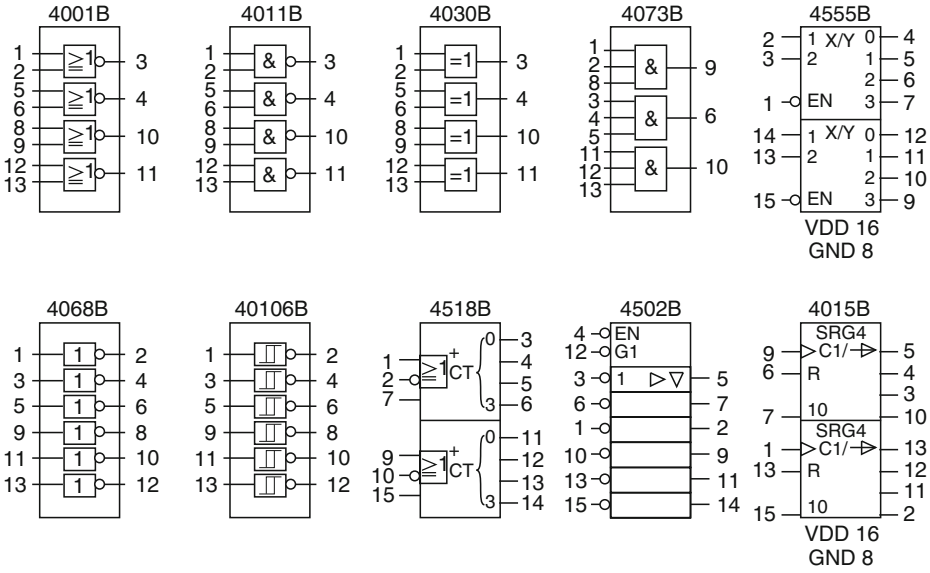


Abb. 51.24 CMOS-Bausteine

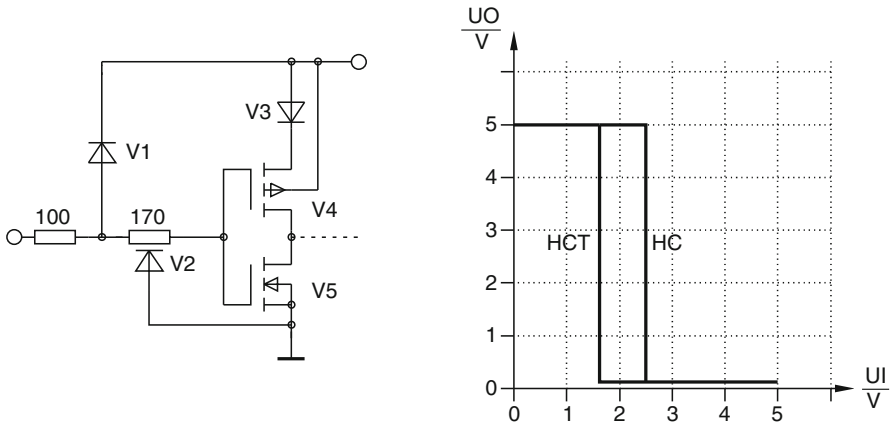


Abb. 51.25 HC-Eingangsschaltung und Übertragungskennlinie

diese Bausteine den 74LS- und 4000-Typen überlegen. Die 74HCT-Serie ist TTL-kompatibel und kann bei einer Betriebsspannung von 5 V jeden LS-Typ ersetzen, da eine zusätzliche Kompatibilität der Spannungspegel vorliegt (Tab. 51.5). Zu beachten ist aber, dass der Ausgangstreiberstrom nur die Hälfte des LSTTL beträgt. Die 40xxHC/HCT-Serie ist pinkompatibel zur 4000-Serie. Abb. 51.25 zeigt die Eingangsschaltung einer HCT-Version und die Übertragungskennlinien bei einer Betriebsspannung von 5 V.

Die beiden Widerstände im Eingang und V1 und V2 dienen als Schutzschaltung bei elektrostatischen Entladungen. Im Gegensatz zu LS-Typen besitzen HC-Typen einen größeren Versorgungsspannungsbereich (4 bis 6 V). Für die Übergänge von „L“ nach „H“ und umgekehrt liegen die Schaltspannungen bei 30 bzw. 70 % der Betriebsspannung. Abb. 51.26 vergleicht die Betriebsspannung und die Pegelbereiche der gängigsten Typen.

51.6.2.4 BICMOS

BICMOS (Bipolar CMOS) vereinigen bipolare mit CMOS-Technik. Durch die CMOS-Technik wird die Leistungsaufnahme wesentlich gesenkt und durch die bipolare Technik die Geschwindigkeit erhöht. Zu dieser Serie zählen die Bausteine 74BCxx und 74ABTxx. Während bei den BC-Typen die Gatterlaufzeiten bei einer Betriebsspannung von 5 V bei 3,8 ns liegen, bewegen sich die Gatterlaufzeiten bei ABT unter 3 ns. Die Bauelemente sind funktions- und pinkompatibel zu LS-TTL oder FAST (74F/74S/74HC). Relativ neu ist die Entwicklung und das Angebot von Niederspannungsschaltkreisen; sogenannte 3,3-V-Logikschaltkreise (Low Voltage). Die Verlustleistung dieser Bausteine gegenüber der 5-V-Technik ist etwa um 1/3 niedriger.

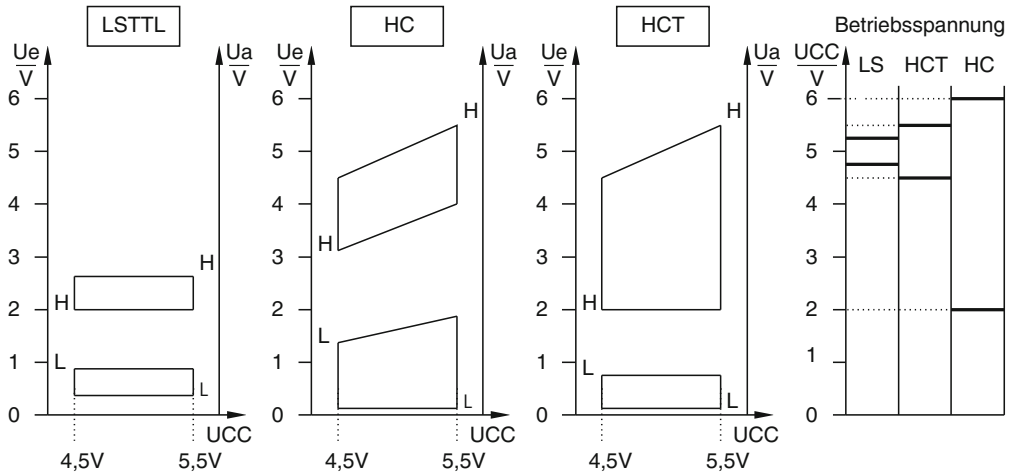
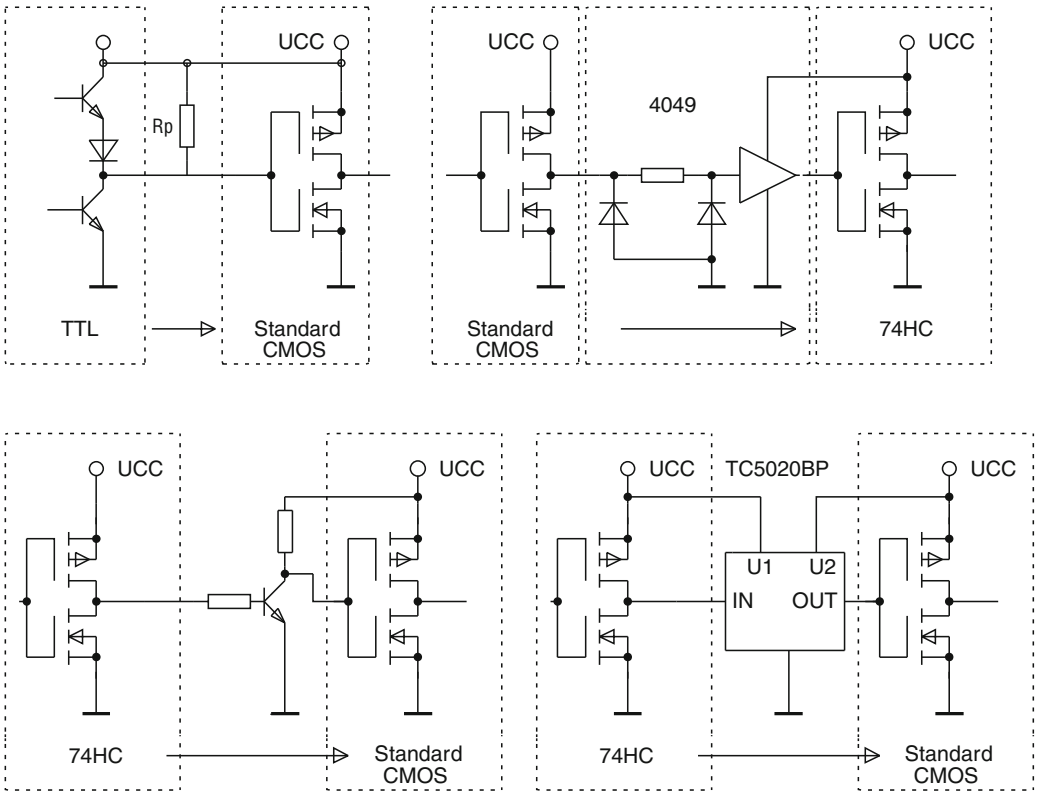


Abb. 51.26 Betriebsspannungen und Pegelbereiche

51.7 Interfaceschaltungen

Interfaceschaltungen passen die Logikfamilien untereinander an. Das ist bei unterschiedlichen Pegelwerten notwendig. Sie verbinden auch Logikschaltungen mit weiteren elektronischen Systemen. Die Verbindung von CMOS zu TTL lässt sich über Puffer realisieren und von TTL nach CMOS über einen Pull-up-Widerstand realisieren. Bei 5 V Betriebsspannung ist auch eine direkte Kopplung möglich (Tab. 51.5). Abb. 51.27 zeigt Beispiele, bei denen die Logik-Familien mit unterschiedlichen Betriebsspannungen betrieben werden.

Der Pull-up-Widerstand Rp beträgt 2 bis 10 kΩ bei der Umsetzung von TTL nach Standard-CMOS. Für die Umsetzung von Standard-CMOS mit einer Betriebsspannung von 6 bis 15 V



Nach Datenblatt der Firma Toshiba

Abb. 51.27 Interface-Schaltungen

Tab. 51.5 Anpassung unterschiedlicher Familien bei 5 V Betriebsspannung

	Zu				
Von	74HC/5 V	74HCT	4000B/5 V	4000B/6-15 V	TTL
74HC/5 V	Direkt	Direkt	Direkt	Puffer	Direkt
74HCT	Direkt	Direkt	Direkt	Puffer	Direkt
4000B/5 V	Direkt	Direkt	Direkt	Puffer	Direkt
4000B/6-15 V	Puffer	Puffer	Puffer	Direkt	Puffer
TTL	Pull-up-Widerstand	Direkt	Pull-up-Widerstand	Puffer	Direkt

nach 74HC ist ein zusätzlicher Umsetzer (4049/50B) beispielhaft eingesetzt. Das 74HC-Standard-CMOS-Interface wurde zum einen mit einem Transistor und zum anderen mit einem Umsetzer (TC5020BP) erreicht.

51.8 Anwendungsspezifische integrierte Schaltungen

51.8.1 Allgemeines

Um eine gewünschte Schaltung zu realisieren, kann bei der Entwicklung digitaler Systeme auf Standardbausteine der unterschiedlichen Logikfamilien zurückgegriffen werden. Mit ihnen werden Baugruppen geschaffen, deren Bausteine durch Signalverbindungen jeweils miteinander verbunden sind. Jeder Baustein benötigt eine eigene Spannungsversorgung. Die Größe der Baugruppe ist jeweils von der Komplexität ihrer Aufgabe abhängig. Soll der Platzbedarf reduziert werden, werden *anwendungsspezifische integrierte Schaltungen* (ASIC = Application Specific IC) gewählt. Sie besitzen eine höhere Integrationsdichte und bei Reduzierung der Systemverlustleistung eine geringere Ausfallrate in Verbindung mit hoher Flexibilität. Anwendungsspezifische integrierte Schaltungen können in zwei Hauptgruppen unterteilt werden, zum einen in die Gruppe der *kundenspezifischen IC's* (Custom IC) mit den Bereichen der *halbkundenspezifischen* (Semi Custom) und den *vollkundenspezifischen* (full custom) Schaltungen und zum anderen in die Gruppe der programmierbaren Logikbausteine (PLD = Programmable Logic Device).

Das für den Anwendungsfall notwendige Systemwissen wird vom Anwender selbst eingebracht oder er nutzt Hersteller-Design-Zentren, die durch Beratung und gezielte Unterstützung eine optimale Lösung anstreben.

51.8.2 Kundenspezifische IC's

Bei den kundenspezifischen IC's wird nach Full-custom- und Semi-custom-Bausteinen unterschieden. Der Full-custom-Baustein wird vom Hersteller nach den Kundenanforderungen entwickelt und hergestellt.

Bei einer relativ langen Entwicklungszeit von über einem Jahr und dem hohen Fertigungsaufwand sind solche Bausteine nur

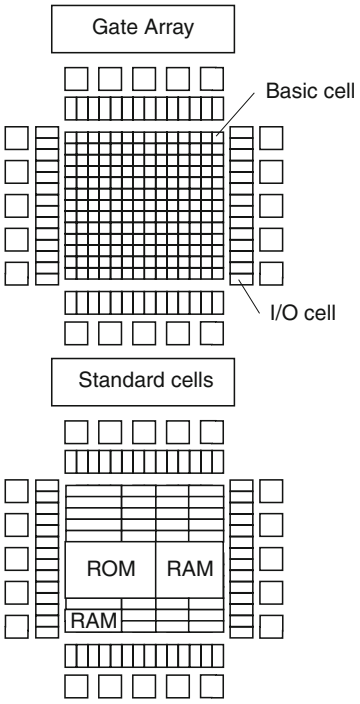


Abb. 51.28 Gate-Array und Standard Cells

für Großkunden mit großen Produktionsmengen wirtschaftlich. Semi-custom-Bausteine umfassen die Typen *Gate-Arrays* und *Cell-Arrays* oder *Standard Cells* (Abb. 51.28). Sie sind von den Herstellern als Gruppen von Gatterbausteinen und Logikzellen aufgebaut.

Ausgangsbasis der Gate-Arrays sind vorgefertigte Grundbausteine aus vielen P- und N-Kanal-MOS-Transistoren, die als Gatter aufgefasst werden können. Nach Wünschen des Kunden werden die Basiszellen zu einer Logikschaltung entsprechend verdrahtet.

Erste Muster für den Kunden sind schon nach wenigen Wochen verfügbar, so dass die Entwicklungskosten gegenüber den vollkundenspezifischen Schaltungen deutlich niedriger liegen. Der Silizium-Ausnutzungsgrad bei einem beliebigen Layout liegt bei etwa 40 %. Bei den Cell-Arrays bedient sich der Entwickler beim Schaltungsentwurf der Zellbibliotheken, die gewünschte Funktionen wie Zähler, Multiplexer oder Speicher beinhalten. Im Gegensatz zu den Gate-Arrays, wo Grundfunktionen erstellt werden, sind diese „Basiszellen" mit größeren Teilaufgaben vorgefertigt. Der Entwicklungsaufwand dieser Bausteine ist größer

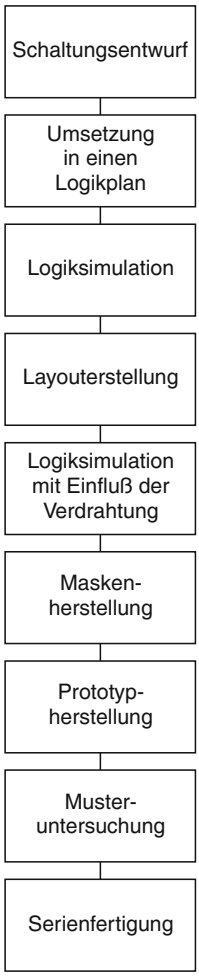


Abb. 51.29 Array-Entwicklung

als bei den Gate-Arrays. Da jedoch die Anordnung der Zellen und auch die Zellbreite vom Entwickler gewählt werden, ergibt sich eine optimal genutzte Chipfläche, die bei hohen Stückzahlen Kostenvorteile mit sich bringen kann. Abb. 51.29 zeigt den Ablaufplan einer Cell-Array-Entwicklung.

51.8.3 Programmierbare Logikbausteine

Programmierbare Logikbausteine (PLD = Programmable Logic Device) sind Standardbausteine, deren Funktion erst durch Software hergestellt wird. Ihre innere Struktur besteht je nach Type aus einer programmierbaren UND- und ODER-Matrixschaltung für Ein- und Ausgänge unter Einbeziehung weiterer Logikbausteine wie beispielsweise Flipflops, Register und Treiberstufen. Der universelle Einsatz *derselben* Bausteine für *unterschiedliche* Funktionen ist möglich. Hinzu kommt, dass die Gesamtschaltung in einem IC-Baustein untergebracht werden kann, wozu sonst mehrere TTL- oder CMOS-Standardbausteine nötig wären. Hieran erkennt man den Vorteil programmierbarer Bausteine gegenüber diskreten Logikbausteinen. Ihre Integration ist höher, sie sind flexibler zu handhaben, und die Schaltungsentwicklung erfolgt durch Software. Damit ergeben sich kurze Entwicklungszeiten bei niedrigen Kosten. Je nach Hersteller existieren unterschiedliche Bezeichnungen programmierbarer Schaltkreise. FPGA (Field programmable Gate Array) sind für Logikschaltungen mit bis zu 10000 Gattern pro Array universell einsetzbar und schnell umzuprogrammieren. Ihre Versorgungsspannung beträgt 5 V. PLA (Programmable Logic Array) eignen sich für Logikschaltungen bis zu 300 Gattern pro Array. Ihr Betriebsspannungsbereich geht von 2 bis 6 V und ihr Stromverbrauch liegt bei 8 mA. EPLD (Erasable PLD) und EEPLD (Electrically EPLD) sind weitere Bezeichnungen für programmierbare Schaltkreise.

Bekannte Vertreter dieser Bausteine sind PAL's und GAL's. Im Gegensatz zu den Gate-Arrays kann der Anwender ihre Funktion durch Programmierung mit entsprechenden Programmiergeräten selbst bestimmen. Hierzu muss die Information in einer JEDEC-Datei vorliegen. Während GAL's mehrmals programmiert werden können, kann dies bei PAL's nur einmal durchgeführt werden. Die Grundstruktur und die Programmierung der unterschiedlichen PLD-Typen sind einander sehr ähnlich.

51.8.3.1 PROM

PROM-Bausteine (Programmable Read Only Memory) werden in Abschn. 50.6.3.1 „Speicherbausteine" beschrieben. An dieser Stelle soll nur auf die von anderen PLD's verschiedene Grundstruktur hingewiesen werden. Das AND-Array (UND-Matrix) ist bei diesen Bausteinen festverdrahtet, während das OR-Array (ODER-Matrix) einmal programmiert werden kann. PLA's besitzen eine programmierbare UND- und ODER-Matrix und PAL's nur eine programmierbare UND-Matrix (Abb. 51.30).

51.8.3.2 PAL

PAL (Programmable Array Logik) sind Bauelemente, die eine Logikmatrix enthalten und vom Anwender programmiert werden können. Mit ihnen können mehrere kleinere Logikschaltungen in einem Gehäuse untergebracht werden. PAL's enthalten eine Matrix (Verbindungen aus Zeilen und Spalten) und logische Funktionsbausteine wie NICHT, UND, ODER und Flipflops. Während sich im Eingang eine *programmierbare Matrix* befindet, ist die nachfolgende *ODER-Matrix* fest verdrahtet. Abb. 51.31 zeigt eine UND-Matrix mit intakten Sicherungen.

Da der UND-Baustein das invertierte und nichtinvertierte Signal erhält, ist Q = „0". Dies wird auch durch die angegebene Gleichung deutlich. Eine programmierte Matrix erkennt man an den defekten Sicherungen. Nicht programmierte (intakte) Brücken (Fuse = Sicherung) sind an dem Kreuz im Schnittpunkt zu erkennen. Im nichtprogrammierten Zustand ist der Ausgang der UND-Verknüpfung immer „0". Die Verknüpfungsausgänge der UND-Matrix (Produktlinien) werden dann der ODER-Matrix zugeführt und bei einigen PAL-Bausteinen mit dem Ausgang verbunden. Abb. 51.32 zeigt hierzu Beispiele.

Die Bausteine 10L8 und 14L8 zeigen, dass die Ausgänge auch invertiert sein können. Bei den PALTypen können verschiedene Ausgangskonfigurationen bei einer allen gemeinsamen UND-Matrix im Eingang unterschieden werden (Abb. 51.33). Man

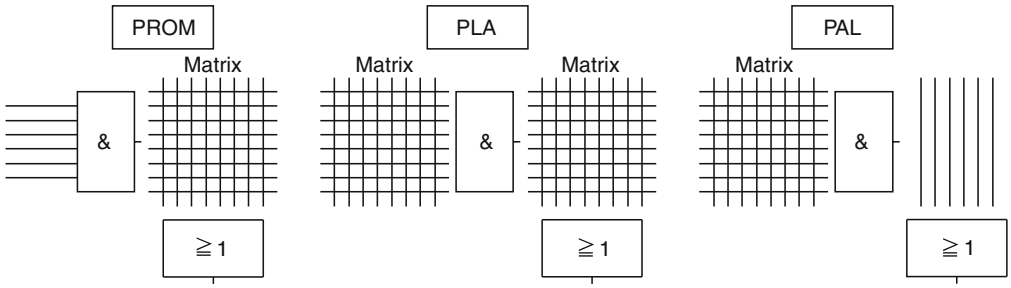


Abb. 51.30 Strukturen unterschiedlicher PLD's

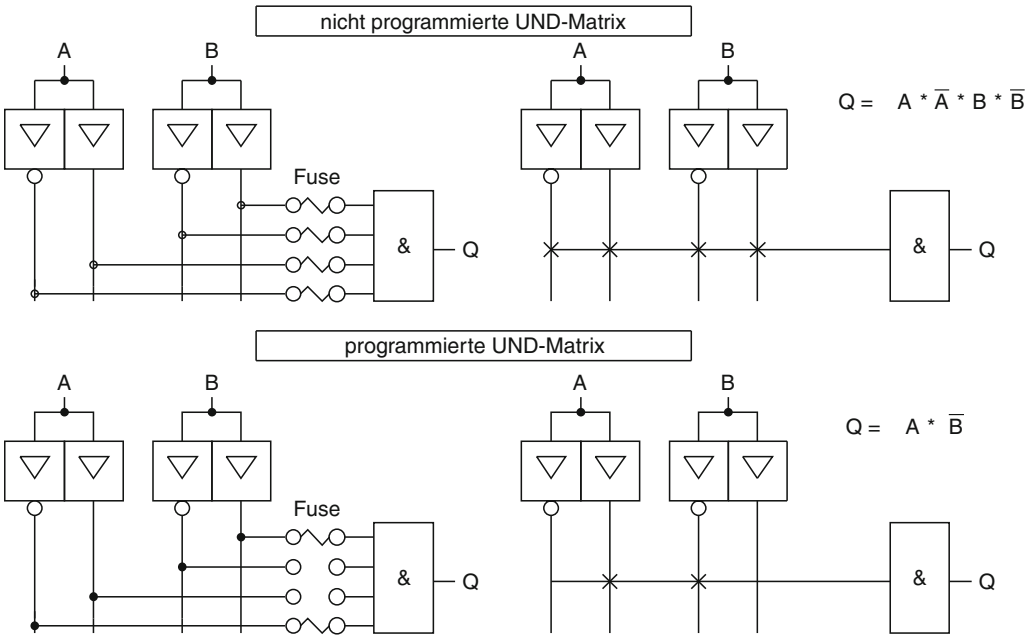


Abb. 51.31 Prinzip und vereinfachte Darstellung der Matrix PAL

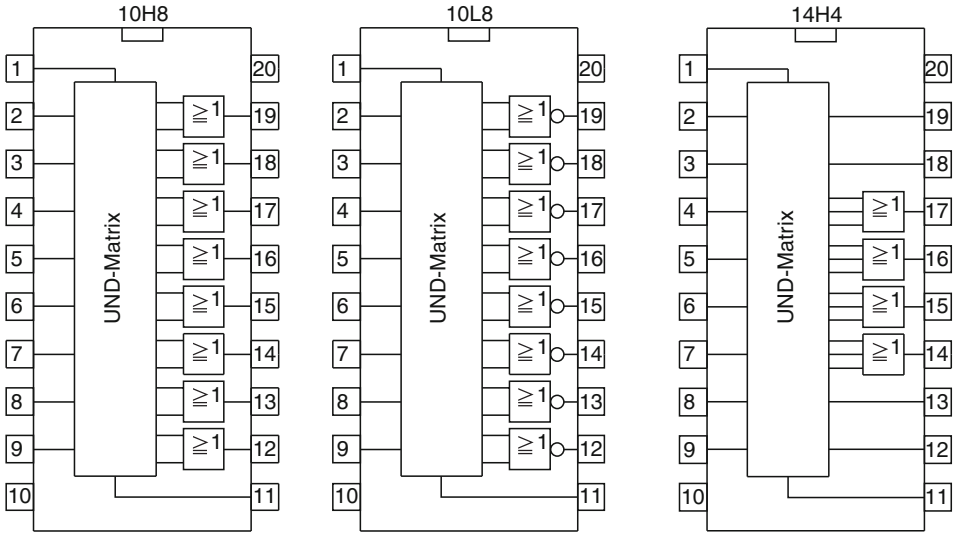


Abb. 51.32 Einfache PAL-Bausteine

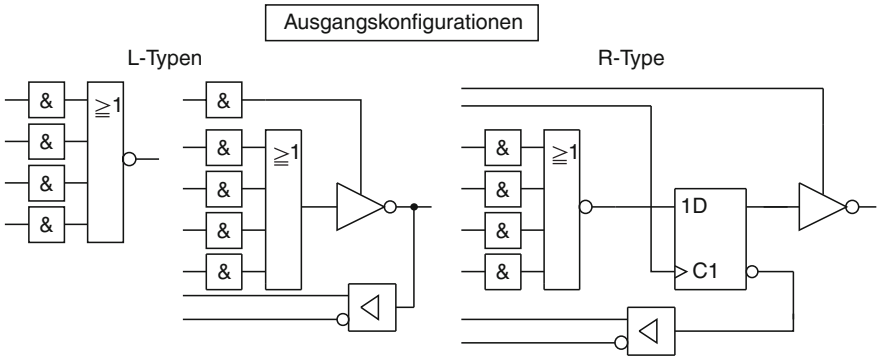


Abb. 51.33 Ausgangskonfigurationen

Tab. 51.6 PAL-Typen und Bezeichnungen

H	UND-ODER (High aktiv)
L	UND-NOR (Low aktiv)
C	UND-ODER
R	UND-ODER-Inverter-Register
X	UND-ODER-EXODER-Register

findet die gewünschten Bauteile anhand der Typenbezeichung (Tab. 51.6). Der Typenname beschreibt die Anzahl der Ein- und Ausgänge durch jeweils eine Zahl, während der Buchstabe die interne Funktion und Struktur des PAL angibt. Die Zahl nach dem Buchstaben gibt die Anzahl der Ausgänge an.

Abb. 51.34 zeigt am Beispiel des Bausteins 16L8 die Grundstruktur von PAL-Bausteinen. Die UND-Matrix besteht aus einunddreißig Spalten. Sieben Zeilen werden jeweils der ODER-Verknüpfung zugeführt und eine Zeile dem Ausgangs-3-state-Treiber.

Um die Programmierung durchzuführen, wird eine JEDEC-Datei (JEDEC = Joint Device Engineering Council, Komitee für Standards in der Datentechnik) erzeugt und einem Programmiergerät zugeführt. Die jeweiligen Verbindungen werden durch die Programmierung aufgetrennt. Eine Zerstörung der sogenannten *„Security Fuse“* verhindert das Auslesen des PAL's. Die Schaltungsbeschreibung selbst erfolgt mit Booleschen Gleichungen, Wahrheitstabellen, Status-(State-)Diagrammen, Grafiken oder in einer höheren Programmiersprache. Normalerweise kann jeder übliche Editor eingesetzt werden. Bei der Schreibweise gibt es je nach Anwendungsprogramm unterschiedliche Darstellungsformen. Darstellungsformen für Verknüpfungen:

Funktion	Darstellung		
NICHT		/	!
UND	*	&	
ODER	+	|	#

Viele Hersteller bieten hierzu Entwicklungskits an, die außer Programmierungsmöglichkeiten zusätzliche Simulationsprogramme beinhalten. Damit lässt sich bereits statisch feststellen, ob der Baustein die gewünschten Funktionen auch tatsächlich, wie geplant, durchführt.

Zwei Beispiele sollen das Programmieren von PAL's veranschaulichen.

Beispiel 1

Ein Decoder soll zur Auswahl dreier Speicher so konzipiert werden, wie das die Schaltskizze Abb. 51.35 zeigt. Zur Verfügung steht ein PAL 12L6. Die Zuordnung der Eingänge und der Ausgänge mit frei wählbaren Namen zu den Anschlüssen des PAL-Bausteins ist beliebig. Damit kann das Platinenlayout je nach Wunsch und Notwendigkeit angepasst werden. Hier wurde die Zuordnung festgelegt.

Zuordnung der Eingangssignale zu den PAL-Anschlüssen:

Programm: Adressdecoder

CS = pin 1; Freigabeanschluss des Decoders
A11 = pin 2; Anschlüsse an den Adressbus
A12 = pin 3
A13 = pin 4
A14 = pin 5
A15 = pin 6

3-state-Ausgang:

CE1.OE = CS
CE2.OE = CS
CE3.OE = CS

Gleichungen:

!CE1 = A15 &A14 & A13 & A12 & !A11
!CE2 = A15 & A14 & !A13 & A12 & A11
!CE3 = A15 & A14 & !A13 & !A12 & !A11 ◄

Mit einem Compiler wird eine JEDEC-Datei erzeugt, aus der ersichtlich ist, an welcher Stelle jeweils eine Sicherung des PAL-Bausteins zerstört bzw. nicht zerstört wurde. Die „0“ gibt die intakte und die „1“ die unterbrochenen Fuse an. Man kann sich auch vorstellen, es sei nur noch dort, wo sich die „0“ befindet, eine Verbindung mit einem Eingang. Abb. 51.36 zeigt einen Auszug aus der JEDEC-Datei ab Adresse 0000. Die

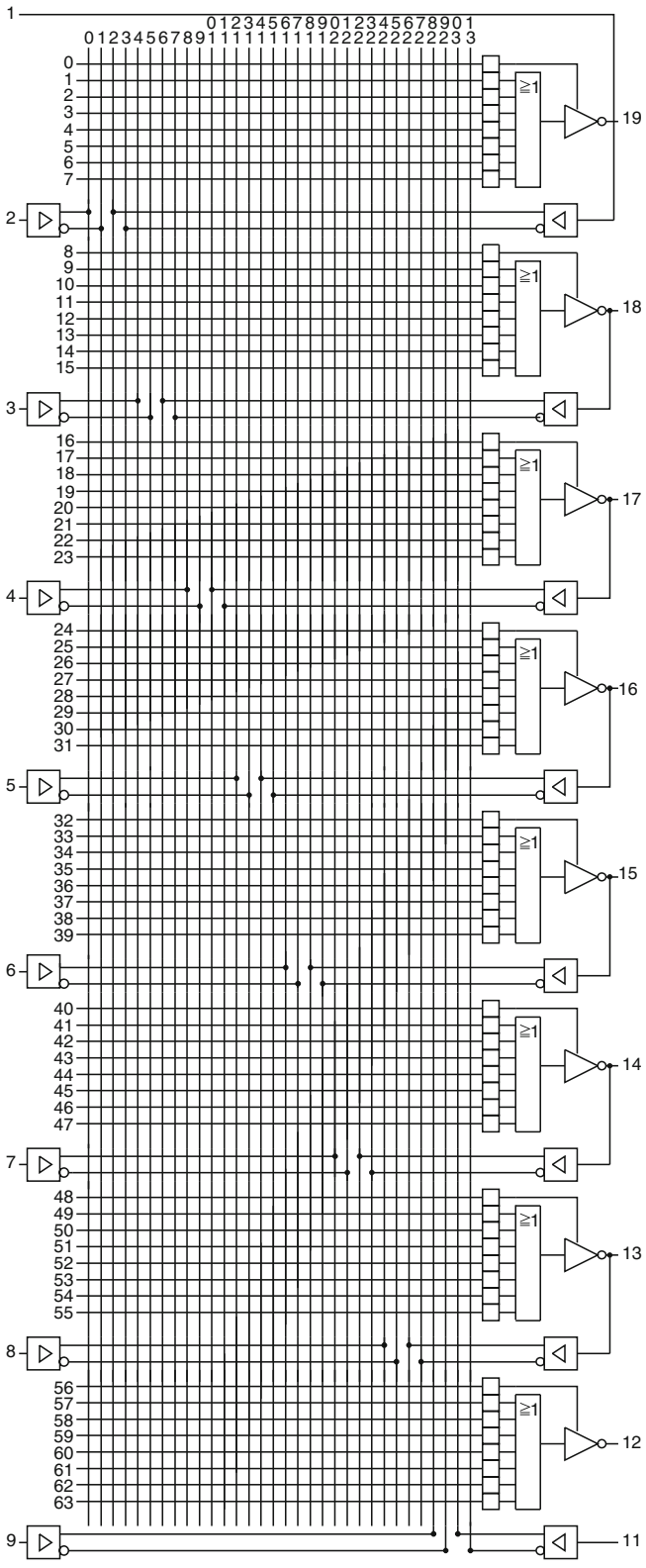


Abb. 51.34 PAL 16L8

Adressen 0032 bis 0063 sind gesondert herausgezogen. Hier sind auch die intakten Sicherungen in den Spalten 1, 4, 8, 12 und 16 gekennzeichnet. Sie stellen die jeweiligen Verbindungen zu den Eingängen dar und sind UND-verknüpft. Mit dem CS-Anschluss wird der Decoder freigegeben (Steuerung der 3-state-Ausgänge).

Abb. 51.37 zeigt dies im Zusammenhang mit den Eingängen für einen Ausgang.

Die Funktionsbeschreibung einer Schaltung kann auch mit einer Wertetabelle erfolgen (Beispiel 2). Zur Überprüfung können Testvektoren zugefügt werden. Die mit einem Compiler erzeugte JEDEC-Datei wird daraufhin softwaremäßig mit einem Simulator getestet. Allen programmierten Zeilen ist eine Adresse zugeordnet.

Beispiel 2

Mit dem PAL 16L8 soll eine 7-Segmentanzeige zur Darstellung in hexadezimaler Form genutzt werden. Im Gegensatz zum Beispiel 1 erfolgt die Eingabe hier mit einer Wertetabelle.

Programm: HEX-Decoder mit 7-Segmentanzeige

Zuordnung der Eingangssignale zu den PAL-Anschlüssen:

A = pin 2
B = pin 3
C = pin 4
D = pin 5

Ausgangssignale des Decoders:

a = pin 19
b = pin 18
c = pin 17
d = pin 16
e = pin 15
f = pin 14
g = pin 13

Wertetabelle:

D	*C*	*B*	*A*	:	*a*	*b*	*c*	*d*	*e*	*f*	*g*
0	0	0	0	:	1	1	1	1	1	1	0
0	0	0	1	:	0	1	1	0	0	0	0
0	0	1	0	:	1	1	0	1	1	0	1
0	0	1	1	:	1	1	1	1	0	0	1
0	1	0	0	:	1	0	1	1	0	1	1
0	1	0	1	:	1	0	1	1	0	1	1
0	1	1	0	:	1	0	1	1	1	1	1
0	1	1	1	:	1	1	1	0	0	0	0
1	0	0	0	:	1	1	1	1	1	1	1
1	0	0	1	:	1	1	1	1	0	1	1
1	0	1	0	:	1	1	1	0	1	1	1
1	0	1	1	:	0	0	1	1	1	1	1
1	1	0	0	:	1	0	0	1	1	1	0
1	1	0	1	:	0	1	1	1	1	0	1
1	1	1	0	:	1	0	0	1	1	1	1
1	1	1	1	:	1	0	0	0	1	1	1

◄

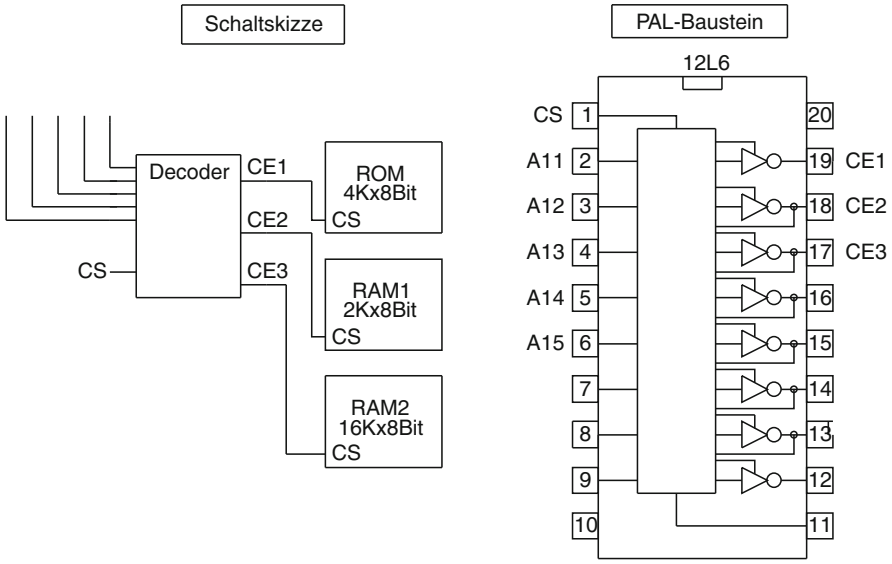


Abb. 51.35 PAL-Baustein als Decoder

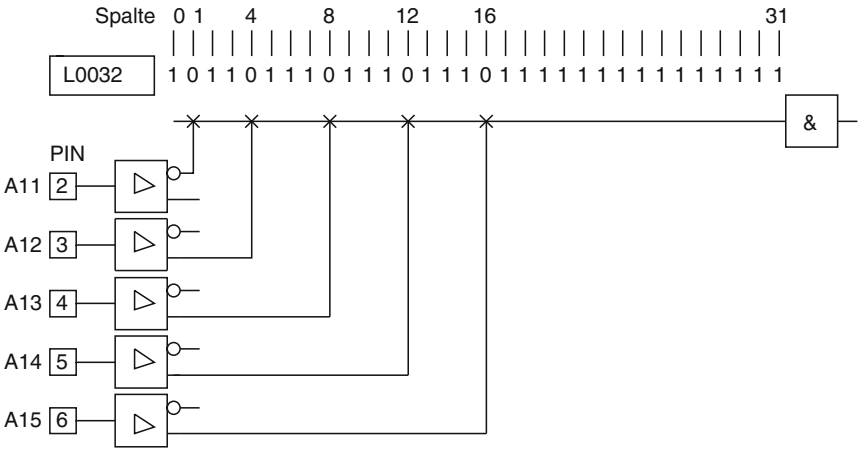


Abb. 51.36 Auszug aus der JEDEC-Datei und Darstellung der UND-Verknüpfung

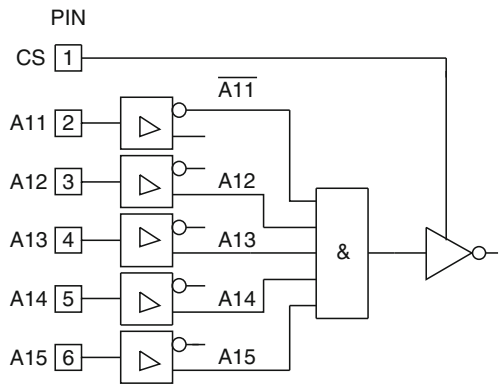


Abb. 51.37 Umsetzung der Gleichung für CE1

Abb. 51.38 zeigt einen Auszug aus der JEDEC-Datei und Abb. 51.39 einen Auszug aus der Fuse-Map. Die Fuse-Map oder Fuse-Matrix, die ebenfalls vom Compiler erzeugt wird, dient zur Dokumentation und zur Kontrolle. Sie beschreibt das logische Verhalten des Bausteins. Die Reihen und Spalten sind wie bei der Bausteinbeschreibung durchnummeriert. 8 Reihen führen jeweils über das ODER-Gatter zum Ausgang. Die Eingänge steuern die Spalten von 0 bis 31.

Die Kreuze in der Fuse-Map kennzeichnen hierbei die Verbindungen zwischen Reihen und Spalten. Nimmt man beispielsweise den Bereich der Reihen 0 bis 7 für den Ausgangsanschluss des Segments „a“, PIN 19, kommt man zu folgender Gleichung:

$$a = /B \cdot /C \cdot D + B \cdot C + /D \cdot C + /D \cdot B + /A$$

```
*L0000  11011111111111111111111111111111
*L0032  11111011101101111111111111111111
*L0064  11110111011111111111111111111111
*L0096  11111111011101111111111111111111
*L0128  11110111111110111111111111111111
*L0160  10111111111111111111111111111111
*L0256  11011111111111111111111111111111
*L0288  01110111111110111111111111111111
*L0320  01111011111110111111111111111111
*L0352  10111111101111111111111111111111
*L0384  11111111011101111111111111111111
*L0512  11011111111111111111111111111111
*L0544  11111111101101111111111111111111
*L0576  11111111011101111111111111111111
*L0608  01111011111111111111111111111111
*L0640  01111111111110111111111111111111
*L0672  11111011111110111111111111111111
*L0768  11011111111111111111111111111111
*L0800  01110111101111111111111111111111
*L0832  11111011111101111111111111111111
*L0864  10111111011111111111111111111111
*L0896  11111101101111111111111111111111
*L0928  10111111111110111111111111111111
*L1024  11011111111111111111111111111111
*L1056  11111111011101111111111111111111
*L1088  11110111111101111111111111111111
*L1120  10110111111111111111111111111111
*L1152  10111111101111111111111111111111
*L1280  11011111111111111111111111111111
*L1312  11111011011110111111111111111111
*L1344  11110111111101111111111111111111
*L1376  11111111101101111111111111111111
*L1408  10111111011111111111111111111111
*L1440  10111011111111111111111111111111
*L1536  11011111111111111111111111111111
*L1568  11111011011110111111111111111111
*L1600  01111111111101111111111111111111
*L1632  11111111101101111111111111111111
*L1664  10110111111111111111111111111111
*L1696  11110111101111111111111111111111
*C9B81
*V1   N0000NNNNNNNHLLLLLLN
*V2   N1000NNNNNNNHHHHLLHN
*V3   N0100NNNNNNNLHLLHLLN
*V4   N1100NNNNNNNLHHLLLLN
```

Abb. 51.38 Auszug aus der JEDEC-Datei (Beispiel 2)

Hierbei gilt folgende Zuordnung:

A = Spalte 0
/A = Spalte 1
B = Spalte 4
/B = Spalte 5
C = Spalte 8
/C = Spalte 9
D = Spalte 12
/D = Spalte 13

Zum gleichen Ergebnis für das Segment „a“ kommt man, wenn die Gleichung aus der Wertetabelle mit Hilfe einer KV-Tabelle vereinfacht wird. Die Software setzt damit nicht nur die Wertetabelle um, sie vereinfacht auch gleichzeitig.

```
                    11  1111 1111 2222 2222 2233
    0123 4567 8901 2345 6789 0123 4567 8901
 0  --X- ---- ---- ---- ---- ---- ---- ---- OE
 1  ---- -X-- -X-- X--- ---- ---- ---- ---- +
 2  ---- X--- X--- ---- ---- ---- ---- ---- +
 3  ---- ---- X--- -X-- ---- ---- ---- ---- +   !pin19
 4  ---- X--- ---- -X-- ---- ---- ---- ---- +
 5  -X-- ---- ---- ---- ---- ---- ---- ---- +
 6  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
 7  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

 8  --X- ---- ---- ---- ---- ---- ---- ---- OE
 9  X--- X--- ---- -X-- ---- ---- ---- ---- +
10  X--- -X-- ---- X--- ---- ---- ---- ---- +
11  -X-- ---- -X-- ---- ---- ---- ---- ---- +   !pin18
12  ---- ---- -X-- -X-- ---- ---- ---- ---- +
13  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
14  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
15  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

16  --X- ---- ---- ---- ---- ---- ---- ---- OE
17  ---- ---- -X-- X--- ---- ---- ---- ---- +
18  ---- ---- X--- -X-- ---- ---- ---- ---- +
19  X--- -X-- ---- ---- ---- ---- ---- ---- +   !pin17
20  X--- ---- ---- -X-- ---- ---- ---- ---- +
21  ---- -X-- ---- -X-- ---- ---- ---- ---- +
22  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
23  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

24  --X- ---- ---- ---- ---- ---- ---- ---- OE
25  X--- X--- -X-- ---- ---- ---- ---- ---- +
26  ---- -X-- ---- X--- ---- ---- ---- ---- +
27  -X-- ---- X--- ---- ---- ---- ---- ---- +   !pin16
28  ---- -X-- X--- ---- ---- ---- ---- ---- +
29  ---- ---- ---- -X-- ---- ---- ---- ---- +
30  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
31  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

32  --X- ---- ---- ---- ---- ---- ---- ---- OE
33  ---- ---- X--- X--- ---- ---- ---- ---- +
34  ---- X--- ---- X--- ---- ---- ---- ---- +
35  -X-- X--- ---- ---- ---- ---- ---- ---- +   !pin15
36  -X-- ---- -X-- ---- ---- ---- ---- ---- +
37  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
38  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
39  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

40  --X- ---- ---- ---- ---- ---- ---- ---- OE
41  ---- -X-- X--- -X-- ---- ---- ---- ---- +
42  ---- X--- ---- X--- ---- ---- ---- ---- +
43  ---- ---- -X-- X--- ---- ---- ---- ---- +   !pin14
44  -X-- ---- X--- ---- ---- ---- ---- ---- +
45  -X-- -X-- ---- ---- ---- ---- ---- ---- +
46  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
47  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

48  --X- ---- ---- ---- ---- ---- ---- ---- OE
49  ---- -X-- X--- -X-- ---- ---- ---- ---- +
50  ---- ---- ---- X--- ---- ---- ---- ---- +
51  ---- ---- -X-- X--- ---- ---- ---- ---- +   !pin13
52  -X-- X--- ---- ---- ---- ---- ---- ---- +
53  ---- X--- -X-- ---- ---- ---- ---- ---- +
54  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
55  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +

56  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX OE
57  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
58  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
59  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +   !pin12
60  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
61  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
62  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
63  XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX +
    |    |    |    |    |    |    |    |    |
 pin2|  3 |  4 |  5 |  6 |  7 |  8 |  9 |
    pin1  18   17   16   15   14   13   11
```

Abb. 51.39 Auszug aus der Fuse-Map (Beispiel 2)

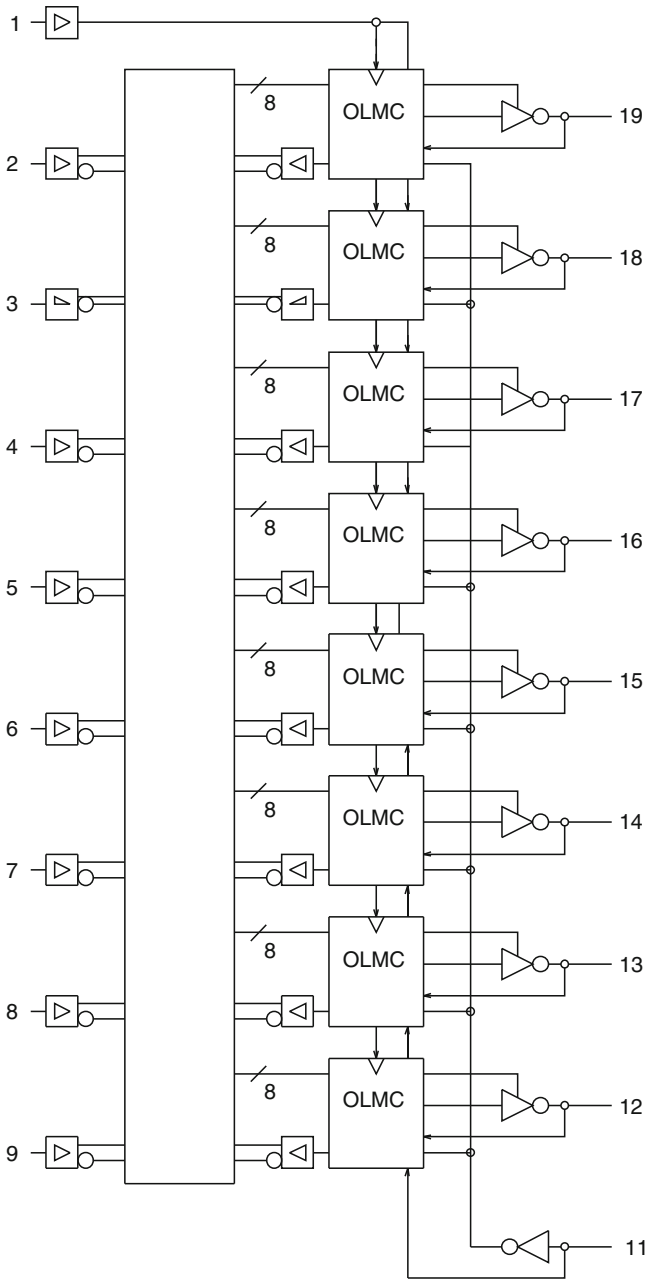


Abb. 51.40 GAL 16V8

51.8.3.3 GAL

GAL's (Generic Array Logic) sind von der Firma Lattice Semiconductor entwickelt worden. Sie können als eine Weiterentwicklung der PAL's aufgefasst werden. Diese Bausteine können PAL's ersetzen und bis zu 100mal elektrisch gelöscht und neu programmiert werden. Nach Lattice wird ein Datenerhalt von 20 Jahren garantiert. Die CMOS-Bausteine besitzen TTL-kompatible Ein- und Ausgänge, und die Verzögerungszeiten liegen je nach Type zwischen 10 und 35 ns. Unbenutzte Eingänge sind wie bei allen MOS-Bausteinen mit +UCC oder GND zu verbinden. Die Programmierung erfolgt mittels geeigneter Programmiergeräte. Das Auslesen der GAL-Funktion kann mit Hilfe einer programmierten Sicherheitszelle verhindert werden (Kopierschutz). GAL's besitzen gegenüber PAL's eine programmierbare Ausgangszelle (OLMC = Output Logic Macro Cell).

Abb. 51.40 zeigt die innere Struktur dieses Bausteins mit seiner Anschlussbelegung. Die programmierbare UND-Matrix ist wie bei den PAL-Bausteinen aufgebaut. Für jeden Ausgang ist eine programmierbare Ausgangszelle vorgesehen.

Abb. 51.41 zeigt eine der OLMC's im Zusammenhang und ihre Programmiermöglichkeiten. Je nach zugeführtem Kontrollwort ergibt sich eine der aufgeführten Ausgangskonfigurationen.

Zur Programmierung von GAL-Bausteinen sind Programmiergeräte erforderlich. Ihre Handhabung ist einfach, da sie nur an einem Computer angeschlossen werden. Der Schaltungsentwurf hängt vor allem von den Möglichkeiten der zugehörigen Software ab. Die eigentliche Programmierung erfolgt wie bei den PAL-Bausteinen. Nach der Eingabe von Boole'schen Gleichungen oder von Wertetabellen mit einem Editor erzeugt ein GAL-Assembler die erforderliche JEDEC-Datei. Die Anschlussbelegung für den Programmierbetrieb unterscheidet sich vom Arbeitsbetrieb nach Abb. 51.42. An den Anschlüssen RAG0 bis RAG5 erfolgt die Adressierung der 64 Zeilen des Bausteins. Die Daten werden seriell mit dem Anschluss CLK dem Dateneingang SDIN zugeführt und zu Kontrollzwecken (verify) dem Datenausgang SDOUT entnommen.

Abb. 51.43 zeigt das Prinzip mit der Zuordnung der Adressen. Am Anschluss P, /V wird dem Baustein mitgeteilt, ob programmiert oder ob Daten ausgelesen werden sollen. Im Programmierbetrieb muss dem Anschluss EDIT eine Spannung von 16,5 V zugeführt werden.

51.8.3.4 pLSI, ispLSI

pLSI-Bausteine (pLSI = programmable Large Scale Integration) und ispLSI (in-system pLSI)-Bausteine der Firma Lattice gehören zu den PLD's, mit denen auch komplexere Schaltungen realisiert werden können. Im Abb. 51.44 ist die Struktur dieser Bausteine beispielhaft dargestellt. Tab. 51.5 gibt eine Übersicht über die Familienmitglieder.

Im GRP (Global Routing Pool) liegen die Verbindungen zu den GLB's (Generic Logic Blocks) und zu den Ein- und Ausgängen. Das Prinzip eines GLB beschreibt Abb. 51.45. Über 18 Eingänge können 20 Produktterme gebildet und den Ausgängen kombinatorisch über AND, OR oder XOR oder über Register zugeführt werden. Bei den Registern kann man zwischen D-, T- und JK-Registern wählen.

Zur Programmierung der ispLSI-Bausteine ist kein externes Programmiergerät erforderlich. Die Pro- grammierdaten werden dem Baustein über ein Interface von der Druckerschnittstelle des PC zugeführt.

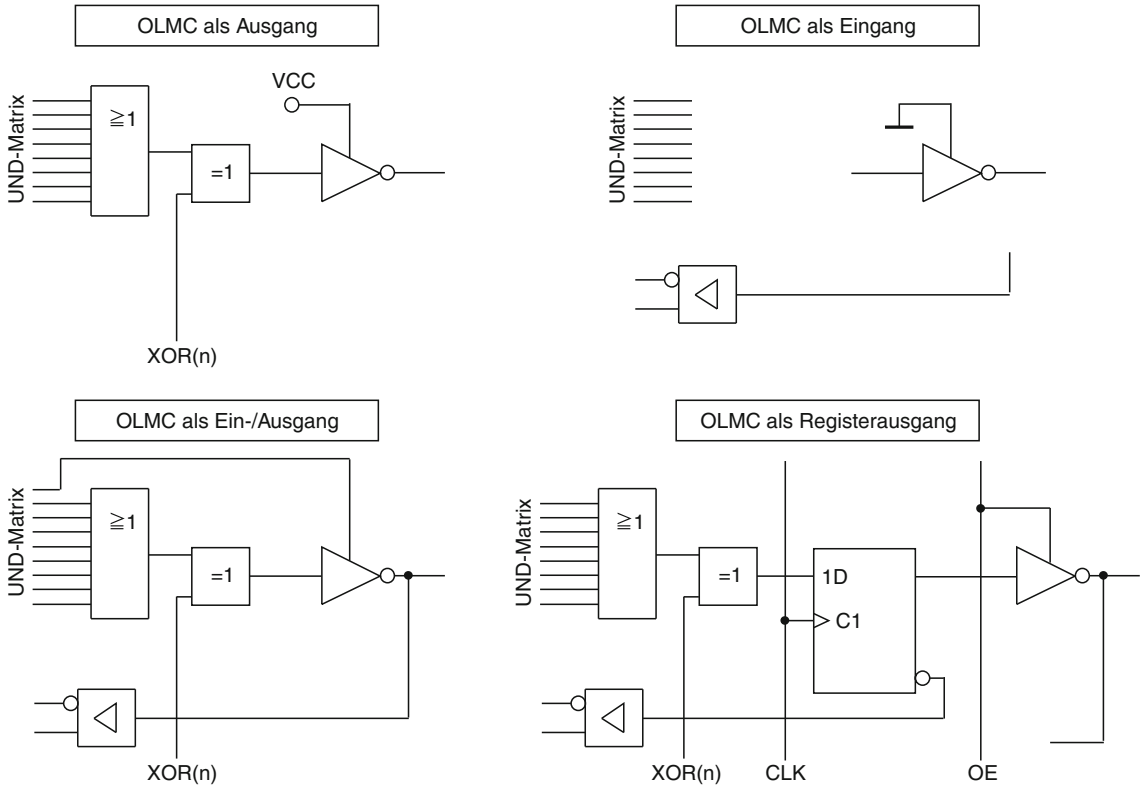


Abb. 51.41 OLMC und ihre Programmiermöglichkeiten

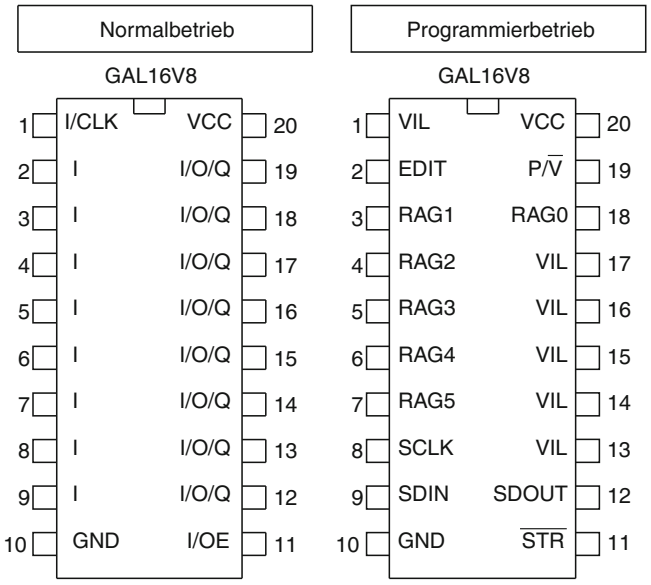


Abb. 51.42 Anschlussbelegung im Arbeits- und Programmierbetrieb

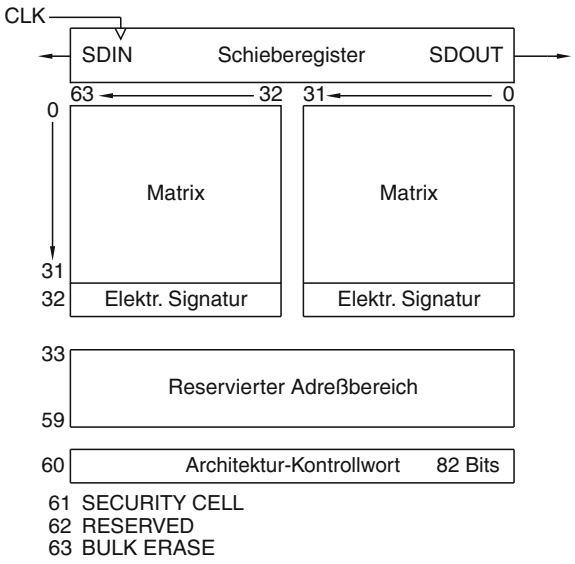


Abb. 51.43 Interne Struktur des GAL 16V8

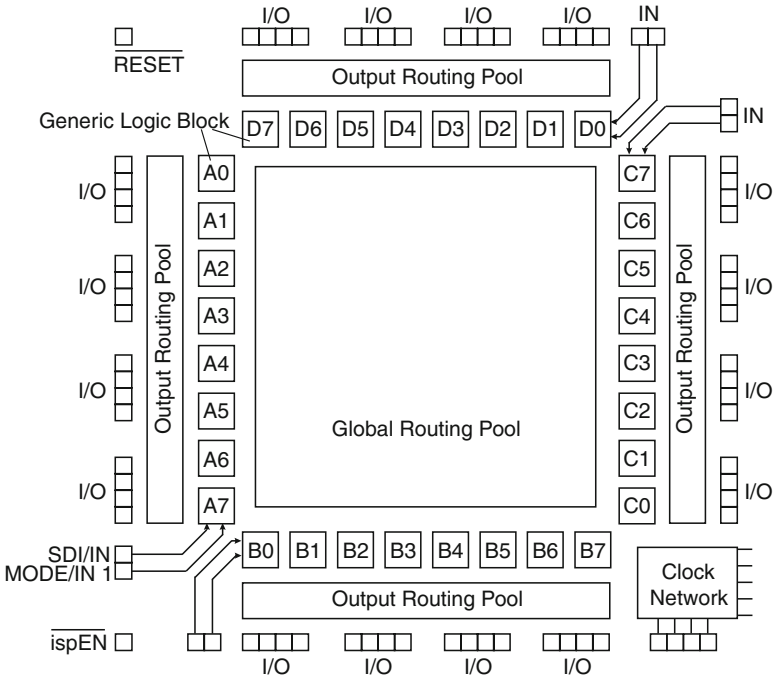


Abb. 51.44 Baustein 1032

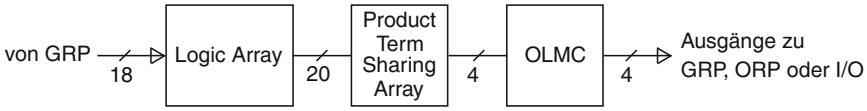


Abb. 51.45 Prinzip GLB

pLSI- und ispLSI-Familie:

Type	1016	1024	1032	1048
f_{max} (MHz)	80	80	80	70
tpd (ns)	15	15	15	20
GLB's	16	24	32	48
Register	96	144	192	288
I/O	36	54	72	106
Pin/Gehäuse	44/PLCC	68/PLCC	84/PLCC	120/PLCC

Beispiel

Schalter ispGDS22

Es handelt sich hierbei um programmierbare Schalter (GDS = General Digital Switch) der Firma Lattice Semiconductor. Sie ersetzen die bekannten mechanischen DIP-Schalter. Die Bezeichnung isp (in-system programmable) bedeutet, dass es sich um einen im System programmierbaren Baustein handelt. Weitere Bausteine dieser Serie sind die IC's ipsGD18 und ipsGD14. Sie unterscheiden sich nur in der Zahl der Anschlüsse und damit der Anzahl der programmierbaren Schalter. ◄

Über eine programmierbare Schaltermatrix (Abb. 51.46) können jeweils Schalterverbindungen von Ax nach Bx hergestellt werden. Hierbei ergeben sind nach Abb. 51.47 verschiedene Konfigurationsmöglichkeiten. Der Baustein erlaubt Signallaufzeiten von 7,5 ns und benötigt einen Strom im Stand-by-Betrieb von 25 mA.

In der Funktionstabelle sind die verschiedenen Programmiermöglichkeiten aufgeführt. Beispielsweise ist der Schalter S nur geschlossen, wenn C0 H-Pegel und C1 L-Pegel besitzt. Durch C0 = „0" wird der 3-state-Treiber, der das Signal vom Multiplexer erhält, gesperrt.

Die Programmierung des Bausteins kann beispielsweise über eine parallele Schnittstelle eines Mikrocontrollersystems oder über den PC und ein Interface direkt erfolgen. Hierbei sind jeweils nur 4 Zuleitungen nötig (Abb. 51.48). Die Programmierdaten müssen mit Hilfe eines Taktes (CLK) seriell am Bausteinanschluss SDIN eingelesen werden. Die Betriebsart wird am Anschluss MODE durchgeführt, und die Kontrolle erfolgt durch Lesen der Daten am PIN SDOUT.

Abb. 51.49 beschreibt über das State-Diagramm die notwendigen Programmierzustände. Die Pegelzustände sind für die Signale MODE und SDIN beschrieben. Um beispielsweise die

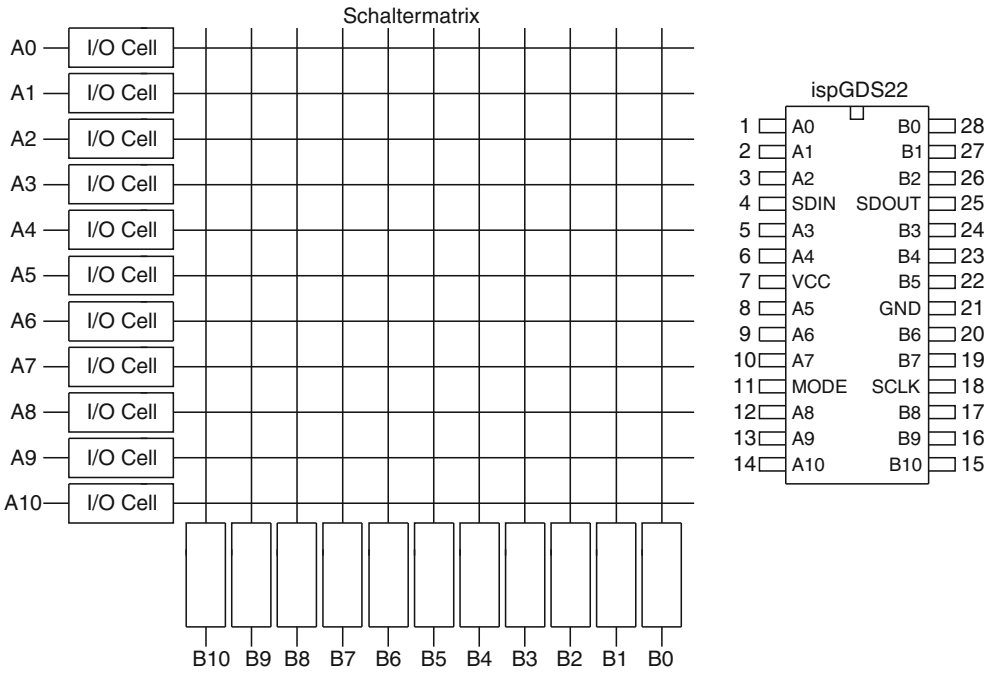
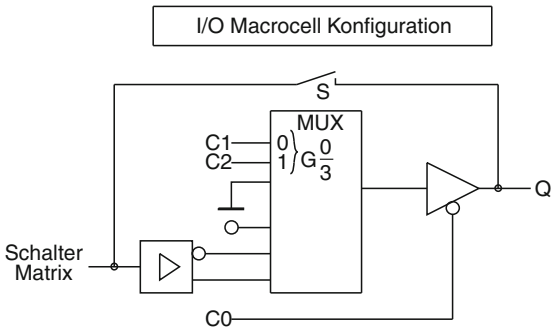


Abb. 51.46 Schaltermatrix und Bausteinanschlüsse des ipsGDC22



Funktionstabelle

C2	C1	C0	Q	Bemerkung
0	0	0	0	(TTL Low)
0	1	0	1	(TTL High)
1	0	0	0	(Aktiv Low)
1	1	0	1	(Aktiv High)
1	0	1	X	(Eingangspegel)

Abb. 51.47 Konfigurationsmöglichkeiten

Identifikationsnummer des Bausteins lesen zu können, wird der Baustein mit MODE = „H“ und „SDIN“ = „L“ unter Zuführung eines Taktes am Anschluss SCLK in den Ausgangszustand (Normal Operation) gebracht. Anschließend wird mit 7 Takten die 8-Bit-Information am Anschluss SDOUT ausgegeben.

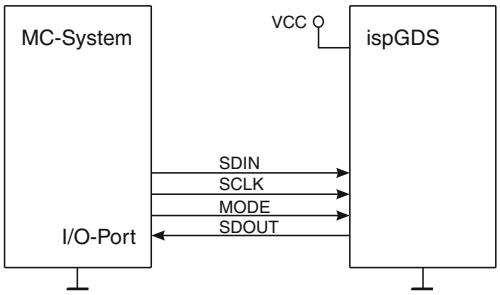


Abb. 51.48 Anschluss an ein MC-System

Die einfache Programmierung dieses Bausteins zeigt das nachfolgende Programm: Die Schalterfunktion ist aus der Zuweisung der PIN's ersichtlich. PIN 20 soll aktiv „H“ und PIN 19 aktiv „L“ sein.

```
title = 'Stupid cross connect test'
"device = ispgds22
"Network
pin 27 = pin 2
pin 26 = pin 3
pin 5 = pin 24
pin 6 = pin 23
pin 20 = h
pin 19 = l
pin 17 = pin 8
pin 16 = pin 9
pin 15 = pin 10
JEDEC file for ispGDS22, created by GASM v1{,}0
```

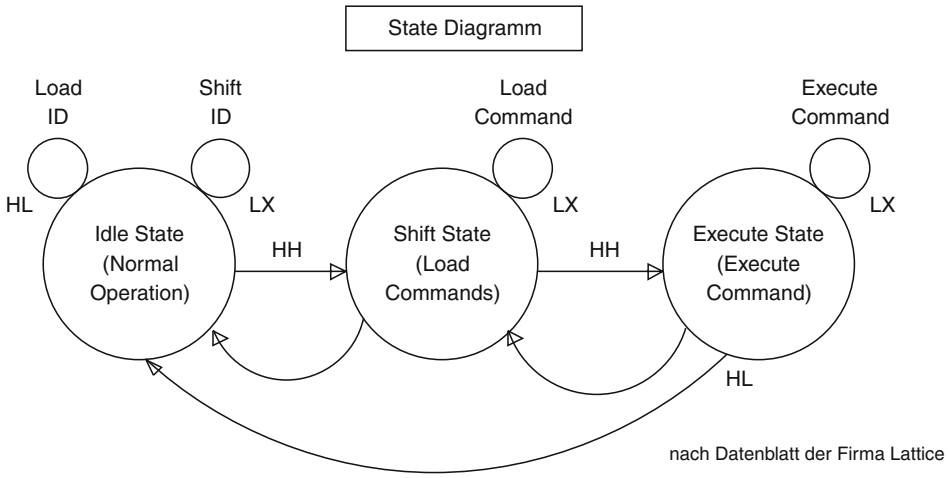


Abb. 51.49 State-Diagramm

```
*
F0*
QP28*QF219
L0000
11111111110
11111111101
11111111011
11111110111
11111101111
11011111111
1011111111
0111111111
1111111111
1111111111
1111111111*
L0121
00000000000000000000000000000000*
L0153
000111000001111100111*
L0175
111001101111100011000
L0197
111111001111111111111*
C1464*
```

Für viele Logik-Compiler besteht die Schaltungsbeschreibung aus einem Deklarationsteil und einer Beschreibung der gewünschten Logik. Im Deklarationsteil wird der gewählte Baustein eingetragen und in einem weiteren Teil erfolgt für die Anschlussbelegung eine gewünschte sinnvolle Namenszuordnung (Definition der Signale). Bei einigen Compilern sind Kenntnisse über die Innenschaltung des Bausteins (Datenbuch) zur Beschreibung erforderlich, um Ein- und Ausgänge nicht zu vertauschen. Andere Compiler geben eine Maske mit der Anschlussbelegung des Bausteins aus. Das nachfolgende Programm dient als Beispiel.

```
title = 'Schalter Test1'
"device = ispgds18
"Network
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
pin 24 = pin 1
```

Nach Aufruf des Assemblers erscheint die nachfolgende Meldung:

GASM: ispGDS Assembler

copyright © 1992 Lattice Semiconductor, written by Guy Townsend

Reading Source File

Fuse Checksum =

JEDEC file successfully written to disk

DOC file successfully written to disk

Der JEDEC file gibt Aufschluss über die jeweiligen Verbindungen, die an der „0“ zu erkennen sind.

JEDEC file for ispGDS18, created on 1/4/1994

Created by GASM v1,0

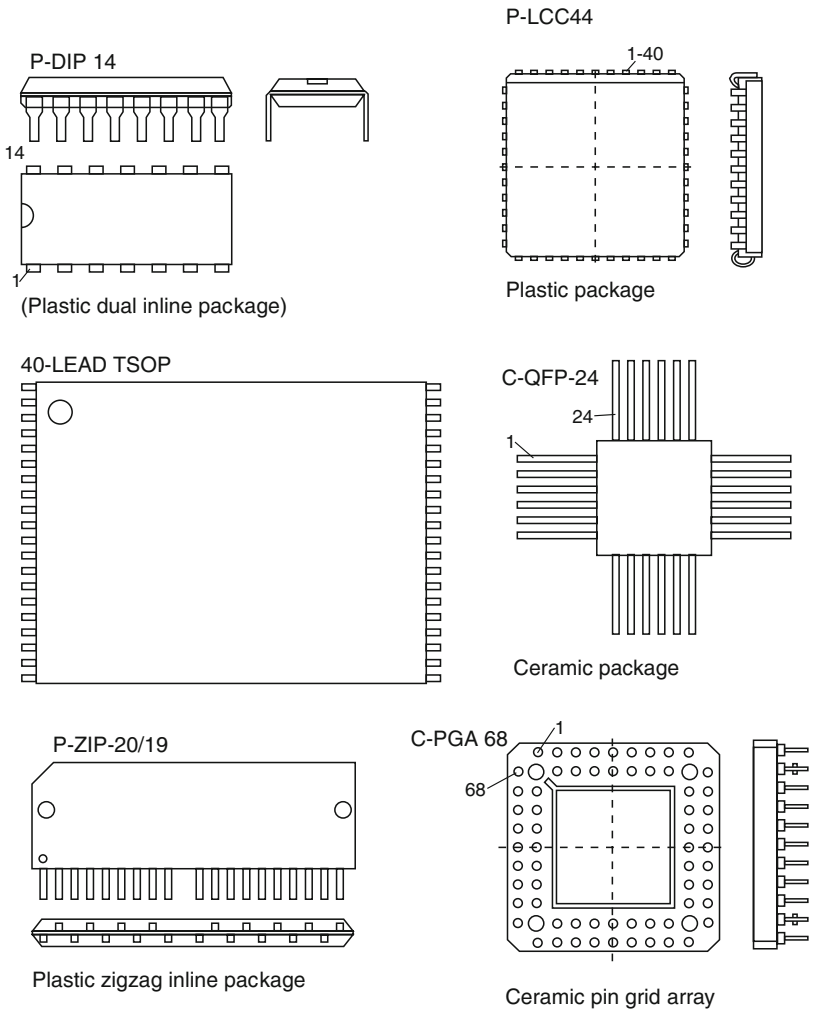


Abb. 51.50 Gehäuseformen

51.9 Gehäuse

Die Hersteller bieten digitale Schaltungen in unterschiedlichen Gehäusen an. Bei den Standardbausteinen verwendet man hauptsächlich das DIL-(Dual-In-Line-)Gehäuse. Der PIN-Abstand beträgt hierbei 1/10 Inch (2,54 mm). Bei Prozessoren und Controllern werden PLCC-(Plastic Leadless Chip Carrier-) Gehäuse eingesetzt. Das nachfolgende Abb. 51.50 gibt eine Übersicht gängiger Gehäuseformen mit ihren Bezeichnungen.

Mikrocomputertechnik

52

Teil VII

Heribert Gierens



52.1 Komponenten eines Mikrocomputers 662
52.2 Mikroprozessoren 662
52.3 Halbleiterspeicher 676
52.4 Peripheriebausteine 689
52.5 Mikrocontroller 708
52.6 Maschinensprache 718
52.7 Befehlsvorrat 726
52.8 Hinweise zur Programmierung und Progammbeispiele 726

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_52

52.1 Komponenten eines Mikrocomputers

Mikrocomputer (Abb. 52.1) bestehen aus hochintegrierten Funktionseinheiten wie der *Zentraleinheit* (CPU = Central-Process-Unit), dem eigentlichen Mikroprozessor, *Speichereinheiten*, wie ROM (Read Only Memory = Nur-Lese-Speicher) und RAM (Random-Access Memory = Schreib-Lese-Speicher), die das auszuführende Programm in Maschinensprache enthalten sowie *Ein- beziehungsweise Ausgabeeinheiten*, die die Verbindung mit peripheren Geräten wie Steuer- und Messeinrichtungen oder Bildschirm, herstellen.

Ein- und Ausgabeeinheiten werden auch als Schnittstellen (Interfaces) bezeichnet. CPU, Speicher und E/A-Einheit ergeben die *Hardware* des Mikrocomputers. Untereinander verbunden sind die Hardwarekomponenten durch Leitungen, denen unterschiedliche Aufgaben zugeordnet sind. Die jeweils parallel liegenden Leitungen werden auch *Bussystem* genannt. Man unterscheidet daher nach *Daten-, Adress-* und *Steuerbus*. Damit der Mikroprozessor Funktionen ausführen kann, ist außerdem *Software* erforderlich, wie beispielsweise das Betriebs- oder Startprogramm.

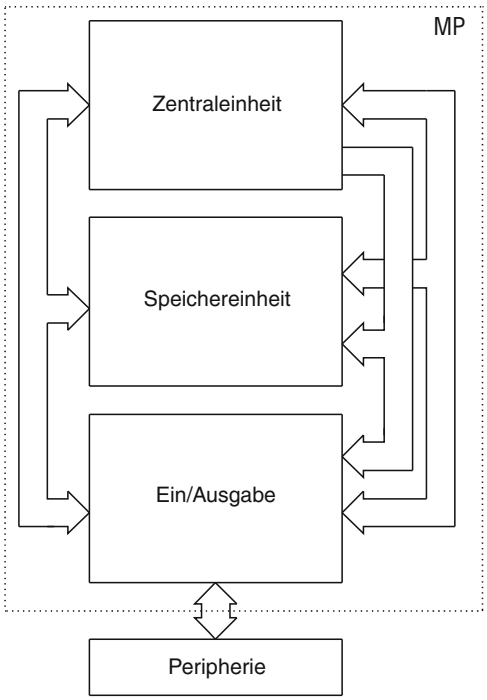


Abb. 52.1 Mikrocomputersystem

52.2 Mikroprozessoren

52.2.1 Allgemeines

Man unterscheidet nach der Architektur der CPU *CISC-Prozessoren* (Complex Instruction Set Computer), *RISC-Prozessoren* (Reduced Instruction Set Computer) und *Signalprozessoren*. Das RISC-Konzept unterscheidet sich von der in PC's (Personal Computer) eingesetzten CISC-CPU, die viele Befehle mit komplexen Funktionen und damit mehreren Maschinenzyklen besitzt, im Wesentlichen durch einen gekürzten, einfacheren Befehlssatz und durch eine schnellere Befehlsabarbeitung, da die meisten Befehle nur einen Maschinenzyklus benötigen. Während RISC- und CISC-Prozessoren nach dem Prinzip der Rechnerarchitektur arbeiten,wie sie von *Von Neumann* entwickelt wurde, besitzen digitale *Signalprozessoren* eine Architektur, bei der die Bussysteme für Daten und Programme getrennt sind. Alle Befehle, die von der CPU ausgeführt werden können, stellen den Befehlssatz (Instruction Set) des Prozessors dar. Der Mikroprozessor kann nur Befehle ausführen, die in binärer Form vorliegen (Maschinensprache). Der von den Prozessoren angesprochene mögliche Adressraum ist von der Anzahl seiner Adressleitungen abhängig. Hieraus ergibt sich die mögliche Informationsmenge (Speicherkapazität), die der Prozessor verarbeiten kann. Der *Adressbus* ist *unidirektional* ausgelegt, da die Adressen nur von der CPU aus übertragen werden, während der *Datenbus*, über den der gesamte Datenverkehr abgewickelt wird, *bidirektional* ausgelegt ist, weil Daten von der CPU und zur CPU gelangen müssen. Eine wichtige Kenngröße ist die Breite des Datenbusses. Man unterscheidet je nach Datenverarbeitungsbreite 8-Bit-, 16-Bit-, 32-Bit- und 64-Bit-Prozessoren. Die Leistungsfähigkeit von Prozessoren wird durch die Zahl möglicher Befehle pro Sekunde (MIPS = Million Instructions Per Second) bestimmt. Dies hängt wesentlich von der Taktfrequenz ab, mit der das System arbeitet. Mit dem *Steuerbus* bestimmt die CPU, mit welchen Baugruppen (z. B. Speicher oder E/A-Bausteine) eine Kommunikation stattfinden soll.

52.2.2 Architektur

Mikroprozessoren besitzen weitgehend gleiche Grundfunktionseinheiten wie *Steuerwerk, Rechenwerk* und *Speicher-* oder *Registerwerk*. Über ein internes Bussystem (Adressen-, Daten- und Steuerbus) erhält der Prozessor Zugang zu den Einheiten des Mikrocomputers. Das Steuerwerk holt einen Befehl von einer externen Speicherstelle zur CPU und decodiert ihn im Befehlsregister. Unter eventueller Einbeziehung von Rechen- und Registerwerk veranlasst das Steuerwerk dann die Ausführung

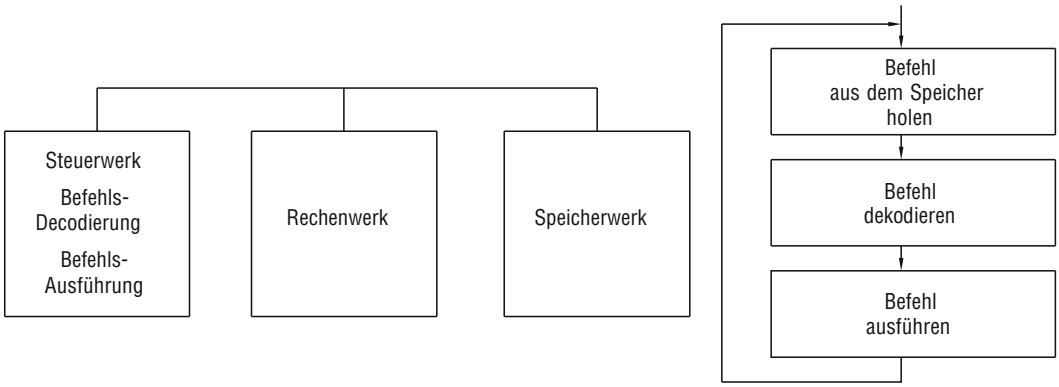


Abb. 52.2 Operationsprinzip

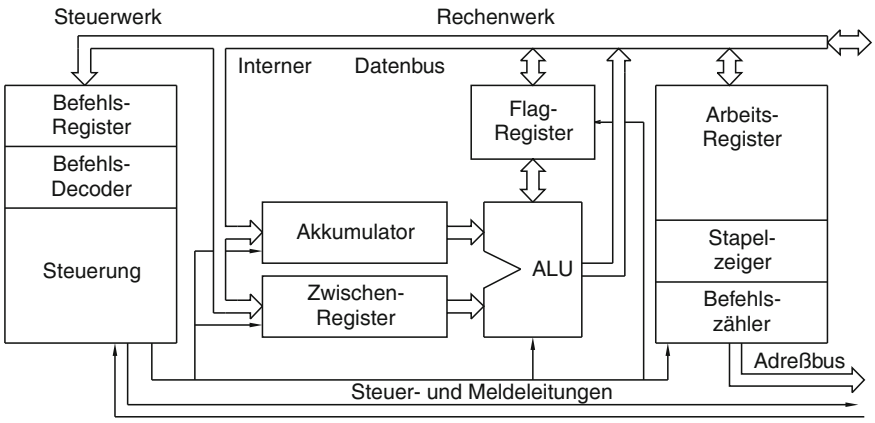


Abb. 52.3 Blockbild einer CPU

des Befehls. Der Mikroprozessor arbeitet dabei in zyklischen Zeitabläufen. Er liest einen Befehl (Fetch), führt die Operation durch (Execute) und holt den nächsten Befehl (Von Neumann Architektur und Operationsprinzip) (Abb. 52.2).

Die Aufgabe des *Steuerwerkes* besteht in der Befehlsannahme durch ein Befehlsregister (Instruction-Register), der anschließenden Decodierung und der Steuerung zur Durchführung des Befehls. Bei der Decodierung wird die eingehende Bit-Folge in ein Mikroprogramm umgewandelt. Ferner muss das Steuerwerk auf externe Signale reagieren. Beispielsweise muss nach einem RESET der Anfangszustand der CPU erreicht werden. Abb. 52.3 zeigt schematisch die wesentlichen Bestandteile eines Prozessors, die nachfolgend beschrieben und am Beispiel der 8-Bit-Prozessoren 8085 und Z80 konkretisiert werden.

Das *Rechenwerk* mit dem Akkumulator und dem Zwischenregister, in denen sich die Operanden befinden, kann arithmetische und logische Operationen mit Hilfe der *ALU* (Arithmetic-Logic-Unit) durchführen. Die ALU besteht im Prinzip aus einem Paralleladdierwerk und einer logischen Einheit, mit der die Verknüpfungen NICHT, UND, ODER und XOR durchgeführt werden können. Das *Kennzeichen-*, *Status-* oder *Flagregister* gibt Aufschluss über die Ergebnisse der Operationen, beispielsweise ob das Ergebnis Null (Zero) war oder ob ein Übertrag (Carry) aufgetreten ist. Es wird so möglich, in Abhängigkeit der Zustandsbits Sprünge zu programmieren (siehe auch Abschn. 52.6 „Maschinensprache“). Dem Rechenwerk ist ein *Speicherwerk* angeschlossen. Dort befinden sich allgemeine *Register* (Arbeitsregister), in denen vorübergehend Zwischenergebnisse abgespeichert werden können. Ein Register besitzt meist eine auf ein Datenwort beschränkte Kapazität (8-, 16- oder 32-Bit).

Register mit bestimmten Aufgaben sind der *Befehlszähler* (Program-Counter) und der *Stapelzeiger* (Stackpointer). Der Befehlszähler zeigt immer auf die nächstfolgende Adresse, die der Prozessor abarbeiten soll. Er erhöht sich nach jedem Adressenaufruf automatisch jeweils um 1 und legt damit die Reihenfolge der Operationen fest. Speicherstelle für Speicherstelle wird somit nacheinander abgearbeitet (Abb. 52.4).

Nur bei Sprungbefehlen, Unterprogrammaufrufen oder Unterbrechungsanforderungen wird der Befehlszähler auf einen neuen Wert gesetzt.

Der Stapelzeiger (siehe auch Abschn. 52.6 „Maschinensprache“) ist ein besonderer Adressenzeiger, der immer auf die Adresse des Stapelspeichers zeigt, in der die letzte Eingabe erfolgte (Ende des Stack). Als Stapelspeicher (Stack oder Kellerspeicher) kann jeder beliebige Teil des Arbeitsspeichers (RAM) verwendet werden. Bei jeder Dateneingabe wird der Stapelzeiger *decrementiert* (erniedrigt), so dass sich der „Stapel“ von einer vorgegebenen Adresse zu immer niedrigeren Adres-

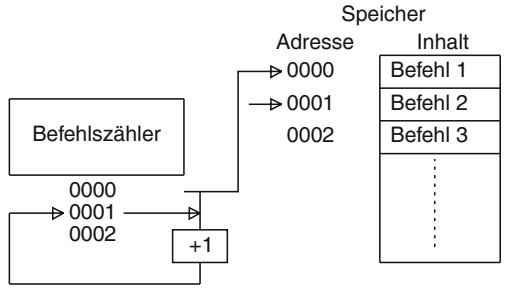


Abb. 52.4 Prinzip Befehlszähler

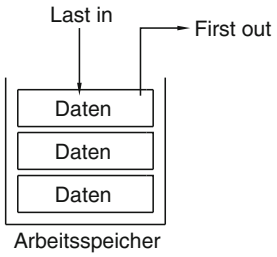


Abb. 52.5 Lifo-Prinzip

sen aufbaut. Die Daten werden in der Reihenfolge, in der sie eintreffen, auf den Stack gebracht und müssen in umgekehrter Reihenfolge zurückgeholt werden (Abb. 52.5). Man bezeichnet ihn daher auch als *LIFO-Speicher* (Last In, First Out).

Aufgabe des Stack ist es, bei *Interrups* (Unterbrechungsanforderungen) oder Unterprogrammaufrufen die Hauptprogrammadresse zu speichern, um nach Rückkehr in das Hauptprogramm bei richtiger Adresse mit dem Programm fortfahren zu können (siehe auch Abschn. 52.6 „Maschinensprache“). Außerdem dient er zur Datensicherung der Arbeitsregister des Hauptprogramms, wenn einzelne Register im Unterprogramm benötigt werden.

Um die genaue zeitliche Folge der Vorgänge (Timing) durchführen zu können, benötigt der Prozessor einen quarzgesteuerten Taktgenerator. Dieser liefert das Bezugssignal für alle Abläufe in einem Mikroprozessorsystem und bestimmt damit die Arbeitsgeschwindigkeit des Mikrocomputers. Bei einigen Prozessoren ist der Generator integriert, so dass nur noch ein externer Quarz angeschlossen werden muss.

52.2.3 Übersicht gängiger Mikroprozessoren

Tab. 52.1 zeigt Beispiele gängiger Mikroprozessoren mit möglichem Adressraum, Daten- und Adressbusbreite und Taktfrequenz. Die Mikroprozessoren einer Familie mit weitergehenden Eigenschaften können Maschinenprogramme ihrer Vorgängertypen ausführen, sie sind also *aufwärtskompatibel*.

Tab. 52.1 Übersicht gängiger Mikroprozessoren

Typ	Adressraum	Datenbus Adress-/Datenbus	Adressbus	Taktfrequenz
6502	64 kB	D0 . . . D7	A0 . . . A15	4 MHz
Z80	64 kB	D0 . . . D7	A0 . . . A15	4 MHz
8080	64 kB	AD0 . . . AD7	A8 . . . A15	4 MHz
8085	64 kB	AD0 . . . AD7	A8 . . . A15	4 MHz
8086	1 MB	AD0 . . . AD15	A16 . . . A19	10 MHz
68000	16 MB	D0 . . . D15	A0 . . . A23	16 MHz
80286	16 MB	D0 . . . D15	A0 . . . A23	16 MHz
68020	4 GB	D0 . . . D31	A0 . . . A31	16 MHz
80386	4 GB	D0 . . . D31	A0 . . . A31	16 MHz
80486	4 GB	D0 . . . D31	A0 . . . A31	25 MHz

Tab. 52.2 Bedeutung der Anschlüsse

AD0 . . . AD7	Adress-/Datenbus
A8 . . . A15	Adressbus
/RD	Ausgangssignal zum Lesen einer Speicherstelle oder eines Portbausteins
/WR	Ausgangssignal zum Schreiben einer Speicherstelle oder eines Portbausteins
ALE	Steuersignal zur Adressenzwischenspeicherung des L-Adressbytes
X1, X2	Quarzanschluss
IO//M	E/A-Bausteine und Speicherunterscheidung
S0, S1	Kontrollsignale und Statussignale
Trap	Interrupteingänge mit festen Verzweigungsadressen
RST7.5, RST6.5, RST5.5	Interrupteingänge mit festen Verzweigungsadressen
INTR	Interrupteingang mit vereinbarter Verzweigungsadresse
/INTA	Interruptbestätigung von INTR
SID	Serieller Eingang
SOD	Serieller Ausgang
HOLD	Eingang zur Abkopplung der CPU (Aufruf externer Geräte)
HLDA	Bestätigung von HOLD durch die CPU
/RESET IN	Rücksetzen (ADR 0000)
RESET OUT	Quittung auf RESET IN
READY	Wartezustand der CPU

52.2.4 8-Bit-Mikroprozessoren

52.2.4.1 8085-CPU

Abb. 52.6 zeigt die Pinbelegung (Anschlussbelegung) und den Registerplan, Abb. 52.7 die Architektur des Prozessors. Tab. 52.2 gibt die Bedeutung der Anschlüsse wieder. Der Prozessor 8085 ist mit den 8-Bit-Arbeitsregistern B, C, D, E, H und L und den 16-Bit-Registern Stapelspeicher, Befehlszähler und Adressenzwischenspeicher ausgestattet. Die Arbeitsregister B/C, D/E und H/L stehen für die Datenspeicherung und bei der Durchführung von Befehlsoperationen zur Verfügung.

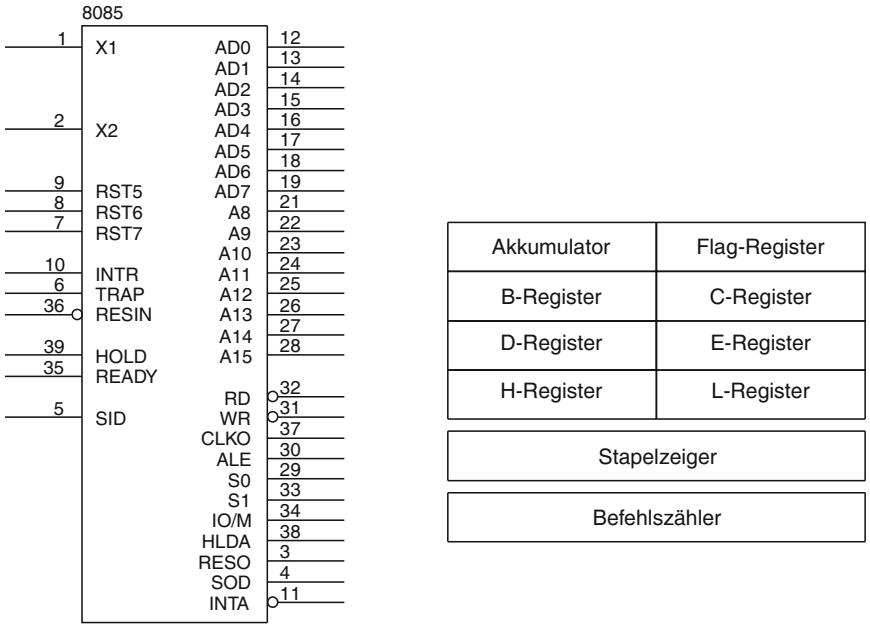


Abb. 52.6 PIN-Belegung und Registerplan der 8085 CPU

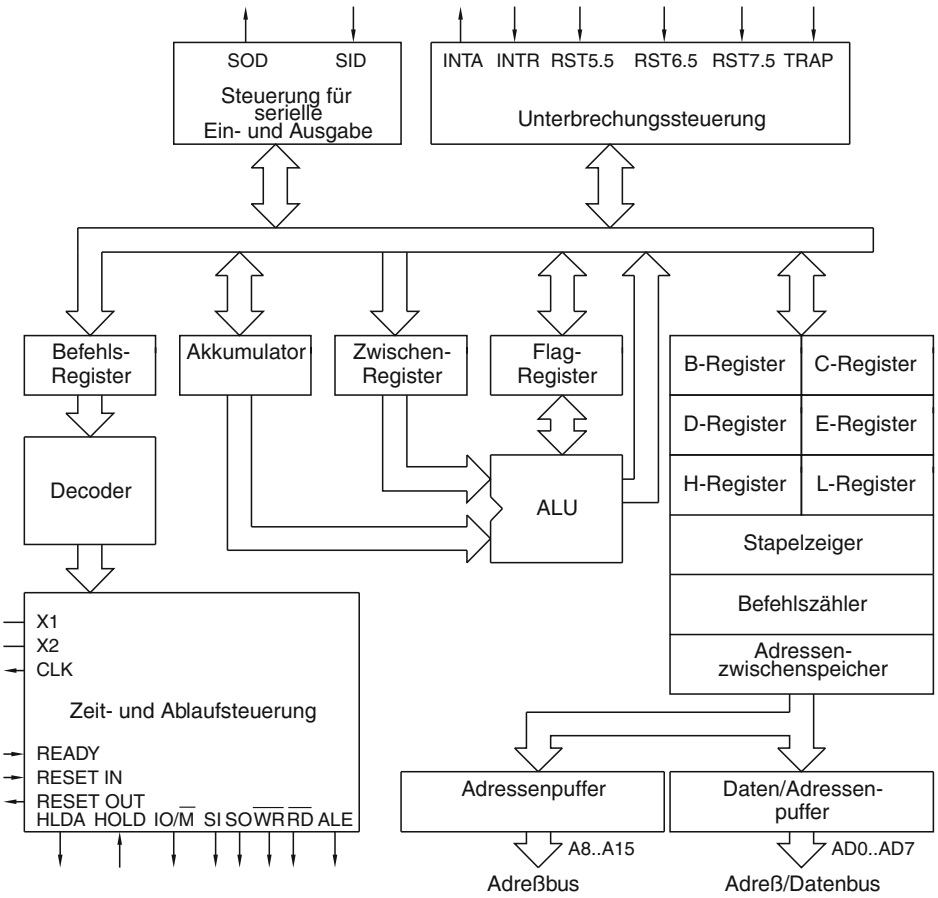


Abb. 52.7 8085-Prozessor CPU

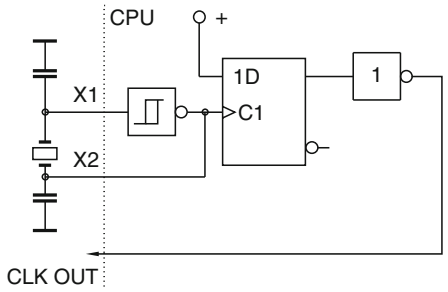


Abb. 52.8 Takterzeugung

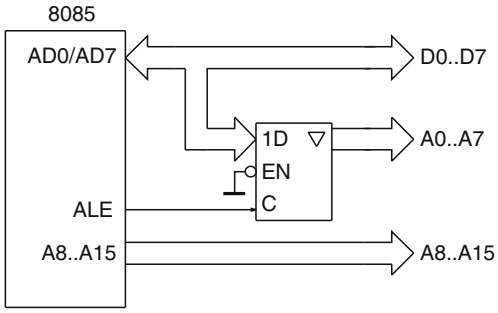


Abb. 52.9 Adressenzwischenspeicher

Akkumulator, Zwischenregister und ALU mit Flagregister und Befehlsregister sind ebenfalls für die Aufnahme von 8 Bit konstruiert. Die Arbeitsregister B/C, D/E und H/L können als Registerpaare (RP) zusammengefasst werden, um 16 Bit aufnehmen zu können.

Die Funktionen innerhalb des Prozessors werden durch die Zeit- und Ablaufsteuerung bestimmt. Der Baustein besitzt hierzu einen internen Generator, der nur von einem zusätzlichen Quarz beschaltet werden muss. Eine Taktgeberschaltung erzeugt die entsprechenden Taktsignale. Der Ausgang CLK gibt die halbe Quarzfrequenz aus (Abb. 52.8).

Der Adressbus ist 16 Bit breit. A0 bis A7 (niederwertiges Adressbyte) wird über den Daten-/Adressbus AD0 bis AD7 im *Zeitmultiplexbetrieb* übertragen. Mit Hilfe der Steuerleitung ALE (Address Latch Enable) können die Adressen zwischengespeichert werden (Abb. 52.9).

Die Anschlüsse A8 bis A15 entsprechen dem höherwertigen Adressbyte des Adressbusses. Mit 16 Adressleitungen wird der mögliche Adressraum $2^{16}\,\text{kB} = 65.536\,\text{kB} = 16\,\text{kB}$ groß. Da der zeitliche Verlauf der Daten und Adressen bei Darstellung jeder Bit-Position unübersichtlich wird, werden hier Daten- und Adresseninhalte schematisch in einem Liniendiagramm zusammengefasst. Abb. 52.10 verdeutlicht dies an einem Beispiel.

Während die Adressen A8 bis A15 konstant bleiben, werden die Adressen A0 . . . A7 und Daten D0 bis D7 umgeschaltet. Adres-

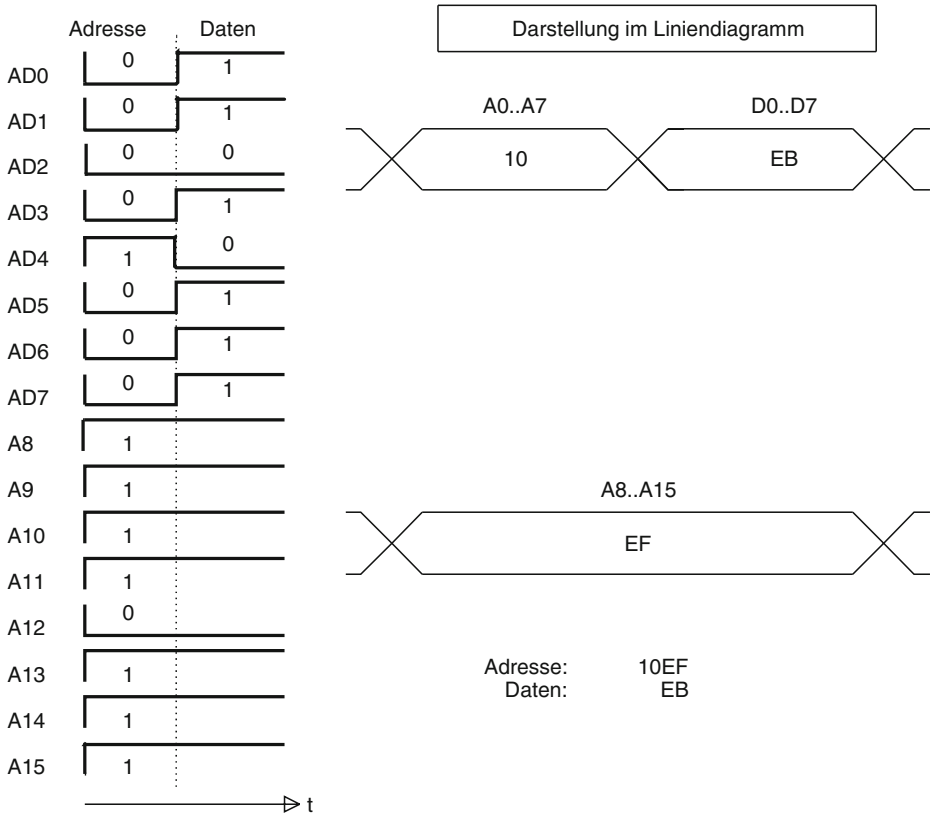


Abb. 52.10 Darstellung der Daten und Adressen

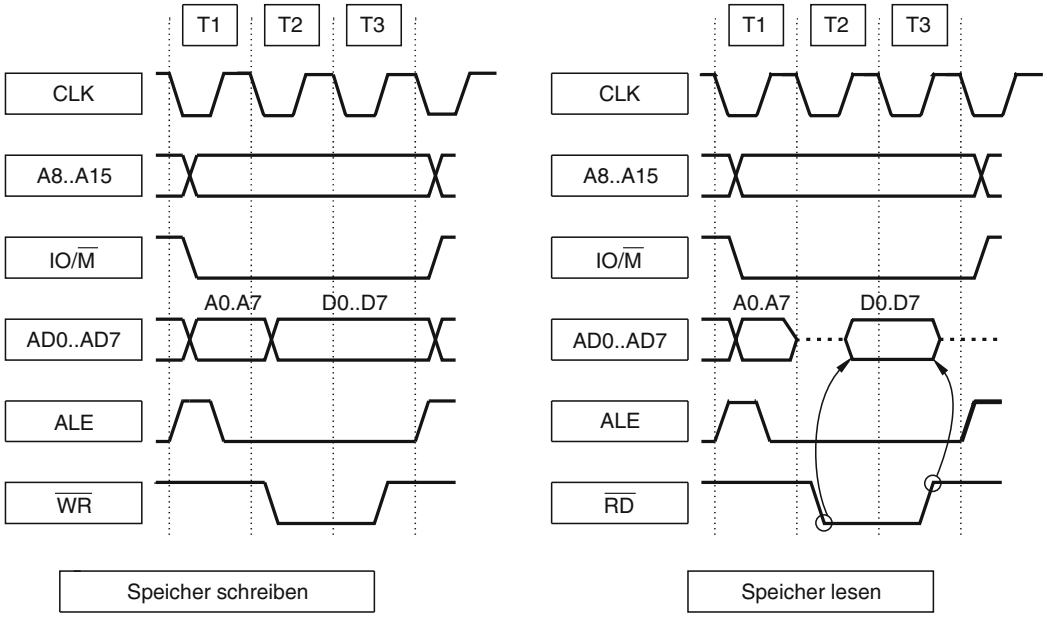


Abb. 52.11 Bustiming

sen und Dateninhalte werden immer in hexadezimaler Form angegeben. Die Darstellung des Zusammenwirkens aller Signale auf dem Bussystem bezeichnet man als *Bustiming*. Abb. 52.11 zeigt die beteiligten Signale während eines Speicherschreib- und Lesezugriffs.

In Datenbüchern findet man häufig zusätzlich Pfeile im Zeitliniendiagramm, die gegenseitige Abhängigkeiten kennzeichnen. Sind weder Daten noch Adressen vorhanden, ist der Bus im Ruhezustand; er ist hochohmig. Dies ist durch die gestrichelte Linie bei den Signalen AD0 . . . AD7 im Lesezyklus dargestellt. Beim Lesezyklus gibt die CPU zunächst die 16-Bit-Adresse auf den Adressbus. Mit dem Taktsignal ALE wird eine Adressenspeicherung A0 bis A7 ermöglicht, und die Leseaufforderung an /RD geht danach auf „L“-Pegel. Der Datenbus wird kurz hochohmig und gibt dann dem angewählten Speicherplatz die Möglichkeit, Daten auf den Datenbus zu bringen. Die Daten stabilisieren sich auf dem Bus und werden gelesen. Mit der ansteigenden Flanke des RD-Signals ist der Vorgang beendet. Beim Schreibzyklus werden die Daten sofort nach dem L-Byte der Adresse von der CPU ausgegeben. Für den Schreib- oder Lesezyklus sind jeweils drei Taktzyklen erforderlich.

Der Steuerbus kann mit seinen Anschlüssen in unterschiedliche Aufgabenbereiche gegliedert werden. Die *System-Kontrollsignale* steuern die Baugruppen wie z. B. RD, WR, IO/M, die an das Mikrocomputersystem unmittelbar angeschlossen sind. Mit Low am Ausgang /RD oder /WR (Low-aktive Ausgänge) teilt die CPU mit, ob sie lesen oder schreiben will. Hierbei ist eine Verknüpfung mit den Adresssignalen notwendig, damit ein bestimmter E/A-Baustein adressiert oder eine bestimmte Speicherzelle beschrieben oder ausgelesen werden kann. Die Unterscheidung zwischen Speicher und I/O-Baustein wird mit der Steuerleitung IO, /M durchgeführt. Bei „L“ an diesem Ausgang wird lesend oder schreibend auf den Speicher und bei „H“ auf einen I/O-Baustein zugegriffen.

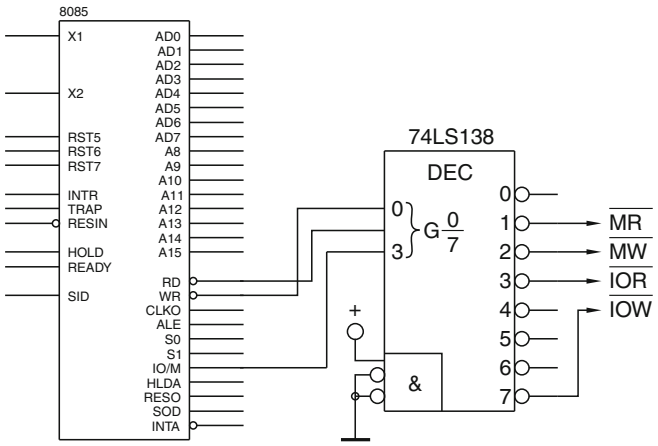
Tab. 52.3 Funktionstabelle 74LS138

E1	E2	E3	A0	A1	A2	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7
1	×	×	×	×	×	1	1	1	1	1	1	1	1
×	1	×	×	×	×	1	1	1	1	1	1	1	1
×	×	0	×	×	×	1	1	1	1	1	1	1	1
0	0	1	0	0	0	0	1	1	1	1	1	1	1
0	0	1	1	0	0	1	0	1	1	1	1	1	1
0	0	1	0	1	0	1	1	0	1	1	1	1	1
0	0	1	1	1	0	1	1	1	0	1	1	1	1
0	0	1	0	0	1	1	1	1	1	0	1	1	1
0	0	1	1	0	1	1	1	1	1	1	0	1	1
0	0	1	0	1	1	1	1	1	1	1	1	0	1
0	0	1	1	1	1	1	1	1	1	1	1	1	0

Zur Unterscheidung zwischen Speicher und I/O-Baustein kann auch ein Decoder eingesetzt werden. Wird der Decoder entsprechend Abb. 52.12 angeschlossen, erhält man an den gekennzeichneten Ausgängen das Signal für „Speicher lesen (/MR)“, „Speicher schreiben (/MW)“, „Ein-Ausgabebaustein lesen (/IOR)“ und „Ein- und Ausgabebaustein schreiben (/IOW)“. Dies lässt sich über die Funktionstabelle (Tab. 52.3) des 74LS138 leicht nachvollziehen.

Die CPU unterscheidet zwischen folgenden Zeitabläufen: Operationsholzyklus, Speicherlese- und -schreibzyklus, E/A-Lese- und Schreibzyklus, Interruptquittung und Bus-Ruhezustand. Allgemein spricht man von Maschinenzyklen, die in unterschiedlichen Verbindungen einen *Befehlszyklus* beschreiben. Der Maschinenzyklus selbst besteht wieder je nach Aufgabe aus einer unterschiedlichen Anzahl von *Taktzyklen*. In Abb. 52.13 ist der Zusammenhang zwischen Takt-, Maschinen- und Befehlszyklus an einem Beispiel aufgezeichnet.

Der erste Maschinenzyklus ist immer ein *Operationscodeholzyklus*. In ihm ist der Operationscode (OP-Code), und damit alle



IO/M	RD	WR	MR	MW	IOR	IOW
0	0	1	0	1	1	1
0	1	0	1	0	1	1
1	0	1	1	1	0	1
1	1	0	1	1	1	0

Abb. 52.12 Steuerung der Schreib- und Lesesignale mit Decoder

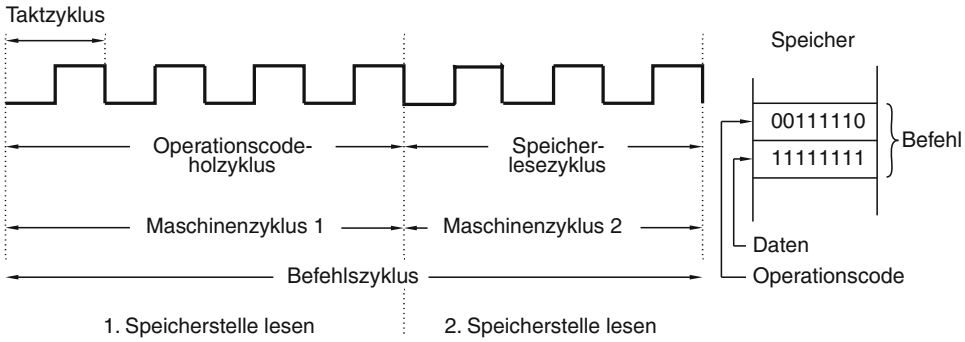


Abb. 52.13 Takt-, Maschinen- und Befehlszyklus

Tab. 52.4 Statussignale

S0	S1	Funktion
0	0	Bus-Ruhezustand
0	1	CPU liest (Speicher, E/A)
1	0	CPU schreibt (Speicher, E/A)
1	1	CPU liest Operationscode

Steuerungsvorgänge für den Prozessor, enthalten. Ob weitere Maschinenzyklen folgen, hängt von dem entsprechenden Befehl ab. Obwohl der Operationscodeholzyklus auch vom Prinzip her ein Speicherlesezyklus ist, besitzt er im Gegensatz hierzu einen Taktzyklus mehr, weil während dieser Zeit die Decodierung im Befehlsregister erfolgen muss. Im aufgeführten Beispiel bekommt der Prozessor mit dem OP-Code (Maschinenzyklus 1), der aus dem Speicher geholt und zur Decodierung dem Befehlsregister zugeführt wird, gesagt, dass er den Inhalt der nachfolgenden Speicherstelle in den Akkumulator laden soll. Folglich befördert der nächste Maschinenzyklus (Maschinenzyklus 2) die Daten zum Akkumulator.

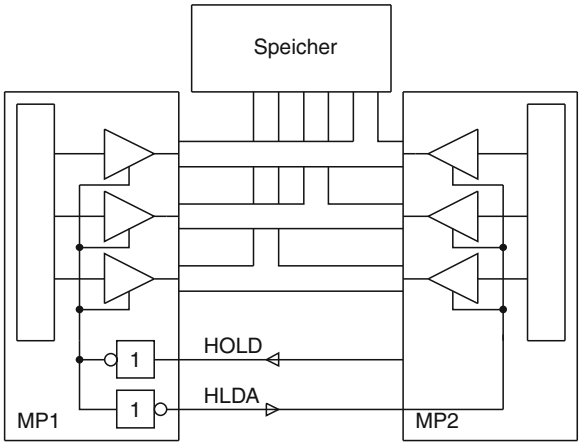
Die *Prozessor-Kontrollsignale* geben Aufschluss über bestimmte Zustände des Prozessors oder geben eine bestimmte Meldung an die CPU. Hierzu gehören die Anschlüsse: S0, S1, RESET IN, RESET OUT, READY. S0 und S1 geben Aufschluss über den Zustand der CPU bei der Befehlsbearbeitung (Tab. 52.4).

Mit RESET IN wird der Programmzähler auf die Adresse 0000 gesetzt, und die Interrupteingänge werden gesperrt. Eine Bestätigung wird an RESET OUT durchgeführt. Damit können weitere Bausteine zurückgesetzt werden. In Abb. 52.20 ist eine übliche RESET-Schaltung mit dargestellt. Beim Einschalten wirkt der Kondensator C als Kurzschluss und ermöglicht durch „L“ einen RESET. Die Diode V sorgt für ein schnelles Entladen des Kondensators, so dass bei kurzzeitigem Aus- und Einschalten des Systems RESET wieder wirksam wird.

Bei READY „L“ geht die CPU in den Wartezustand, bis READY wieder „H“ wird. Dieser Anschluss dient zur Steuerung externer Geräte, die langsamer als die CPU sind. Der READY-Eingang kann auch zur schrittweisen Überprüfung von Programmen genutzt werden.

Bus-Kontrollsignale ermöglichen einem anderen Prozessor oder auch Massenspeichern Zugriff auf das CPU-eigene Bussystem (DMA, Direct Memory Access = direkter Speicherzugriff). Das Bussystem wird von der CPU abgekoppelt. Hierzu zählen die Anschlüsse HOLD und HLDA. Abb. 52.14 zeigt das Prinzip.

Wünscht der MP2 (Mikroprozessor) einen Speicherzugriff, so gibt er eine Meldung „H“ an den Anschluss HOLD des MP1. Dadurch wird das interne Bussystem des Prozessors vom externen Bussystem getrennt (hochohmig). Der MP2 erhält gleichzeitig eine Quittung über die Leitung HLDA mit „H“, um



S0	S1	Funktion
0	0	Bus-Ruhezustand
0	1	CPU liest (Speicher,E/A)
1	0	CPU schreibt (Speicher,E/A
1	1	CPU liest Operationscode

Abb. 52.14 Prinzip DMA

sein internes Bussystem an den externen Bus anzuschließen. In der Praxis werden hier DMA-Kontroller eingesetzt, die externen Geräten das direkte Arbeiten mit dem CPU-eigenen Arbeitsspeicher ermöglichen.

Ein Prozessor muss auf äußere Ereignisse (Impulse, Pegel), wie auf die Meldung von Sensoren, Tastern oder externer Geräte reagieren können. Bei deren Ansprache muss ein Programm eine gewünschte Reaktion auslösen. Eine Möglichkeit besteht darin, alle Sensoren und Geräte softwaremäßig nacheinander abzufragen. Man spricht dann von der *Polling*-Methode. Dies ist allerdings sehr zeitintensiv. Effektiver ist der Aufruf eines Unterprogramms bei der Reaktion eines Sensors. Erreichen kann man das durch eine Programmunterbrechung (Interrupt). Hierunter versteht man den direkten Zugriff externer Geräte auf die CPU. Wird ein Interrupteingang aktiviert, bearbeitet der Prozessor noch den gerade begonnenen Befehl und verzweigt dann in die Interruptroutine, zum Interrupt-Programm, um auf die Anforderung des Interrupts einzugehen. Der 8085 besitzt fünf Interrupteingänge. Zu den Unterbrechungs-Kontrolleitungen gehören die Unterbrechungseingänge (Interrupteingänge) TRAP, RST7.5, RST6.5, RST5.5 mit unterschiedlicher Priorität, um bei gleichzeitiger Aktivierung aller Eingänge eindeutige Zuordnungen zu gewährleisten, und der Eingang INTR mit dem Quittungssignal INTA.

Da gleichzeitige Anforderungen mehrerer Geräte vermieden werden müssen, ist bei vielen Prozessoren ein *Prioritätsdecoder* oder *Interruptcontroller* erforderlich, weil sie nur einen Interrupteingang besitzen. Abb. 52.15 zeigt für den 8085 beide Möglichkeiten auf.

Beim 8085 ist die Priorität hardwaremäßig festgelegt. Tab. 52.5 führt die Hardwareeinsprungadressen und die Prioritäten auf.

Des Weiteren ist der Tabelle zu entnehmen, mit welchem Pegel der Interrupteingang angesprochen werden muss (Triggerung) und welche Interrupteingänge softwaremäßig abgeschaltet werden können. Die Eingänge RST6.5, RST5.5 und INTR müssen mit Pegeln angesprochen werden. Bis zur Annahme durch die CPU muss der Pegel bei den genannten Eingängen anliegen.

Tab. 52.5 Interrupt

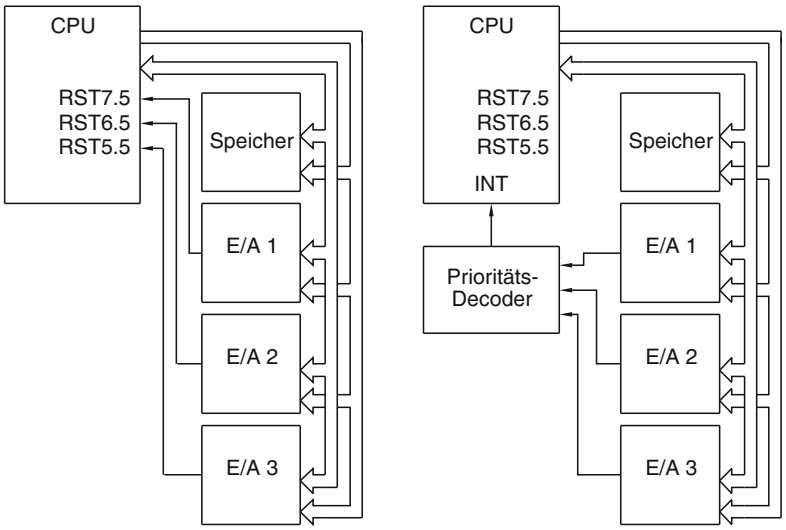
Name	Priorität	Adresse	Eigenschaft	Triggerung
TRAP	1	24	Nicht maskierbar	„L“ → „H“
RST7.5	2	3C	Maskierbar	„H“
RST6.5	3	34	Maskierbar	„H“
RST5.5	4	2C	Maskierbar	„H“

Der Eingang RST7.5 reagiert auf die ansteigende Flanke und speichert die Anforderung in einem Flipflop.

Von der Softwareseite besteht für einige Interrupts die Möglichkeit, diese zu maskieren. Hierunter versteht man, Interrupts zu sperren oder zuzulassen. Die Programmierung läuft über den Inhalt des Akkumulators in Verbindung mit dem Befehl SIM (Set Interrupt Mask) ab. Abb. 52.16 zeigt die vom Hersteller angegebenen Bitpositionen.

Mit den Befehlen EI (Enable Interrupt) können alle Interrupts zugelassen und mit DI (Disable Interrupt) alle gesperrt werden (Ausnahme: TRAP). Das Prinzip verdeutlicht Abb. 52.17 am Beispiel des RST7.5-Eingangs. Nur wenn durch den Befehl SIM die Maske für RST7.5 gesetzt und durch den Befehl EI die Zulassung der Interrupts ermöglicht wurde, kann bei anliegendem Signal an RST7.7 eine Interruptannahme durch die CPU erfolgen, und der Programmzähler ruft die Adresse 3C auf.

Der Interruptvorgang wird in Abb. 52.18 erläutert. Tritt im laufenden Hauptprogramm ein Interrupt auf, wird der gerade begonnene Befehl noch bearbeitet. Der Programmzählerinhalt wird zum Stack gebracht, und der Programmzähler zeigt auf die festgelegte Einsprungadresse, um das Interrupt-Programm ausführen zu können. Das Interrupt-Programm selbst befindet sich im Arbeitsspeicher, da sich zwischen den Interrupteinsprungadressen – hier steht in der Regel nur ein Sprungbefehl zum Arbeitsspeicher – nur wenige freie Speicherplätze befinden (siehe auch Tab. 52.7). Ein Interrupt-Programm ist wie ein Unterprogramm zu behandeln. Registerinhalte sind zu „retten“ und auf den Stack zu bringen. Den Abschluss bildet der Befehl RET (Return). (Programmbeispiele siehe Abschn. 52.6 „Maschinensprache“.)



Name	Priorität	Adresse	Eigenschaft	Triggerung
TRAP	1	24	nicht maskierbar	⌐ und H bis Abfrage
RST7.5	2	3C	maskierbar	H bis Abfrage (Speicherung)
RST6.5	3	34	maskierbar	H bis Abfrage
RST5.5	4	2C	maskierbar	H bis Abfrage
INTR	5			

Abb. 52.15 Interruptmöglichkeiten

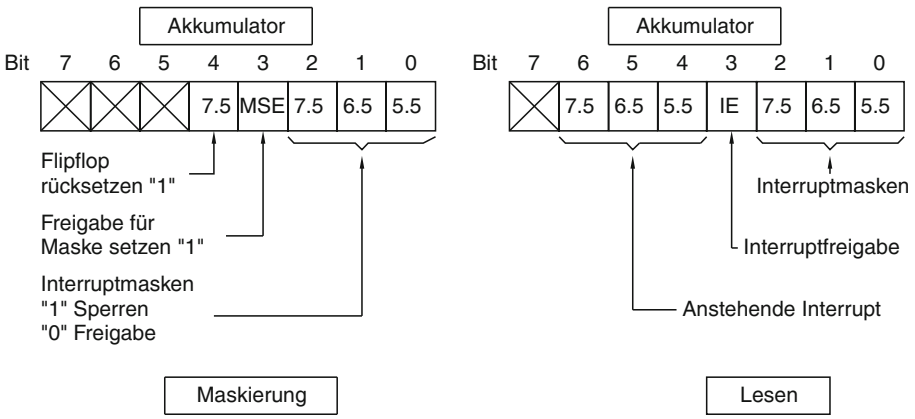


Abb. 52.16 Interruptmaske setzen und lesen

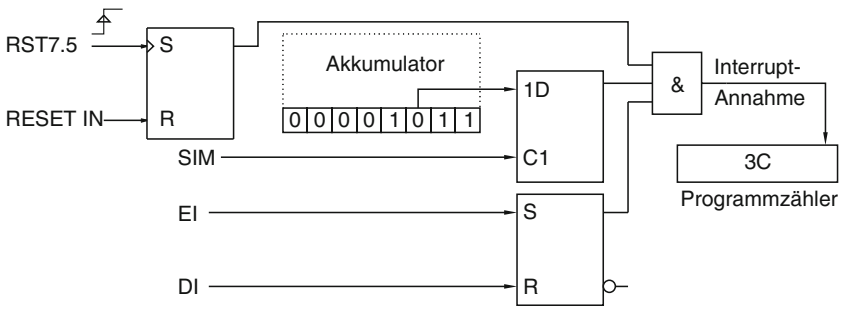


Abb. 52.17 RST7.5

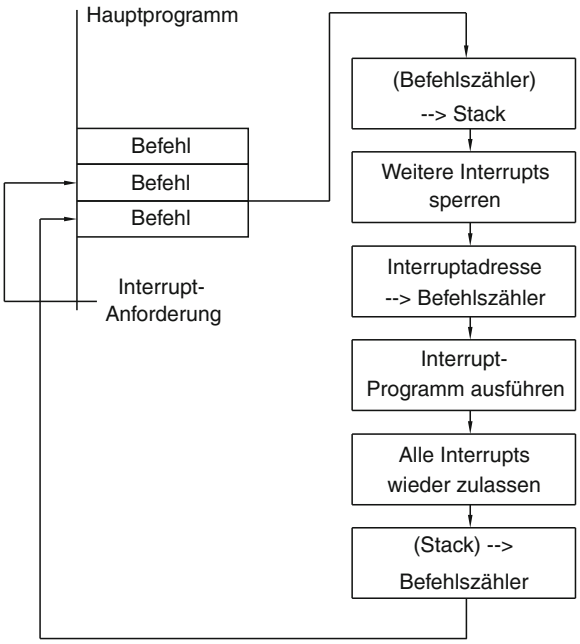


Abb. 52.18 Interruptvorgang

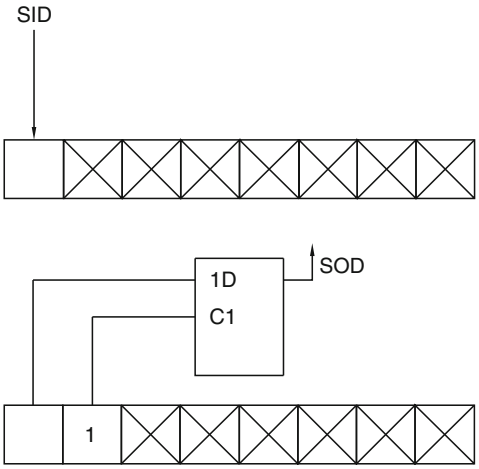


Abb. 52.19 Bitposition im Akkumulator für den seriellen Ein- und Ausgang

Für die serielle Datenübertragung besitzt der Prozessor 8085 einen eigenen Ein- und Ausgang (SID = Seriell In Data, SOD = Seriell Out Data). Hierbei wird die Information Bit für Bit zeitlich nacheinander übertragen. Die Steuerung erfolgt über ein Programm und der serielle Datenfluss über den Akkumulator.

Mit dem Befehl RIM gelangt das Bit am SID-Anschluss nach Bit 7 des Akkumulators. Von dort ist eine weitere Verarbeitung möglich. Zur Datenausgabe wird ebenfalls Bit 7 des Akkumulators benutzt. Hierzu muss der Ausgang SOD mit dem Befehl SIM freigegeben worden sein (Abb. 52.19).

Um das Zusammenwirken mit Peripheriebausteinen darzustellen, ist in Abb. 52.20 ein 8085-Minimalsystem mit einem RAM (256 Byte), EPROM (2 kB), Zeitgeber und E/A-Ports abgebildet.

Tab. 52.6 Bedeutung der Anschlüsse CPU

A0 . . . A15	Adressbus
D0 . . . D7	Datenbus
/RD	Lesen
/WR	Schreiben
/RFSH	A0 . . . A7 führen Refreshadresse (für dynamische Speicher)
/HLT	Software-Halt, die CPU wartet auf ein Interruptsignal
/WAIT	Wartesignal (CPU beginnt mit Wartezyklen)
/INT	Interrupteingang
/NMI	Nicht maskierbarer Interrupt (ADR 0066)
/MREQ	Speicheranforderung. Bei „L“ an diesem Anschluss steht auf dem Adressbus eine Adresse für einen Speicherzugriff an
/IOREQ	E/A-Bausteinanforderung
/BUSRQ	Busanforderung externer Geräte. Das interne Bussystem wird in den hochohmigen Zustand gebracht, damit externe Geräte Adress-, Daten- und Steuerbus nutzen können
/BUSAK	Busanforderungsbestätigung. Hiermit wird externen Geräten bestätigt, dass Adress-, Daten- und Steuerbus benutzt werden darf
/RESET	Rücksetzen
/M1	Statusleitung (Maschinenzyklus 1)

Tab. 52.7 Peripheriebausteine für ein Z80-System

Z80-PIO	Parallel-E/A-Baustein
Z80 CTC	Zähler und Zeitgeberbaustein
Z80 DMA	Baustein für direkten Speicherzugriff
Z80 SIO	Serieller E/A-Baustein

Die Speicherkapazität der Bausteine lässt sich aus der Anzahl der Adressleitungen ermitteln. Die Auswahl der Bausteine erfolgt mit den Adressleitungen A11 und A12, so dass ein Speicherplan aufgestellt werden kann (Abb. 52.21). Die Portadressen können mit Hilfe der Beschreibung der Bausteine 8755 und 8156 (siehe Peripheriebausteine) gefunden werden.

52.2.4.2 Beispiel Z80 CPU

Der Z80 besitzt die 6 Arbeitsregister B, C, D, E, H, L, die jeweils 8-Bit-Operationen oder als Registerpaare BC, DE und HL 16-Bit-Operationen ermöglichen, einen Akkumulator und das Flagregister als Hauptregister-Block. Die gleiche Anzahl Register sind als Zweitregister-Block vorhanden (Abb. 52.22). Außer Befehlszähler und Stapelzeiger sind die Spezialregister IX und IY als Indexregister, das I-Register (Interrupt-Vektor) und R-Register (Refresh-Speicher), der ein direktes Arbeiten mit dynamischen Speichern ohne zusätzlichen Bausteinaufwand ermöglicht, vorhanden. Die 16-Bit-Index-Register dienen zur Bearbeitung von Tabellen und Datenverschiebungen im Arbeitsspeicher.

Als Taktgenerator wird ein externer Oszillator eingesetzt. Abb. 52.23 zeigt die Z80 CPU und deren Anschlussbelegung mit Taktgenerator und Treiberstufen. Im Gegensatz zur 8085 CPU sind 16 Adressleitungen und 8 Datenleitungen herausgeführt. In Tab. 52.6 werden die Funktionen der jeweiligen Anschlüsse des Prozessors beschrieben.

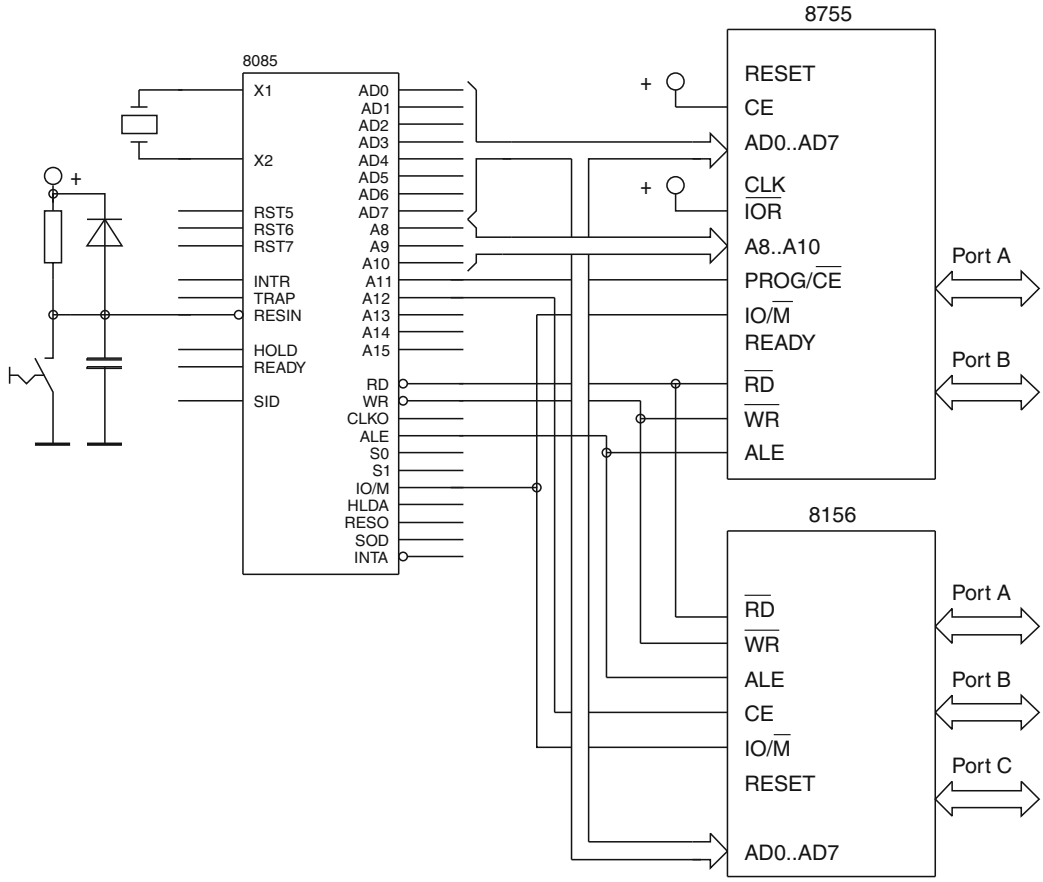


Abb. 52.20 Minimalsystem nach Datenbuch Fa. Siemens

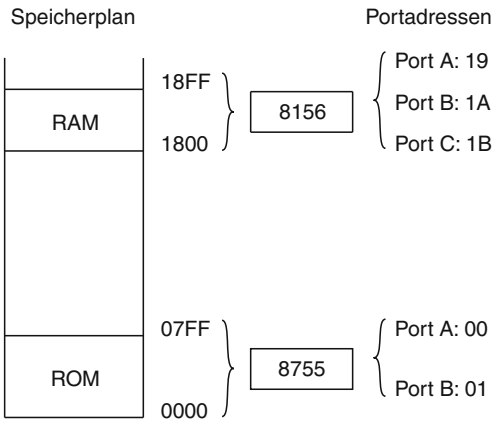


Abb. 52.21 Adressen der Bausteine 8156 und 8755

Der Z80 verfügt über den Befehlssatz des 8085 hinaus über weitere Befehle und zusätzliche Adressierungsarten wie indizierte und relative Adressierung (siehe auch Abschn. 52.6 „Maschinensprache“). Der Operationscode ist mit dem Code der 8085 CPU identisch. Unterschiede bestehen in der Mnemonik (Assembler).

52.2.5 16-Bit-Prozessoren

In Personalcomputern ist die 8086-Familie sehr verbreitet (8086, 80286, 80386, 80486).

52.2.5.1 8086/80286

Die Prozessoren besitzen einen gemultiplexten Daten-Adressbus (AD0 . . . AD15) und die Adressleitungen A16 . . . A19. Damit steht ein Adressbereich von 1 MB ($2^{20} = 1.048.576$) zur Verfügung. Der 8086 kann in den Betriebsarten *MIN*- und *MAX-Mode* eingesetzt werden. Die Auswahl erfolgt mit dem Anschluss MN/MX. Im MIN-Mode werden die Steuersignale für Speicher und E/A-Bausteine vom Prozessor erzeugt, und im MAX-Mode werden komplexere Systeme mit einem zusätzlichen Steuerbaustein zur Systemsteuerung (Buscontroller) betrieben. Abb. 52.24 zeigt die Pinbelegung beider Prozessoren und Tab. 52.8 die Bedeutung der Anschlüsse. Tab. 52.9 und 52.10 beschreiben Steuerfunktionen.

Das Prinzip des Prozessors beschreibt Abb. 52.25, und der Registersatz ist im Abb. 52.26 wiedergegeben.

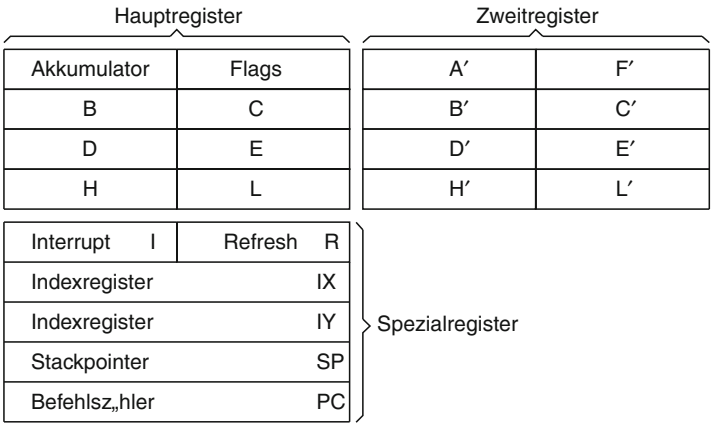


Abb. 52.22 Registerplan der Z80 CPU

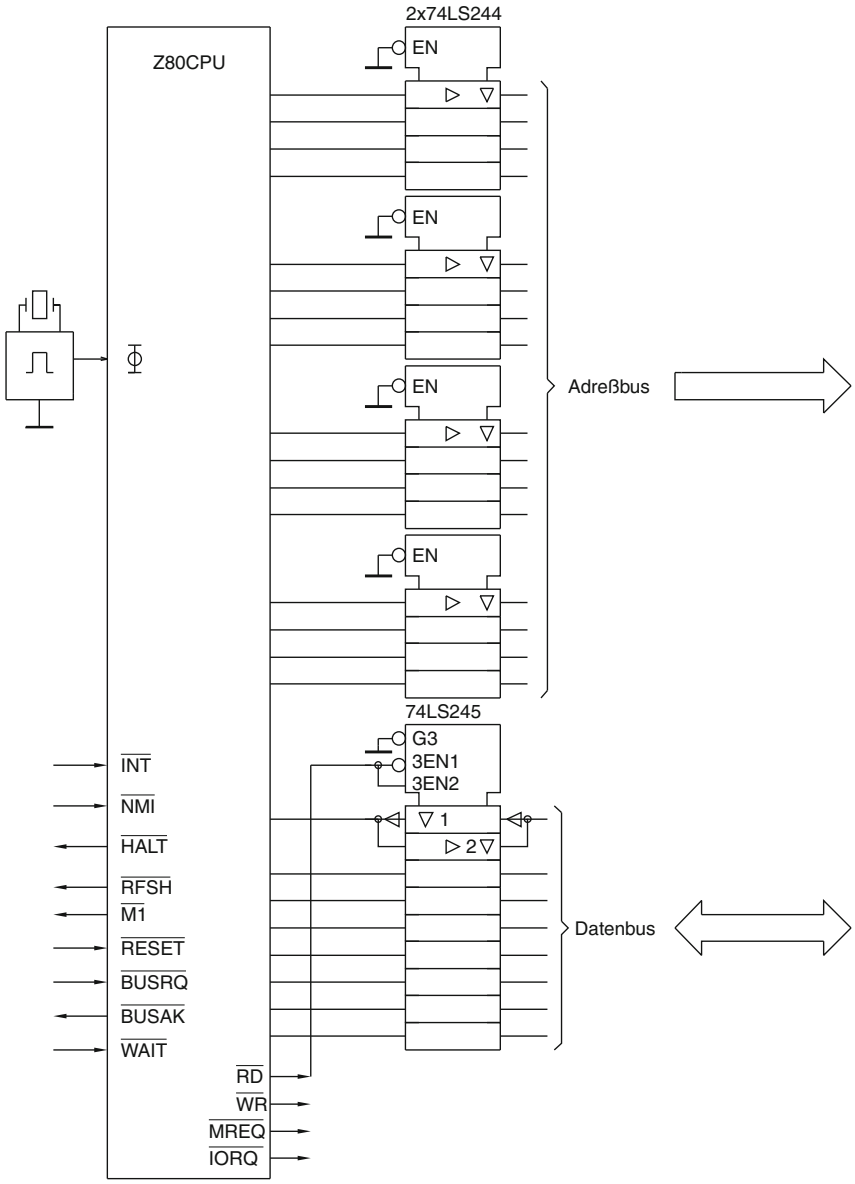


Abb. 52.23 Z80-System

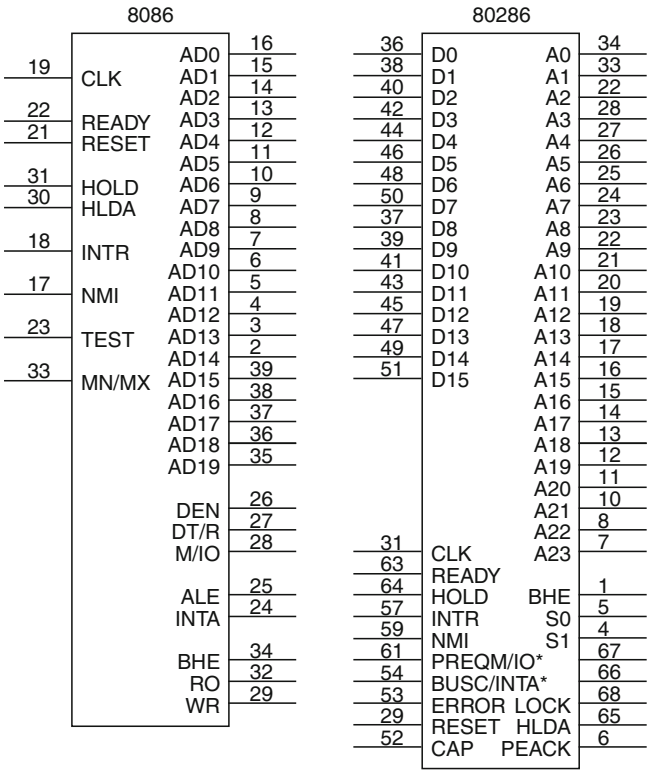


Abb. 52.24 CPU 8086 und 80286

Tab. 52.8 Bedeutung der Anschlüsse

AD0 . . . AD15	Adress-/Datenbus	
RD	Read	Lesen
WR	Write	Schreiben
ALE	Adress Latch Enable	Adresszwischenspeicherung
INTR	Interrupt Request	Interruptabfrage
INTA	Interrupt Acknowledge	Interruptbestätigung
NMI	Non Mascable Interrupt	Nicht maskierbarer Interrupt
TEST	Test	Überprüfung des Wartezyklus
RESET	Reset	Rücksetzen
CLK	Clock	Takt
MN/MX	Min/Max	Minimum-/Maximum-Mode
BHE/S7	Bus High Enable/Status	Freigabe von L- oder H-Adressteil
DEN	Data Enable	Busfreigabe (Datenübertragung)
DT/R	Data Transmit/Receive	Datenübertragung/-empfang
M/IO	Memory/IO	Speicher und E/A-Zugriff
HOLD	Aufruf externer Geräte (DMA)	
HLDA	Bestätigung auf HOLD (DMA)	
LOCK	Angabe über Sperrung der CPU	
S0, S1, S2	Statusinformation	

Tab. 52.9 Steuerfunktionen

BHE	A0	Funktion
0	0	Gesamtes Wort
0	1	Oberes Byte von/zu Adresse
1	0	Unteres Byte von/zu Adresse
1	1	Keine Funktion

Tab. 52.10 Statusinformation

S2	S1	S0	Funktion
0	0	0	Interruptbestätigung
0	0	1	E/A-Lesen
0	1	0	E/A-Schreiben
0	1	1	Halt
1	0	0	Codezugriff
1	0	1	Speicher lesen
1	1	0	Speicher schreiben
1	1	1	Passiv

In der Befehlseinheit werden die von der Befehlswarteschlange (Instruction Queue) gelieferten Befehle decodiert. Bei Befehlsausführung kann die Befehlswarteschlange parallel mit dem nachfolgenden Maschinencode geladen werden.

Damit steuert diese Einheit alle Vorgänge des Systems. Hauptbestandteil der Ausführungseinheit ist die aus 8-Bit-Systemen bekannte ALU.

Die vier Hauptregister der Execution Unit A, B, C und D sind aus je zwei 8-Bit-Registern (L = Low und H = High) zusammengesetzte 16-Bit-Register. Sie dienen zur Aufnahme von arithmetischen und logischen Operanden. Die Breite der übrigen Register beträgt 16 Bit. Das SP-Register (Stack Pointer) und BP-Register (Base Pointer) können für Stack- und 16-Bit-Operationen verwendet werden. SI- und DI-Register sind Indexregister (Source Index = Quelle, Destination = Ziel).

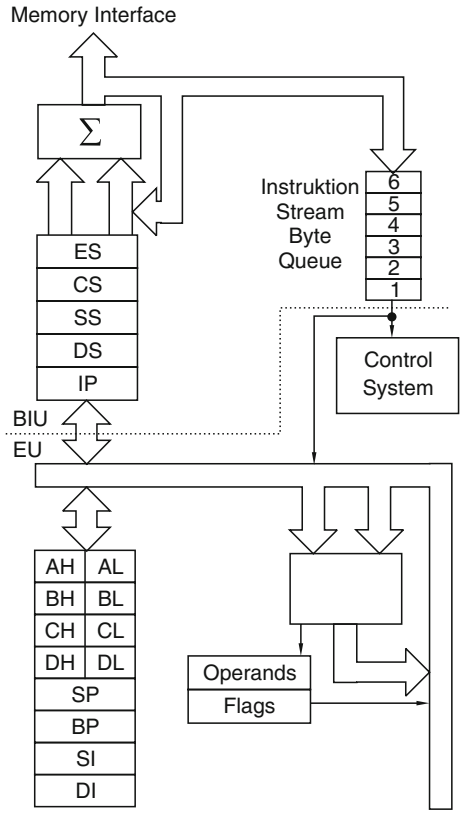


Abb. 52.25 Blockbild der 8086 CPU

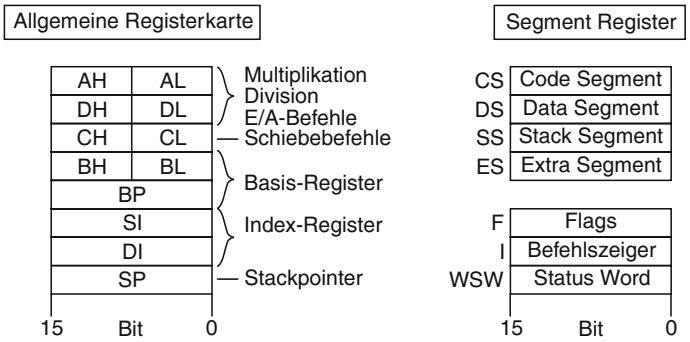


Abb. 52.26 Registersatz und Funktion

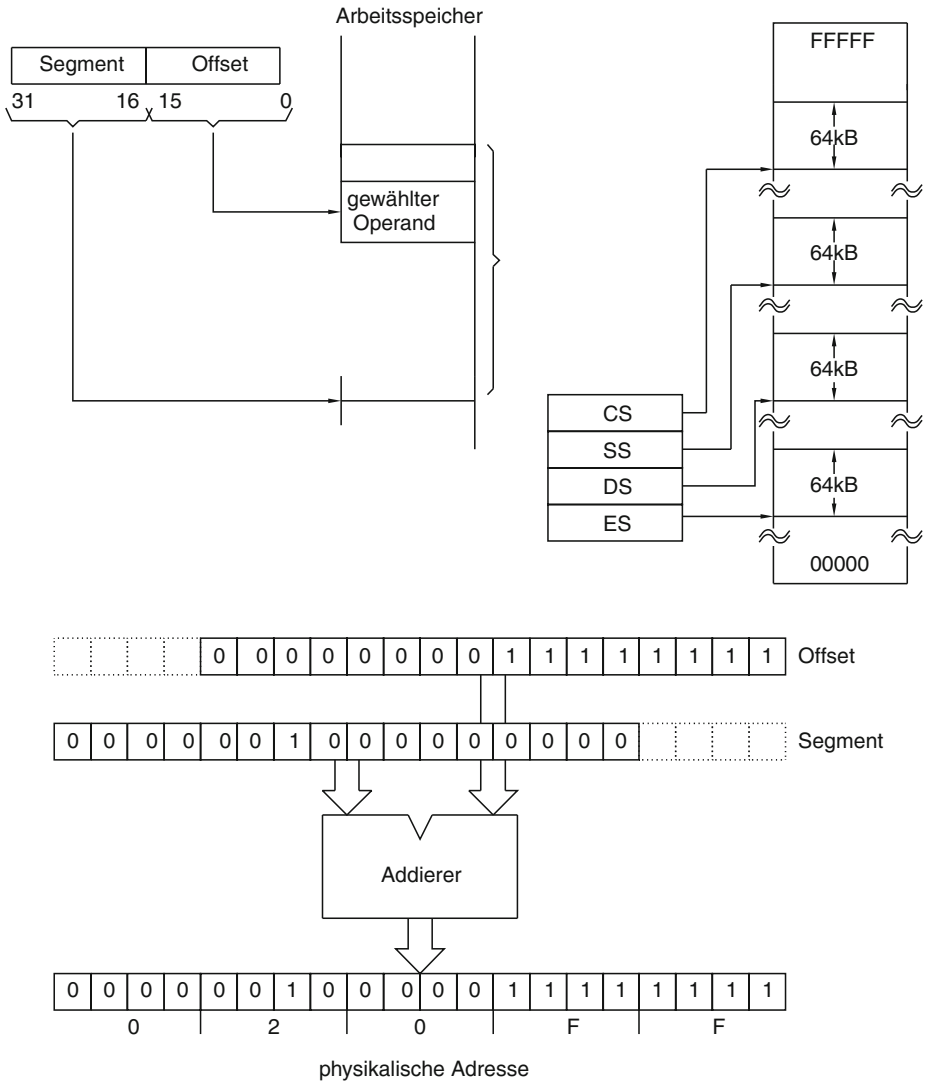


Abb. 52.27 Bildung der physikalischen Adresse

52.2.5.2 Adressenbildung

In der BIU befinden sich die Segmentregister ES (Extra-Segment), CS (Code-Segment), SS (Stack Segment) und DS (Daten-Segment). Ihre Aufgabe ist es, den Speicherbereich über Segmente in vier Bereiche bis zu 64 kB zu verwalten (Real Mode). Die Anfangsadressen stehen jeweils in den Segmentregistern. Mit ihrer Hilfe und dem Befehlszeiger IP (Offset) wird die physikalische Adresse (absolute Adresse) nach Abb. 52.27 gebildet. Hierbei wird die 16-Bit Offsetadresse als L- und die

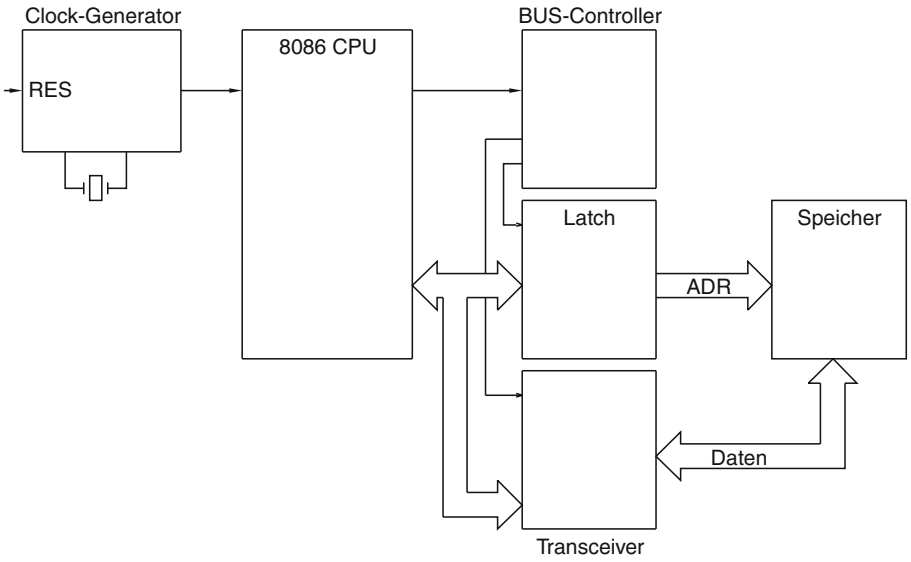


Abb. 52.28 8086-System

16-Bit-Segmentadresse als H-Teil einer fiktiven 20-Bit-Adresse angesehen. Das gewählte Segment gibt die Anfangsadresse des Segments an, und der Offset zeigt auf den gewählten Operanden. Ihre Addition ergibt somit die physikalische 20-Bit-Adresse.

Ein Mikroprozessorsystem mit 16-Bit-Prozessoren besitzt außer den bekannten Speicher- und Ein- und Ausgabeeinheiten spezielle Peripheriebausteine, die jeweils eine besondere Aufgabe übernehmen, wie beispielsweise der Buscontroller (siehe auch Peripheriebausteine). Abb. 52.28 zeigt schematisch ein 8086-System. Der Transceiver ist vergleichbar mit dem Zweirichtungstreiber in 8-Bit-Systemen. Über den Eingang T wird die Datenrichtung festgelegt, und mit /OE kann der BUS hochohmig geschaltet werden.

52.3 Halbleiterspeicher

52.3.1 Allgemeines

Speicher dienen zur Speicherung und Sicherung von Daten und Programmen. Der Speicherbaustein muss den Prozessor mit Daten beliefern, und der Prozessor muss Daten darin ablegen können. Zur Auswahl eines bestimmten Speicherplatzes, einer Adresse, werden dem Baustein Adressleitungen zugeführt. Die Daten gelangen über den Datenbus zum und vom Speicher, wenn der Prozessor die Schreib-Lese-Leitung aktiviert und die Speicherzelle ausgewählt wurde. Um eine Unterscheidung verschiedener Speicher durchführen zu können, besitzt der Baustein Aktivierungs- oder Auswahleingänge. Abb. 52.29 zeigt die Anschlüsse eines 1 K × 8 Bit-Schreib-Lese-Speichers.

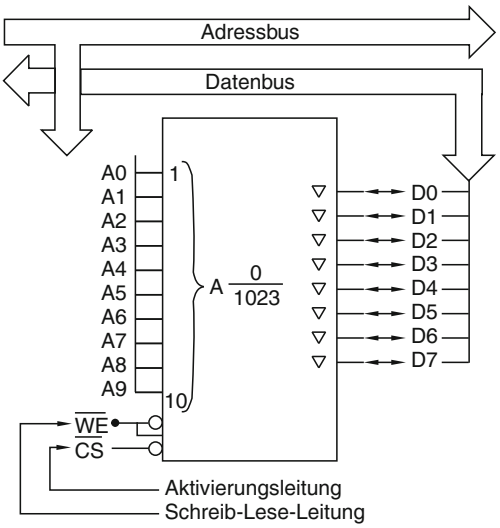


Abb. 52.29 Anschlüsse an einem Speicherbaustein

52.3.2 Kenndaten und Technologie

Wichtige Kenndaten eines Speichers sind die Speicherkapazität (Capacity) in Verbindung mit der Angabe über die Speicherorganisation, die Zugriffszeit (Access Time), die von der Ansprache des Speichers vergeht bis zur Verfügbarkeit der Daten und die Leistungsaufnahme im aktiven und Ruhe-Betrieb (Standby). Die Speicherkapazität gibt die Informationsmenge an, die der Speicher aufnehmen kann. Sie wird üblicherweise in Kbit, KByte oder MByte angegeben. Die Kenndaten selbst

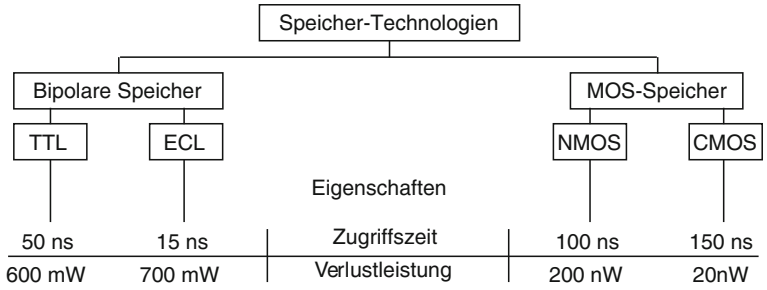


Abb. 52.30 Speichertechnologien und Eigenschaften

Tab. 52.11 Anschlussbedeutungen

A0 . . . An	Adressleitungen; über diese Leitungen wird der Speicherplatz angewählt
D0 . . . Dn	Datenleitungen; über diese Leitungen gelangen Daten vom oder zum Speicher
CE, CE1, CE2, CS	Chip Enable, Chip Select; über diese Leitung(en) kann der Speicherbaustein angewählt werden
WE, WR	Write Enable, Write; diese Leitung wird von der CPU aktiviert, wenn Daten in den Speicher geschrieben werden sollen
RD	Read; diese Leitung wird von der CPU aktiviert, wenn Daten vom Speicher gelesen werden sollen
O0 . . . On	Output, Datenausgänge
OE	Output Enable; mit dieser Leitung werden die Ausgänge eines Speichers freigegeben
ALE	Adress Latch Enable; an diesem Anschluss kann die CPU bei einem gemultiplexten Adress-/Datenbus die Adresse zwischenspeichern
RAS	Row Adress Select, Zeilenauswahlleitung
CAS	Column Adress Select, Spaltenauswahlleitung
Vpp	Programming Voltage; an diesem Anschluss wird die Programmierspannung angelegt
VCC	Betriebsspannungsanschluss
GND	Ground, Masseanschluss

sind von der Technologie, also dem Herstellungsverfahren des Speichers, abhängig. Hier unterscheidet man bipolare Speicher, deren Grundelemente der TTL- oder ECL-Technologie entsprechen, von unipolaren Speichern, deren Grundelemente in MOS-Technologie gefertigt sind.

Bipolare Speicher besitzen zwar eine geringere Zugriffszeit, haben aber eine höhere Leistungsaufnahme sowie eine geringere Integrationsdichte als unipolare Speicher (siehe Abb. 52.30).

Die heutigen großen Speicherkapazitäten werden mit CMOS-Speichern erreicht. Für den Entwickler spielt das Preis-Leistungsverhältnis und notwendige zusätzliche Steuerbausteine noch eine bedeutende Rolle.

52.3.3 Bedeutung der Anschlüsse

Tab. 52.11 fasst die Anschlussbedeutungen, die bei Speichern auftreten, zusammen.

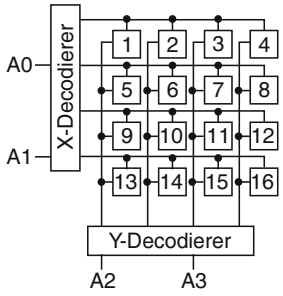


Abb. 52.31 Speichermatrix

52.3.4 Organisation und Aufbau

Die Speicherzellen in einem Speicherbaustein sind matrixförmig angeordnet, um die Anzahl der Auswahlleitungen möglichst gering zu halten. Die Ansprache der Speicherzelle ist nur möglich, wenn jeweils gleichzeitig eine Zeilenleitung und eine Spaltenleitung, oder auch X- und Y-Leitung, aktiv sind. Durch die Anordnung innerhalb der Matrix ist jede Speicherstelle eindeutig bestimmt. Die Auswahl der Speicherstelle in der Matrix erfolgt über einen Zeilen- und Spaltendecoder, der jeweils über Adressleitungen angesteuert wird. Die Anzahl der Adressleitungen ist von der Anzahl der Speicherstellen abhängig.

Beispiel

Für einen 16-Bit-Speicher (siehe Abb. 52.31) sind normalerweise 16 Auswahlleitungen erforderlich. Durch den matrixförmigen Aufbau reduzieren sich die Auswahlleitungen auf 8, da 4 Zeilen- und 4 Spaltenleitungen benötig werden. Mit Hilfe jeweils eines 4-aus-1-Decoders können die 4 Leitungen von zwei Adressen angesprochen werden. Damit ergeben sich für den 16-Bit-Speicher insgesamt 4 Adressleitungen.

Um n Speicherstellen zu adressieren, benötigt man 2^n Adressleitungen. ◄

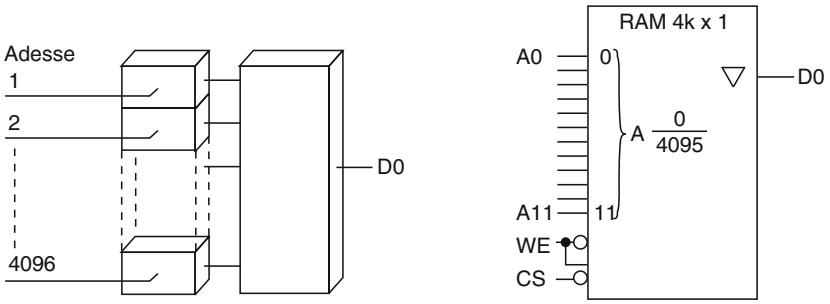


Abb. 52.32 Prinzip und Schaltbild eines bitorganisierten Speichers

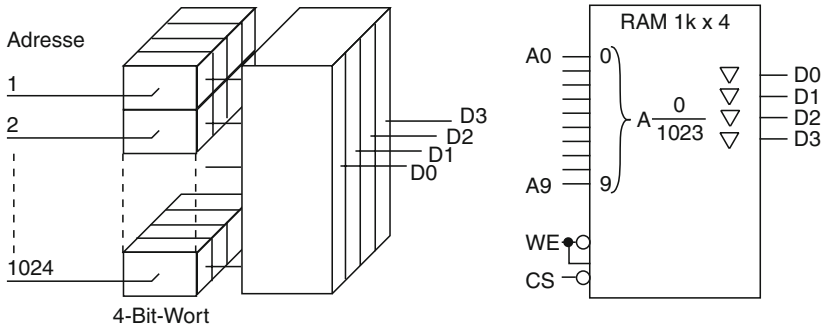


Abb. 52.33 Prinzip eines wortorganisierten Speichers und Schaltbild

Beispiel

Ein Speicher mit der Organisation 4 K × 8 Bit besitzt 4096 Speicherstellen mit einer Wortbreite von 8 Bit. Damit sind 12 Adressleitungen nötig.

$$\text{Adressleitungen} = \frac{\ln \text{Speicherkapazität}}{\ln 2} = \frac{\ln 4096}{\ln 2} = 12 \tag{52.1}$$

◀

52.3.4.1 Bitorganisierter und wortorganisierter Speicher

Der *bitorganisierte* Speicher speichert pro Adresse nur ein Bit. Er ist somit unabhängig von der Datenwortbreite. Beim *wortorganisierten* Speicher steht in jeder Adresse ein Datenwort. Damit sind n-Bitstellen gemeint, die gleichzeitig von einer Adresse angesprochen werden. Wortorganisierte Speicher, deren Wortbreite der Datenbreite des Mikroprozessors entsprechen, werden häufig für Arbeitsspeicher bei kleineren Mikroprozessor- oder Mikrocontrollersystemen eingesetzt (z. B. 8-Bit-Systemen).

Ein Speicher mit der Kapazität von 4096 Bit kann als bitorganisierter Speicher (4 K × 1 Bit) (Abb. 52.32) oder als wortorganisierter Speicher (1 K × 4 Bit) (Abb. 52.33) aufgebaut sein (1 KBit = 2^{10} Bit). Beim bitorganisierten Speicher ist jeder Speicherzelle eine Adresse zugeordnet, während der wortorganisierte Speicher pro Wort eine Adresse besitzt. Legt man die Anzahl der Speicherstellen zugrunde, dann besitzen beide dargestellten Speicher die gleiche Speicherkapazität.

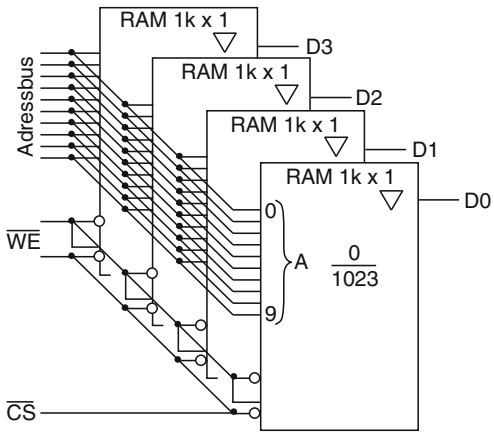
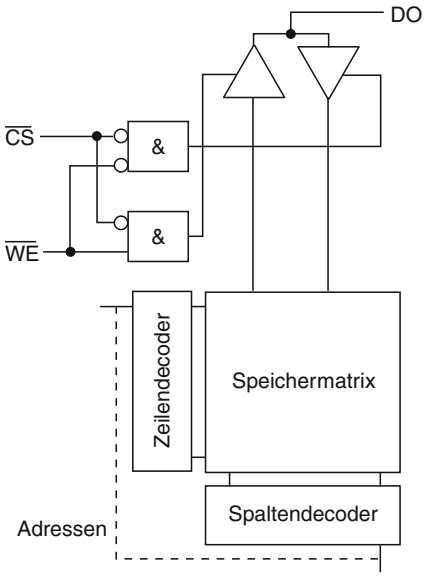


Abb. 52.34 Vom bit- zum wortorganisierten Speicher

Ist beim bitorganisierten Speicher eine bestimmte Wortbreite erforderlich, müssen verschiedene Speicher mit der gleichen Adresse angesprochen werden. Für ein 4-Bit-Datenwort müssen 4 Speicherbausteine adressenmäßig parallel geschaltet und mit /CS und /WE gleichzeitig angesprochen werden (Abb. 52.34).

Zur Aus- und Eingabe von Daten ist eine interne Steuerschaltung notwendig, die je nach Speicher unterschiedlich sein kann. Sie erfolgt über die Steuereingänge WE (Write Enable = Schreibfreigabe) und CS (Chip Select = Bausteinauswahl) und ist häufig „Low"-aktiv (Abb. 52.35).



CS	WE	Funktion
0	0	Einschreiben der Daten
0	1	Auslesen der Daten
1	x	Baustein hochohmig

Abb. 52.35 Prinzip der Datensteuerung

Mit /CS oder /CE wird der einzelne Speicherbaustein ausgewählt. Beschrieben und ausgelesen wird er durch Pegelumschaltung am WE- oder RD- und /WR-Eingang.

52.3.4.2 Speicher mit Adressenzwischenspeicher

Außer Schreib-Leseeingang und Bausteinauswahl gibt es Speicher, die einen weiteren Steuereingang ALE (Adress-Latch-Enable) besitzen und damit ein internes Adressenzwischenregister. Sie können direkt an den gemultiplexten Daten-Adressbus eines Mikrocontrollers oder Mikroprozessors angeschlossen werden. Speicher und Prozessor sparen so Anschlussleitungen. Die Adressen und Daten werden hier zeitlich nacheinander auf gleichen Leitungen übertragen. Mit ALE übergibt der Prozessor seine Adresse einem Adressenzwischenspeicher im Speicherbaustein und liefert oder holt anschließend die Daten. Speicher ohne Adressenzwischenregister können von einem Prozessor mit gemultiplextem Daten-Adressbus nur über ein externes Zwischenregister angesteuert werden. Im Abb. 52.36 ist ein Speicher mit und ohne Steuereingang ALE an einem Mikroprozessor angeschlossen.

52.3.5 Zeitverhalten

Damit der Speicher seine Funktion durchführt, muss der Prozessor über den Steuerbus zeitliche Bedingungen einhalten

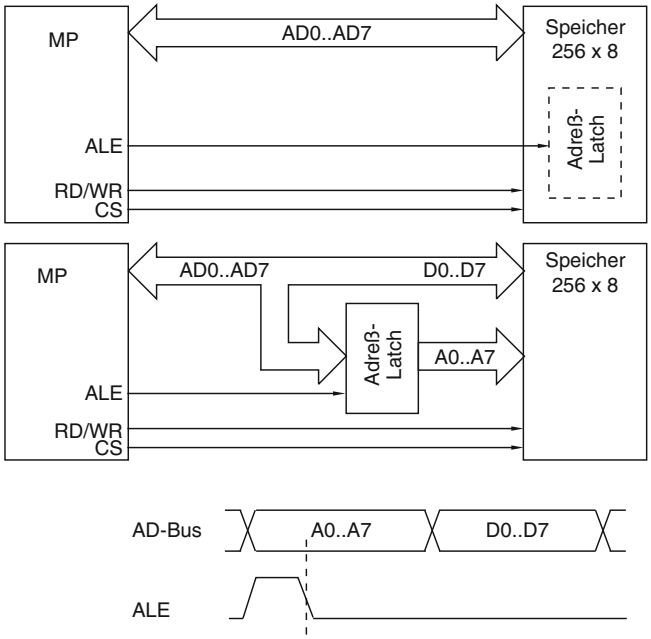


Abb. 52.36 Speicher mit und ohne Adressenzwischenspeicher

(Abb. 52.37). Beispielsweise muss zunächst die gewünschte Adresse auf dem Adressbus anliegen, dann muss der Baustein, der diese Adresse besitzt, mit dem Auswahlsignal /CS angewählt werden. Des Weiteren müssen Daten vom Datenbus angeboten werden, um sie danach mit dem Schreibsignal in den Speicher zu bringen.

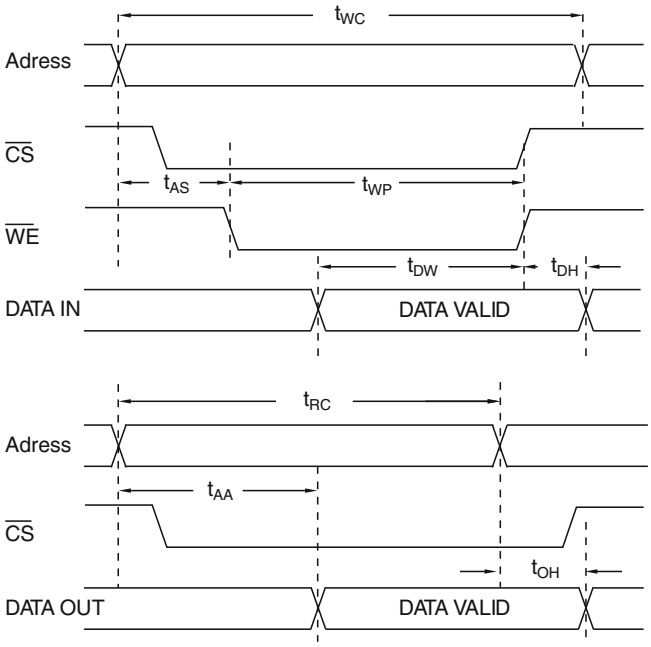


Abb. 52.37 Zeitverhalten auf den Bus- und Steuerleitungen

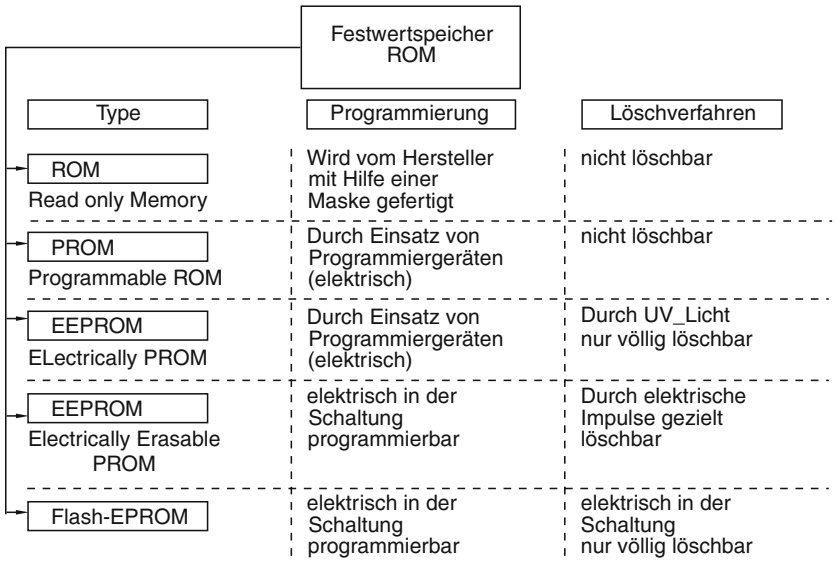


Abb. 52.38 Festwertspeicher

Damit die gewünschten Daten den Speicher oder die Daten des Speichers den Prozessor erreichen, müssen bestimmte zeitliche Bedingungen eingehalten werden. Zwischen dem Anliegen der Adresse (t_{WC} = Write Cycle Time) und dem Einschreiben der Daten ist eine Zeitspanne t_{AS} (Adress Setup Time) erforderlich, weil sonst die Daten eine falsche Speicherstelle erreichen. Die Zeit, die benötigt wird, um das Einschreiben der Daten in den Speicher zu gewährleisten, ist die Mindestdauer des Schreibimpulses (/WE) und wird t_{WP} (Write Pulse Width) genannt. Zur korrekten Datenübertragung müssen die Daten auf dem Datenbus stabil sein. Diese Zeit, in der sich die Daten auf dem Datenbus (Data Valid = Daten gültig) eingependelt haben, nennt man t_{DW} (Data Valid Time). Mit t_{DH} (Data Hold Time) wird die Zeit nach Ablauf des Schreibimpulses bezeichnet, solange noch Daten und Adressen anliegen. Beim Lesezyklus t_{RC} (Read Cycle Time) stabilisieren sich die Adressen auf dem Bus in der Zugriffszeit t_{AA} (Adress Access Time). Anschließend können die Daten gelesen werden.

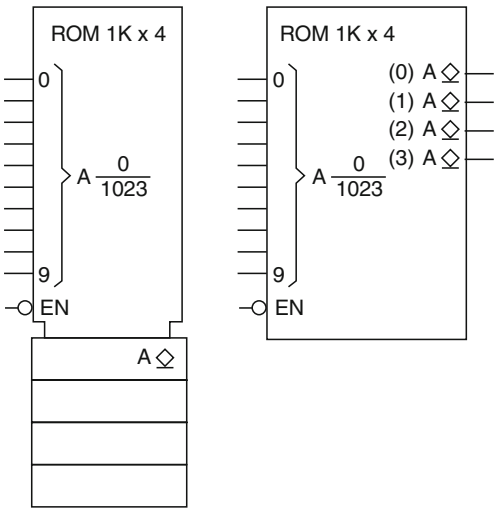


Abb. 52.39 Schaltbild eines ROM und vereinfachte Darstellung

52.3.6 Speichertypen

Es werden zwei Hauptgruppen von Speichertypen unterschieden: Schreib-Lesespeicher und Festwertspeicher. Während Schreib-Lesespeicher als Arbeitsspeicher dienen, liegen in Festwertspeichern häufig benötigte konstante Daten vor, wie beispielsweise Tabellen oder Teile eines Betriebssystems (Firmware).

52.3.6.1 Festwertspeicher

ROM (Read Only Memory = Nur-Lese-Speicher) sind Speicher, die im Normalbetrieb nicht beschrieben, sondern nur ausgelesen werden. Es sind *Festwertspeicher* (Nonvolatile Read-Only-Memory = nicht flüchtige Speicher), da sie ihre Information ohne anliegende Betriebsspannung beibehalten. Daten und Programme gehen bei Ausfall der Spannungsversorgung nicht verloren. Abb. 52.38 gibt eine Übersicht über Festwertspeicher.

52.3.6.1.1 Masken-ROM

Für Standardanwendungen kann der Hersteller ROM-Bausteine mit Hilfe einer Maske fertigen. Man spricht dann von einem Masken-ROM. Der Inhalt des Speichers ist nicht mehr veränderbar. Abb. 52.39 zeigt das Schaltbild eines ROM und die vereinfachte Darstellung. Speicher

52.3.6.1.2 PROM

ROM, die mit Hilfe von Programmiergeräten programmiert werden können, sind PROM (Programmable Read Only Memory)

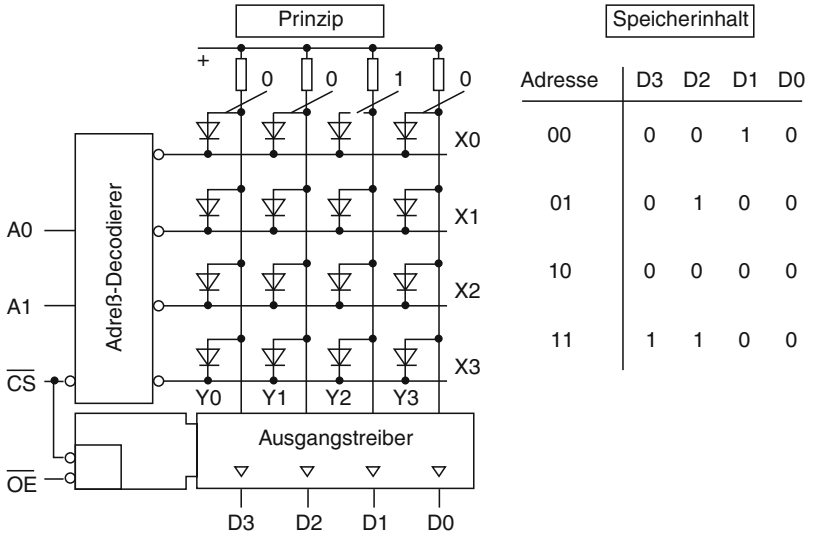


Abb. 52.40 Speicher mit Diodenmatrix

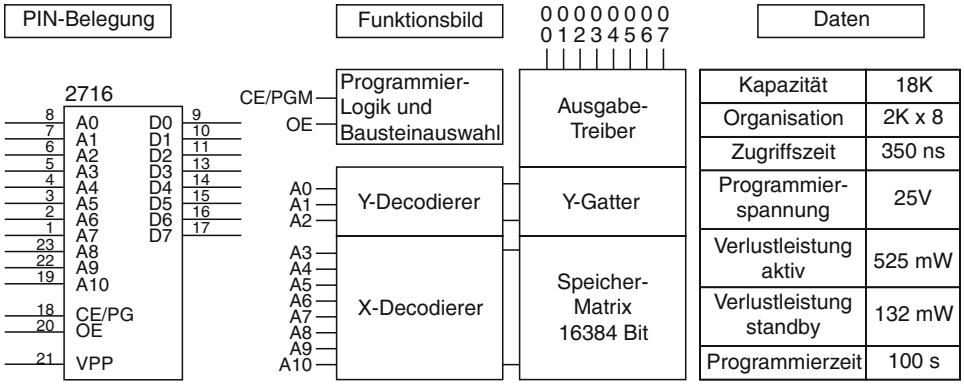


Abb. 52.41 EPROM

oder *fusible* PROM (Sicherungs-PROM). Der Hersteller oder Anwender kann diese Bausteine nach eigenem Wunsch programmieren. Aufgebaut sind solche Festwertspeicher in Form einer Dioden- oder Transistormatrix (Abb. 52.40). Beim Programmiervorgang werden durch einen Stromstoß leitende Verbindungen abgeschmolzen. Einmal programmierte PROM behalten ihre Daten und sind nicht mehr löschbar.

Die Ausgänge D0 ... D4 können in Abhängigkeit der Adresse die einprogrammierten Daten ausgeben. Die Leitungen X0 ... X3 nennt man Wortleitungen, die Leitungen Y0 ... Y3 werden Bit-Leitungen genannt. Mit 2 Adressleitungen lassen sich 4 Wortleitungen ansprechen. Der Eingang CS (Chip Select) dient zur Anwahl des Speichers, und mit OE (Output Enable) werden die Ausgangstreiber aktiviert.

52.3.6.1.3 EPROM

Ein EPROM (Erasable PROM) ist ein programmierbarer und löschbarer Festwertspeicher. Der Speicherinhalt kann durch Löschen und erneutes Beschreiben geändert werden. Beim Abschalten der Betriebsspannung gehen die Daten nicht verloren. Die Funktion einer Speicherzelle beruht auf Veränderung der Leitfähigkeit eines MOS-Transistors mit einem schwebenden (nicht angeschlossenen) Gate (FAMOS = Floating Gate Avalanche Injection MOS). Im nicht programmierten Zustand sperrt der Transistor. Leitend wird der Transistor durch den Programmiervorgang, bei dem dem Gate mit Hilfe des Lawineneffekts (Avalance-) eine Ladung aufgeprägt wird (Abb. 52.41).

EPROM-Bausteine besitzen auf ihrer Oberseite ein Glasfenster, um ein Löschen zu ermöglichen. Dies geschieht durch Bestrahlung mit UV-Licht mit definierter Wellenlänge und Strahlungsdichte. Die Dauer der Bestrahlung ist vom Bausteintyp abhängig (ca. 30 Min.). Der Löschvorgang löscht den gesamten Speicherinhalt. Eine unprogrammierte oder gelöschte Speicherstelle besitzt immer „1“ als Information.

Zur Programmierung werden entsprechende Programmiergeräte benötigt. Die notwendige Programmierspannung ist vom Bausteintyp abhängig und höher als die Betriebsspannung (Tab. 52.12). Nach dem Programmiervorgang ist das Löschfenster zuzukleben, um kein Außenlicht eindringen zu lassen.

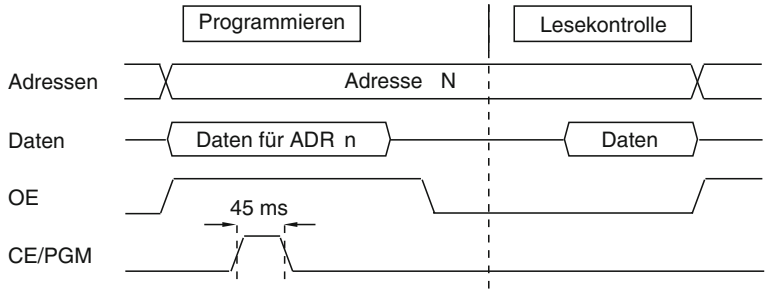
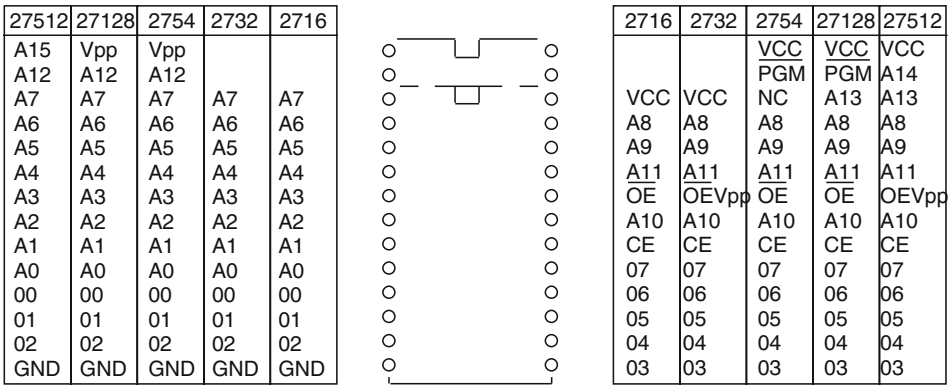


Abb. 52.42 Programmierzyklus



27512	27128	2754	2732	2716
A15	Vpp	Vpp		
A12	A12	A12		
A7	A7	A7	A7	A7
A6	A6	A6	A6	A6
A5	A5	A5	A5	A5
A4	A4	A4	A4	A4
A3	A3	A3	A3	A3
A2	A2	A2	A2	A2
A1	A1	A1	A1	A1
A0	A0	A0	A0	A0
00	00	00	00	00
01	01	01	01	01
02	02	02	02	02
GND	GND	GND	GND	GND

2716	2732	2754	27128	27512
		$\overline{\text{VCC}}$	$\overline{\text{VCC}}$	VCC
		PGM	PGM	A14
VCC	VCC	NC	A13	A13
A8	A8	A8	A8	A8
A9	A9	A9	A9	A9
$\overline{\text{A11}}$	$\overline{\text{A11}}$	$\overline{\text{A11}}$	$\overline{\text{A11}}$	A11
$\overline{\text{OE}}$	OEVpp	$\overline{\text{OE}}$	$\overline{\text{OE}}$	OEVpp
A10	A10	A10	A10	A10
CE	CE	CE	CE	CE
07	07	07	07	07
06	06	06	06	06
05	05	05	05	05
04	04	04	04	04
03	03	03	03	03

Abb. 52.43 Anschlüsse verschiedener EPROM-Typen

Tab. 52.12 EPROM-Beispiele

Type	Kapazität	Organisation	Zugriffszeit	Leistungsaufnahme	
Programmierspannung					
2716	16 K	2 K × 8	350 ns	500/125 mW	25 V
2732	32 K	4 K × 8	250 ns	625/175 mW	21 V
2764	64 K	8 K × 8	250 ns	500/200 mW	21 V
27C256	256 K	32 K × 8	120 ns	150/0,5 mW	12,5 V
27C512	512 K	64 K × 8	120 ns	150/0,5 mW	12,5 V
27C100	1 M	128 K × 8	120 ns	150/0,5 mW	12,5 V

(Nach Datenblatt der Firma Intel)

Nachteil: Es ist nur möglich, den Gesamtbaustein zu löschen. Dazu muss er aus der Schaltung herausgenommen werden.

Beschreibung der Betriebsarten:

- *Lesezyklus*
  Nach dem Anlegen einer gültigen Adresse wird der Baustein mit CE = „0“ aktiviert. Anschließend kann mit OE = „0“ der Dateninhalt der Adresse auf die Datenleitungen geschaltet werden. Danach wird mit OE = „1“ und CE = „1“ der Datenausgang hochohmig.
- *Standby*
  Im nicht aktiven Zustand (standby) werden interne Stufen des Bausteins abgeschaltet. Dadurch vermindert sich die Verlustleistung. Der Wartezustand wird durch eine „1“ am Eingang CE erreicht. Die Ausgänge sind in dieser Betriebsart hochohmig, unabhängig vom Eingang OE.
- *Löschen*
  Das Löschen geschieht mit ultraviolettem Licht. Nach dem Löschen liegt in jeder Speicherzelle die Information „1“.
- *Programmieren*
  Die Adresse der Speicherstelle, die programmiert werden soll, muss anliegen. Mit OE (Output En- able) werden dann die Ausgangstreiber gesperrt und die Programmierspannung von 25 V angelegt. Während sich die Daten auf dem Datenbus befinden, wird über den Eingang CE/PGM (Chip Enable) der Programmierimpuls zugeleitet. Die Einhaltung der Adressenreihenfolge ist nicht notwendig. Allerdings müssen gewisse zeitliche Bedingungen wie die Impulslänge am Eingang CE eingehalten werden.

Abb. 52.42 zeigt das Impulsdiagramm der Programmierung mit anschließender Lesekontrolle. Nach der Programmierung eines Datenwortes sollte immer eine Kontrolle der programmierten Bits erfolgen, um festzustellen, ob die gewünschten Daten im Speicher vorhanden sind (Verify Check). Hierbei darf bei diesem Baustein die Spannung am Anschluss $V_{pp} = 25\,V$ angeschaltet bleiben.

EPROM-Programmiergeräte können, da die Speicherbausteine standardisiert sind, sowohl 24polige als auch 28polige EPROMs aufnehmen (Abb. 52.43).

52.3.6.1.4 EEPROM

EEPROM (Electrically Erasable PROM), oft auch als E^2PROM bezeichnet, haben eine geringere Speicherdichte als EPROM's.

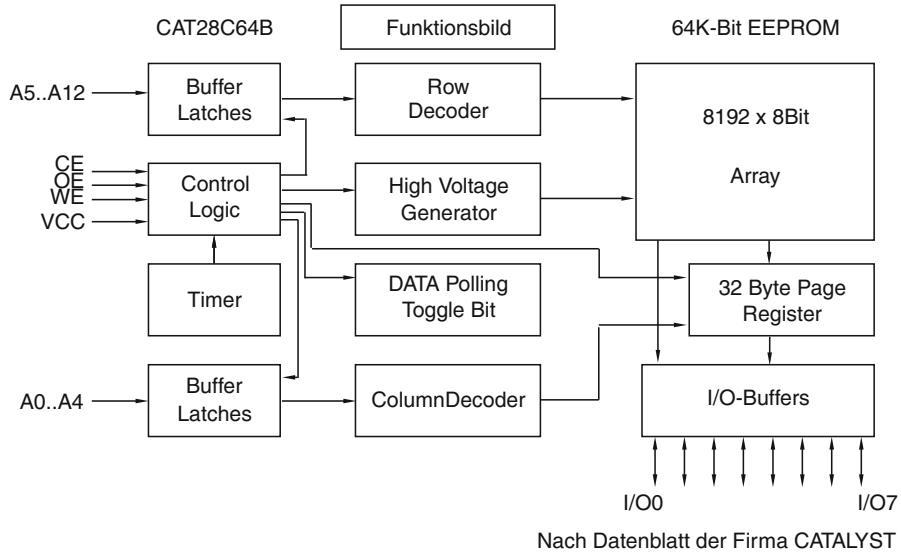


Abb. 52.44 Funktionsbild eines EEPROM

Tab. 52.13 EEPROM-Beispiele

Type	Kapazität	Organisation	Zugriffszeit
HN58064	64 K	8 K × 8	250 ns
HN58C256	256 K	32 K × 8	200 ns
HN58C1001	1 M	28 K × 8	200 ns

(Nach Datenblatt der Firma Hitachi)

Ihr Speicherprinzip beruht, wie bei den EPROM-Typen, auf dem *„Floating Gate"* eines MOS-Transistors. EEPROM's können Wort für Wort elektrisch gelöscht und neu programmiert werden. Allerdings ist die Zahl der möglichen Löschund Programmiervorgänge (ca. 10.000 Programmierzyklen) begrenzt. War der Datenerhalt vor einiger Zeit noch mit bis zu 10 Jahren angegeben, so liegen heutige Angaben bei bis zu 100 Jahren. EEPROM's verlieren ihre Daten bei Betriebsspannungsausfall nicht. Der Vorteil gegenüber EPROM's liegt darin, dass die Daten der einzelnen Speicherstellen gezielt in der Schaltung verändert werden können, ohne den Baustein zu entnehmen. Der schnelle Löschvorgang innerhalb weniger Millisekunden ist ein weiterer Vorteil. Ein EEPROM benötigt eine Programmierspannung von 5 V. Dies stimmt allerdings nur für die externe Zuführung. Im Inneren des Bausteins befindet sich ein Spannungsgenerator, der die Programmierspannung erheblich erhöht.

EEPROM's werden als parallele oder serielle Speicher mit verschiedenen Betriebsspannungen wie 1,8, 2,5, 2,7 und 5 V angeboten. Tab. 52.13 führt einige Typen mit unterschiedlichen Kapazitäten auf. Abb. 52.44 zeigt das Funktionsbild eines EEPROM.

52.3.6.1.5 Flash-EPROM

Flash-EPROM (Flash-Memory) verbinden die Eigenschaften eines RAM, EPROM und EEPROM.

Im Gegensatz zur EPROM's, die nur durch UV-Licht gelöscht werden können, sind Flash-EPROM's elektrisch und in der Schaltung löschbar. Der Baustein kann insgesamt oder sektorweise gelöscht werden. Alle Bits müssen – abgesehen von neueren Typen – vor dem Löschen zuerst auf „0" gesetzt werden. Die gelöschten Speicherstellen besitzen „1" als Information. Bei der Programmierung verhalten sich Flash-EPROM's wie EEPROM's. Sie können in der Schaltung Byte für Byte beschrieben werden. Allerdings erlauben diese Bausteine eine höhere Zahl von Programmierzyklen (bis zu 100.000 Zyklen). Die Programmierzeit liegt etwa bei 1/10 der Zeit eines EPROM. Die längste Programmierzeit benötigen EEPROM's. Adressen und Daten werden zwischengespeichert (Adress Latch, Data Latch). Die gewünschten Betriebsarten wie Löschen und Programmieren realisiert man über ein Kommandoregister. Im Signature Mode (Hersteller-Kennzeichnung) befindet sich unter der Adresse 0000_{16} der Hersteller-Code und unter Adresse 0001_{16} der Baustein-Code. Die Hersteller geben in der Regel den Lösch- und Programmier-Algorithmus in den entsprechenden Datenbüchern mit an. Flash-EPROM lassen sich kostengünstiger als EEPROM herstellen. Abb. 52.45 zeigt als Beispiel das Funktionsbild und die Anschlussbelegung eines 512 K × 1 Bit-CMOS-Flash-Speichers. Mit den beiden TSOP-Gehäusevarianten lassen sich Platinenlayouts vereinfachen.

52.3.6.2 Schreib-Lesespeicher

Die Daten dieses Speichers können über eine Adresse beliebig oft gelesen oder eingeschrieben werden. Beim Ausfall der Betriebsspannung sind die Daten des Speichers allerdings verloren. Man spricht daher auch von flüchtigen Speichern (Volatile RAM). Eine Übersicht verschiedener Schreib-Lese-Speicher zeigt Abb. 52.46.

52.3.6.2.1 SRAM

Statische RAM werden in bipolarer oder MOS-Technik ausgeführt. Sie arbeiten nach dem Prinzip einer bistabilen Kippstufe

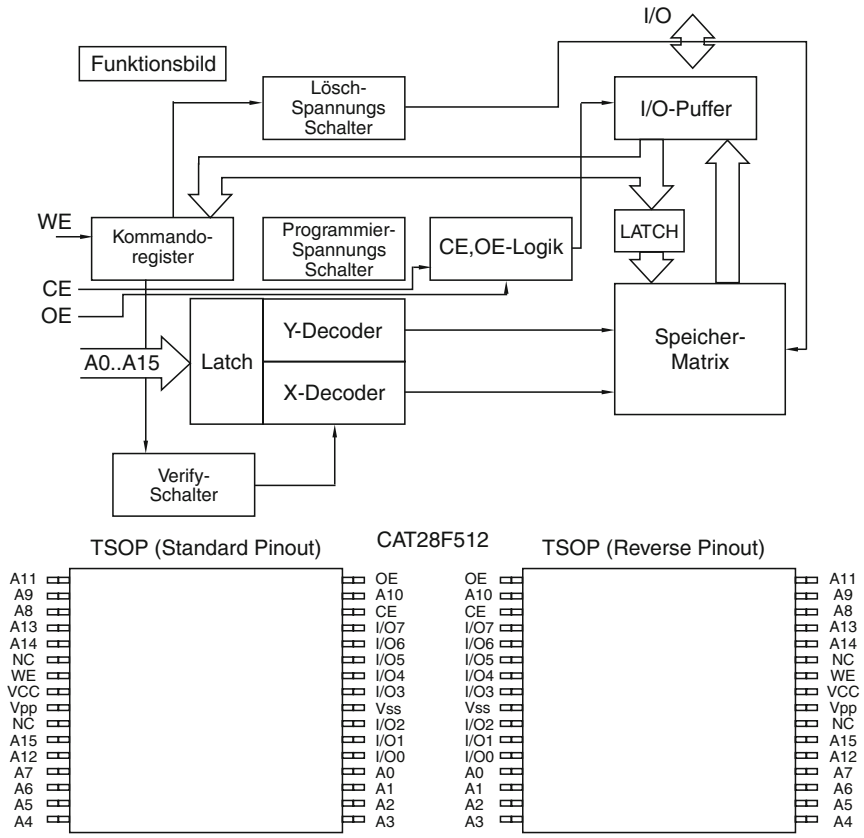


Abb. 52.45 Pin-Belegung und Funktionsbild eines Flash-Bausteins

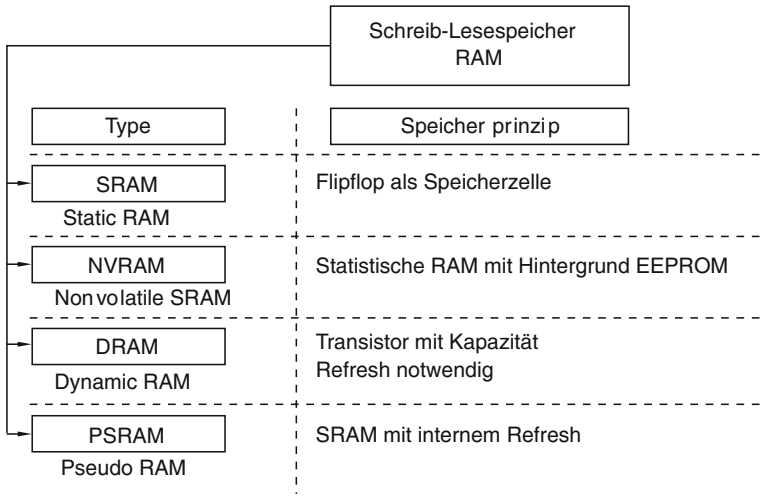


Abb. 52.46 Schreib-Lesespeicher

(Abb. 52.47). SRAM werden mit unterschiedlicher Kapazität und Organisation angeboten (Tab. 52.14). Aus der Tabelle ist zu erkennen, dass diese Speicher mit ihrer kurzen Zugriffszeit recht schnell sind.

Abb. 52.48 zeigt das Funktionsbild eines Schreib-Lese-Speichers. Die Auswahl der 4-Bit-Speicherstelle erfolgt über die Matrix. Bei einer 64 × 64-Speichermatrix wird die Spaltenmatrix in 16 × 4 Spalten zur Ansteuerung unterteilt. Damit ergeben sich für den Spaltendecoder 4 Adressleitungen. Die Verbindung zum Datenbus erfolgt über die Ein- und Ausgabeschaltung. Die Steuerung der Bausteinauswahl geschieht hier über den Steuereingang /CS. Mit RD, /WR kann die Steuerung für Lesen oder Schreiben erfolgen. Datenein- und Datenausgabe erfolgt über *Tristate-Treiber*, über die es möglich ist, den Ausgang hochohmig zu schalten.

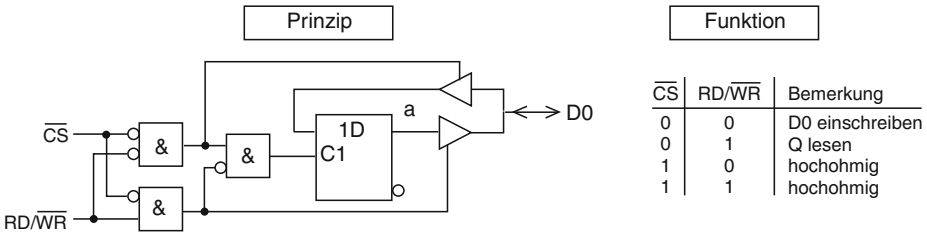


Abb. 52.47 Prinzip einer Speicherzelle

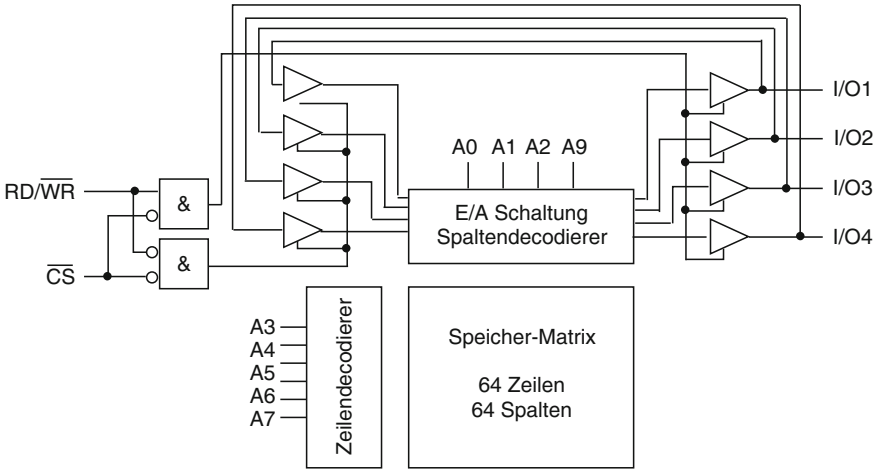


Abb. 52.48 Funktionsbild eines SRAM-Bausteins

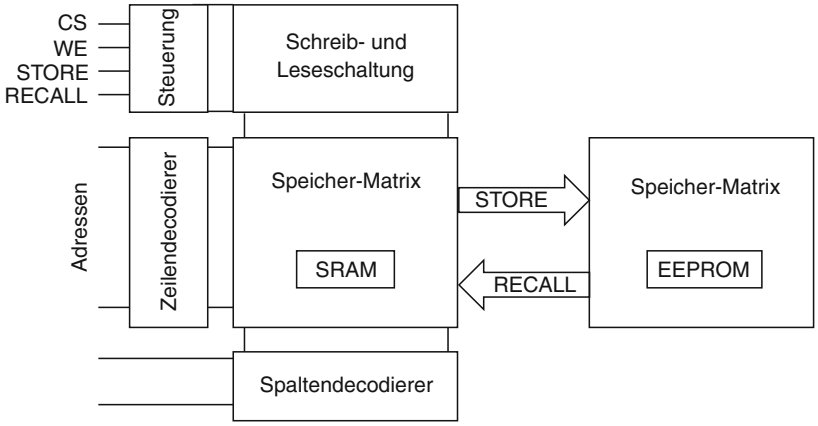


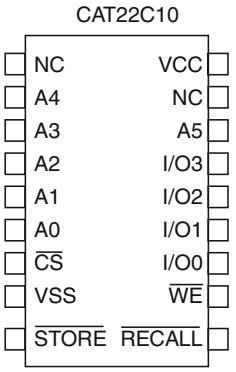
Abb. 52.49 Prinzip eines NVRAM

52.3.6.2.2 NVRAM

Das NVRAM (Nonvolatile SRAM) ist ein statisches RAM mit der Fähigkeit, über die Steuerleitung *STORE* seine Daten in ein spiegelbildgleiches EEPROM zu kopieren. Damit sind die Daten nicht mehr flüchtig. Sie bleiben bei Ausfall der Betriebsspannung erhalten. Die EEPROM-Daten können jederzeit über die Steuerleitung *RECALL* in das SRAM zurückgeholt werden. Abb. 52.49 zeigt das Funktionsbild eines NVRAM, und Abb. 52.50 erläutert die Steuerung der beiden unterschiedlichen Betriebsarten.

52.3.6.2.3 DRAM

Dynamische RAM-Speicherzellen (Dynamic RAM's) bestehen im Prinzip aus einem Transistor und einer Kapazität. Dadurch ergeben sich kleine Verlustleistungen bei großen Integrationsdichten (Abb. 52.51). Der Transistor ist der Schalter, der die Kapazität auflädt oder entlädt, um den Informationszustand „1" oder „0" zu erhalten. Da sich die Kapazitäten in kurzer Zeit über Sperrschichten und Dielektrikum entladen, verliert der Speicher seine Information, und in bestimmten Zeitabständen muss ein Auffrischzyklus erfolgen. Beispielsweise kann nach 2 ms der



Betriebsarten

	CS	WE	RECALL	STORE
Standby	1	x	1	1
Speicher lesen	0	1	1	1
Speicher schreiben	0	0	1	1
EEPROM nach RAM	x	1	0	1
EEPROM nach RAM	1	x	0	1
RAM nach EEPROM	x	1	1	0
RAM nach EEPROM	1	x	1	0

Abb. 52.50 NVRAM und Betriebsarten

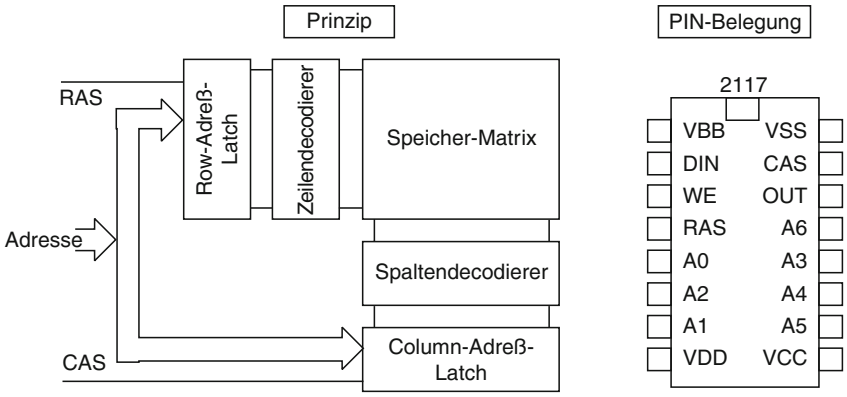


Abb. 52.51 Prinzip eines DRAM

Tab. 52.14 SRAM-Beispiele

Type	Kapazität	Organisation	Zugriffszeit
HM6287P	64 K	64 K × 1	70 ns
HM6264	64 K	8 K × 8	150 ns
HM6207	256 K	256 K × 1	45 ns
HM6208	256 K	64 K × 4	45 ns
HM62256	256 K	32 K × 8	150 ns
HM621100	1 M	1 M × 1	35 ns
HM628128	1 M	128 K × 8	120 ns
HM621664	1 M	64 K × 16	25 ns
HM621400	4 M	1 M × 4	35 ns

Tab. 52.15 DRAM-Beispiele

Type	Kapazität	Organisation	Zugriffszeit
21256	256 K	256 K × 1	150 ns
21464	256 K	64 K × 4	150 ns
21010	1 M	1 M × 1	130 ns
21044	4 M	1 M × 4	130 ns

(Nach Datenblatt der Firma Intel)

gesamte Speicher aufgefrischt werden (Blockrefresh). Ein zusätzlicher Takt leitet den Zyklus ein. Während dieser Zeit kann die CPU nicht auf den Speicher zugreifen. Die Refreshzeiten (refresh = auffrischen) der einzelnen Speicher sind unterschiedlich. Der Betrieb dynamischer RAM's hat eine besondere Adressierung, wie das folgende Beispiel zeigt.

Der Speicher 2117 besitzt eine Speicherkapazität von 16 K × 1 Bit. Die Speichermatrix selbst ist in Reihen und Spalten angeordnet. Um alle Speicherplätze anzusprechen, werden 14 Adressleitungen benötigt. Damit man dennoch mit einem 14-DIL-Gehäuse auskommt, werden die Adressen in zwei Gruppen aufgeteilt (Zeilenadresse = Row-Adress und Spaltenadresse = Column-Adress), die zeitlich nacheinander mit den jeweiligen Steuerleitungen RAS (Row Address Strob = Reihenadressenzugriff) und CAS (Column Adress Strobe = Spaltenadressenzugriff) den Baustein ansprechen. Zunächst wird dem Baustein über den Adressbus die Zeilenadresse zugeführt und im Latch gespeichert. Anschließend gelangt die Spaltenadresse zum Column-Latch. Den zeitlichen Zusammenhang eines Lese- und Schreibzyklus beschreibt Abb. 52.52. DRAM-Speicher sind deutlich langsamer als SRAM-Speicher, wie aus der nachfolgenden Tab. 52.15 hervorgeht. Überlegen sind sie statischen Speichern allerdings durch eine höhere Integrationsdichte und geringerer Verlustleistung.

52.3.6.2.4 PSRAM

PSRAM (Pseudo SRAM) sind nach außen hin wirkende statische RAM-Bausteine mit einer inneren Struktur und notwen-

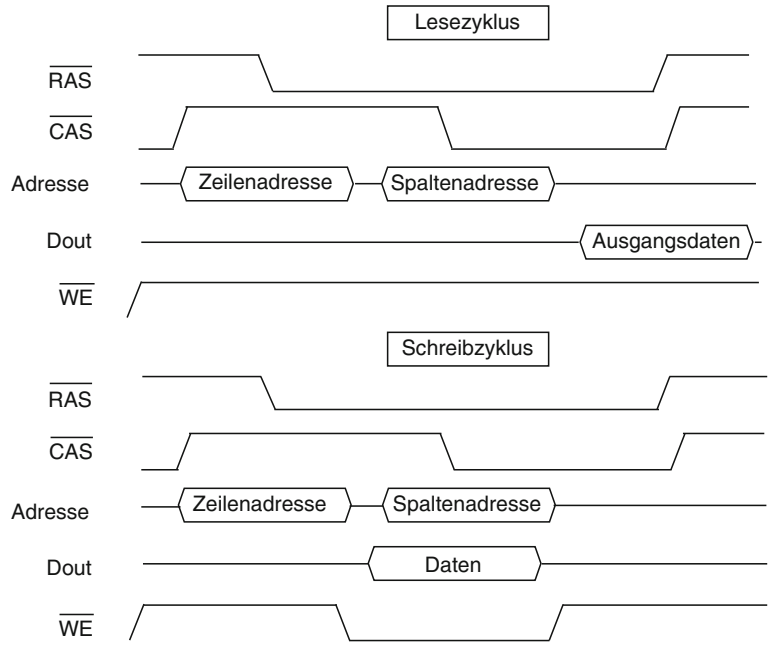


Abb. 52.52 Lese- und Schreibzyklus eines DRAM

Tab. 52.16 PSRAM-Beispiele

Type	Kapazität	Organisation	Zugriffszeit
HM65256	256 K	32 K × 8	150 ns
HM658128	1 M	128 K × 8	120 ns
HM658512	4 M	512 K × 8	120 ns

(Nach Datenblatt der Firma Hitachi)

digen Hilfsschaltungen eines dynamischen RAM. Die Refreshschaltung ist in diesem Baustein integriert (Tab. 52.16).

52.3.7 Speichererweiterung

Der Zentralspeicher oder Hauptspeicher eines Mikrocomputers besteht aus mehreren Speicherbausteinen (RAM, ROM). Die Erweiterung des Speicherraumes wird durch Parallelschalten aller Daten- und Adressleitungen erreicht. Um das gleichzeitige Ansprechen mehrerer Bausteine zu verhindern, sind Auswahlbedingungen erforderlich.

Beispiel 1 (siehe Abb. 52.53)

An einem 8-Bit-Prozessor sollen 4 Speicher mit einer Kapazität von 2 K × 8 Bit angeschlossen werden. Da alle Speicher mit ihren Adressleitungen A0 bis A10 an dem Adressbus liegen, würden alle Speicher bei Ausgabe einer Adresse im Bereich bis $07FF_H$ gleichzeitig angesprochen werden. Um dies zu verhindern, wird mit Hilfe der Auswahleingänge CS jedem Speicher ein bestimmter Speicherbereich zugeordnet. Es werden dazu die Adressleitungen benutzt, die nicht zu den Adressenleitungen gehören, die an den Speichern angeschlossen sind. Da bei diesem Verfahren nicht alle Adressleitungen des Systems in Gebrauch sind, können Speicher mit mehreren Adressen angesprochen werden. Beispielsweise wird mit der ADR 8000H dieselbe Speicherstelle wie mit der ADR 0000_H angesprochen. Man spricht auch von *unvollständiger Decodierung*.

Häufig werden *Decodierer* zur Unterscheidung der Speicher eingesetzt (Abb. 52.54). Der Decodierer wird dann von Adressleitungen, die nicht an die einzelnen Speicher angeschlossen sind, gesteuert. Um eine Mehrfachansprache von Speicherstellen zu vermeiden, sollte ein Decodierer gewählt werden, der einen Auswahleingang besitzt. Mit Hilfe einer einfachen Logik können alle Adressen in die Decodierung mit einbezogen werden. Man spricht dann von *vollständiger Decodierung* (Tab. 52.17). Die Anfangs- und Endadressen der einzelnen Speicher werden jeweils aufgeführt. Da die Speicher eine Speicherkapazität von 2 K × 8 Bit besitzen, sind jeweils 11 Adressleitungen (A0 ... A10) vorhanden ($2^{11} = 2048 = 2$ K). Zur Unterscheidung der Speicher (Decodierung) werden die zwei nachfolgenden Adressleitungen A12 und A13 benutzt. Alle übrigen Adressleitungen werden über eine ODER-Verknüpfung zur Deaktivierung bzw. Aktivierung des Decoders eingesetzt. Damit sind die Speicherplätze eindeutig bestimmt und ergeben den in Abb. 52.54 dargestellten Speicherplan. ◀

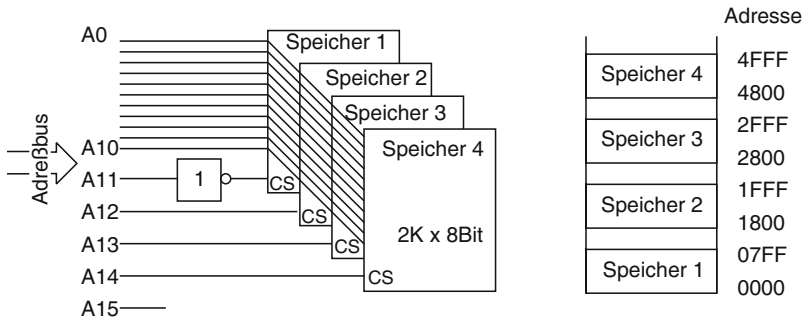


Abb. 52.53 Unvollständige Decodierung und Speicherplan

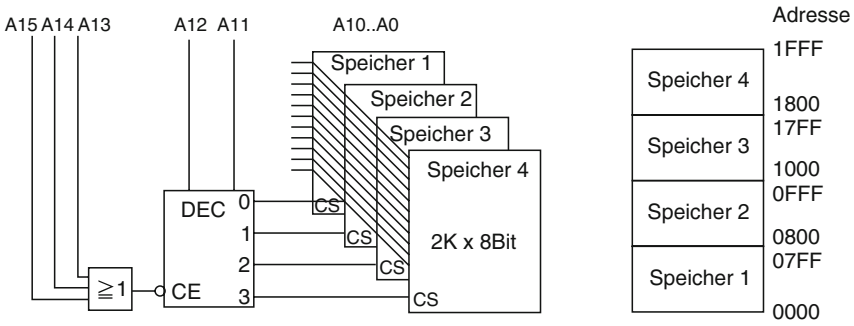


Abb. 52.54 Vollständige Decodierung

Tab. 52.17 Vollständige Decodierung

Adreßleitung

	A15	A14	A13	A12	A11	A10	A9	A8	A7	A6	A5	A4	A3	A2	A1	A0
Speicher 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
			[0]				[7]				[F]				[F]	°
	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Speicher 2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
																°
	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Speicher 3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
																°
	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1
Speicher 4	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
																°
	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Grundbereich

Codierung — Speicher 1 bis 4

Deaktivierung des Zentralsspeichers — bei Adressen > 1FFF

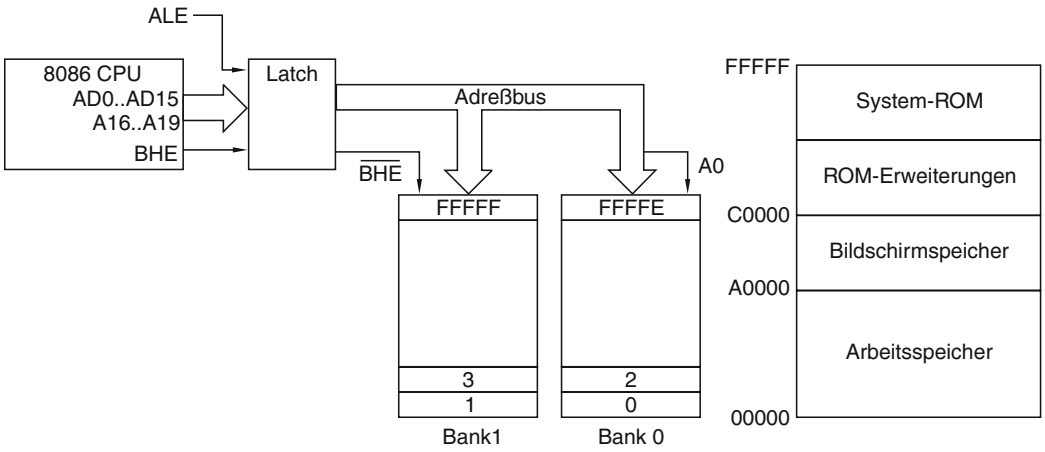


Abb. 52.55 Blockbild eines 1-MByte-Speichers und SPeicherorganisation

52.3.8 Zentralspeicher

Der Zentralspeicher eines Mikroprozessorsystems ist sein Hauptspeicher. Soll ein solcher Zentralspeicher für ein Mikroprozessorsystem entworfen werden, muss die Größe des gewünschten Speichers bekannt sein. Anschließend können gängige Speicherbausteine ausgewählt, ein Speicherplan aufgestellt und der Adressdecodierer geplant werden.

Beispiel

Ein Zentralspeicher für einen 8-Bit-Prozessor soll mit seinem 64-K-Speicherraum voll ausgenutzt werden. Als Speicherbausteine stehen 16 K × 8-Bit-Speicher zur Verfügung. Zu jedem Speicher gehen somit 14 Adressleitungen (A0 . . . A13). Die Adressleitungen A14 und A15 können dann zur Decodierung der vier Speicherbausteine herangezogen werden.

Der Zentralspeicher eines PC mit 8086/80286 CPU und seinem 1-MB-Adressbereich ist physikalisch in zwei Speicherblöcke mit je 512 kB aufgeteilt (Abb. 52.55). Die Auswahl wird über die Adresse A0 (Bank 0) und die Steuerleitung/BHE (Bank 1) vorgenommen. Die Datenleitungen sind so aufgeteilt, dass D0 bis D7 Bank 0 und D8 bis D15 Bank 1 verbinden, während die Adressleitungen A1 bis A19 an beiden Speicherblöcken anliegen.

Der Adressbereich ist in einen RAM- und ROM-Bereich unterteilt. Dem eigentlichen Arbeitsspeicher sind max. 640 kB zugeordnet und dem Bildschirmspeicher die nachfolgenden 128 kB. Das BIOS (Basic Input/Output Services) liegt mit ebenfalls 128Kbyte bis zur Adresse FFFFF. Der zwischen Bildschirmspeicher (Video-RAM) und BIOS liegende Bereich ist für ROM-Erweiterungen wie beispielsweise das BIOS des Festplattenkontrollers reserviert. Aus diesem Speicherbereich werden ebenfalls 64 kB als Expansionsspeicher (EMS) verwendet, um über *Bank-Switching* die Adressierung eines größeren Speicherbereiches zuzulassen. Prozessoren ab 80286 können mehr als 1 MB adressieren. Der Speicher oberhalb 1 Mbyte wird *Extendet Memory* genannt und kann durch zusätzliche Speicherkarten realisiert werden (Erweiterungsspeicher). Mit Hilfe initialisierbarer Steuerprogramme kann dieser Speicherbereich zum Speichern von Daten oder zur Verwendung von Programmen genutzt werden.

Bei der Speicherverwaltung der 80 × 86 unterscheidet man die Betriebsarten Real Mode und Protected Mode. Im Real Mode wird ein Speicherbereich von 1 MB verwaltet, da nur die Adressleitungen A0 bis A19 verwendet werden. Damit sind in dieser Betriebsart die Prozessoren mit dem 8086 kompatibel. Im Protected Mode werden alle Adressleitungen und damit der Speicherbereich oberhalb von 1 MB (Extended Memory = Erweiterungsspeicher) ausgenutzt. Der 80286 kann somit einen Speicherbereich von 16 MB adressieren. ◄

52.4 Peripheriebausteine

52.4.1 Allgemeines

Peripheriebausteine sind *Schnittstellenbausteine*, die periphere Geräte wie Monitor, Tastatur, Drucker oder Disketten-Laufwerke mit dem Bussystem des Prozessors verbinden. Man unterscheidet einfache (nichtprogrammierbare) Bausteine, die ihre Funktion durch die hardwaremäßige Beschaltung in Verbindung mit den Ein- und Ausgabebefehlen des Prozessors durchführen und *programmierbare* Bausteine, die zunächst durch Software initialisiert werden müssen. Bausteine mit besonderen Aufgaben sind Treiberstufen, Taktgeneratoren, Zeitgeberbau-

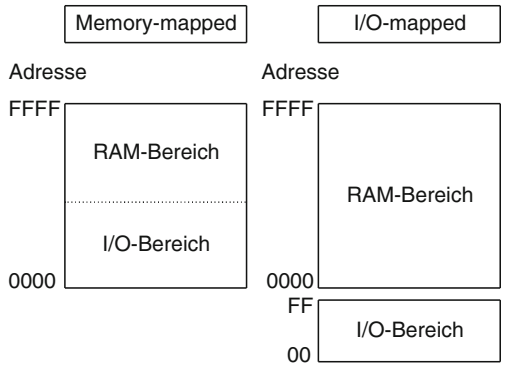


Abb. 52.56 Adressierungsarten und Adressenaufteilung

steine (TIMER) und Steuerbausteine für den direkten Speicherzugriff (DMA = Direct Memory Access).

Die *Schnittstelle* wird *Interface* genannt. Sie passt die Signale der Schnittstellenbausteine an die Signale, die der Prozessor liefern oder empfangen kann, an. Damit ergibt sich für jeden Baustein eine Bus- und eine Geräte-Schnittstelle.

Generell kann die Datenein- und -ausgabe mit dem Lesen und Beschreiben eines Speicherbausteins verglichen werden. Hat die Adresse den gleichen Aufbau wie die Speicheradresse, spricht man von „*Memory-mapped-Adressierung*“. Der E-/A-Baustein wird hierbei als Teil des Speichers angesehen. Um Speicher und E-/A-Baustein zu unterscheiden, schafft man mit Adressdecodern unterschiedliche Adressbereiche. Bei einem 16 Bit breiten Adressbus liegen dann E-/A- Baustein und Speicher in einem Adressbereich von 0000 bis FFFF.

Wird die Unterscheidung zu den Speichern mit Signalen wie IO/M oder IORQ des Steuerbusses durchgeführt, spricht man von „*I/O-mapped-Adressierung*“. Hierbei wird bei IN-Befehlen oder OUT-Befehlen eine 8-Bit-Adresse auf dem Low-Adressbus übertragen. Unter Ausnutzung aller Adressen ergibt sich hierbei ein Adressbereich von 00 bis FF. Abb. 52.56 stellt die beiden Adressierungen gegenüber.

52.4.2 BUS-Treiber

Jeder Bausteineingang, der an die CPU angeschlossen ist, belastet die CPU-Ausgänge. Dabei wird zwischen statischem und dynamischem Lastverhalten unterschieden. Beim dynamischen Lastverhalten spielen die parasitären Kapazitäten, also die Eigenkapazität der Bauteile und Kapazitäten, die durch den Schaltungsaufbau bedingt sind, eine große Rolle, weil hierdurch Signale verformt werden. Da auch die statische Belastbarkeit der CPU gering ist, dürfen nicht beliebig viele Bauelemente an den Prozessor angeschlossen werden. Um ein Fehlverhalten der CPU auszuschließen, werden BUS-Treiber zur Lasterweiterung und Signalformung eingesetzt. Handelsüblich sind Einrichtungs- (unidirektional) und Zweirichtungstreiber (bidirektional). Ihr Einsatz ist abhängig von der Datenflussrichtung. Während Einrichtungstreiber beispielsweise als Adressenbustreiber eingesetzt werden, dienen Zweirichtungstreiber zur Datenübertragung. Mit einem Eingang (OE = Output Enable) können die Treiber in einen hochohmigen Zustand (Tristate, siehe auch Abschn. 42.1 „Integrierte Schaltungen“) versetzt werden. Der Zweirichtungstreiber besitzt neben diesem Freigabeeingang einen weiteren Eingang zur Umschaltung der Datenflussrichtung (DIR = Direction). Abb. 52.57 zeigt einen Mikroprozessor mit Treiberstufen und erläutert über eine Funktionstabelle das Prinzip der Umschaltung des Zweirichtungstreibers. Mit dem Eingang OE = „0“ wird der Baustein aktiviert. Bei DIR = „1“ gelangen die Daten von A0 nach B0; bei DIR = „0“ von B0 nach A0. Eine „1“ am Steuereingang des Tristatetreibers veranlasst diesen durchzuschalten, ansonsten ist der Ausgang hochohmig.

52.4.3 Einfache E-/A-Bausteine für den parallelen Betrieb

E-/A-Bausteine nennt man auch Portbausteine, da sie das „Tor“ zur Außenwelt sind. Sie werden direkt an den Datenbus angeschlossen und können als Ein- oder Ausgabebaustein funktionieren. Zur Auswahl der Bausteine ist ein Adressdecoder nötig, der aus einer simplen Logik aufgebaut sein kann. Bringt man die Steuersignale WR und RD mit in die Decodierung des Adressdecoders ein, kann die gleiche Adresse für Ein- und Ausgabeport benutzt werden, weil ein Zugriff mit RD nur lesend und mit WR nur schreibend erfolgt. Mit dem Eingabebefehl gelangen die Daten vom Eingabeport zum Akkumulator und mit dem Ausgabebefehl vom Akkumulator zum Ausgabeport. Der Datenfluss läuft immer über den Akkumulator. Da die Eingabedaten für den Portbaustein nur kurzzeitig anliegen brauchen, um von der CPU angenommen zu werden, lassen sich auch Tristate-Treiber als Eingabebausteine benutzen. Anwendungen für die Eingabe sind von Hand oder durch Maschinen betätigte Schalter, Taster oder Optokoppler. Mechanische Taster prellen oft. Der Prozessor erkennt dann unter Umständen mehrere Impulse. Dies muss durch Software oder Hardware ausgeschlossen werden.

Um einen eindeutigen H-Pegel bei geöffneten Schaltern zu erhalten, sind jeweils Widerstände gegen +UCC gelegt (Pull-up-Widerstände). Ihre Werte liegen üblicherweise zwischen 1 kΩ und 10 kΩ. Die Ausgabedaten stehen nur kurzzeitig während des Ausgabe-Befehls an der Peripherie an. Da die Daten bis zur nächsten Ausgabe anstehen sollen, müssen sie gespeichert werden. Dies erreicht man mit D-Flipflops. Abb. 52.58 zeigt als Beispiel einen Eingabebaustein mit Decoder. Mit dem Befehl „IN FF“ gelangen die anliegenden Daten zum Akkumulator.

Abb. 52.59 zeigt die Innenschaltung des Ausgabebausteins mit dem entsprechenden Adressdecodierer. Mit dem Befehl „OUT FF“ werden die Daten des Akkumulator zum Ausgangsport kopiert.

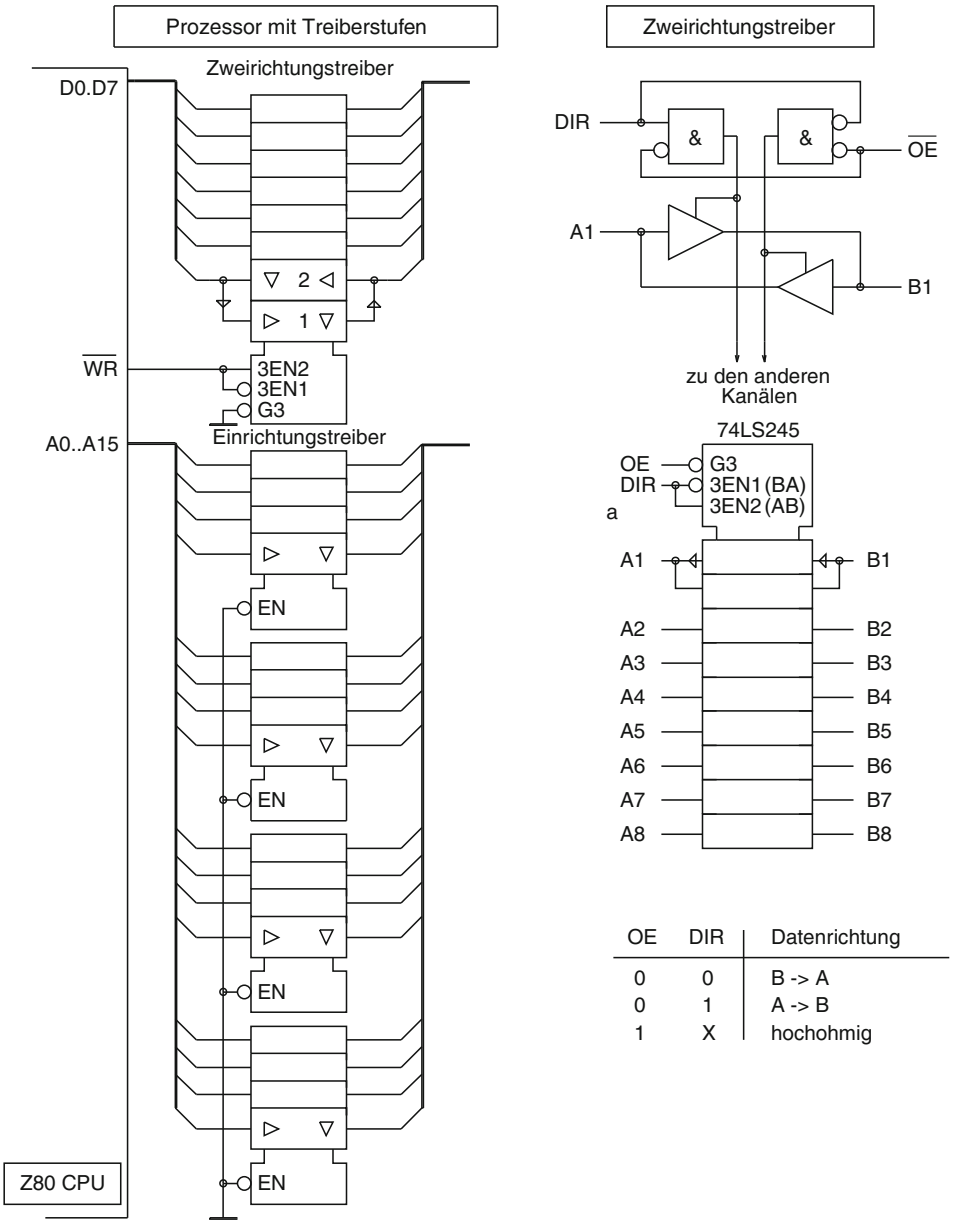


Abb. 52.57 Prozessor mit Treiberstufen und Zweiruchtungstreiber

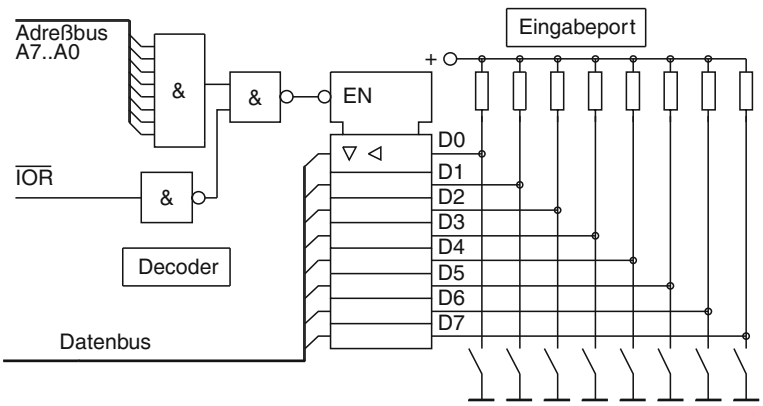


Abb. 52.58 Einfaches Eingabeport mit Decoder

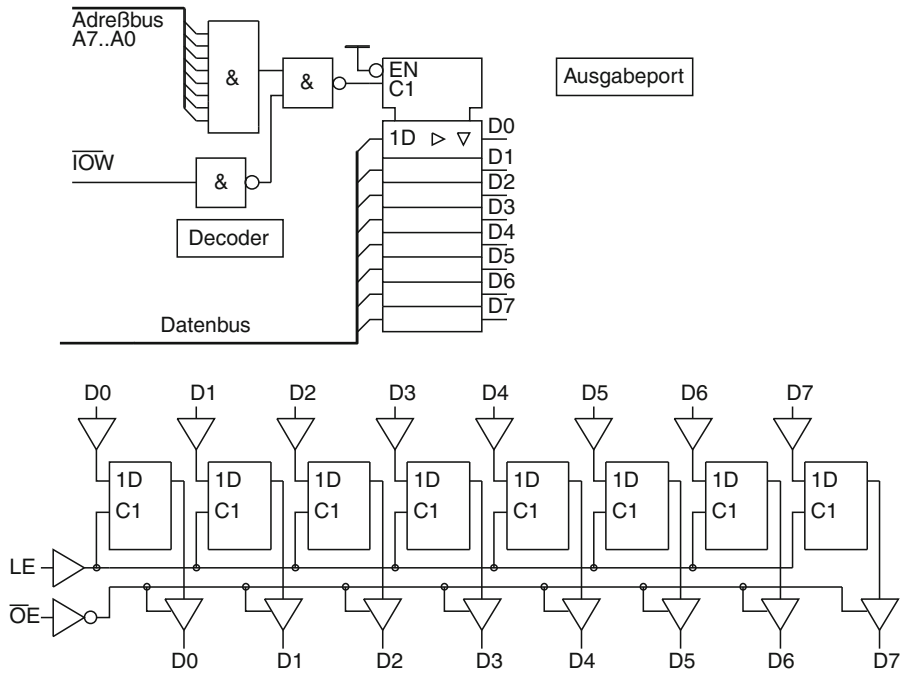


Abb. 52.59 Ausgabeport mit Decodierer und Portbaustein

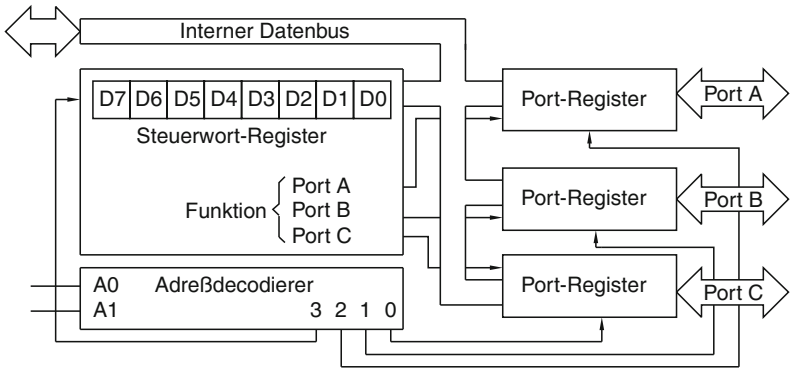


Abb. 52.60 Prinzip eines programmierbaren Portbausteins

52.4.4 Programmierbare Schnittstellenbausteine

Ein Kennzeichen dieser Bausteine (Abb. 52.60) sind 16 bzw. 24 programmierbare Ein-/Ausgabeleitungen, wobei 8 Leitungen jeweils zusammengefasst ein Port darstellen.

Die Funktion dieser Bausteine ist nicht allein hardwaremäßig festgelegt. Über ein *Kommando-* oder *Steuerwortregister* kann die gewünschte Funktion und Betriebsart softwaremäßig eingegeben werden. Damit können Ports als Eingabe- oder Ausgabeports arbeiten. Des Weiteren können Betriebsarten gewählt werden, bei denen Portleitungen als Steuer- oder Meldeleitungen definiert sind. Die einmal gewählte Funktion bleibt solange erhalten, bis der Inhalt des Steuerwortregisters überschrieben wird oder eine Rückstellung erfolgt.

Ports und Steuerwortregister können über Adressen angesprochen werden. Besitzt ein Baustein ein zusätzliches Zustands- oder Statusregister, über das Zustände der Ports abgefragt werden, kann seine Adresse mit der Adresse des Kommandoregisters übereinstimmen. Die Unterscheidung der beiden Register erfolgt dann über die Steuerleitungen /RD und /WR. Um den Baustein zu initialisieren und die gewünschte Funktion zu speichern, muss zunächst das Steuerwort in das Steuerwortregister gebracht werden. Komplexere Bausteine, wie der 8156, besitzen zusätzlich einen Zeitgeber und Speicher. Sie benötigen dann mehr als zwei Adressleitungen.

Häufig eingesetzte Bausteine sind der 8255 von INTEL und Z80-PIO von ZILOG. Die Adressierung dieser Bausteine erfolgt wie bei den Speichern über einen Auswahleingang /CS (Chip Select). Das Signal für CS kann im einfachsten Fall aus dem Signal IO/M bzw. IORQ und einem Adresssignal gebildet werden.

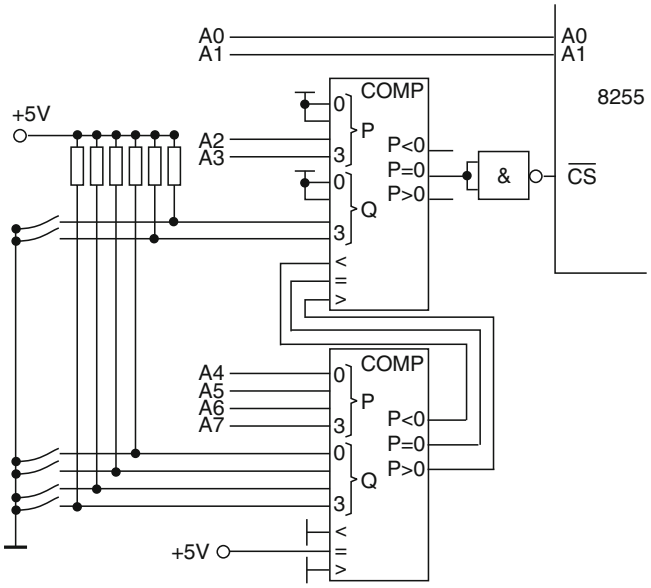


Abb. 52.61 Adressdecoder für I/O-Bausteine

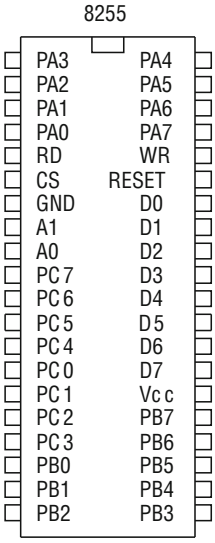


Abb. 52.62 PIN-Belegung des 8255

Tab. 52.18 Bedeutung der Anschlüsse

Anschlüsse	Funktion
D0 ... D7	Datenbus
RESET	Rücksetzungseingang
CS	Bausteinauswahl
RD	Leseeingang
WR	Schreibeingang
A0, A1	Adressleitungen
PA0 ... PA7	Port A
PB0 ... PB7	Port B
PC0 ... PC7	Port C
VCC	+5 V Betriebsspannung

Häufiger findet man Adressdecoder aus Vergleicherstufen. Der Ausgang dieser Stufen wird nur dann „1“, wenn die Bitmuster am Eingang A mit denen an Eingang B übereinstimmen. Mit Hilfe von Jumpern (Steckbrücken) oder DIL-Schaltern ist dann jede Adresse von 00 bis FF wählbar. Abb. 52.61 stellt ein Beispiel für einen Adressdecoder dar. Die Anschlussbelegung des Portbausteins erkennt man im Abb. 52.62. Abb. 52.63 zeigt das Funktionsbild des 8255.

Der Baustein besitzt 3 TTL-kompatible Ports mit je 8 Leitungen oder Kanälen. Port C kann in 2 Gruppen zu je 4 Leitungen unterteilt werden. Hierbei gibt es allerdings nur die Möglichkeit, die L-Bits, das sind die Leitungen PC0 bis PC3, und die H-Bits mit den Leitungen PC4 bis PC7 zusammenzufassen. Die Steuerlogik bestimmt mit dem Kommando- oder Steuerwort, ob ein Port als Eingabe- oder Ausgabeport arbeitet. Das Format dieses Steuerwortes sehen Sie im Abb. 52.64. Hierbei kommt jeder Bitstelle eine bestimmte Bedeutung zu. D1 entscheidet beispielsweise über Ein- oder Ausgabe von Port B, und D4 ordnet die Eigenschaften dem Port A zu (Tab. 52.19).

Der 8255 unterscheidet die Betriebsarten 0, 1 und 2.

In der Betriebsart 0 können die Ports jeweils als Ein- oder Ausgabe programmiert werden, wobei das Port C noch in zwei 4-Bit-Leitungsgruppen unterteilt ist. Die Eingabedaten gelangen ohne Zwischenspeicherung zum Datenbus, während bei der Ausgabe der Daten eine Zwischenspeicherung erfolgt. Die im Abb. 52.64 aufgeführten Beispiele gelten für diese Betriebs-

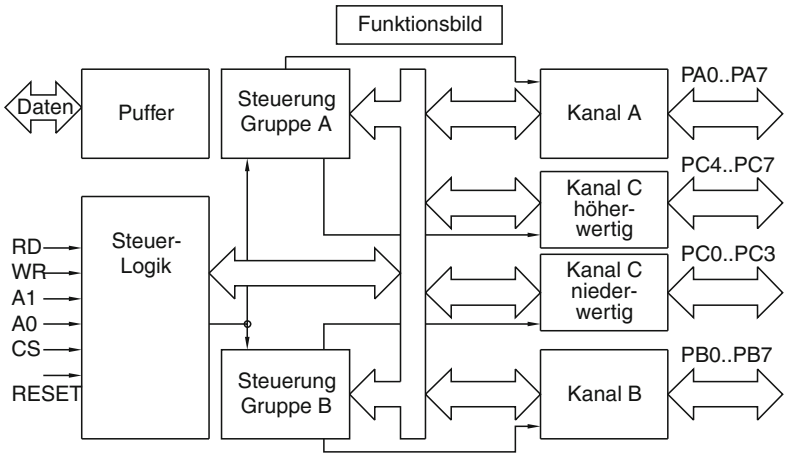


Abb. 52.63 Funktionsbild des 8255

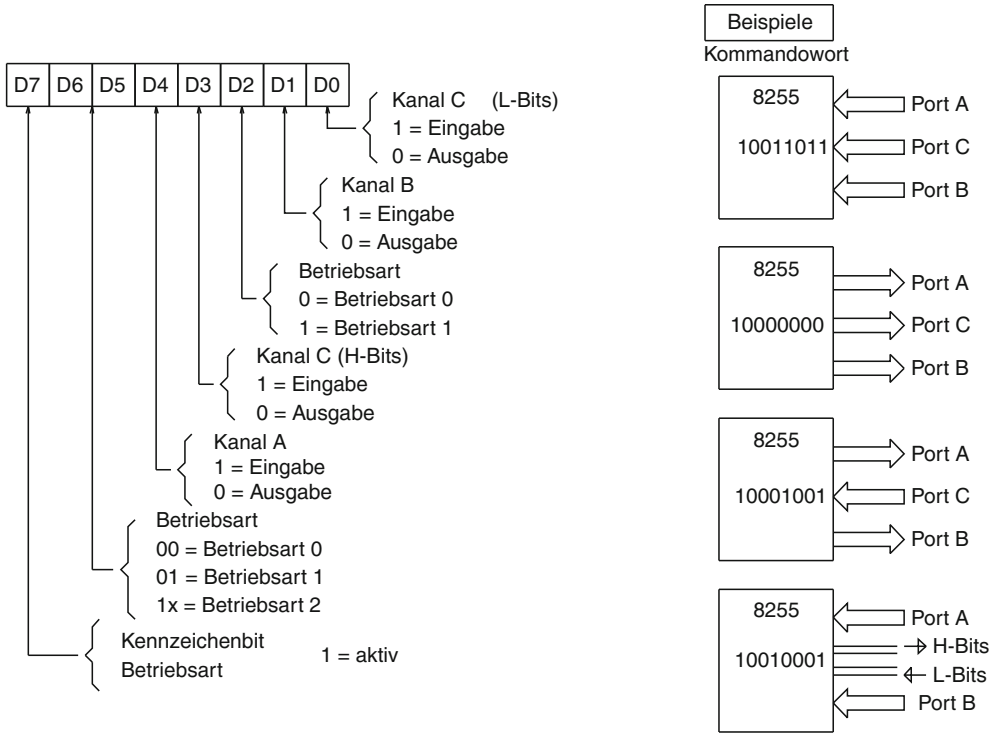


Abb. 52.64 Format des Steuerwortes und Beispiele

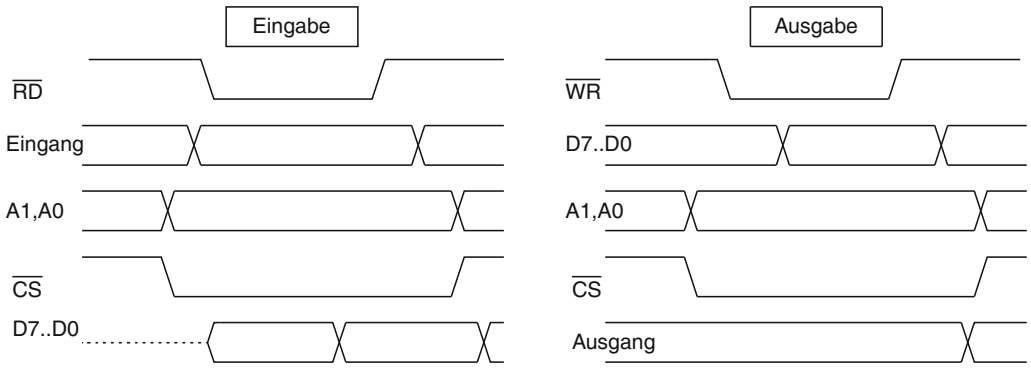


Abb. 52.65 Impulsdiagramm für die Betriebsart 0

Tab. 52.19 Definition der Steuersignale in der Betriebsart 1 und 2

STB	Strobe	„0“ ermöglicht es, Daten in den Eingangs-zwischenspeicher zu laden
IBF	Input Buffer Full	„1“ zeigt an, dass Daten geladen wurden
OBF	Output Buffer Full	wird „0“, wenn Daten geschrieben wurden
ACK	Acknowledge	Quittungseingang
INTR	Interrupt	Unterbrechungsanforderung

art. Das zeitliche Zusammenwirken der Signale für die Ein- und Ausgabe gibt Abb. 52.65 wieder. Mit einem L-Pegel am Eingang /WR können Daten oder Steuerworte in den Baustein geschrieben und mit einem L-Pegel am Eingang RD können Daten vom Baustein gelesen werden. Während dieser Zeit muss die gültige Adresse an A0 und A1 anliegen. Angewählt wird der Baustein mit der Bausteinauswahlleitung /CS. Ein H-Pegel am Eingang RESET setzt alle internen Register zurück.

Bei der Betriebsart 1 (getastete Ein-Ausgabe) kann Port A oder Port B als Eingabe oder Ausgabe benutzt werden. Das Port C dient mit je drei Leitungen zur Steuerung (Abb. 52.66). In dieser Betriebsart ist *Handshake-Betrieb* möglich (Abb. 52.67). Diese Betriebsart wird verwendet, wenn Geräte, die nicht synchronisiert sind, Daten austauschen sollen. Sobald die Eingabedaten den Portbaustein erreicht haben, kann über ein Interrupt der CPU eine Meldung gemacht werden. Über eine Interruptroutine werden die Daten eingelesen. Die genaue Funktion verdeutlicht das Prinzip im Abb. 52.68 in Verbindung mit dem gleichzeitig dargestellten Impulsdiagramm.

Input Buffer Full (IBF) ist zunächst „0“. Damit wird dem Eingabegerät ein Zugang zu dem Eingabebaustein ermöglicht. Der

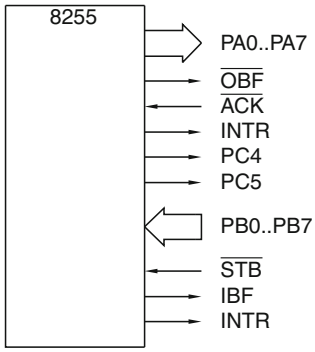


Abb. 52.66 Steuerleitungen für die getastete Ein-Ausgabe

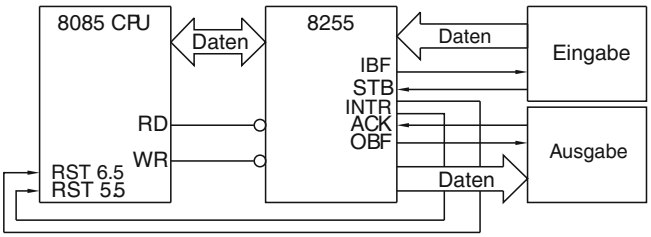


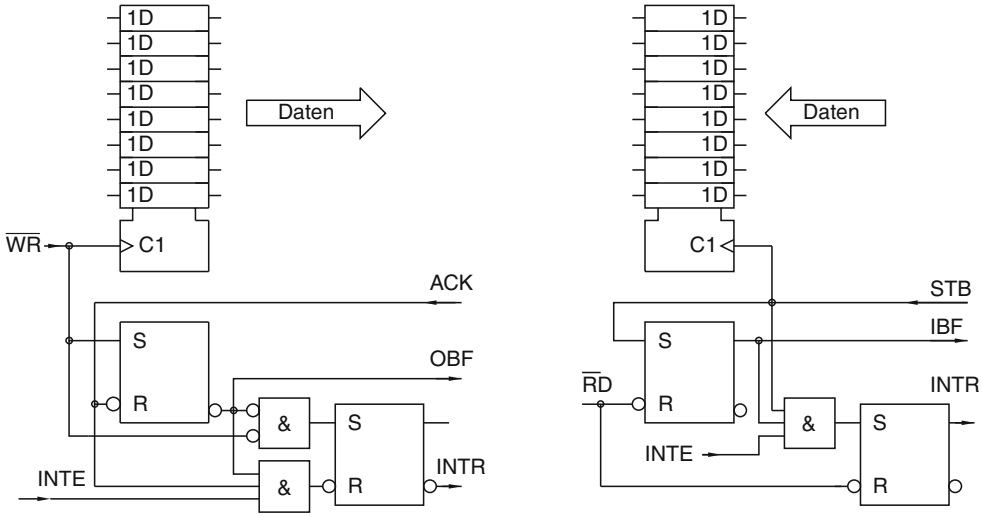
Abb. 52.67 8255 in der Betriebsart 1

Baustein 8255 zeigt an, dass er Daten aufnehmen kann. Stellt das Eingabegerät nun Daten zur Verfügung, wird dies dem Baustein durch STB = „0“ mitgeteilt, und die Daten werden übernommen. Gleichzeitig wird IBF = „1“ gesetzt und dem Eingabegerät signalisiert, dass keine weiteren Daten übertragen werden. Über INTRB = „1“ wird die Meldung von der Ankunft der Daten an die CPU übermittelt und kann über eine Interruptroutine von ihr bearbeitet werden. Beim Lesen der Daten werden INTRB und IBF wieder auf „0“ gesetzt (Tab. 52.15).

Sind die Ausgabedaten vom Ausgabegerät übernommen worden, kann dies wiederum über ein Interrupt der CPU mitgeteilt werden. Die CPU sendet Daten zum Baustein und gibt eine Meldung über OBF = „0“ zum Ausgabegerät, um die Ankunft neuer Daten mitzuteilen. Das Ausgabegerät kann die Daten übernehmen und mit ACKA = „0“ bestätigen und anschließend über INTRA neue Daten von der CPU anfordern.

Bei der Betriebsart 2 erfolgt der Datenaustausch nur über einen gemeinsamen Datenbus. Die Adressen steuern in Verbindung mit /WR und /RD die Auswahl der Ports oder des Steuerwortregisters.

Voraussetzung für die Programmierung des Bausteins sind Kenntnisse der Registeradressen und der Steuerlogik. Tab. 52.20 zeigt die Zusammenarbeit der Steuersignale und Adressen mit



Impulsdiagramm für die Betriebsart 1

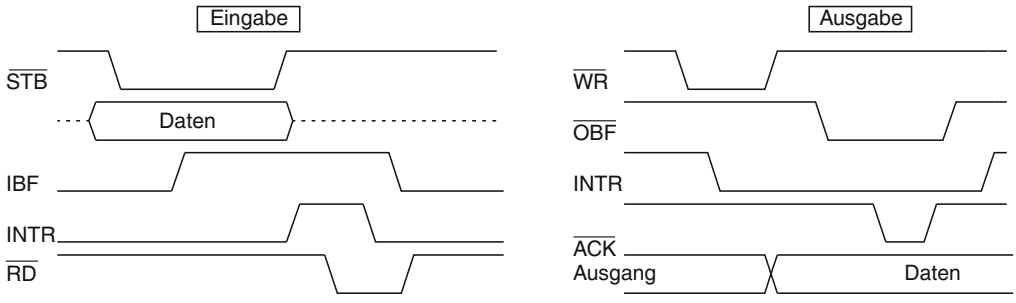


Abb. 52.68 Funktionsprinzip der Betriebsart 1

Tab. 52.20 Zusammenarbeit der Steuersignale und Adressen und ihre Auswirkung auf die Datenrichtung

/CS	/RD	/WR	A1	A0	Funktion
0	0	0	0	0	Kanal A → Datenbus
0	0	0	0	1	Kanal B → Datenbus
0	0	0	1	0	Kanal C → Datenbus
0	1	1	0	0	Datenbus → Kanal A
0	1	1	0	1	Datenbus → Kanal B
0	1	1	1	0	Datenbus → Kanal C
0	1	1	1	1	Datenbus → Steuerlogik
1	0		×	×	Hochohmig
0	×	×	1	1	Ungültige Bedingung
0	1	1	×	×	Hochohmig

ihrer Auswirkung auf die Datenrichtung. Zur Verdeutlichung sind die Adressen der Register zusätzlich extra aufgeführt. Die Initialisierung erfolgt immer so, dass zunächst das entsprechende Kommandowort in den Akkumulator geladen und anschließend der Inhalt des Akkumulators über einen OUT-Befehl mit Angabe der Steuerwortadresse zum Kommandoregister kopiert wird. Danach können die Ports mit ihren Adressen angesprochen werden. Arbeitsregister für Ein- und Ausgabe ist immer der Akkumulator. Das nachfolgende Programm zeigt beispielhaft eine einfache Byteeingabe in der Betriebsart 0 (siehe auch Abschn. 52.6 „Maschinensprache“).

Beispiel

Im Hauptprogramm wird über das C-Register bestimmt, von welchem Port die Daten geholt werden sollen. Die eingelesenen Daten sollen dann zu der Adresse gebracht werden, auf die das HL-Register zeigt. Das nachfolgende Listing zeigt eine mögliche Lösung:

```
; Unterprogramm DATIN für 8255 und 8085 CPU
; Byte-Eingabe über ein Port in Abhängigkeit des C-Registers
; Das gewünschte Port wird im Hauptprogramm
; über das Register C angewählt:
; Ist der Inhalt des Registers C = 00, dann wird Port A gewählt
; Ist der Inhalt des Registers C = 01, dann wird Port B gewählt
; Ist der Inhalt des Registers C = 02, dann wird Port C gewählt
; Die Eingabedaten werden zu der Adresse gebracht,
; auf die das HL-Register zeigt
;
; Programmbeginn ist die Adresse 8100
;
; ************************************************

ORG 8100
ADRPA EQU 00 ;Adresse Port A
ADRPB EQU 01 ;Adresse Port B
ADRPC EQU 02 ;Adresse Port C
ADRKR EQU 03 ;Adresse Kommandoregister
KW EQU 9B ;Kommandowort
;alle Port auf Eingabe
8100 F5 DATIN: PUSH PSW ;(AF) zum Stack
8101 C5 PUSH B ;(BC) zum Stack
8102 3E 9B MVI A,KW ;8255 initialisieren
8104 D3 03 OUT ADRKR ;
8106 79 MOV C,A ;Wahl zum Akkumulator
8107 FE 01 CPI 01 ;(A) -- 1, Flag setzen
8109 FA 15 81 JM PORTA ; -> Port A
810C CA 1A 81 JZ PORTB ; -> Port B
810F DB 02 IN ADRPC ;Port C Eingabe
8111 C3 1C 81 JMP RUECK ;
8114 DB 00 PORTA: IN ADRPA ;Port A Eingabe
8116 C3 1C 81 JMP RUECK ;
8119 DB 01 PORTB: IN ADRPB ;Port B Eingabe
811B 77 RUECK: MOVE M,A ;(A) -> (@HL)
811C C1 POP B: ;Stack -> BC
811D F1 POP PSW ;Stack -> AF
811E C9 RET ;zum Hauptprogramm zurück
```

◄

52.4.5 Zeitgeberbausteine

Zeitgeberbausteine (TIMER) dienen zur Anpassung zeitbedingter Programmabläufe zwischen dem Prozessor und peripheren Geräten. Sie enthalten meist voneinander unabhängige Zähler und können wie ein programmierbarer E/A-Baustein über ein Steuerwortregister für verschiedene Betriebsarten programmiert werden.

Beispiel

Der 8253 besitzt zur Steuerung die Siganaleingänge /WR, /RD und /CS. Als Adressleitungen sind A0 und A1 zugeführt. Der Datenbus D0 bis D7 ist 8 Bit breit. Der Baustein verfügt über drei voneinander unabhängige Zähler mit jeweils drei externen Anschlüssen: Signalausgang OUT, Takteingang CLK und Steuereingang GATE (Abb. 52.69).

Die Adressleitungen A0 und A1 sind häufig direkt mit dem Adressbus des MP-Systems verbunden. Mit ihrer Hilfe erfolgt die Auswahl eines der Zähler und die Adressierung des Steuerwort-Registers zur Betriebsart-Auswahl (MODE). Angewählt wird der Baustein über die Auswahlleitung /CS. Dieser Anschluss wird zur Adressenbildung über einen Adressdecodierer gesteuert (Abb. 52.70). Tab. 52.21 beschreibt die Datenrichtung in Abhängigkeit von Adress- und Steuersignalen. ◄

Die einzelnen Zähler sind voreinstellbare Abwärtszähler für eine 16-Bit-Wortbreite, die mit dem Signal des CLK-Eingangs getaktet werden. Die Taktfrequenz darf max. 2 MHz betragen. Abb. 52.71 zeigt das Prinzip eines der Zähler.

Das Zählregister kann über den Datenbus mit einem Wert geladen werden, der bei Zählbeginn in den eigentlichen Zähler gelangt. Der Zählerzwischenspeicher enthält den jeweiligen Zählerstand und kann softwaremäßig abgefragt werden. Mit Hilfe des Steuerwortregisters kann die Zählart (dual oder dezimal) und die Betriebsart (Betriebsart 0 bis 5) festgelegt werden. Abb. 52.72 gibt das Format des Steuerwortes und die Adressen der Zähler an. Mit Bit 7 und Bit 6 kann einer der Zähler gewählt werden, mit Bit 1, 2 und 3 wird die Betriebsart festgelegt.

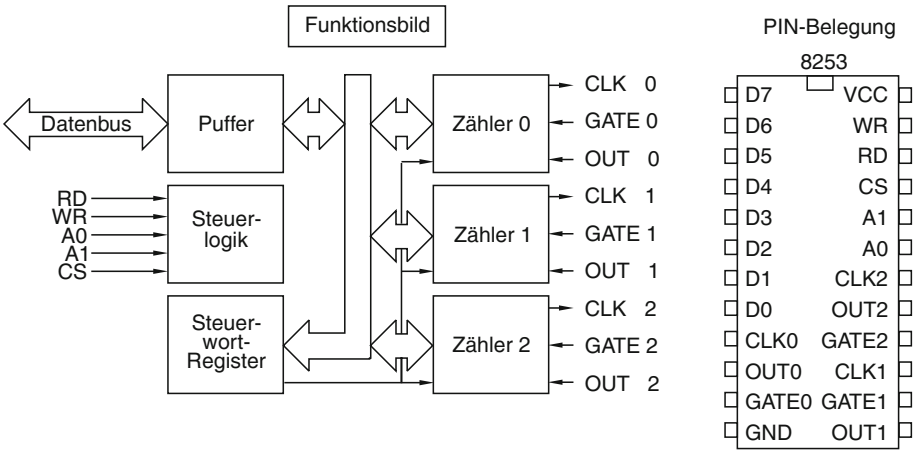


Abb. 52.69 Funktionsbild und Anschlussbelegung des 8253

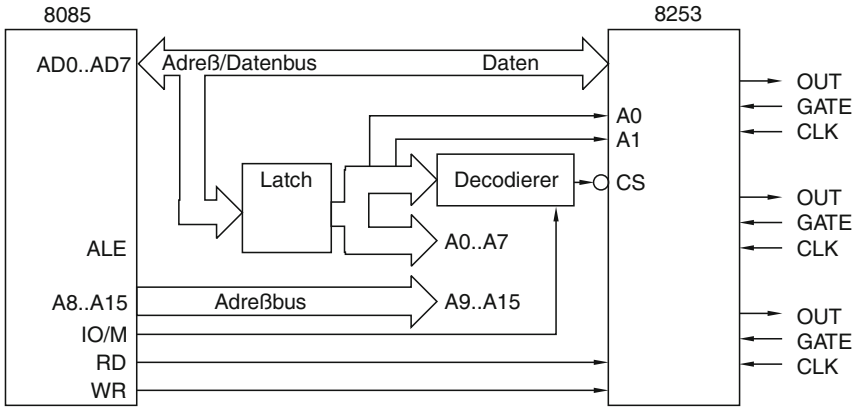


Abb. 52.70 Baustein mit 8085 CPU

Tab. 52.21 Datenrichtung in Abhängigkeit von Adress- und Steuersignalen

/CS	/RD	/WR	A1	A0	Funktion
0	1	0	0	0	Zähler 0 laden
0	1	0	0	1	Zähler 1 laden
0	1	0	1	1	Zähler 2 laden
0	1	0	1	1	Steuerwort laden
0	0	1	0	0	Zähler 0 lesen
0	0	1	0	1	Zähler 1 lesen
0	0	1	1	0	Zähler 2 lesen
0	0	1	1	1	Datenbus hochohmig
1	×	×	×	×	Baustein nicht angesprochen
0	1	1	×	×	Datenbus hochohmig

Bit 0 legt den Zählmodus fest und Bit 4 und 5 haben Einfluss auf das L- und H-Byte beim Lesen oder Beschreiben des Bausteins. Die Betriebsarten unterscheiden sich durch verschiedene Ausgangsimpulse.

In der Betriebsart Mode 0 wird der Zähler über das Zählregister mit einem Anfangswert geladen, während der Ausgang OUT zu Beginn der Zählung „0“ ist. Mit jedem Takt wird der Zähler dekrementiert und gibt bei Erreichen des Wertes Null eine „1“ am Ausgang aus. Man spricht bei dieser Betriebsart auch von Interrupt bei Zählernulldurchgang, weil sich die positive Flanke am OUT-Ausgang zum Auslösen von Unterbrechungsanforderungen gut eignet. Abb. 52.73 zeigt die beteiligten Signale beispielhaft für einen Zähleranfangswert = 4.

Der Ausgangspegel bleibt so lange erhalten, bis der Zähler neu geladen wird. Der Zählvorgang selbst kann am Anschluss GATE mit „0“ unterbrochen werden und mit „1“ fortgeführt werden (Abb. 52.73, 52.74).

In der Betriebsart Mode 1 kann eine monostabile Kippstufe nachgebildet werden. Nach dem Laden des Zählers wird er durch „1“ am Anschluss Gate gestartet. Der Ausgang wird dann „0“, bis der Zählerstand Null erreicht ist.

In der Betriebsart Mode 2 kann ein Zähler als Teiler oder Generator arbeiten. Nach dem Erreichen des Nullwertes beginnt sofort ein erneuter Zählvorgang.

In der Betriebsart Mode 3 arbeitet der TIMER als Rechteckgenerator.

In der Betriebsart Mode 4 spricht man auch vom softwaregesteuerten Signal, weil der Ausgang beim Erreichen des Nullwertes für die Dauer einer Taktperiode von „1“ auf „0“ geht.

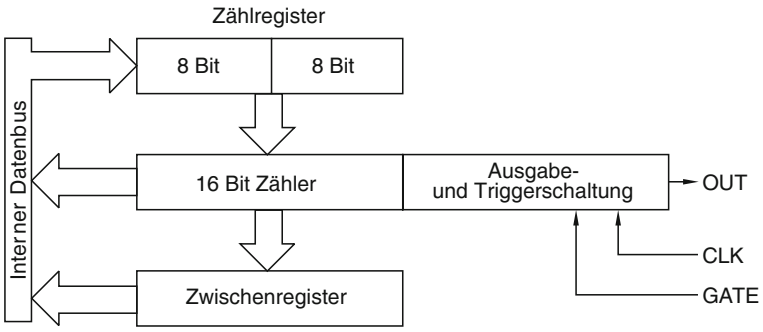


Abb. 52.71 Prinzip eines Zählers

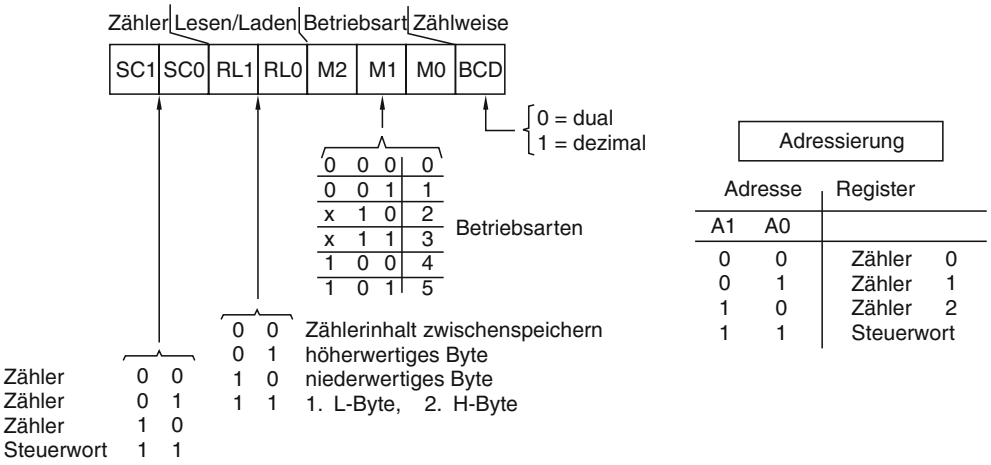


Abb. 52.72 Format des Steuerwortes und Adressen der Register

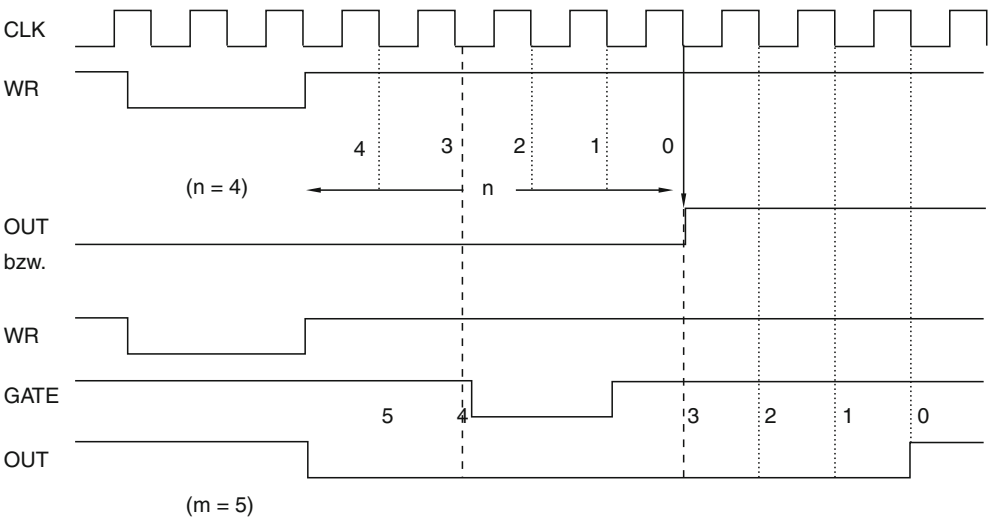


Abb. 52.73 Signalverlauf in der Betriebsart 0

In der Betriebsart Mode 5 spricht man von einem hardwaregesteuerten Signal, weil der Zähler erst mit der ansteigenden Flanke des GATE-Signals startet. Wie in der Betriebsart Mode 4 geht der Ausgang beim Erreichen des Nullwertes für die Dauer einer Taktperiode von „1“ auf „0“.

Eine mögliche Anwendung ist die, den TIMER für Ereigniszählungen einzusetzen. Man versteht darunter, dass der Zähler kontinuierlich bis zum Eintreffen einer Bedingung (eines Ereignisses) zählt. Hierbei ist es notwendig, den Zählerstand abzufragen und in Abhängigkeit von diesem Wert den Prozes-

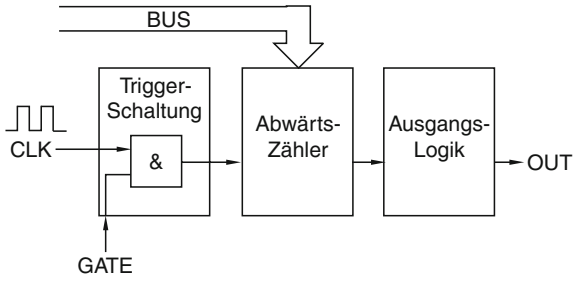


Abb. 52.74 Blockbild eines Zählers

sor Entscheidungen treffen zu lassen. Der Zählerinhalt aller drei Zähler kann ausgelesen werden. Es gibt zwei Möglichkeiten, den Zählerstand zu erfahren. Einmal kann der gewünschte Zähler über A0 und A1 addressiert werden und mit /RD zur CPU gelangen. Wenn RL0 und RL1 auf „1“ gesetzt sind, handelt es sich um einen 16-Bit-Wert. Der Lesebefehl muss dann zweimal erfolgen. Mit dem ersten Befehl wird das LSB (niederwertiges Byte) und mit dem zweiten das MSB (höherwertiges Byte) eingelesen. Während dieser Zeit muss aber das CLK-Signal abgeschaltet sein, beispielsweise durch den GATE-Anschluss, um einen gültigen Zählerinhalt zu bekommen.

Eine zweite Möglichkeit besteht darin, den Zählerinhalt in den Zwischenspeicher zu bringen und zu lesen, ohne den eigentlichen Zählvorgang zu unterbrechen. Hierbei wird der entsprechende Zähler über das Steuerwort mit SC0 und SC1 ausgewählt, und RL0 und RL1 werden auf „0“ gesetzt. Die übrigen Bits sind hierbei ohne Bedeutung. Eine einfache Programmsequenz zeigt dies beispielhaft für den Zähler 1.

Beispiel

Lesen des Zählerzustands

```
MVI A, 40 ;Zähler 1 und Zählerinhalt zwischenspeichern
OUT 03 ;Steuerwort-Register, Wert zwischenspeichern
IN 01 ;(Zähler 1, L-Byte) -> Akkumulator
MOV L,A ;(Akkumulator) -> L-Register "retten"
IN 01 ;(Zähler 1, H-Byte) -> Akkumulator
MOV H,A ;(Akkumulator) -> H-Register "retten"
;Der Zählerstand liegt jetzt im Registerpaar H vor
```

Damit der Zähler seine Tätigkeit aufnimmt, muss er mit einem Zähleranfangswert geladen werden. Hierzu ist es notwendig, das entsprechende Steuerwort ins Steuerwort-Register zu schreiben und anschließend den gewünschten Zählwert einzubringen. Von Bedeutung ist nur die zuvor festgelegte Reihenfolge durch das Steuerwort mit RL0 und RL1. ◄

52.4.6 Programmierbarer E/A-Baustein mit Speicher und Zeitgeber

Neben programmierbaren E/A-Bausteinen werden solche mit zusätzlichen integrierten Baustufen wie TIMER und/oder Speicher angeboten. Beispiele hierfür sind der 8155 und 8755. Der 8755 besitzt ein internes ROM von 2 K × 8 Bit und ein 8-Bit-Port PA und PB, wobei im Gegensatz zum 8255 die einzelnen Bitleitungen als Aus- oder Eingang programmiert werden können.

Beispiel

8155/8156

Der Unterschied zwischen dem 8155 und dem 8156 besteht lediglich im Bausteinauswahlanschluss CE. An diesem Eingang wird der 8155 mit „0“, der 8156 mit „1“ angesteuert. Der Baustein besitzt 2 TTL-kompatible Ports (A und B) mit je 8 Leitungen und ein Port C mit 6 Leitungen. Die Speicherkapazität des RAM beträgt 256 × 8 Bit. Ferner ist ein 14-Bit-Zähler integriert. Der direkte Anschluss an den AD-Bus der 8085 CPU oder eines Mikrocontrollers ist wegen des vorhandenen Adressenregisters (Adress-Latch) möglich (Abb. 52.75). Die Bedeutung der Anschlüsse werden in der Tab. 52.22 erklärt. Alle weiteren technischen Daten zeigt Abb. 52.77.

Für parallele Ein- und Ausgaben besitzt der Baustein drei Ports: Port A und Port B mit 8 und Port C mit 6 Leitungen. Port C kann durch das Kommando-Register so programmiert werden, dass an den Anschlüssen Steuersignale für die parallele Ein- und Ausgabe mit *Quittungsbetrieb* zur Verfügung stehen. In der Betriebsart ALT4 arbeiten dann die beiden anderen Ports im *Handshake-Betrieb*, während drei Leitungen von Port C dem Port A und 3 Leitungen dem Port B zugeordnet sind. Ein Beispiel für den Anschluss peripherer Geräte im Quittungsbetrieb (Handshaking) zeigt Abb. 52.76. Die Steuerleitungen können in der Betriebsartentabelle mit ALT4 verglichen werden. In den Betriebsarten ALT1 und ALT2 arbeiten alle Ports als einfache Ein- und Ausgabeports. Für die Adressierung der Ports, des Kommando-Registers und des Zeitgebers kommt der Baustein mit zwei Adressleitungen nicht mehr aus, daher dienen A0, A1 und A2 zur Adressierung. ◄

Die Programmierung erfolgt über das Kommando-Register. Eine einfache Programmsequenz zeigt den Programmiervorgang zur Ausgabe über das Port A.

Beispiel

```
MVI A,01 ;Portbaustein initialisieren
;PA = Ausgang
OUT 00 ;Adresse des Kommando-Registers
MVI A,FF ;Daten -> Akkumulator
OUT 01 ;Datenausgabe an Port A
;am Port A liegen jetzt alle Ausgänge auf "1"
```

Neben dem Kommando-Register besitzt der 8155 ein Zustandsregister, wo der Zustand des Zeitgebers und der Zustand der Steuerleitungen abgefragt werden können. Die Adresse ist die des Kommando-Registers. Das ist möglich,

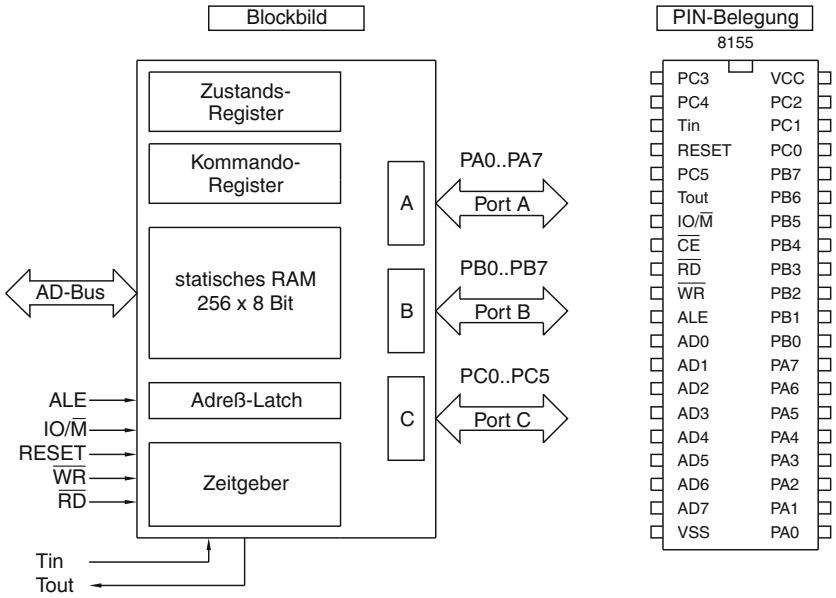
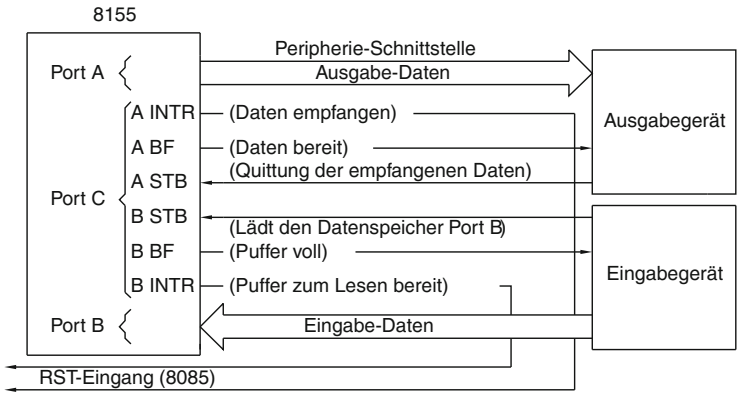


Abb. 52.75 Blockbild und Anschlussbelegung des 8155



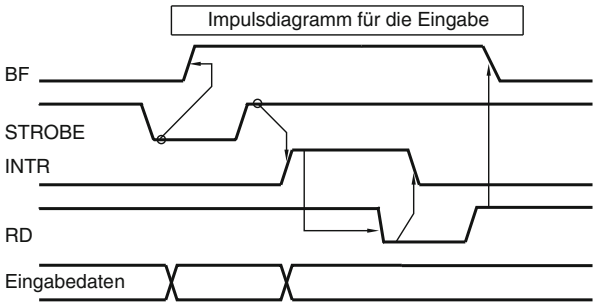


Abb. 52.76 Baustein im Quittungsbetrieb

weil in das Kommando-Register nur eingeschrieben werden kann und das Zustands-Register nur gelesen wird. Der Zeitgeber des Bausteins ist ein 14-Bit-Abwärtszähler, der die Impulse am Eingang Tin zählt und sich nach Ende der Zählung betriebsartenabhängig verhält, d. h. am Ausgang TIMER OUT einen Rechteckimpuls ausgibt, wenn das Ende der Zählung erreicht ist. Die Bedeutung der Bits für den Zeitgeber sind ebenfalls im Bild dargestellt. ◄

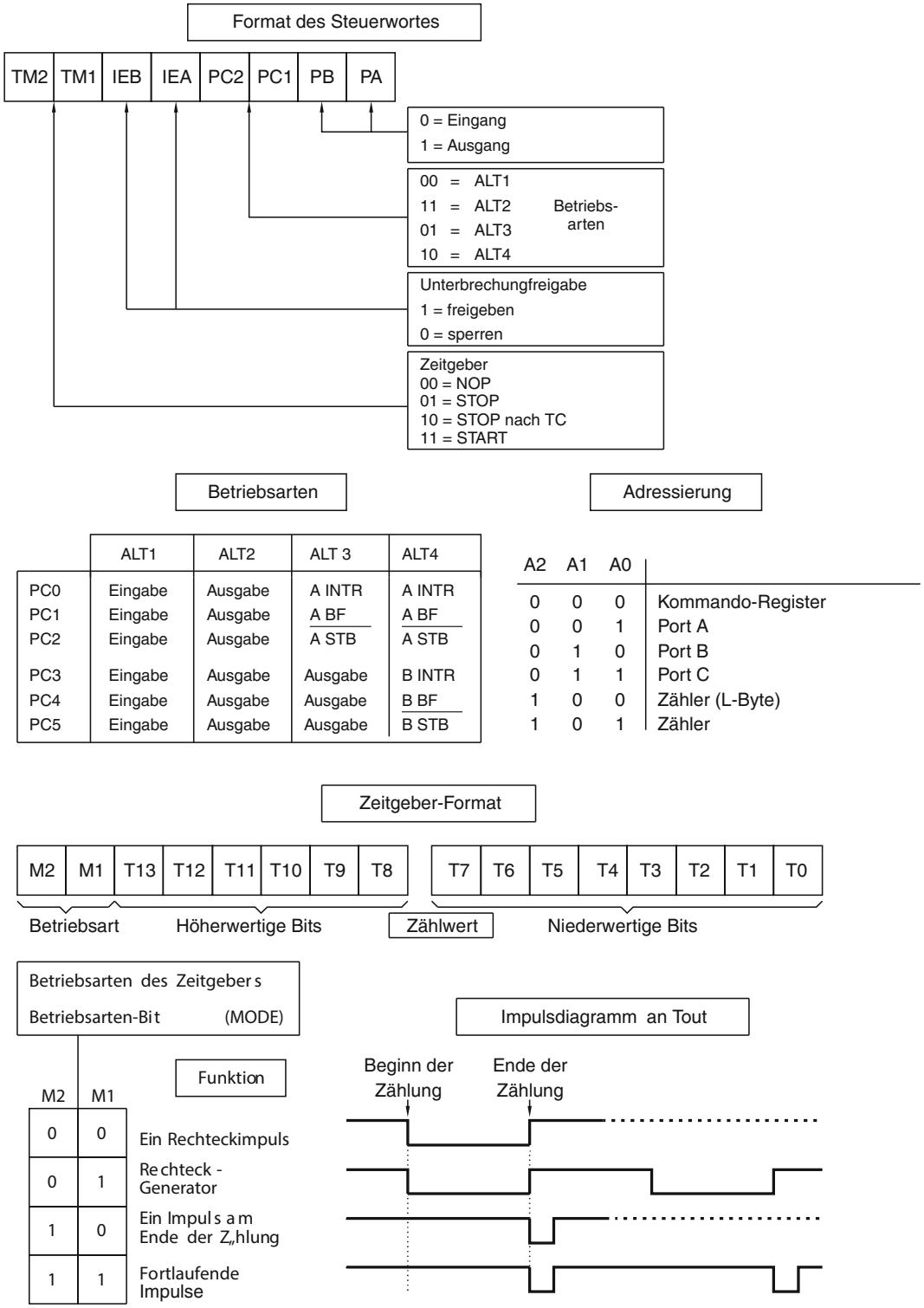


Abb. 52.77 Technische Daten des 8155

Tab. 52.22 Bedeutung der Anschlüsse

PIN	Funktion	Anmerkungen
D0 . . . D7	Datenbus	
AD0, AD7	Adressen-/Datenbus	Die Adresse wird mit der abfallenden Flanke des Signals ALE gespeichert
ALE	Adress Latch Enable	Adressenspeicher-Freigabe
RESET	Rücksetzeingang	Eingang High-aktiv; ein H-Pegel an diesem Eingang setzt alle Register zurück und bringt die Ports in die Betriebsart 0 auf Eingabe
IO, /M	E/A-Speicherwahl	
CE	Baustein-Auswahl	Eingang High-aktiv
/RD	Lese-Eingang	Eingang Low-aktiv
/WR	Schreib-Eingang	Eingang Low-aktiv
PA0 . . . PA7	Port A	Bit 0 . . . Bit 7
PB0 . . . PB7	Port B	Bit 0 . . . Bit 7
PC0 . . . PC5	Port C	Bit 0 . . . Bit 6
TIMER IN	Zeitgebereingang	
TIMER OUT	Zeitgeberausgang	
VCC	Betriebsspannung +5 V	
GND	Masse 0 V	

Beispiel

Der Portbaustein 8755 besitzt als Speicher ein EPROM (elektrisch programmierbarer und UV-löschbarer Speicher) mit einer Speicherkapazität von 2 K × 8 Bit. Zur Bausteinauswahl sind die beiden Anschlüsse /CE1 (low-aktiv) und CE2 (high-aktiv) anzusprechen. Als Ports besitzt der Baustein das Port A und das Port B mit jeweils 8 Leitungen oder Kanälen (Abb. 52.78). Im Gegensatz zu vielen anderen Portbausteinen kann hier jeder Kanal als Eingabe- oder Ausgabekanal arbeiten. Die Funktion wird durch die entsprechende Adressierung vorgenommen. Der Ein- oder Ausgabezustand jeden Kanals bestimmt das Datenrichtungs-Register (DDR). Die Funktion eines Kanals beschreibt Abb. 52.79. ◀

52.4.7 Eingabe-Ausgabe-Bausteine für den seriellen Betrieb

52.4.7.1 Allgemeines

Bausteine für den seriellen Betrieb übernehmen die Daten wortweise vom Prozessor und wandeln sie in eine kontinuierliche Folge serieller Informationen um. Des Weiteren kön-

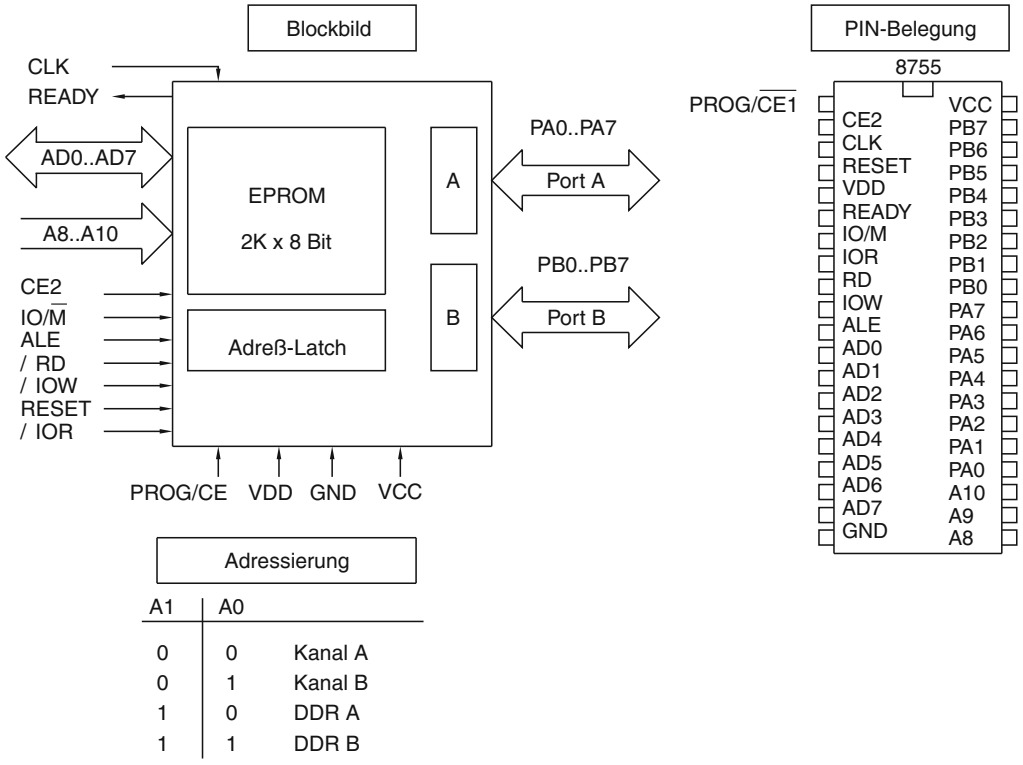


Abb. 52.78 E/A-Baustein

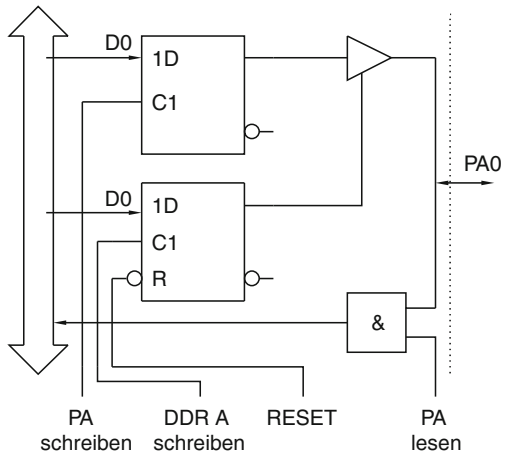


Abb. 52.79 Datenrichtungs-Register mit Datenleitung

nen Daten seriell empfangen werden und für den Prozessor in ein Datenwort parallel umgewandelt werden. Wesentlicher Bestandteil sind Parallel-Serienwandler und Serien-Parallelwandler. Abb. 52.80 zeigt das Prinzip der seriellen Übertragung mit einem E/A-Baustein. Der Aufwand, die Realisierung dieser Wandler mit logischen Schaltungen vorzunehmen, ist unwirtschaftlich. Man bedient sich daher USART's (universell asynchron receiver/transmitter = Universeller Asynchron-Empfänger Sender) oder SIO's (serial input/output = Serieller E/A-Baustein).

52.4.7.2 USART

Beim Senden gelangen die parallel gelieferten Daten vom Prozessor über Zwischenregister (Latch) und Parallel-Serienwandler auf die Sendeausgangsleitung. Beim Empfang müssen die seriellen Daten wieder in parallele Worte rückgeführt werden. Damit der Prozessor hinsichtlich des Zeitpunktes, wann Daten von ihm gesendet bzw. empfangen werden sollen, flexibler ist, wird ein zusätzliches Sende- und Empfangsregister eingebaut. Der Sender fügt noch zusätzliche Bits in den Datenstrom ein, damit der Prozessor sich nicht mit den Besonderheiten der speziellen Übertragungsverfahren wie beispielsweise Start- und Stopbits beschäftigen muss. Ähnlich muss der Empfänger Zeichen, die nicht zum Dateninhalt gehören, entfernen. Hierzu dient eine Steuerung des Taktstromes (Abb. 52.81). Die eigentliche Bausteinfunktion wird durch die Software festgelegt.

Beispiel

Der USART 8251 ist ein universeller Synchron/Asynchron-Sender/Empfänger-Baustein und kann durch den Prozessor für die gebräuchlichen Datenübertragungsverfahren programmiert werden. Abb. 52.82 gibt die Anschlussbelegung wieder und zeigt das Blockbild des Bausteins. Tab. 52.23 beschreibt die Bedeutung der Anschlüsse.

Der Baustein kann im Synchron- oder Asynchronbetrieb eingesetzt werden. Die Betriebsarten unterscheiden sich

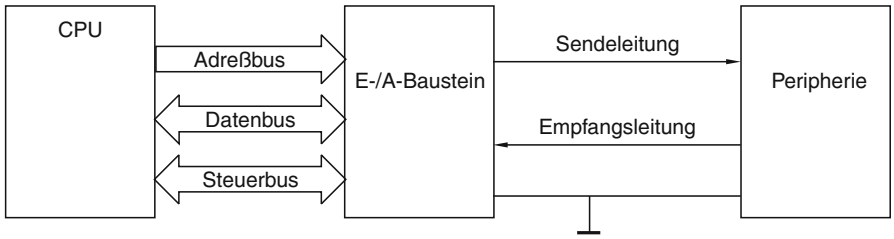


Abb. 52.80 Prinzip der seriellen Übertragung mit E/A-Baustein

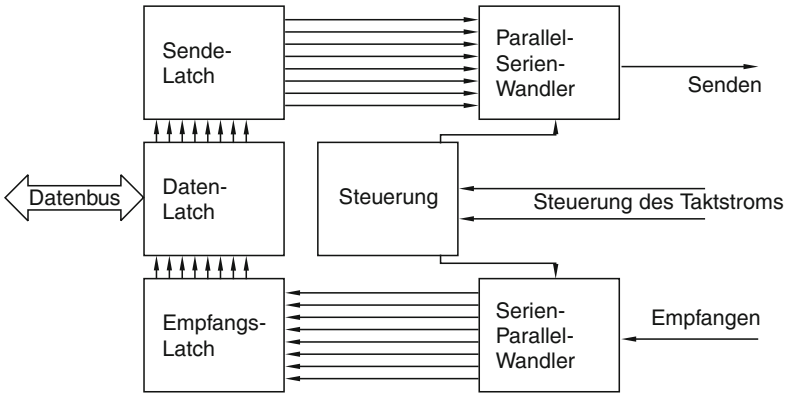


Abb. 52.81 Prinzip eines USART

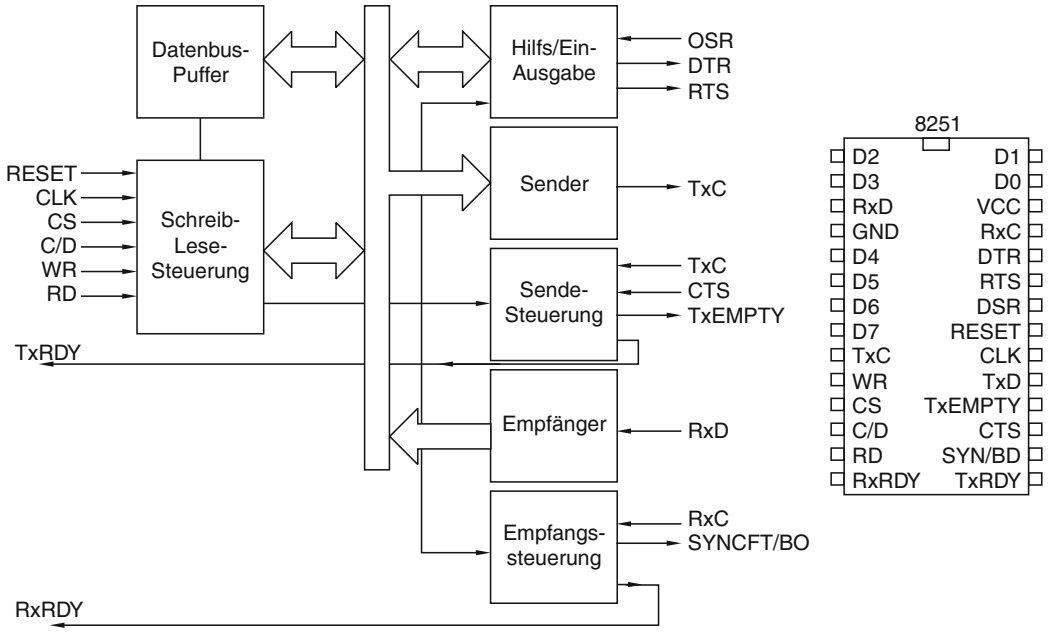


Abb. 52.82 Blockbild des 8251 und Anschlussbelegung

im Aufbau des Übertragungszeichens und der Synchronisation. Beim Synchronbetrieb bestimmt das periphere Gerät die Übertragungsrate. Beim Asynchronbetrieb werden die Daten mit vorgegebener Taktfrequenz übertragen. Ein interner Frequenzteiler kann mit dem Mode-Wort auf die Teilverhältnisse 1, 16 oder 24 eingestellt werden. Mit zugeführtem Sendetakt und programmiertem Teilerverhältnis wird die Übertragungsgeschwindigkeit (Baudrate 1 Byted = 1 Bit/s) bestimmt. Der Baustein lässt Übertragungsgeschwindigkeiten bis zu 64 kBd zu. Jede Übertragung beginnt mit einem Start-Bit und endet je nach Programmierung mit dem Paritybit und mit bis zu 2 Stop-Bits. Die Wortübertragung beginnt stets mit D0. Die Länge der zu übertragenden Worte ist zwischen 5 und 8 Bit programmmierbar. Der Sendevorgang über die Leitung TxD wird durch den Sendetakt TxC ausgelöst. Steht kein Sendezeichen an, hat der TxD-Ausgang H-Pegel. Der Baustein 8251 empfängt die seriellen Daten über die Leitung RxD. RxD liegt normalerweise auf H-Pegel. Bei fallender Flanke an diesem Eingang und Startbits erfolgt der Empfangsvorgang. Die Daten gelangen zum Datenbus-Puffer und können vom Prozessor abgeholt werden. Der Datenempfang wird durch den Anschluss RxRDY mit einem H-Pegel gemeldet.

Zum Handshaking ist bei Empfangssteuerung der Ausgang RxRDY zuständig. SYNDET/BD hat unterschiedliche Funktionen, je nach programmierter Betriebsart. Die Anschlüsse DSR, DTR und RTS dienen zur Modemsteuerung. ◄

Treten Fehler bei der Übertragung auf, wird dies im Statusregister (Tab. 52.24) gekennzeichnet.

Tab. 52.23 Bedeutung der Anschlüsse

PIN	Funktion	
D0 ... D7	Datenbus	
RESET	Rücksetzen	
CLK	Systemtakt	Clock
CS	Bausteinauswahl	Chip Select
C/D	Kennzeichnung der Daten als Steuer- bzw. Statuswort/Zeichen	Control/Data
RD	Daten oder Zustand lesen	Read
WR	Daten oder Steuerinformation schreiben	Write
DSR	Betriebsbereitschaft	Data Set Ready
DTR	DE-Einrichtung betriebsbereit	Data Terminal Ready
RTS	Sender einschalten	Request to Send
TxD	Serielle Sendedaten	Transmit Data
TxC	Sendetakt	Transmitter Clock
CTS	Senderfreigabe	Clear to Send
TxEMPTY	Senderegister leer	Transmitter Empty
TxRDY	Sender kann Daten annehmen	Transmitter Ready
RxD	Serielle Empfangsdaten	Receive Data
RxC	Empfangstakt	Receiver Clock
RxREADY	Empfänger kann Daten liefern	Receiver Ready
VCC	Betriebsspannung 5 V	
GND	Masse 0 V	

Tab. 52.24 Übertragungsfehler

Parity-Error	Paritätsfehler
Overrun-Error	Überlauffehler
Framing-Error	Sperrschrittfehler
Break-Detection	Break-Erkennung

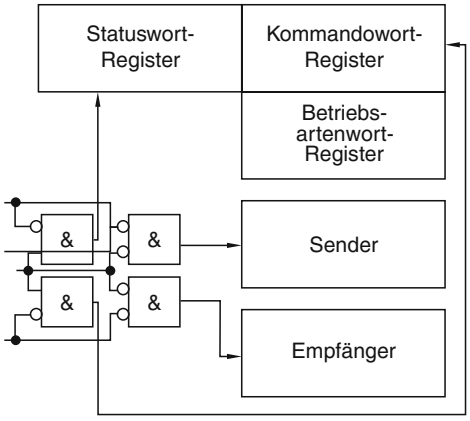


Abb. 52.83 Prinzip der Adressierung

Die Programmierung erfolgt wie bei den anderen E-/A-Bausteinen durch Laden des Kommandoregisters vom Prozessor über den Datenbus. Mit dem Steuerwort für MODE werden die allgemeinen Betriebseigenschaften wie Synchron- oder Asynchronbetrieb, Zeichenlänge und Paritätsart festgelegt. Mit dem Kommandowort wird die Sende- und Empfangssteuerung beeinflusst (Abb. 52.83). Die Reihenfolge „Betriebsart festlegen“ und anschließend „Kommandowort eingeben“ muss eingehalten werden, da beide Codes mit der gleichen Adresse Anschluss C/D = „1“ übergeben werden. Das Statusregister kann bei C/D = „1“ gelesen werden (Tab. 52.25). Dadurch kann der Prozessor den Zustand des Bausteins zu jeder Zeit abfragen (Abb. 52.84 und 52.85).

Tab. 52.25 Register-Adressierung

C/D	RD	WR	
1	0	1	Statuswort-Register
1	1	0	Kommandowort-Register

52.4.8 Bausteine mit Sonderfunktionen

Einige Mikroprozessorsysteme besitzen außer den bekannten Speicher- und Ein- und Ausgabeeinheiten spezielle Peripheriebausteine, die jeweils eine besondere Aufgabe übernehmen. Als Beispiel wurde ein 80286-System gewählt (Abb. 52.86). Außer dem bekannten Adress-Latch und dem Transceiver, einem Zweirichtungstreiber und dem Prozessor gehören als weitere Bausteine ein *Taktgenerator*, ein *Bus-Controller* und ein *Interrupt-Controller* zum System.

Der Taktgenerator 82284 (Abb. 52.87) besteht aus einem Oszillator (Osz. 9), der für den Systemtakt verantwortlich ist, einer Synchronisationseinheit, einer RESET-Schaltung und dem PCLK-Generator. Der Taktgenerator selbst kann durch einen externen Generator gesteuert werden, oder der interne Oszillator wird nur noch mit einem Quarz versehen (Anschluss X1 und X2), der dann die Frequenz bestimmt. Der Anschluss F/C (Frequency/Crystal select) dient zur Auswahl der Takterzeugung. Das PCLK-Signal (Peripheral Clock), das über ein Flipflop (PCLK-Generator) vom CLK-Signal erzeugt wird, ist für die Synchronisation peripherer Geräte verantwortlich. Die Erzeugung des /READY-Signals ist von der entsprechenden Steuerung abhängig (/ARDYEN, /ARDY, /SRDYEN, /SRDY). Das Signal dient zur Steuerung langsamerer Peripheriegeräte und

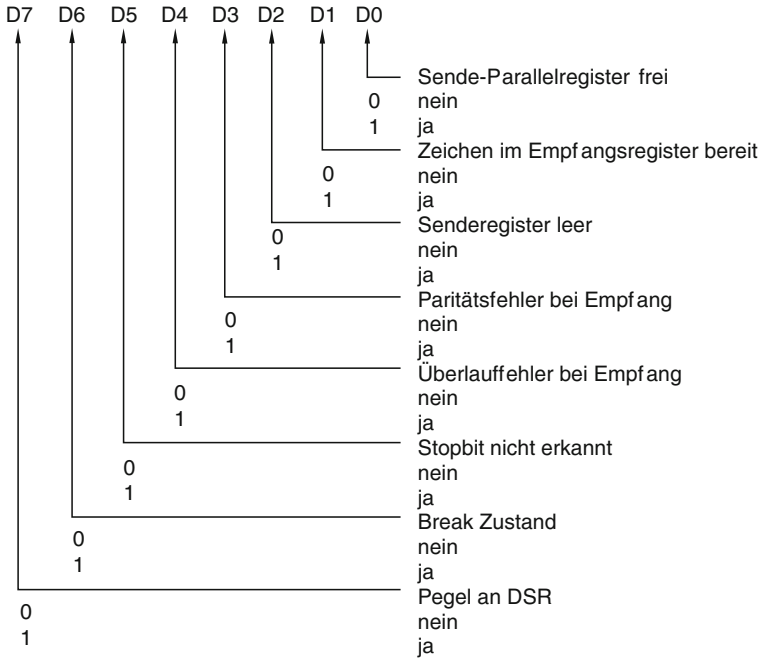


Abb. 52.84 Statuswort

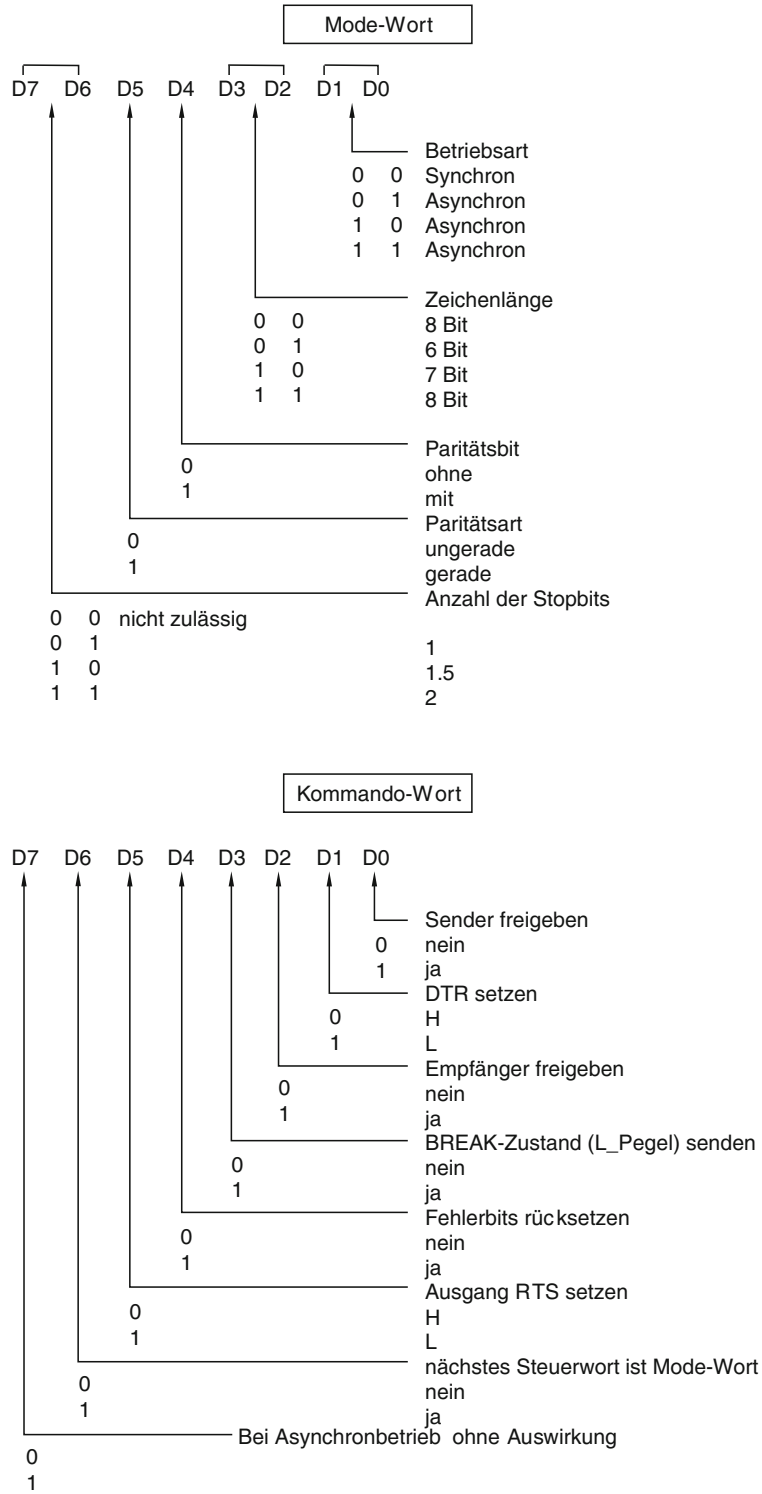


Abb. 52.85 MODE- und Kommandowort

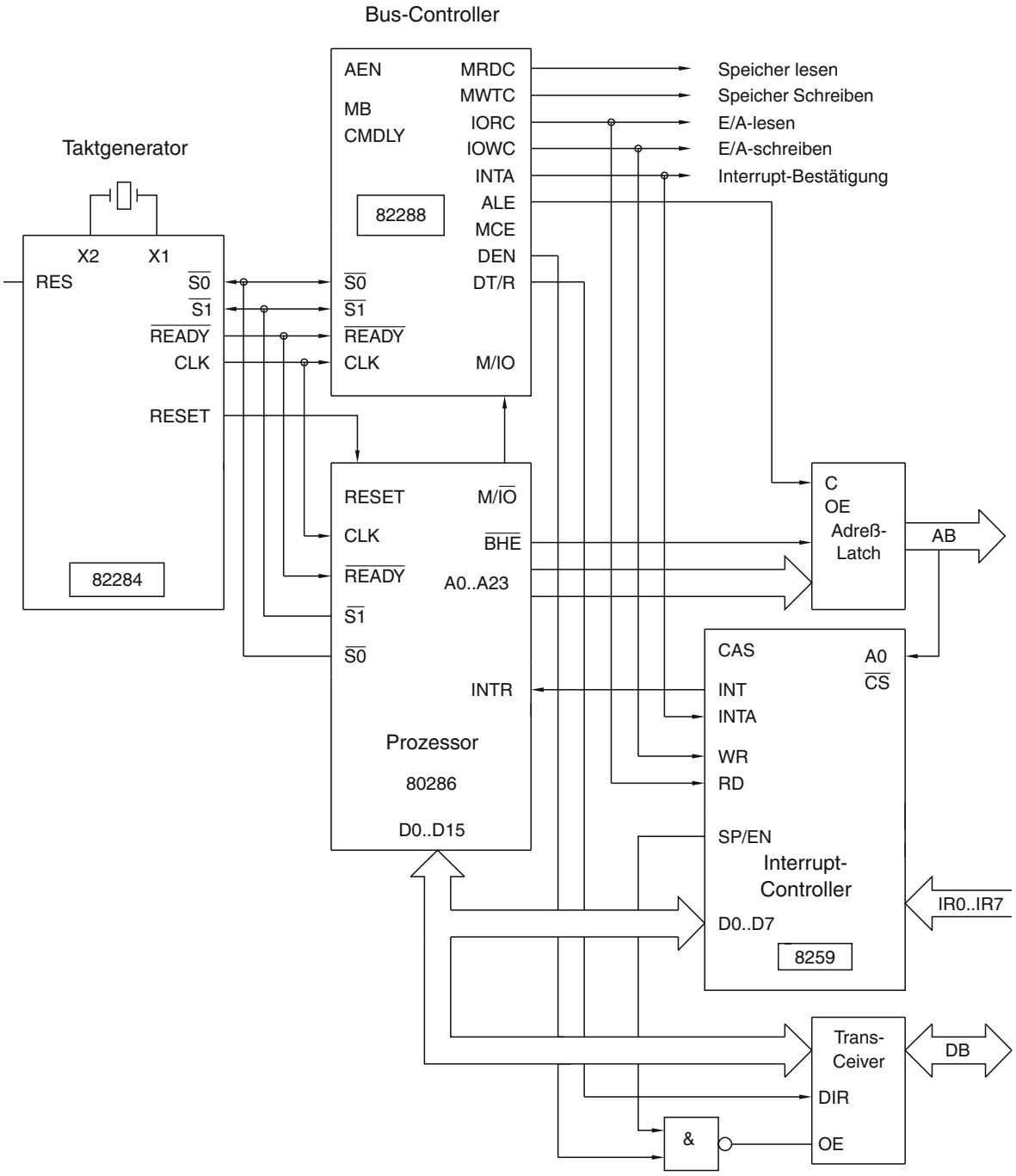


Abb. 52.86 Mikroprozessorsystem mit 80286 CPU

fügt Wartezustände (Wait States) in einen Schreib- oder Lesezyklus ein. Am /RES (Reset in) kann das Reset-Signal erzeugt werden.

Der Buscontroller 82288 (Abb. 52.88) decodiert die Statussignale (S0, S1 und M/IO) des Prozessors und erzeugt mit dem CLK-Signal die Kommando- und Kontrollsignale (Tab. 52.26).

Interrupt-Controller sollen bei einer Unterbrechungsanforderung eines externen Gerätes am Prozessor einen Interrupt aus-

Tab. 52.26 Kommando- und Kontrollsignale

Kommandosignale	Kontrollsignale
INTA	DT/R
IORC	DEN
IOWC	ALE
MRDC	MCE
MWTC	

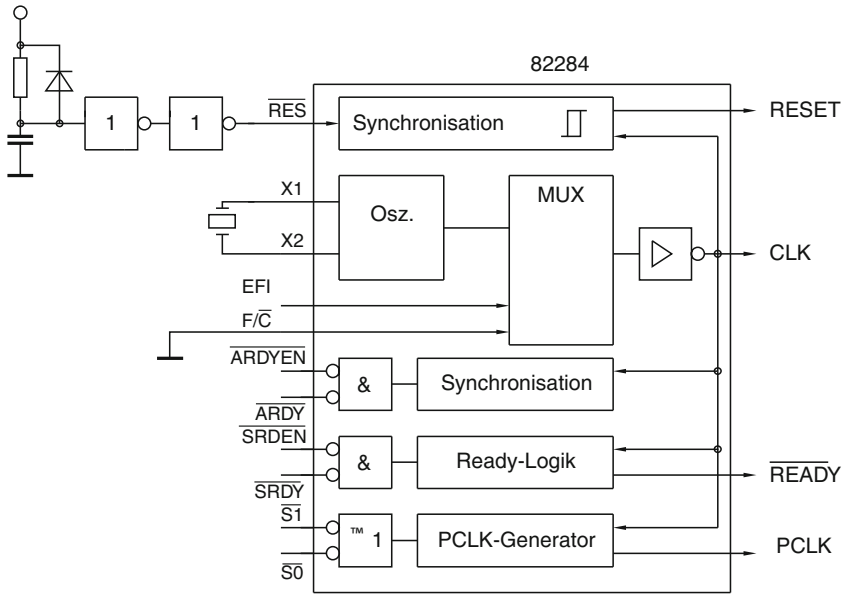


Abb. 52.87 Taktgenerator 82284

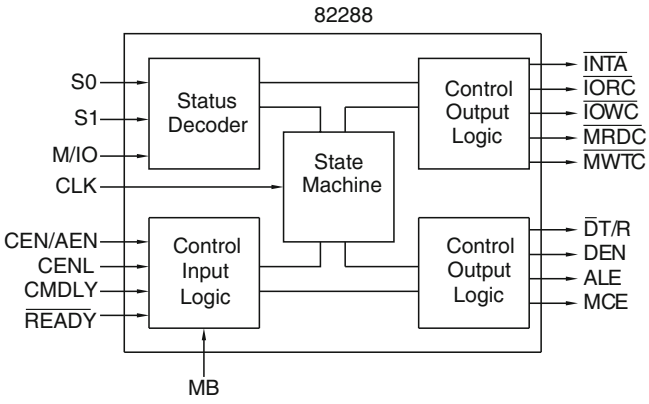


Abb. 52.88 Bus-Controller

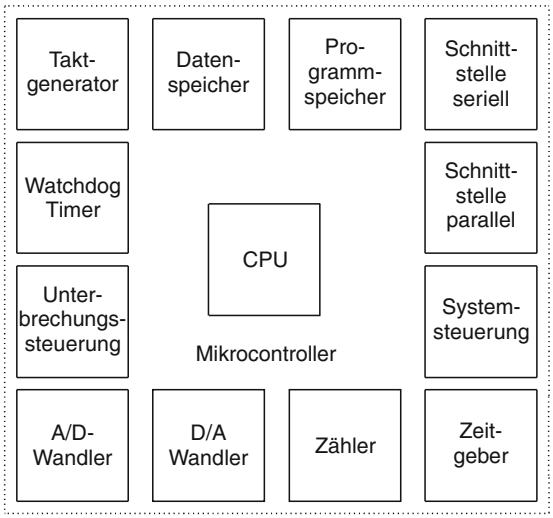


Abb. 52.89 Komponenten eines Mikrocontrollers

lösen. Der Baustein 8259 im Abb. 52.86 besitzt 8 Eingänge für Interruptquellen (IR0 . . . IR7). Der Prioritätsdecoder hat die Aufgabe, bei gleichzeitiger Anforderung mehrerer Quellen eine Reihenfolge der Bearbeitungssequenzen festzulegen.

52.5 Mikrocontroller

52.5.1 Allgemeines

Mikrocontroller (MC) sind Bauelemente, bei denen auf einem Chip alle erforderlichen Bausteine eines Mikroprozessorsystems integriert sind. Hierzu gehören je nach Type der eigentliche Rechenkern mit Multiplikations- und Divisionseinheit und Taktoszillator, die Systemsteuerung und die Interrupteinheit. Die Kommunikation mit externen Baugruppen wird über parallele und serielle Schnittstellen ermöglicht. Für Mess-, Steuerungs- und Regelzwecke sind *Analog-Digital-Wandler* (AD-Wandler), *Digital-Analog-Wandler* (DA-Wandler) und *Zähler* integriert. Ein *Watchdog-Timer* („Wachhund“) kontrolliert den gesamten Arbeitsablauf. MC's besitzen eigene Programmspeicher, die als ROM oder EPROM ausgeführt sein können und als Arbeitsspeicher ein internes RAM (siehe auch Abschn. 52.3 „Halbleiterspeicher“). Weiter besteht die Möglichkeit, das MC-System mit externen Speichern zu erweitern. Abb. 52.89 zeigt in einer Übersicht Komponenten eines Mikrocontrollers. Je nach Verwendungszweck sind bei den unterschiedlichen Typen nicht alle Funktionsstufen enthalten, dafür sind andere Stufen mehrfach vorhanden. Um die Gehäusegröße und damit PIN-Zahl in Abhängigkeit der Funktionsaufgaben möglichst gering zu halten, sind viele Anschlüsse für Doppelfunktionen vorgesehen.

Tab. 52.27 Mikrocontroller

Type	ROM (kB)	RAM (B)	E/A 8-Bit-Port	ADC/E 8-Bit	Clock (MHz)	TIMER
8031	–	128	4	–	12	2
8051	4	128	4	–	12, 16	2
8052	8	256	4	–	12, 16, 20	3
80C52	8	256	4	–	12	3
80512	4	128	6 + 1E	8	12	2
80513	12	256	4	–	12, 16	3
80515	12	256	6 + 1E	8	12, 16	3
80C517	12	256	7 + 1,5E	12	12, 16	4

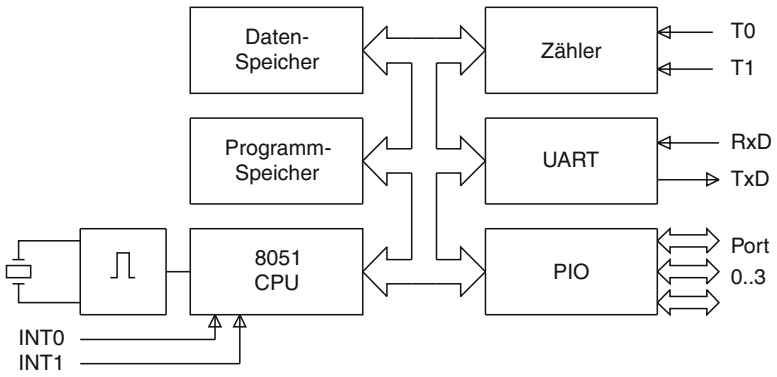


Abb. 52.90 Vereinfachtes Blockbild

Aus der Vielzahl der Einsatzgebiete sind die Schwerpunkte Datenverarbeitung, Industrieelektronik, Mess-, Steuerungs- und Regelungstechnik, Unterhaltungselektronik und Kraftfahrzeugtechnik zu nennen. Weitverbreitet sind 4-Bit- und 8-Bit-Versionen. Mikrocontroller verdrängen den klassischen Prozessor aus einigen Anwendungsbereichen, weil „Ein-Chip-Systeme“ dort wirtschaftlicher zum Einsatz kommen. Einige Anwendungen, wie beispielsweise in der Automobilindustrie, werden erst durch MC's ermöglicht. Bei größeren Stückzahlen sind MC's mit ihrer integrierten Peripherie kostengünstiger als Prozessoren mit externen Peripheriebausteinen. Der Platzbedarf verringert sich erheblich, und die Betriebssicherheit steigt auf Grund der kleineren Anzahl von Fehlerquellen. Die Leistungsfähigkeit der MC's liegt einmal in der relativ hohen Verarbeitungsgeschwindigkeit, da die meisten Befehle mit nur einem Befehlszyklus (z. B. 1 ms) auskommen, der flexiblen Adressierungsarten und den zahlreichen Bitverarbeitungsmöglichkeiten. Beispielsweise können logische Befehle mit einzelnen Bits durchgeführt oder einzelne Portbits gesetzt oder rückgesetzt werden, ohne andere Bits zu beeinflussen (siehe auch Abschn. 52.6 „Maschinensprache“).

Werden in einem System mehrere Mikrocontroller eingesetzt, spricht man von einem *Multi-Controller-System*, bei dem aus Verwaltungsgründen MC's als *Master* oder *Slave* arbeiten. Der Master übernimmt hierbei die Organisation der Datenübertragung. Der entsprechende Einsatz wird den Bausteinen durch Software mitgeteilt. Die Kommunikation untereinander erfolgt über eine serielle Schnittstelle.

Tab. 52.27 führt beispielhaft Mikrocontroller der 80x1x-Familie auf, die sich weltweit als Standard durchgesetzt hat. Bei den Mikrocontrollern dieser Serie werden zwei Versionen unterschieden. Man spricht von *ROM-Versionen*, wenn der Anwender über das interne ROM verfügen kann und von *ROM-losen Versionen*, wenn das interne ROM abgeschaltet ist. Der bis dahin interne Bereich ist dann als externer Bereich ansprechbar. Weitere Unterschiede liegen in der Anzahl der E/A-Ports und der Anzahl der Zeitgeber.

52.5.2 8-Bit-Mikrocontroller

52.5.2.1 Funktionsbeschreibung des MC 8051

Der MC 8051 besitzt einen Daten- und Programmspeicher. Eine Kommunikation mit peripheren Geräten ist über die serielle Schnittstelle (UART) und über parallele Schnittstellen (Port 0 bis 3) möglich. Um auf externe Geräte zu reagieren, sind Interrupteingänge vorhanden. Der TIMER ist als Zähler oder Zeitgeber programmierbar. Hauptbestandteil ist die 8051 CPU, der eigentliche Prozessor (Abb. 52.90). Bei dem Mikrocontroller 8051 können eine Reihe von Anschlussleitungen für unterschiedliche Verwendungen genutzt werden. Erkennbar sind die Doppelbelegungen anhand der PIN-Zuordnung, wie in Abb. 52.91 dargestellt. In der Tab. 52.28 ist die Bedeutung der jeweiligen Anschlüsse aufgeführt.

Das Funktionsbild des Bausteins gibt Abb. 52.92 wieder. Die CPU besteht aus dem Befehlszähler (PC), dem Befehlsdecoder der ALU (Rechen- und Logikeinheit) und der Steuerlogik. Der Progammzähler steuert den Programmablauf, und der Befehlsdecoder sorgt für die Befehlsausführung (siehe auch

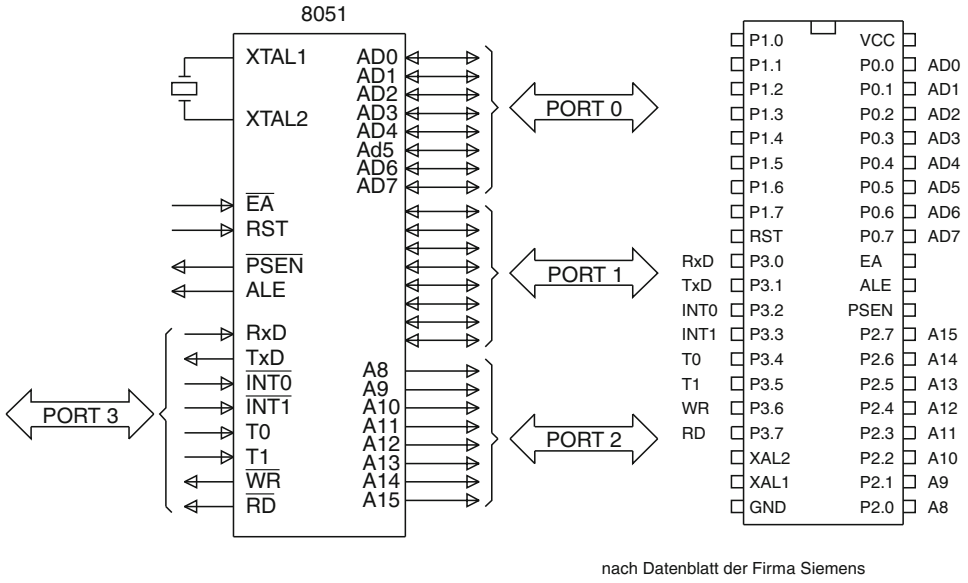


Abb. 52.91 Anschlussbelegung beim MC 8051

Tab. 52.28 Bedeutung der Anschlüsse 8051

Bezeichnung	Funktion
AD0 . . . AD7	Adress-Datenbus
RST	Rücksetzen bei H-Pegel
VPD	Bei Anschluss Ruhestrom für internes RAM
RxD	Serielle Schnittstelle
TxD	Serielle Schnittstelle
/INT0	(Interrupt)/Unterbrechungseingang
/INT1	(Interrupt)/Unterbrechungseingang
T0	(Timer 0)/Zähler/Zeitgeber 0
T1	(Timer 1)/Zähler/Zeitgeber 1
/WR	(Write)/Schreiben
/RD	(Read)/Lesen
XAL2	Quarzanschluss, Externe Taktmöglichkeit
XAL1	Quarzanschluss
/E/A	External Access/Externer Programmspeicherzugriff (bei L-Pegel); bei H-Pegel: 4 K intern, ab 4 K extern
/PSEN	Program Store Enable/Externe Programmspeicherfreigabe
ALE	Adress Latch Enable/Signal zur Adressenzwischenspeicherung
PX.0 . . . PX.7	Port (0, 1, 2, 3) digitale E/A, 8 Bit

Abschn. 52.2 „Mikroprozessoren“). Die ALU des 8051 kann folgende Operationen durchführen: Addition, Subtraktion, Multiplikation, Division, Vergleichen, Rotieren. Außerdem können die logischen Operationen UND, ODER, Exclusiv ODER und Bitmanipulationen durchgeführt werden. Als Arbeitsregister dienen Akkumulator, die Register R0 bis R7 mit ihren vier Bänken (siehe auch Abb. 52.102) und das B-Register. Die Bänke können mit dem Alternativ-Registersatz der Z80 CPU verglichen werden und die Register R0 bis R1 mit den von den 8-Bit-Prozessoren bekannten Registern B, C, D, E, H und L. Im Gegensatz zu den genannten Registern wird beim 8051 eine Adresse angesprochen, und die jeweiligen Registersätze werden durch Umschaltung aktiviert. Wie bei der Standard-CPU sind Stapelzeiger (Stack-Pointer), der hier bei einem RESET auf Adresse 07 gesetzt wird, der Programmzähler (PC = Program Counter) und das vom Mikroprozesser bekannte Kennzeichen oder Flagregister (hier: PSW = Programm Status Word) vorhanden. Die von einem herkömmlichen 8-Bit-Prozessor abweichenden zusätzlichen Baustufen sind im Blockbild Abb. 52.92 besonders gekennzeichnet. Der Datenzeiger (DPTR = Data Pointer) zeigt auf die als nächstes zu bearbeitende Adresse des Datenspeichers.

Zur Steuerung interner Funktionen dienen Spezial-Funktions-Register (SFR = Special-Function-Register). Im Funktionsbild sind sie mit dem Block SFR und Zähler gekennzeichnet. Die genaue Aufteilung dieses Blockes zeigt Abb. 52.93. Tab. 52.29 erläutert die entsprechenden Kürzel der Registernamen.

52.5.2.2 Ein-/Ausgabeeinheit

Die 4 E/A-Ports mit einer 8-Bit-Breite arbeiten bidirektional (in beiden Richtungen) und können jeweils einzeln unabhängig voneinander als Eingabe- oder Ausgabekanal arbeiten. Abb. 52.94 verdeutlicht dies an Hand der Portstruktur eines Kanals. Die Ausgabeinformation („0“ oder „1“) wird mit dem Schreibsignal dem D-Flipflop übergeben und steht somit dem Portausgang zur Verfügung. Der Zustand des D-Flipflops kann über Lesen des D-Flipflops (Latch lesen) jederzeit abgefragt werden. Soll eine Information eingelesen werden, wird das D-Flipflop gesperrt. Mit dem Lesesignal (PIN lesen) gelangt das anliegende Bit zum Datenbus. Bei einem RESET wird der Portanschluss auf Eingang geschaltet.

Port 0, 2 und 3 besitzen Doppelfunktionen. Beispielsweise wird das Port 0 zur Adressierung externer Speicher mit verwendet. Port 0 arbeitet dann als Adress-/Datenbus. Abb. 52.95 zeigt den Anschluss eines externen Speichers an den Baustein.

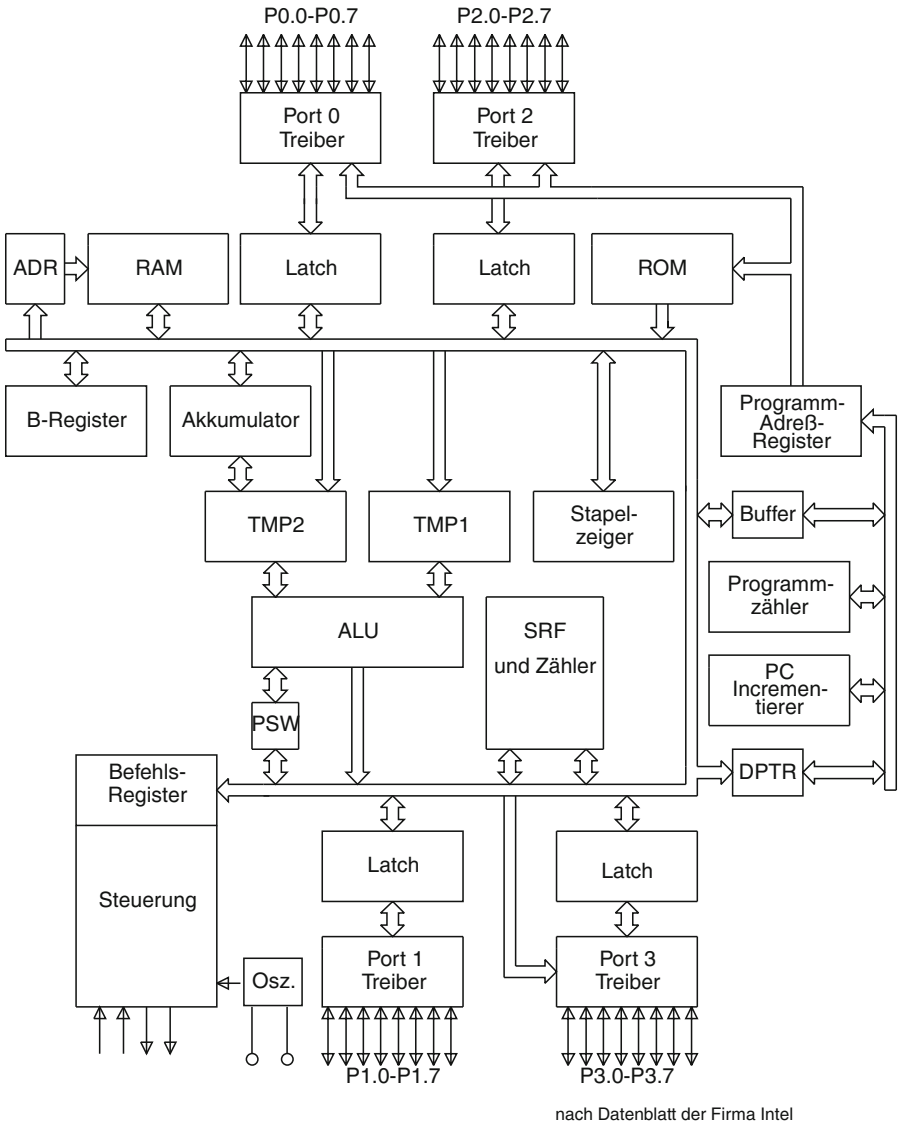


Abb. 52.92 Funktionsbild

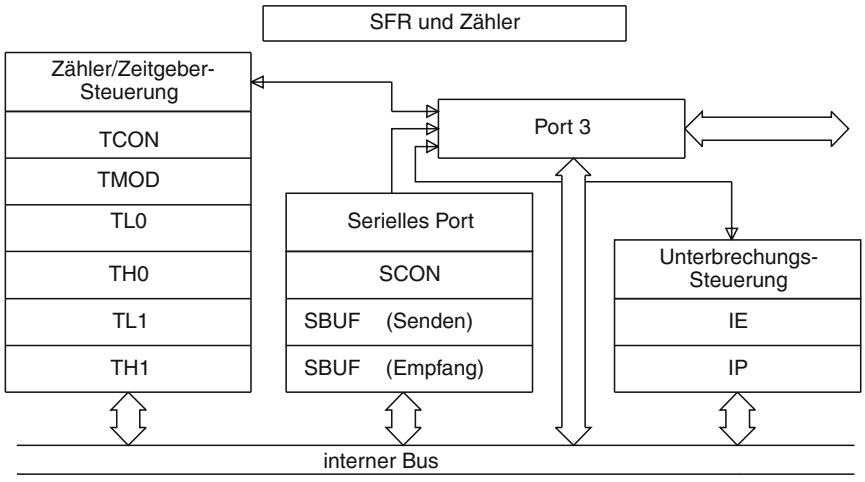


Abb. 52.93 Special-Function-Register und Zähler

Tab. 52.29 Special-Function-Register

Bezeichnung	Funktion	
P0	Port 0	
SP	(Stack-Pointer)/Stapelzeiger	
DPL	(Data-Pointer)/Datenzeiger	L-Byte
DPH	(Data-Pointer)/Datenzeiger	H-Byte
TCON	(Timer/Counter) Zeitgeber/Zähler	Control
TMOD	(Timer/Counter) Zeitgeber/Zähler	Betriebsart
TL0	(Timer/Counter) Zeitgeber/Zähler 0	L-Byte
TL1	(Timer/Counter) Zeitgeber/Zähler 1	L-Byte
TH0	(Timer/Counter) Zeitgeber/Zähler 0	H-Byte
TH1	(Timer/Counter) Zeitgeber/Zähler 1	H-Byte
P1	Port 1	
SCON	(Serial Interface Control) Status/Kontroll-Register	
SBUF	(Serial Data Buffer)/Serielles Pufferregister	
P2	Port 2	
IE	(Interrupt Enable)/Interruptmaske/-freigabe	
P3	Port 3	
IP	(Interrupt Priority)/Interruptpriorität	
PSW	Program Status Word	
ACC	Accumulator B	B-Register

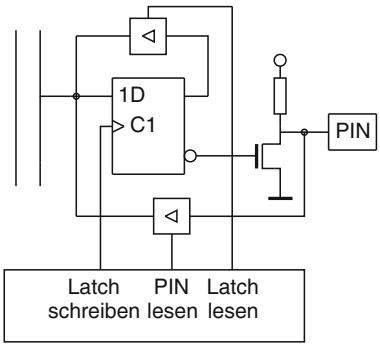


Abb. 52.94 Portstruktur eines Kanals

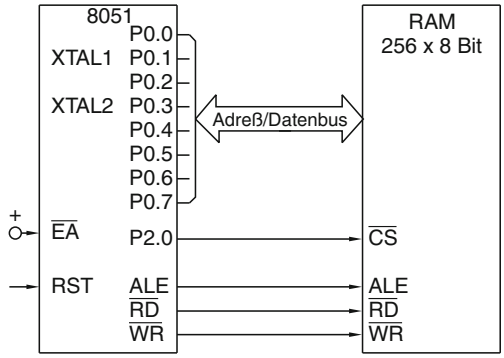


Abb. 52.95 Anschluss an einen externen Speicher

Da der dargestellte Speicher einen internen Adressenzwischenspeicher besitzt (Eingang ALE), ist kein zusätzlicher Zwischenspeicher erforderlich. Besitzt der externe Speicher eine größere Kapazität als 0,25 k, wird Port 2 ebenfalls in die Adressierung mit einbezogen. Hier wird die Bausteinauswahl mit P2,0 durchgeführt. Port 3 kann neben Ein- und Ausgabe von Daten die im Anschlussplan angegebenen Funktionen durchführen.

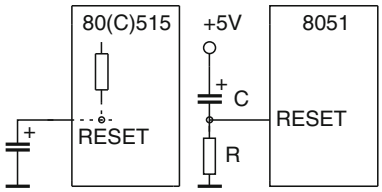


Abb. 52.96 RESET-Schaltungen

52.5.2.3 RESET-Schaltung

Damit die CPU nach dem Einschalten in einen definierten Anfangszustand gebracht wird, ist eine RESET-Schaltung notwendig (Abb. 52.96). Die CPU arbeitet dann immer ab Adresse 0000 des Speichers. Der Baustein wird durch einen „H"-Pegel am Anschluss RST zurückgesetzt. Beim Einschalten der Betriebsspannung muss der Anschluss mindestens für die Zeit von 1 ms H-Pegel besitzen. Bei der dargestellten Schaltung wirkt der Kondensator bei Anlegen der Versorgungsspannung wie ein Kurzschluss und legt zunächst „H"-Pegel an den RESET-Eingang. Nach Aufladung des Kondensators liegt „L"-Pegel an. Der MC 80(C)515 benötigt nur einen Kondensator am RESET-Anschluss, weil ein interner Widerstand vorhanden ist.

52.5.2.4 Taktgenerator

Der interne Taktgenerator des MC 8051 braucht nur noch mit dem Quarz ergänzt zu werden, um seine Funktion ausführen zu können. Die Quarzfrequenz liegt in einem Bereich bis zu 12 MHz. Der Baustein kann auch über einen externen Taktgenerator gesteuert werden. Es wird dann XTAL1 auf Masse gelegt und XTAL2 über einen Treiber angeschlossen (Abb. 52.97). Der Taktgenerator erzeugt alle benötigten Taktsignale des Systems.

52.5.2.5 Stromaufnahme

Um die Stromaufnahme zu reduzieren, besitzt die CMOS-Version, beispielsweise der MC 80C51, zwei Betriebsarten. Im *Power-down-Mode* wird der interne Taktgenerator abgeschaltet und der MC damit in einen inaktiven Zustand gebracht. Die im internen RAM gespeicherten Daten bleiben erhalten. Mit einem Hardware-Reset kann diese Betriebsart beendet werden. Im *Idle-Mode* wird der Taktgenerator nur von der CPU getrennt und versorgt alle anderen Komponenten wie Zähler und serielle

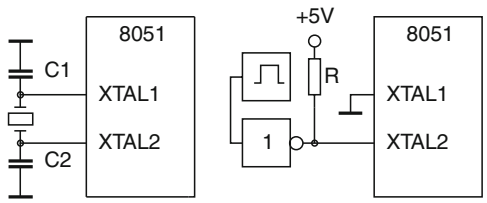


Abb. 52.97 Taktgenerator

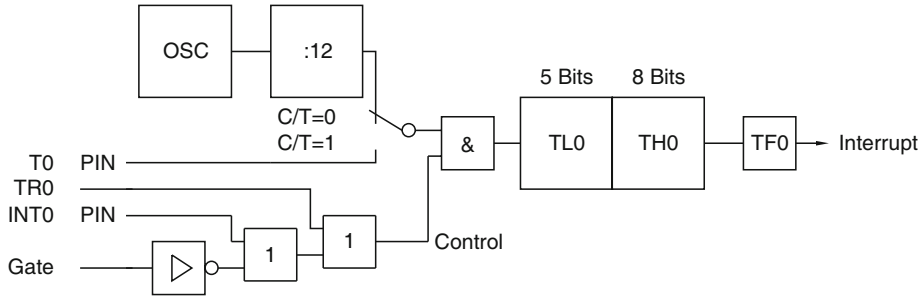


Abb. 52.98 TIMER in der Betriebsart 0

Tab. 52.30 Stromaufnahme *I* (mA)

	VCC					
	Active Mode			Idle Mode		
Frequenz	4 V	5 V	6 V	4 V	5 V	6 V
0,5 MHz	1,6	2,2	3	0,6	0,9	1,2
3,5 MHz	4,3	5,7	7,5	1,1	1,6	2,2
8,0 MHz	8,3	11	14	1,8	2,7	3,7
12 MHz	12	16	20	2,5	3,7	5

Schnittstelle weiterhin mit Taktsignalen. Diese Betriebsart kann über ein *Reset* oder aber auch über ein *Interrupt* beendet werden. Die unterschiedliche Stromaufnahme wird in Tab. 52.30 dargestellt.

52.5.2.6 TIMER

Der MC 8051 verfügt über 2 interne 16-Bit-Zähler T0 und T1, die als Zähler (Counter) oder Zeitgeber (Timer) programmiert werden können. Man spricht von einem Zeitgeber, wenn der interne Takt benutzt wird. Mit Zähler/Zeitgeber können Ereignisse gezählt, Zeitintervalle oder Pulszeiten gemessen und Baudraten für die serielle Schnittstelle erzeugt werden. Die jeweilige Arbeitsweise wird in den entsprechenden SFR's, dem TMOD-Register (Timer Modus) und TCON-Register (Timer Control) festgelegt. Hiermit kann beispielsweise der Zähler gestartet oder angehalten werden. Auch können in der Betriebsart 2 beide Zähler als 16-Bit-Zähler verwendet werden. In der Betriebsart 3 wird der Timer automatisch bei einem Überlaufen nachgeladen. Man spricht hier auch von der Betriebsart „Auto Reload". Ein gerade vorhandener Zählerstand lässt sich aus den SFR T1 bzw. T2 ablesen. Die Zähler/Zeitgeber sind Aufwärtszähler, die ein Überlauf-Kennzeichen erzeugen, beispielsweise beim 16-Bit-Zähler von FFFF nach 0000. Die Überlaufkennzeichen sind Bestandteil des TCON-Registers (Bit 7: TF1, Bit 5: TF0). Diese Kennzeichen dienen auch als Unterbrechungs-kennzeichen (Timer Interrut Flags) und werden durch ein Programm abgefragt. Abb. 52.98 zeigt das Prinzip eines 13-Bit-Zählers/Zeitgebers in der Betriebsart (Mode) 0.

Über TR0, INT0(PIN) und Gate wird der Zeitgeber über eine TOR-Schaltung (das UND-Gatter wirkt als Schalter) aktiviert. Die Frequenz des Zeitgebers ist vom Quarz abhängig. Da jeder Maschinenzyklus den TIMER incrementiert und jeder Maschinenzyklus aus 12 Taktperioden besteht, erfolgt die im Bild dargestellte Teilung durch 12. Im Abb. 52.99 sind die Bitanordnung der zuständigen Register und die verschiedenen Betriebsarten aufgeführt.

TCON-Register

TF1	TTR1	TF0	TR0	IE1	IT1	IE0	IT0

TMOD-Register

GATE	C/T	M1	M0	GATE	C/T	M1	M0
TIMER 1				TIMER 2			

Betriebsarten

M1	M2	Funktion
0	0	13-Bit-Zähler/Zeitgeber
0	1	16-Bit-Zähler/Zeitgeber
1	0	8-Bit-Zähler/Zeitgeber
1	1	Timer 0 TL0 ist 8-Bit Zähler oder Zeitgeber TH0 ist 8-Bit Zeitgeber
1	1	Timer 1 Zähler stoppt

Abb. 52.99 Zähler-Betriebsarten und Register

52.5.2.7 Unterbrechungssystem

Dem MC 8051 stehen 5 Interruptquellen (Interrupt = Unterbrechung) mit 2 Prioritätsebenen zur Verfügung. Nach Abb. 52.100 sind dies die beiden Eingänge INT0 und INT1, die beiden Zähler und eine interne serielle Schnittstelle. Nur das Interrupt der seriellen Schnittstelle wird durch Software ausgelöst, alle übrigen sind Hardware-Interrupts. Die Interrupteingänge INT0 und INT1 können je nach Programmierung flanken- oder pegelgesteuert sein. Nach einer Interruptanforderung wird das gerade laufende Programm unterbrochen, und es erfolgt ein Sprung zum Interruptprogramm. Über eine Maske können Interrupts einzeln zugelassen oder gesperrt oder über die Interruptfreigabe alle Quellen freigegeben oder alle gesperrt werden. Liegen mehrere Interruptanforderungen gleichzeitig an, gilt die Prioritätenreihenfolge. Bei gewünschten Interrupts wer- den die entsprechenden Einsprungadressen des internen Programmspeichers angewählt (siehe auch Abb. 52.102).

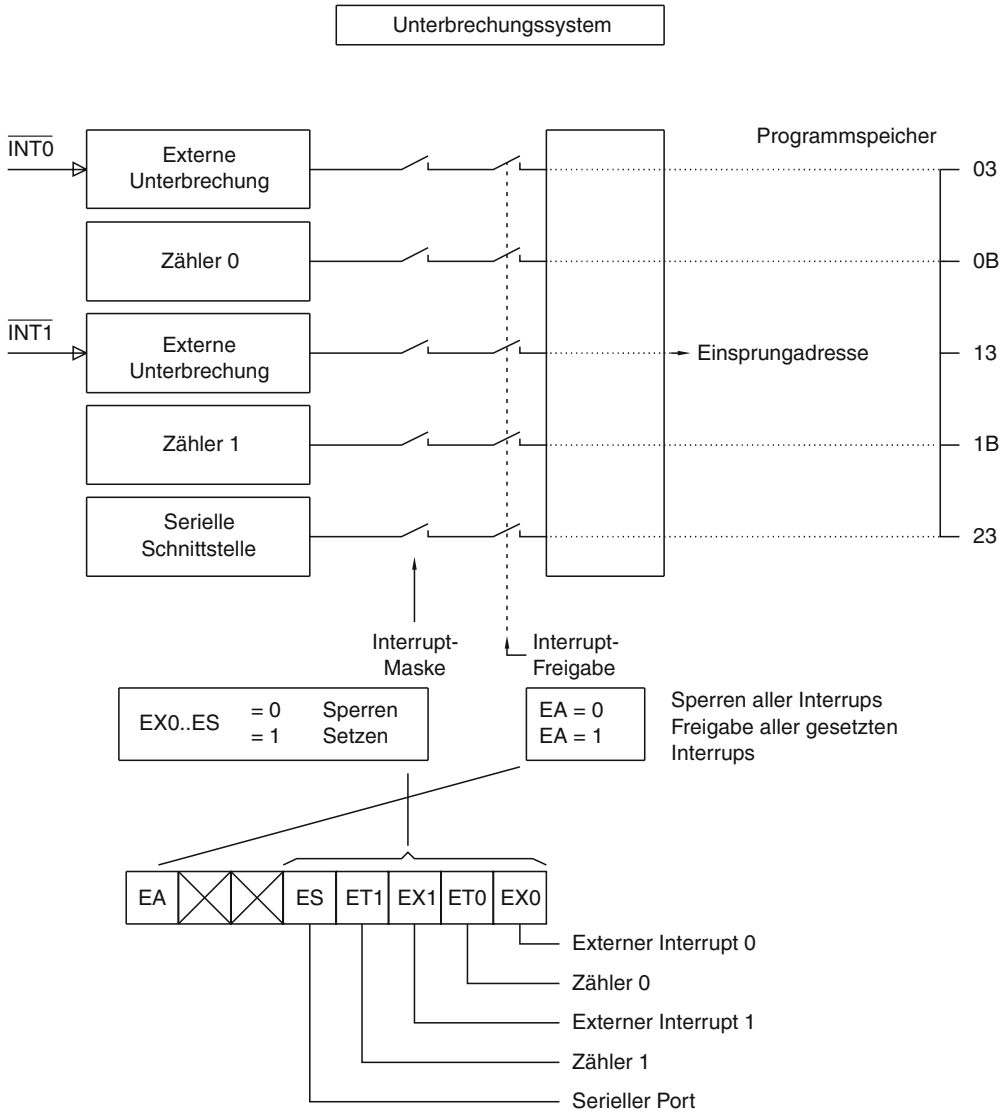


Abb. 52.100 Unterbrechungssystem

52.5.2.8 Speicher

Als Speicher sind ein 128 bzw. 256 Byte RAM und 4 oder 8 kB ROM integriert. Der 8052AH-BASIC besitzt zusätzlich intern einen *Basicinterpreter* mit 8 kB EPROM. Mikrocontroller unterscheiden nach *Daten-* und *Programmspeicher* (Abb. 52.101). Befehle stehen daher nur im Programmspeicher. Der MC 8051 kann als 8-Bit-Mikrocontroller höchstens jeweils 64 KByte adressieren.

Der MC besitzt als Programmspeicher ein internes 4KB ROM. Liegt der /EA-Anschluss auf „1“, wird von dort aus gelesen. Bei /EA „0“ wird der *externe* Programmspeicher angesprochen. Im internen Programmspeicher sind Einsprungadressen reserviert, wie der internen Speicheraufteilung zu entnehmen ist (Abb. 52.102). Eine Gesamtübersicht verschafft Tab. 52.31. Der *interne* Datenspeicherbereich liegt von 00 bis FF, wobei sich bis Adresse 1F vier Registerbänke mit jeweils acht Registern R0 . . . R7 befinden. Dieser Speicherbereich ist *direkt* und *indirekt* adressierbar (siehe Abschn. 52.6 „Maschinensprache“). Der sich anschließende *bitadressierbare* Bereich (jedes Bit ist einzeln ansprechbar) ist direkt nur bis Adresse 2F adressierbar. Der weitere Datenspeicherbereich ist von einer Anzahl Registern, die teilweise ebenfalls bitadressierbar sind, belegt. Es sind Register mit speziellen Funktionen, wie der Stackpointer SP mit der Adresse 81. Die entsprechende Adressenaufteilung von 80 bis 7F sind in der Tab. III-34 zu finden. Der 64 KByte externe Datenspeicher ist nur *indirekt* adressierbar.

52.5.2.9 Serielle Schnittstelle

Alle Mikrocontroller der 8051-Familie verfügen über mindestens eine serielle Schnittstelle, mit der im Asynchron- oder Synchronbetrieb eine serielle Datenübertragung durchgeführt werden kann. Nach Abb. 52.103 besteht die chipinterne Schnitt-

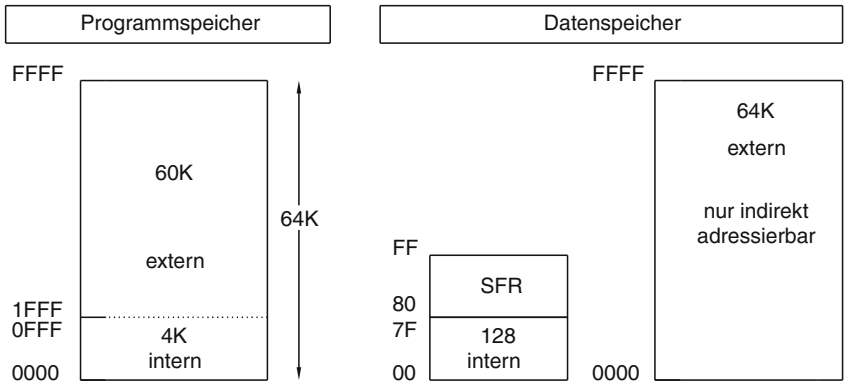


Abb. 52.101 Speicherplan

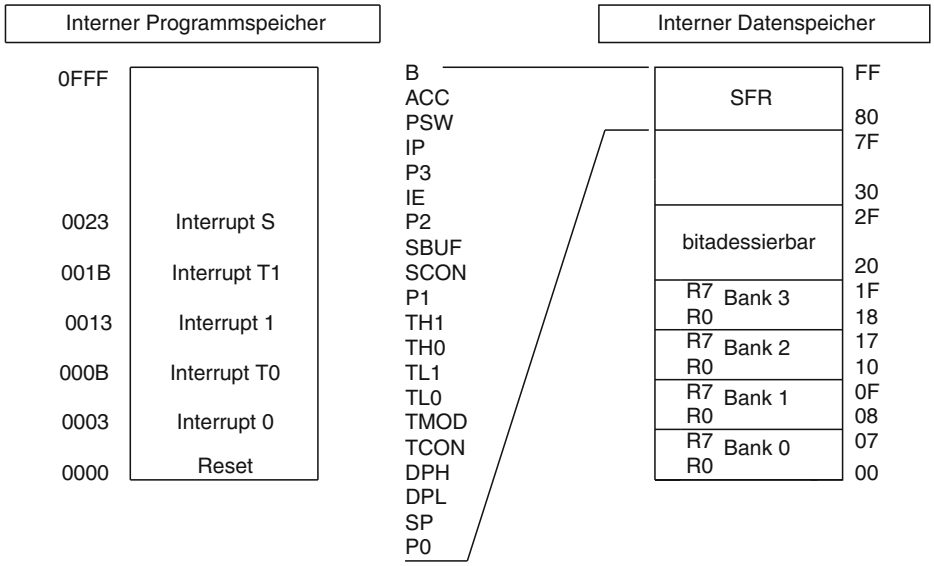


Abb. 52.102 Speicheraufteilung

stelle aus dem Datenregister SBUF und dem Kontroll- oder Statusregister SCON. Das Datenregister SBUF besitzt ein Sende- und Empfangsregister, auf das lesend oder schreibend zugegriffen wird. Die Eingangsdaten gelangen über den Anschluss RxD in ein Schieberegister, wo eine Umwandlung des seriellen Datenstroms in eine parallele Information erfolgt. Nach Ausmaskierung der nicht zur Information gehörenden Bits wie Start- und Stopbit gelangen die Daten ins Empfangsregister. Die Ausgangsdaten gelangen über das Senderegister und über das SCON-Register zum Ausgang TxD. Mit dem Kontroll-/Statusregister SCON kann die gewünschte Betriebsart (Mode 0, 1, 2, 3) gewählt werden. Weiter enthält es die Steuerbits für Empfänger- und Senderinterrupt. In der Betriebsart 0 werden die Daten über RxD (P3,0) gesendet oder empfangen, und an TxD (P3,1) wird ein Sendetakt mit dem Tastverhältnis 1 : 1 bei einer festen Frequenz von $f_{osz}/12$ ausgegeben.

Abb. 52.104 zeigt die Bedeutung der Bitanordnung im SCON-Register und die Bitzuordnung zur Einstellung der einzelnen Betriebsarten.

52.5.3 16-Bit-Mikrocontroller

Abb. 52.105 zeigt die Architektur des MC SAB80C515/535 als vereinfachtes Blockbild. Der Baustein besitzt außer den 6 Ports (= 48 Kanäle) zur digitalen Ein- und Ausgabe einen 8-Bit gemultiplexten A/D-Wandler-Eingang. Das 256 Byte interne RAM kann extern bis zu 64 kB erweitert werden und ist bei Betriebsspannungsausfall über einen Anschluss ruhestromversorgt. Im Gegensatz zum ...515 hat der ...535 keinen internen Programmspeicher. Seine Programme müssen in einem externen ROM untergebracht werden. Außerdem ist im ...535 ein Watchdog-Timer integriert. Seine Aufgabe ist es, bei Störungen einen Reset auszulösen. Es handelt sich hierbei im Prinzip um einen 16-Bit-Zähler, der mit jedem Maschinenzyklus inkrementiert wird. Bei Überlauf erfolgt ein interner Reset, und ein Statusbit wird gesetzt, das durch eine Abfrage die Resetursache kennzeichnet. Im Programmablauf wird der Watchdog-Timer jedesmal vor dem Überlauf zurückgesetzt. Weitere Besonderheiten dieses Bausteins sind *Capture*-, *Compare*- und *Reload*-Funktionen des Zählers/Zeitgebers 2.

Tab. 52.31 Register- und Adressenübersicht

	Bit-Adressen								Byte-Adresse
SFR	Bit 7 (MSB)	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)	
B	F7	F6	F5	F4	F3	F2	F1	F0	F0
ACC	E7	E6	E5	E4	E3	E2	E1	E0	E0
PSW	D7	D6	D5	D4	D3	D2	D1	D0	D0
IP	BF	BE	BD	BC	BB	BA	B9	B8	B8
P3	B7	B6	B5	B4	B3	B2	B1	B0	B0
IE	AF	AE	AD	AC	AB	AA	A9	A8	A8
P2	A7	A6	A5	A4	A3	A2	A1	A0	A0
SBUF									99
SCON	9F	9E	9D	9C	9B	9A	99	98	98
P1	97	96	95	94	93	92	91	90	90
TH1									8D
TH0									8C
TL1									8B
TL0									8A
TMOD									89
TCON	8F	8E	8D	8C	8B	8A	89	88	88
DPH									83
DPL									82
SP									81
P0	87	86	85	84	83	82	81	80	80

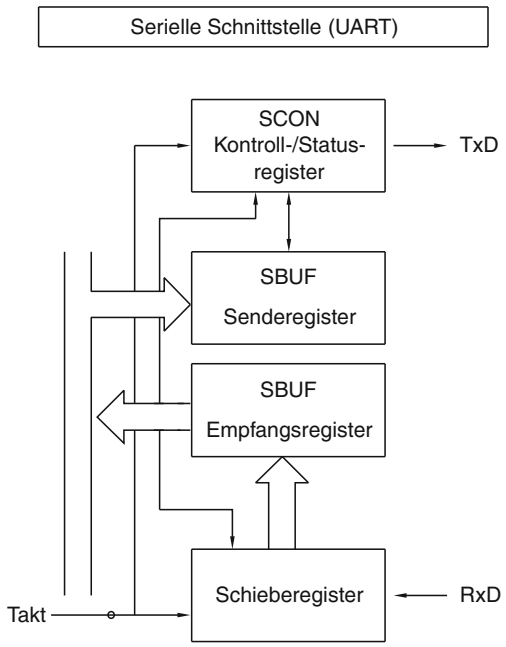


Abb. 52.103 Serielle Schnittstelle

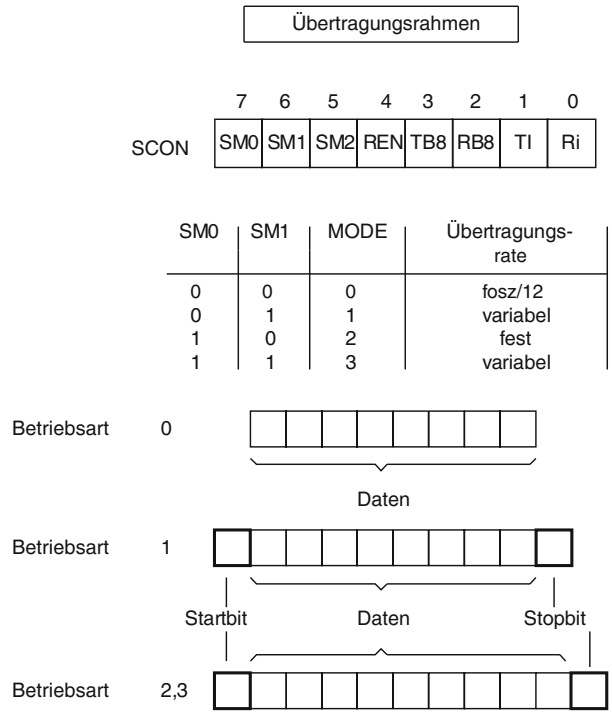


Abb. 52.104 Bitzuordnung im SCON-Register und Betriebsarten

Bei der Capture-Funktion (capture = einfangen) sind Zeitmessungen realisierbar, weil durch ein externes Signal ein Zwischenspeichern des Zählerstandes in ein entsprechendes SFR ausgelöst wird. Bei der Compare-Funktion (compare = vergleichen) kann eine Meldung (z. B. Interrupt) erfolgen, wenn der Inhalt des laufenden Timers 2 mit dem des Compare-Registers übereinstimmt. Mit der Reload-Funktion wird die Zykluszeit der Timer variiert. Hiermit lässt sich dann eine PWM (Pulsweitenmodulation) durchführen. Solche Signale können verwendet werden, um beispielsweise Schritt- oder Gleichstrommotoren steuern zu können.

Die Port-Kanäle können digital oder analog über die gemultiplexten Eingänge genutzt werden. Der Baustein verfügt über

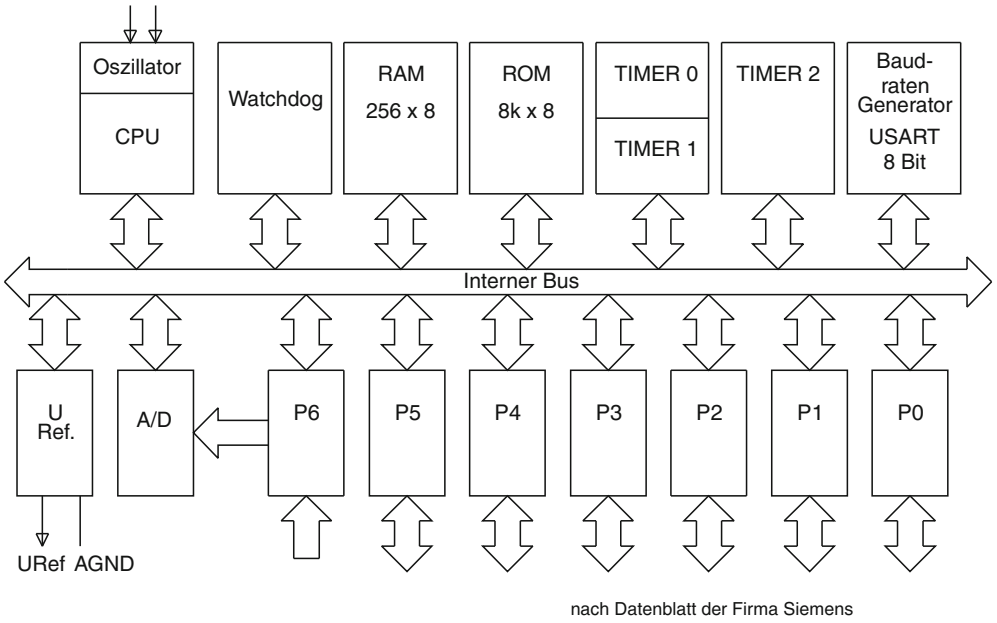


Abb. 52.105 Blockbild des SAB 80C515/C535

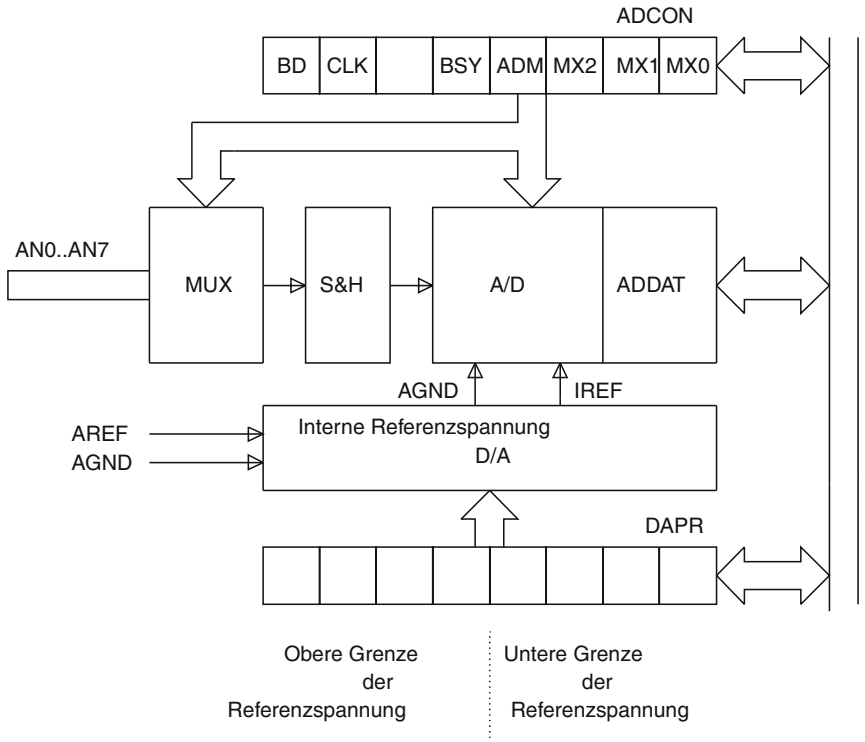


Abb. 52.106 AD/DA-Wandler

eine programmierbare Referenzspannungsquelle. Der Vorteil in der Programmierbarkeit der Referenzspannung liegt einmal in der eventuell gewünschten unterschiedlichen Referenzspannung pro Kanal und in der veränderbaren Wandlerauflösung.

Der Wandler selbst arbeitet nach dem Prinzip der *sukzessiven Approximation* mit einer Wandlungszeit von 15 ms. Abb. 52.106 zeigt das Blockbild des A/D-Wandlers mit dem SFR DAPR (D/A-Program-Register). Im L-Nibble ist die untere Grenze der Referenzspannung und im H-Nibble die obere Grenze definiert. Die Wahl des Analogeingangs ist im SFR ADCON festgelegt (A/D-Controll-Register), in dem auch festgelegt ist, ob nur eine oder fortlaufende Wandlungen stattfinden sollen.

52.6 Maschinensprache

52.6.1 Allgemeines

Soll ein Mikrocomputer Anweisungen ausführen, müssen die dazu erforderlichen Befehle in *binärcodierter Form* im Speicher vorliegen. Jeder in dieser Form zum Mikroprozessor gebrachte Befehl gelangt zum Befehlsregister, wird dort interpretiert und veranlasst die CPU, die dazugehörende Operation auszuführen. Da der Mikroprozessor diese Anweisungen direkt in Operationen umsetzen kann, spricht man von *Maschinensprache*. Der *Maschinencode* (Mikrocode, Operationscode), nach denen der Mikroprozessor seine Operationen durchführt, ist hardwaremäßig festgelegt und bei den Prozessoren unterschiedlich.

Alle Befehle, die der Mikroprozessor ausführen kann, stellen seinen *Befehlssatz* (Instructions Set) dar. Er wird in den Datenbüchern der Hersteller ausführlich beschrieben. Um in Maschinensprache zu programmieren, muss der Programmierer den Prozessor kennen. Hierbei ist vor allem die Kenntnis des Registerplans und des Befehlssatzes der CPU wesentlich. Als Beispiel dient der Prozessor 8085. Ein Auszug des Befehlssatzes befindet sich in Tab. 52.31 und 52.32. Abb. 52.107 zeigt den Registerplan der 8085 CPU.

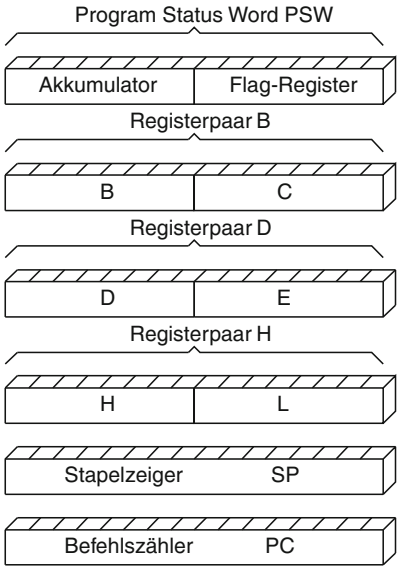


Abb. 52.107 Registerplan der 8085 CPU

Hier sind alle Register zusammengefasst, auf die man als Programmierer Einfluss üben kann. Die Register Akkumulator bis L-Register sind 8-Bit-Register. Der Stapelzeiger und der Befehlszähler sind 16-Bit Register. Damit ist die Bit-Zahl gemeint, die das jeweilige Register aufnehmen kann. Die Register B, C, D, E und H, L können zusammengefasst werden. Sie arbeiten dann als Registerpaare B, D und H. Der Name des Registerpaares gibt an, wo sich jeweils die höherwertigen Bits befinden. Akkumulator und Flag-Register (Kennzeichen-Register) fasst man unter dem Namen PSW (Progamm Status Word) zusammen.

52.6.2 Maschinencode

Für alle Befehle eines Mikroprozessors existiert eine Struktur unterschiedlicher Bits, die eine bestimmte Systematik beinhaltet. Daher lassen sich aus dem Format des Maschinencodes und aus der Kenntnis der Registertabelle weitere Befehle herleiten. Der Registername oder die Registerkennung und der Operationscode sind hardwaremäßig festgelegt.

Beispiel

Der Befehl „INR D“ (Increment D) bedeutet, dass sich der Registerinhalt nach Befehlsausführung um 1 erhöht. Der Operationscode lautet $00sss100_2$ (Abb. 52.108). Für die Platzhalter sss kann jetzt die Registerkennung 010_2 = D-Register eingefügt werden, und man erhält den Maschinencode 00010100_2. ◄

Beispiel

Die Bitfolge 10110sss2 bedeutet, dass der Akkumulator mit einem Register eine Oder-Verknüpfung durchführen soll. Mit welchem Register dies zu geschehen hat, ist wieder durch die Platzhalter sss (Quellenregister) symbolisiert. Damit lautet der Befehl ORA B: 10110000_2 und ORA C: 10110001_2.

Die Z80 CPU besitzt bei gleichen Befehlen den gleichen Maschinencode. Die MC's 8085 und Z80 arbeiten mit Informationseinheiten von 8 Bit (1 Byte). ◄

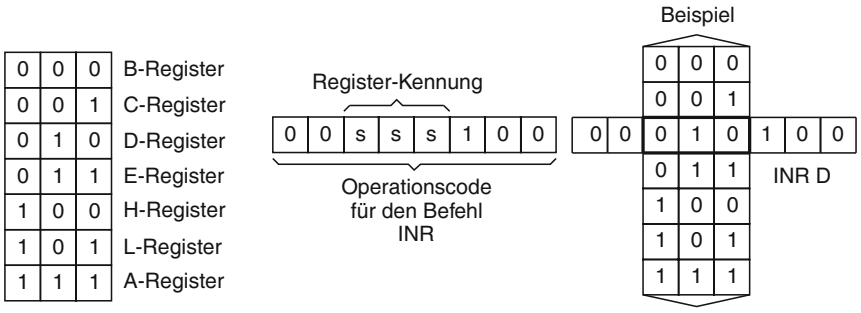


Abb. 52.108 Registertabelle und Herleitung von Befehlen

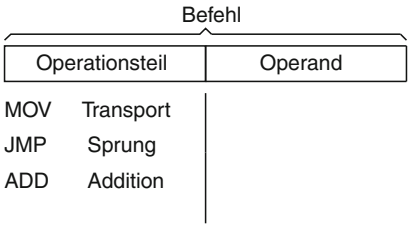


Abb. 52.109 Aufbau eines Befehls

52.6.3 Befehlsaufbau

Ein Befehl besteht aus einem Operationsteil und einem Operanden (Abb. 52.109). Das erste Byte des Befehls ist immer der Befehlscode (Operationscode).

Hiermit wird dem Prozessor mitgeteilt, welche Operation ausgeführt werden soll. In weiteren Bytes können sich Daten oder Adressen befinden. Soll der Prozessor Daten vom Akkumulator zu einer Speicherstelle bringen, dann ist dies mit einem Byte nicht mehr zu codieren, da die Adresse des Speichers schon 16 Bit ausmacht. Folglich wird dem Prozessor im ersten Byte mitgeteilt, dass er Daten zu einer Speicherstelle zu kopieren hat; in den beiden folgenden Bytes wird dann die Adresse angegeben. Unterschieden werden dem Aufbau nach 1-Byte-, 2-Byte- und 3-Byte-Befehle (Abb. 52.110).

Der 1-Byte-Befehl bedeutet, dass vom Akkumulatorinhalt der Inhalt des Registers B subtrahiert wird. Das Ergebnis wird wieder zum Akkumulator geschrieben. Im 2-Byte-Befehl werden die Daten 11111111_2 zum Akkumulator gebracht, und im 3-Byte-Befehl gelangen die Daten des Akkumulator zur Speicherstelle $1000\ 0010\ 0000\ 0000_2$. Bei Angabe der Adresse ist wesentlich, dass die Reihenfolge genau eingehalten ist. Zunächst muss dem Befehlsregister das L-Byte (niederwertigsten 8 Bit) und dann das H-Byte (höherwertigsten 8 Bit) zugeführt werden.

52.6.4 Befehlsdarstellung

Heute ist es nicht mehr üblich, in binärcodierter Form zu programmieren. Wenn von Maschinenprogrammen gesprochen wird, meint man die *Assemblerdarstellung* der Befehle. Die *Mnemoniks* (Symbole) der Operationscodes (Assembler) lassen sich mit einiger Übung leicht merken, da sie sich aus den Befehlsfunktionen ableiten lassen. Die Darstellung der Daten und Adressen erfolgt hierbei in hexadezimaler Form.

Beispiel

Mnemonik/ Assembler	Kommentar	
MVI A,FF	Kopiere die Daten 11111111_2 zum Akkumulator	(8085)
LD A,FF	Kopiere die Daten 11111111_2 zum Akkumulator	(Z80)
STA 8200	Kopiere den Inhalt des Akkumulators zur Adresse	(8085)
	$1000\ 0010\ 0000\ 0000_2$	

Um Schreibarbeit zu sparen und Programme übersichtlich zu halten, verwendet man bei Kommentaren häufig eine Kurzschreibweise. Alles, was in der Klammer steht, bedeutet dann den Inhalt eines Registers oder einer Adresse. Ein Pfeil gibt die Datenrichtung an.

Beispiel:

(A) → B bedeutet: Akkumulatorinhalt nach B bringen

(8200) → A bedeutet: Inhalt der Adresse 8200 zum Akkumulator bringen.

Die Umsetzung der Mnemonik und der Hexform in die Binärform des Maschinencodes erfolgt mit einem Übersetzerprogramm (Assembler). ◄

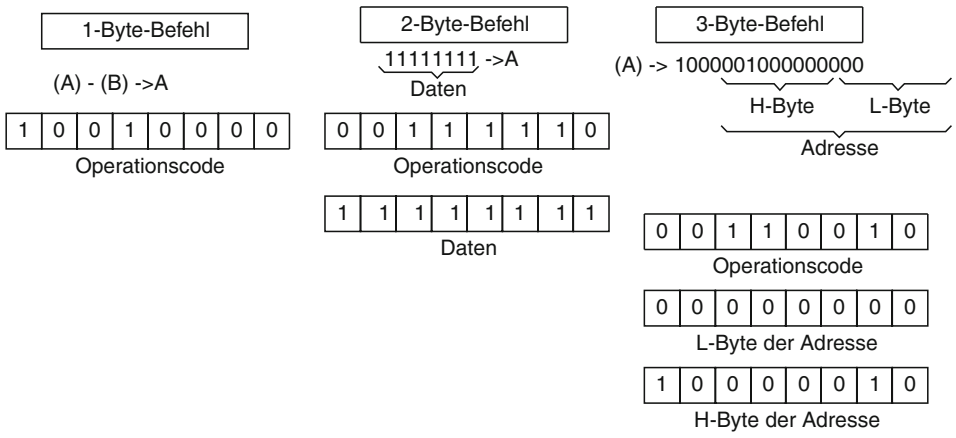


Abb. 52.110 Befehlsformate

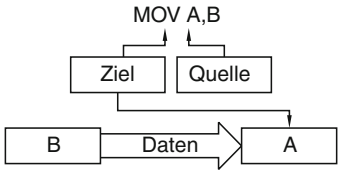


Abb. 52.111 Assemblerschreibweise

52.6.5 Befehle

Die Befehle eines Mikroprozessors lassen sich unter dem Gesichtspunkt der Befehlsfunktionen oder nach Adressierungsarten ordnen.

52.6.5.1 Befehlsfunktionen

Bei den *Datentransport-Befehlen* (Kopierbefehlen) werden Daten *unmittelbar* in ein Register eingegeben oder Daten werden zwischen den Registern oder zwischen Speicher und Registern kopiert. Bei den Transportbefehlen wird zwischen verschiedenen *Adressierungsarten* unterschieden. In der Assemblerschreibweise ist es üblich, zunächst das Ziel und dann die Quelle anzugeben. MOV A,B bedeutet, dass Daten von B nach A transportiert werden sollen (Abb. 52.111).

Mit Hilfe der *arithmetischen Befehle* werden Daten oder Registerinhalte miteinander addiert, subtrahiert, inkrementiert (um 1 erhöht) oder dekrementiert (um 1 vermindert).

Beispiel

Die Inhalte der Speicherstellen A000 (Inhalt $= 03_H$) und A001 (Inhalt $= 11_H$) werden addiert und das Ergebnis zur Speicherstelle A002 gebracht. Nach dem Programmablauf befindet sich in der Speicherstelle A002 der Inhalt 14_H bzw. $0001\ 0100_2$.

Adresse	Operations	Assembler	Kommentar
0000	3A 00 A0	LDA A000	(A000) → A
0003	47	MOV B,A	(A) → B „retten“
0004	3A 01 A0	LDA A001	(A001) → A
0007	80	ADD B	(A) + (B) → A
0008	32 02 A0	STA A002	(A) → A002
000B	76	HLT	Halt

Mit den *Schiebe-Befehlen* (Rotations-Befehle) kann der Inhalt vom Akkumulator links oder rechts jeweils um eine Bitstelle verschoben werden. Dadurch gelangt Bitstelle 7 bei Links-Rotation zur Bitstelle 0 und Bitstelle 0 bei Rechts-Rotation zur Bitstelle 7. Hierbei wird auch gleichzeitig das Carrybit verändert. Eine weitere Befehlsart lässt die Rotation nur über das Carrybit zu. Abb. 52.112 führt die verschiedenen Befehle auf.

Mit Hilfe der *logischen Befehle* können zwischen dem Akkumulator und den Registern oder mit Konstanten logische Verknüpfungen wie UND, ODER und EXKLUSIV-ODER durchgeführt werden. Durch diese Befehle wird das Flag-Register beeinflusst.

Mit den *Sprungbefehlen* wird der Ablauf eines Programms geändert. Der Prozessor arbeitet normalerweise Adresse für Adresse nacheinander ab. Es können unbedingte oder bedingte Sprünge programmiert werden. Ein *unbedingter Sprung* enthält im Befehl selbst eine Adresse und wird *immer* ausgeführt. ◀

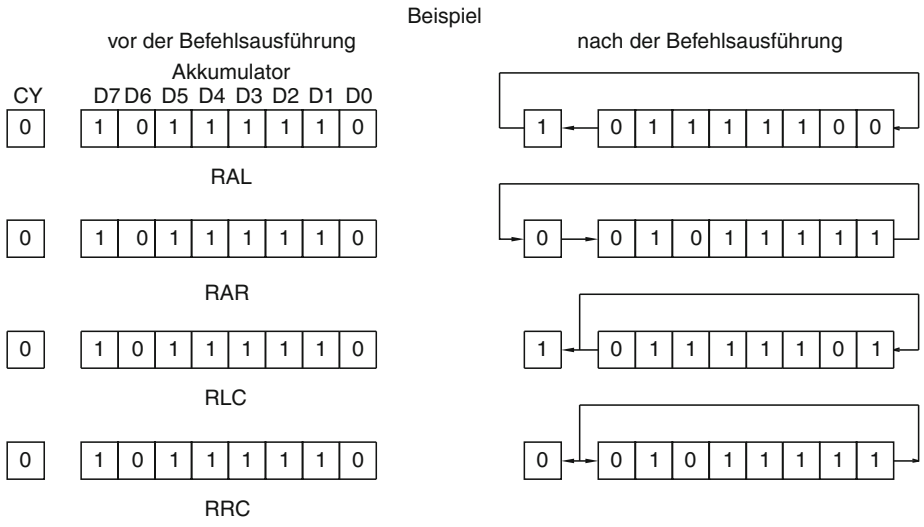


Abb. 52.112 Rotationsbefehle



Abb. 52.113 Bitanordnung im Flagregister

Beispiel

JMP 8200 ;Springe zur Adresse 8200

Der *bedingte Sprung* ist von einer Bedingung abhängig. Hierbei ist zwischen bejahenden und verneinenden Sprüngen zu unterscheiden. ◀

Beispiel

JNZ ;Springe, wenn Z nicht "1" gesetzt ist.
JZ ;Springe, wenn Z "1" gesetzt ist. ◀

Die Bedingung überprüft der Mikroprozessor im *Flagregister* (Kennzeichen-Register). In Abb. 52.113 befindet sich die Anordnung der Bits. Die Bitstellen (Flags) werden bei bestimmten Operationen wie arithmetischen und logischen Befehlen gesetzt.

Beispiele

Das Carry-Flag wird bei arithmetischen Operation gesetzt oder zurückgesetzt. Flagregister

$$\begin{array}{ll} (A) & = \ 1010\ 1111_2 \\ (B) & = \ 0110\ 0000_2 \\ \hline (A) + (B) = 1\ 0000\ 1111_2 & \text{Cy wird auf „1“ gesetzt} \end{array}$$

Ist das Ergebnis bestimmter Operationen „0“, dann wird das Zero-Flag gesetzt. Stack

$$\begin{array}{ll} (A) & = 1011\ 1111_2 \\ (B) & = 1011\ 1111_2 \\ \hline (A) - (B) = 0000\ 0000_2 & \text{Z wird auf „1“ gesetzt} \end{array}$$

$$\begin{array}{ll} (A) & = \ 1010\ 0001_2 \\ (B) & = \ 0101\ 1111_2 \\ \hline (A) + (B) = 1\ 0000\ 0000_2 & \text{Z und Cy wird auf „1“ gesetzt} \end{array}$$

Das Sign-Flag (Vorzeichen) bezieht sich auf das MSB (Most Significant Bit = höchstwertigstes Bit) des Akkumulators bei Interpretation des Datenwortes in *Zweierkomplementdarstellung*.

$$(A) = 0011\ 1111_2 \quad S = \text{„0“}$$
$$(A) = 1000\ 1101_2 \quad S = \text{„1“}$$

Das Hilfscarry-Flag wird durch arithmetische und logische Verknüpfungen bei Entstehen eines Übertrags von Bit 3 nach Bit 4 gesetzt.

Das Parity-Flag gibt eine „1“-Meldung, bei gerader Parität der 1-Bit-Stellen im Akkumulator und eine „0“-Meldung bei ungerader Parität

$$(A) = 1010\ 1111_2 \quad P = \text{„1“}$$
$$(A) = 1010\ 1101_2 \quad P = \text{„0“}$$ ◀

Beispiel

Abfrage des Zero-Flag

Bei einfachen Zählschleifen wird der Inhalt eines Registers so lange dekrementiert, bis der Registerinhalt Null ist.

Adresse	Operationscode	Assembler	Bemerkung
0000	3E FF	MVI A,FF	FF → A
0002	3D	DCR A	(A) − 1 → A
0003	C2 02 00	JNZ 0002	Null?
0006	76	HLT	Halt

◀

Unterprogramm- und Stapelspeicherbefehle: Programmteile oder Programme, die innerhalb eines Hauptprogramms häufiger benötigt werden, brauchen nur einmal programmiert werden. Solche Programme nennt man *Unterprogramme*. Wie bei Sprungbefehlen gibt es bedingte und unbedingte Unterprogrammbefehle. Ein unbedingter Unterprogrammaufruf wird mit dem Befehl CALL erreicht. Der Rücksprung zum Hauptprogramm erfolgt mit dem Befehl RET. Mit Aufruf des Unterprogramms wird die nachfolgende Hauptprogrammadresse zum Stack gebracht, um beim Rücksprung aus dem Unterprogramm das Hauptprogramm dort fortsetzen zu können. Oft werden im Unterprogramm Register benötigt, die im Hauptprogramm schon Daten übertragen bekommen haben und deren Inhalte sich nicht während des Unterprogrammablaufs ändern dürfen. Zur „Rettung“ dieser Daten wird dann ebenfalls der *Stapelspeicher* eingesetzt.

Der Stapelspeicher (Stack) selbst ist Teil des Arbeitsspeichers. Er ist bei Programmierung von Unterprogrammen im Hauptprogramm stets zu initialisieren. Hiermit ist gemeint, dass der Stapelzeiger mit einer Adresse geladen wird. Dabei kann jeder beliebige Teil des Speichers als Stack verwendet werden.

Da der Stapelspeicher mit der Adresse des Stackpointer beginnend bei jedem Aufruf je die zwei niederwertigeren Adressen mit Daten füllt, ist es zweckmäßig, ihn an die höchste Adresse des Arbeitsspeichers zu legen. Aus Abb. 52.114 geht das Prinzip hervor. Hier wird der Stackpointer auf die Adresse 27FF gebracht. Nach Aufruf des Unterprogramms mit CALL steht der Stackpointer auf der Adresse 27FD. Mit den Befehlen PUSH PSW und PUSH B im Unterprogramm werden die Inhalte der Register auf den Stack gebracht, sie werden für das Hauptprogramm „gerettet“. Der Stackpointer zeigt auf die Adresse 27F9. Im Unterprogramm kann jetzt mit den Registern A, B und C gearbeitet werden, weil am Ende des Unterprogramms durch POP

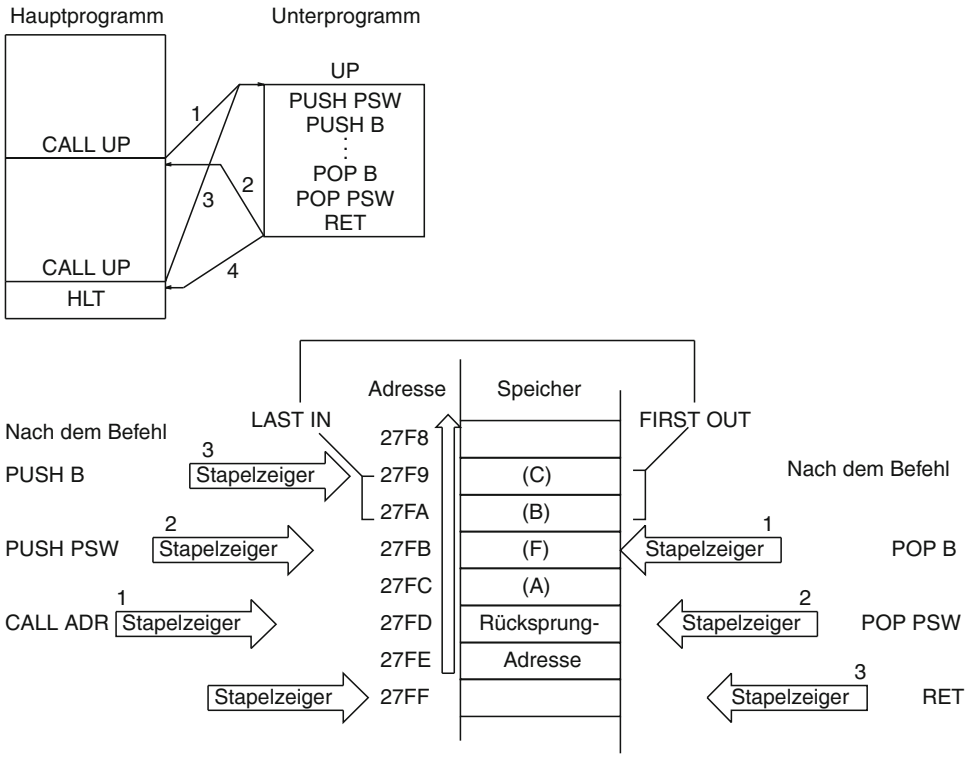


Abb. 52.114 Stack-Prinzip

B und POP PSW der Originalzustand der Register B und A wieder hergestellt wird. Mit POP B wird beispielsweise der Inhalt der Adresse 27F9 zum C-Register und der Inhalt der Adresse 27FA zum B-Register gebracht. Bei der Programmierung ist, wie auch das Beispiel zeigt, darauf zu achten, dass der zuletzt zum Stack gebrachte Registerinhalt als erstes vom Stack zurückgerufen wird (Last In-First Out), weil sonst Registerinhalte vertauscht werden.

Ein- und Ausgabe-Befehle: Sollen Daten von einer E/A-Einheit zur CPU oder von der CPU zur E/A-Einheit geleitet werden, wird das mit IN (Eingabebefehl) und OUT (Ausgabebefehl) durchgeführt. Hierbei muss hinter dem IN- und OUT-Befehl die jeweilige Adresse der E/A-Einheit (des Ports) angegeben werden.

Beispiel

IN FF: Die Daten gelangen von der Portadresse FF zum Akkumulator
OUT FF: Die Daten werden vom Akkumulator zur Portadresse FF gebracht.

Bei Einsatz eines programmierbaren Portbausteins ist dieser zunächst zu initialisieren. ◄

Interrupt-Befehle dienen zur *Unterbrechungssteuerung* (siehe auch Abschn. 52.2 „Mikroprozessoren“). Durch einen Impuls oder einen Pegel an einem Interrupteingang des Prozessors wird das laufende Programm nach Abarbeitung des gerade in Arbeit befindlichen Befehls unterbrochen. Es erfolgt ein Sprung zu einer vom Hersteller hardwaremäßig festgelegten Adresse, zur Interrupt-Routine. Der Befehlszählerstand wird auf dem Stack gespeichert, um nach dem Rücksprung aus der Routine mit dem Programm fortfahren zu können.

Die Anzahl der Interrupteingänge sind bei den Prozessoren unterschiedlich. Der MC 8085 besitzt die fünf Interrupteingänge TRAP, RST7.5, RST6.5, RST5.5 und INTR. Hierbei wird zwischen *maskierbaren* und *nichtmaskierbaren* Interrupteingängen unterschieden. Mit dem Befehl SIM (Set Interrupt Mask) und Verwendung des Akkumulatorinhalts können die maskierbaren Interrupts gesperrt oder freigegeben werden, wenn allgemein Interrupts mit dem Befehl EI (Enable Interrupt) zugelassen werden. Mit DI (Disable Interrupt) kann die Zulassung der Interrupts verweigert werden. Aus Abb. 52.115 wird die Wirkung der Befehle sichtbar. Eine Interruptannahme erfolgt nur, wenn mit SIM die entsprechende Maske gesetzt wurde und mit EI die Freigabe erfolgt. Nicht von diesen Befehlen betroffen ist der TRAP-Interrupt (nicht maskierbar). Er wird bei einer Anforderung immer wirksam und besitzt Vorrang vor allen anderen Interruptzugriffen.

Ermöglicht die Interrupt-Maske Zugang verschiedener Geräte zur CPU und erfolgen Interruptanforderungen gleichzeitig, tritt die vom Hersteller festgelegte Priorität ein. In Abb. 52.116 ist die Priorität und die Lage der Hardware-Verzweigungen dargestellt. An der jeweiligen Adresse kann nur eine kleine Routine geschrieben werden, oder es steht dort ein Sprungbefehl zum

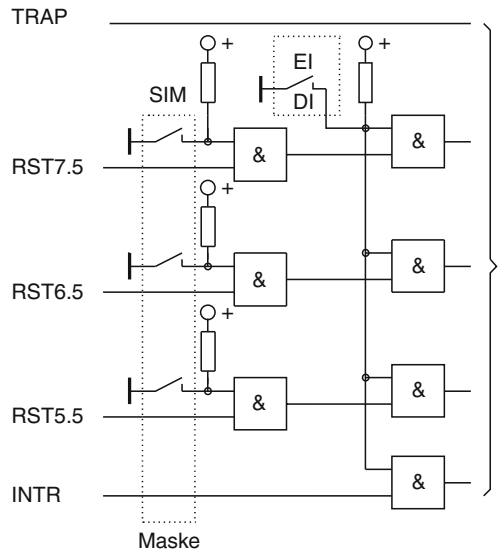


Abb. 52.115 Interruptannahme

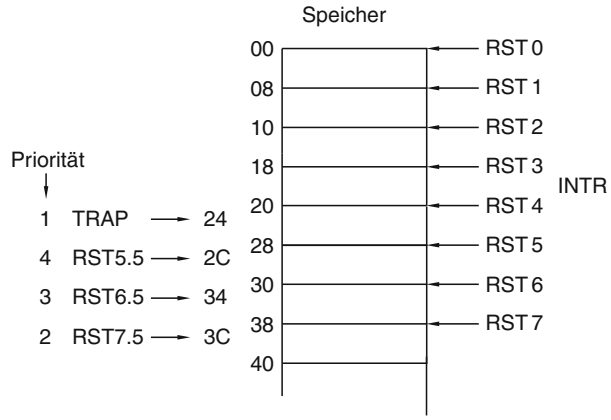


Abb. 52.116 Interruptadressen und Priorität

Arbeitsspeicher. Wie ein Unterprogramm wird die Routine mit RET abgeschlossen.

Mit dem Befehl RIM (Read Interrupt Mask) wird der Akkumulator geladen. Damit können die Maske und tatsächliche Unterbrechungen abgefragt werden, wie die Anordnung der Bits Abb. 52.117 zeigt. Der INTR-Interrupteingang dient als Vektorinterrupt. Bei seiner Auslösung zeigt er nach Abfrage INTA (Interrupt Acknowledge) des Datenbusses auf eine der 8 Hardwareadressen (Abb. 52.117).

52.6.5.2 Adressierungsarten

Bei den Transportbefehlen lassen sich verschiedene Adressierungsarten (Abb. 52.118) unterscheiden. Die *implizite Adressierung* enthält keine Operanden-Adresse, wie das beispielsweise bei dem Kopieren der Daten von einem zum anderen Register der Fall ist. Hier ist im Befehl implizit (inbegriffen) die Funktion gegeben. Man bezeichnet diese Adressierungsart auch als Registeradressierung. Bei der *unmittelbaren* (immediate)

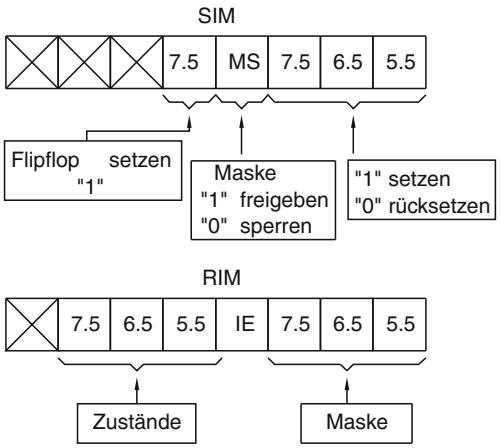


Abb. 52.117 Bit-Anordnung für SIM- und RIM-Befehl

Adressierung enthält der Befehl die zu verarbeitende Konstante. Hierbei kann es sich um eine 8-Bit- oder 16-Bit-Konstante handeln. Man spricht von direkter Adressierung, wenn die Speicherstelle im Befehl benannt wird. Die 16-Bit-Adresse folgt direkt hinter dem Operationscode. Bei der *indirekten Adressierung* wird die Adresse, die angesprochen werden soll, zuvor in einem Registerpaar abgelegt (siehe Beispiele Abb. 52.119).

Der MC Z80 kennt außerdem noch die relative und indizierte Adressierung. Bei der relativen Adressierung erhält der Sprungbefehl eine durch eine Konstante festgelegte Distanzadresse, um von einer absoluten Adresse unabhängig zu sein. Die indizierte Adressierung legt das Indexregister und die Adressen fest. Beispiel: LD A, (IX + 0009) Sprung um (IX) + 9 Adressen.

52.6.6 Befehlszyklus und Befehlszeiten

52.6.6.1 Befehlszyklus

Alle Befehle, die ein Mikroprozessor durchführen soll, stehen im Speicher. Das Zusammenwirken von CPU und Speicher zeigt Abb. 52.120 an einem 2-Byte-Befehl. Beim Einschalten wird der Programmzähler immer auf die Adresse (ADR) 0000 gestellt, die Steuerung der CPU wird mit RD (Read) aktiviert, und der Inhalt der ADR 0000 wird zur CPU gebracht. Dort erfolgt die Decodierung. Der Programmzähler erhöht sich automatisch um 1 und legt die ADR 0001 auf den Adressbus. Der Inhalt dieser Adresse wird von der CPU wieder über den Datenbus gelesen. Mit der Information 3E erhält die CPU den Befehl, das nächste Byte zum Akkumulator zu bringen. Diese Befehlsbearbeitung erfordert eine gewisse Zeit. Das Einlesen und Ausführen eines Befehls bezeichnet man als *Befehlszyklus* (instruction cycle). Er wird in bestimmte Zeiteinheiten mit definierten Vorgängen, den *Operationsschritten* (states), aufgeteilt; jeder Operationsschritt wiederum in mehreren Taktzyklen. Der erste Operationszyklus ruft immer einen Befehl auf und wird daher *Befehlsabrufzyklus* (Opcode Fetch) genannt. Weitere Operationszyklen können der Speicher-Lesevorgang (Memory Read) und der Speicher-Einschreibvorgang (Memory Write) oder Ein-/Ausgabeschnittstelle lesen oder schreiben sein.

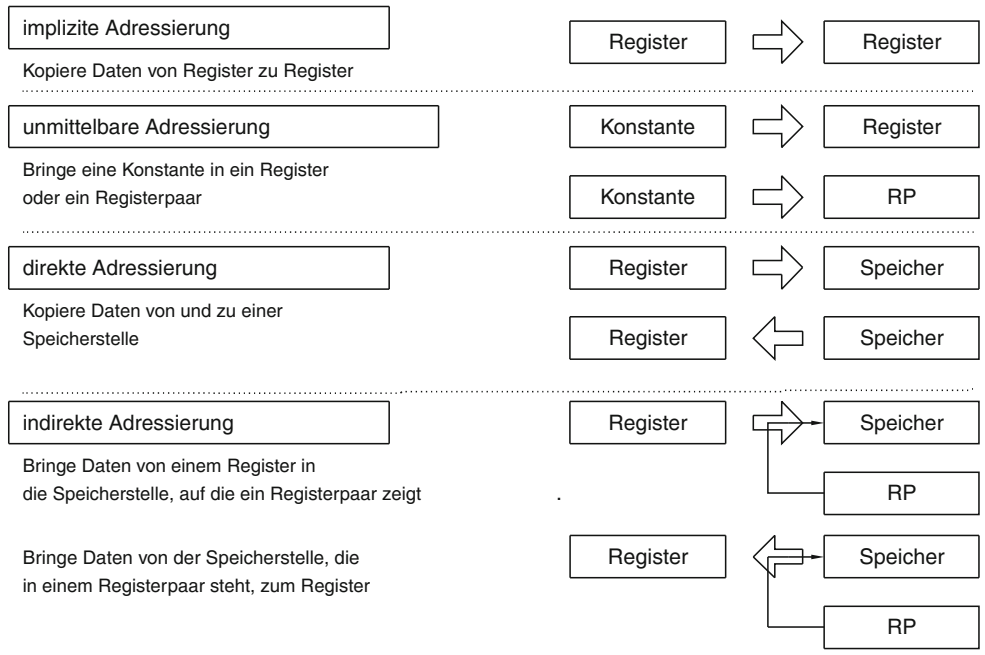


Abb. 52.118 Übersicht: Adressierungsarten

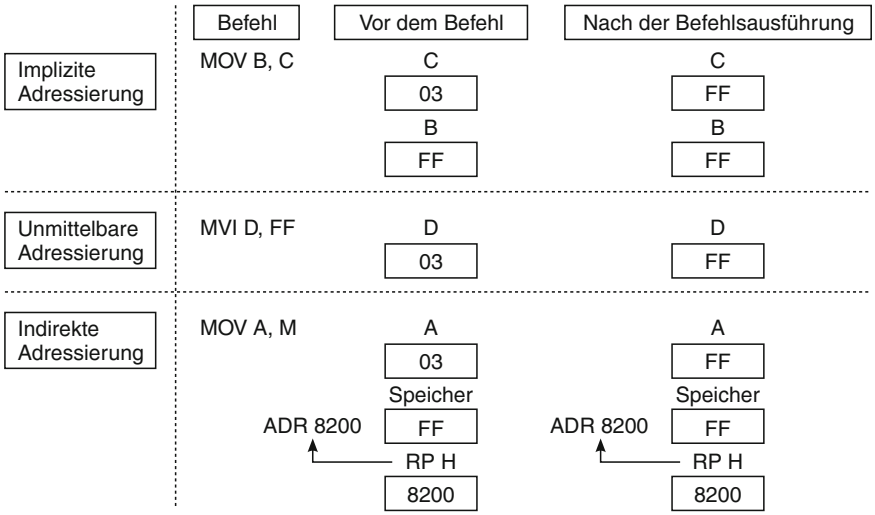


Abb. 52.119 Beispiele

Die Operationszyklen werden auch als *Maschinenzyklen* bezeichnet. Abb. 52.121 führt dies beispielhaft für den Befehl STA 8200 an. Mit diesem Befehl wird der Inhalt des Akkumulators zur Adresse 8200 kopiert.

52.6.6.2 Befehlszeiten

Die Befehlszeit (Dauer eines Befehls) ist zum einen von der Taktdauer und zum anderen von der Anzahl der Takte pro Befehl abhängig. Hieraus lässt sich die Zeit, die ein Programm benötigt, berechnen.

Beispiel

Bei einem Taktzyklus von 0,5 µs ergibt sich mit $7 + (15 \cdot 5) - 3 = 79$ Takten eine Zeitdauer von 39,5 µs für die in Abb. 52.122 dargestellte Zeitschleife. Soll die Zeitschleife genau 40 µs dauern, kann ein NOP-Befehl eingefügt werden. ◀

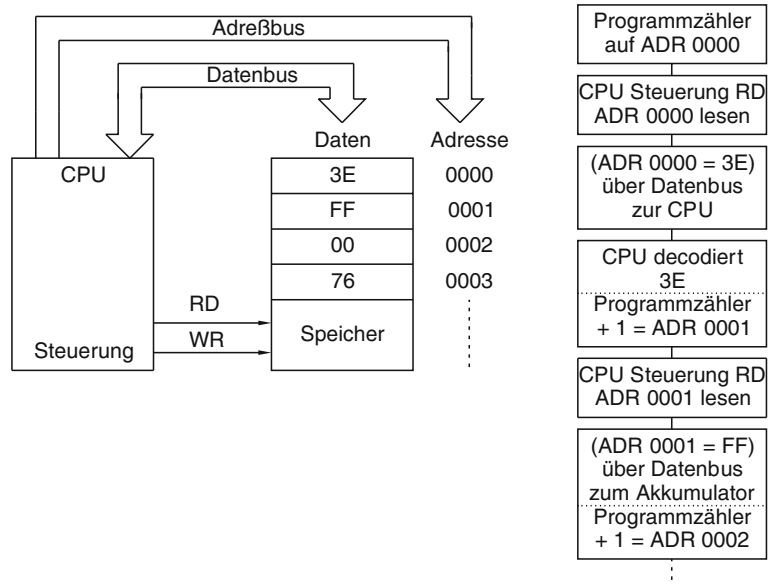
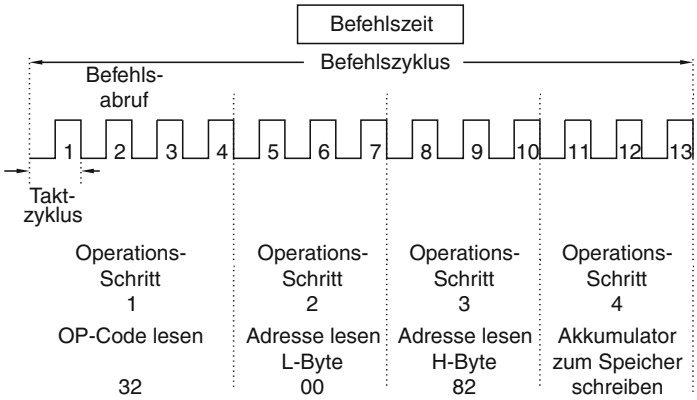


Abb. 52.120 Zusammenwirken von Speicher und CPU



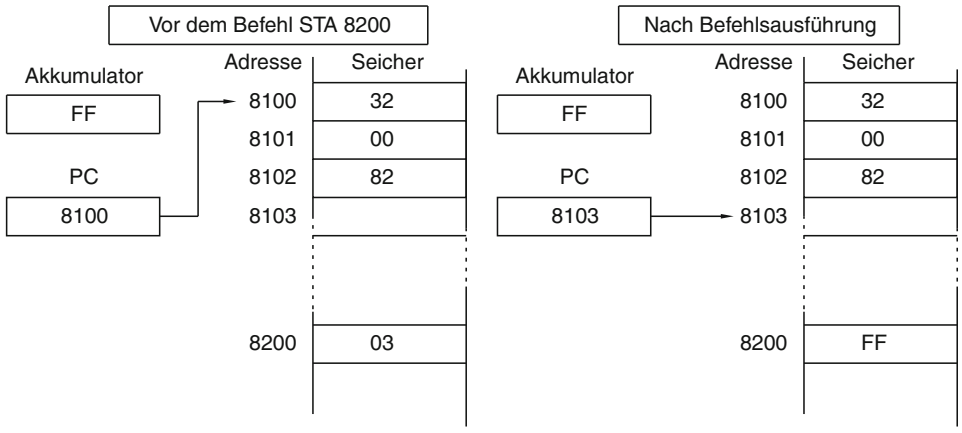
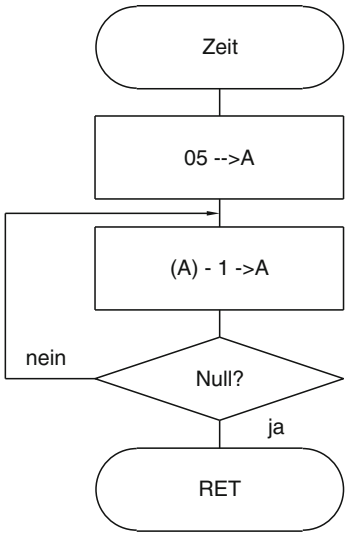


Abb. 52.121 Befehlszyklus



		Anzahl der Takte
Zeit:	MVI A,05	7 Takte
NEXT:	DCR A	5 Takte
	JNZ NEXT	10 Takte bei Rücksprung
	RET	sonst 7 Takte

Abb. 52.122 Zählschleife

52.7 Befehlsvorrat

Der Befehlsvorrat eines Prozessors enthält alle Befehle, die er ausführen kann. Die nachfolgende Tab. 52.32 gibt einen Auszug aus dem Befehlsvorrat des MC 8085 und die Anzahl der Operationszyklen an. Tab. 52.33 informiert über den Operationscode einiger Befehle.

52.8 Hinweise zur Programmierung und Progammbeispiele

Größere Programme werden in überschaubare Blöcke zerlegt und in Bibliotheken als Module gespeichert. Sie lassen sich dann für andere Programme nutzen. Beispiele für immer wieder benötigte Module sind Zählschleifen, Tabellenverarbeitung, Rechenprogramme, Interruptverarbeitung und Ein- und Ausgaben.

Programme werden mit einem Textverarbeitungsprogramm (Editor) geschrieben und anschließend assembliert. Der Assembler übersetzt die Mnemonik in den Maschinencode und überprüft das Programm auf Syntaxfehler. Weiterhin stellt das Assemblerprogramm Hilfen zur Programmierung bereit. Dazu gehören z. B. Zuweisungen von Marken (Labels) und Anweisungen wie ORG (Origin = Anfang) und EQU (Equate = gleichsetzen). EQU weist dem Namen der Marke einen Operanden zu, und mit ORG wird die Anfangsadresse des Programms festgelegt.

Beispiel 1

Für eine langsame Ausgabeeinheit wird eine Zeitschleife benötigt. Daher soll ein Unterprogramm mit dem Namen TIME programmiert werden. Im Hauptprogramm wird die Zeitdauer in das Registerpaar BC gebracht.

```
Unterprogamm: TIME
    ORG 0100
TIME  PUSH PSW ;AF zum Stack "retten"
    PUSH B   ;BC zum Stack {\glqq}retten{\grqq}
LOOP: DCX B    ;Registerpaar BC - 1 -> BC
    MOV A,B  ;Flagbeeinflussung

    ORA C
    JNZ LOOP ;Springe, wenn nicht Null, nach LOOP
    POP B    ;vom Stack zurückholen
    POP PSW  ;AF vom Stack zurückholen
 RET         ;Zurück zum Hauptprogramm
```

Eine Flagbeeinflussung muss hier vorgenommen werden, weil beim Inkrementieren eines Registerpaares keine Flags gesetzt werden, also auch nicht bei Null.

Adresse	OP-Code	Marke	Assembler	Bemerkung
0100	F5		PUSH PSW	(AF) → Stack
0101	C5		PUSH B	(BC) → Stack
0102	0B	LOOP:	DCX B	(BC) – 1 → BC
0103	78		MOV A,B	(B) → A
0104	B1		ORA C	(C) OR (A) → A
0105	C2 02 01		JNZ LOOP	(BC) = 0?
0108	C1		POP B	(Stack) → BC
0109	F1		POP PSW	(Stack) → AF
010A	C9		RET	Zum Hauptprogramm

◀

Tab. 52.32 Befehlsvorrat

Assembler	Bedeutung	Operationszyklen	Taktzyklen	Flags
Transportbefehle				
MOV D,S	Move Data; Übertrage Daten	1	4	–
MVI K	Load Data Immediate Into Register; Lade Daten unmittelbar in das Register	2	7	–
LDA ADR	Load Akku From Memory; Bringe den Inhalt einer Adresse zum Akku	4	13	–
STA ADR	Store Akkumulator From Memory; Speichere Akkuinhalt direkt zum Speicher			
LXI KK	Load 16 Bit Into Register Pair; Lade 16 Bit in ein Registerpaar			
Arithmetische und logische Befehle				
INR D	Increment Register; Inkrementiere Registerinhalt	1	4	Z,S,P,HC
DCR D	Decrement Register; Dekrementiere Registerinhalt	1	4	Z,S,P,HC
ADD S	Add Register; Addiere Registerinhalt zum Akkumulator	1	4	Z,S,P,HC
ADC S	Add Register With Carry; Addiere mit Übertrag	1	4	Z,S,P,HC
SUB S	Subtract Register; Subtrahiere Registerinhalt vom Akkumulator	1	4	Z,S,P,HC
CMP	Compare Register; Vergleiche Register mit Akkumulator (der Akkumulatorinhalt ändert sich hierbei nicht)	1	4	Z,S,P,HC
ADI K	Add Immediate; Addiere direkt zum Akkumulator	2	7	Z,S,P,Cy
ACI K	Add With Carry; Addiere mit Carry zum Akkumulator	2	7	Z,S,P,Cy
SUI K	Subtract Immediate; Subtrahiere direkt zum Akkumulator	2	7	Z,S,P,Cy
SBI K	Subtract With Borrow; Subtrahiere direkt vom Akkumulator mit Borger	2	7	Z,S,P,Cy
CPI K	Compare With Immediate Data; Vergleiche direkt	2	7	Z,S,P,Cy
CMA	Complement With Akkumulator; Komplementbildung	1	4	-STC
STC	Set Carry; Carryflag setzen	1	4	Cy
INX RP	Increment; Inkrementiere Registerpaar	1	6	–
DCX RP	Decrement; Dekrementiere Registerpaar	1	6	–
ANA S	AND Register; UND-Verknüpfung mit Akkumulator	1	4	Z,S,P,Cy
XRA S	EXCLUSIV-OR; XOR-Verknüpfung mit Akkumulator	1	4	Z,S,P,Cy
ORA S	ÔR Register; ODER-Verknüpfung mit Akkumulator	1	4	Z,S,P,Cy
ANI K	UND-Verknüpfung mit Konstante	2	7	Z,S,P,Cy
XRI K	XOR-Verknüpfung mit Konstante	2	7	Z,S,P,Cy
ORI K	ODER-Verknüpfung mit Konstante	2	7	Z,S,P,Cy
Rotationsbefehle				
RLC	Rotate Akku Left	1	4	Cy
RAL	Rotate Akku Left Through Carry	1	4	Cy
RRC	Rotate Akku Right	1	4	Cy
RAR	Rotate Akku Right Through Carry	1	4	Cy
Sprungbefehle				
JMP ADR	JUMP; direkter Sprung	3	10	–
JZ ADR	Jump On Zero; Springe, wenn Z-Flag gesetzt	3	10	–
JNZ ADR	Jump If Not Zero	3	10	–
JM ADR	Jump On Minus; Springe, wenn S-Flag gesetzt	3	10	–
JC ADR	Jump On Carry	3	10	–
JPO ADR	Jump On Parity Odd; Springe bei ungerader Parität	3	10	–
JPE ADR	Jump On Parity Even; Springe bei gerader Parität	3	10	–
Unterprogrammaufruf und Stapelspeicher				
CALL ADR	Unterprogrammaufruf	5	18	–
RET	Return, Rücksprung zum Hauptprogramm	3	10	–
PUSH B	BC zum Stack schreiben	3	12	–
PUSH D	DE zum Stack schreiben	3	12	–
PUSH H	HL zum Stack schreiben	3	12	–
PUSH PSW	AF zum Stack schreiben	3	12	–
POP B	BC vom Stack holen	3	10	–
POP D	DE vom Stack holen	3	10	–
POP H	HL vom Stack holen	3	10	–
POP PSW	AF vom Stack holen	3	10	–
E/A-Befehle				
IN	Eingabe zum Akkumulator	3	10	–
OUT	Ausgabe vom Akkumulator	3	10	–

Tab. 52.32 (Fortsetzung)

Assembler	Bedeutung	Operationszyklen	Taktzyklen	Flags
Interruptbefehle				
EI	Enable Interrupts; Interrupts freigeben	1	4	–
DI	Disable Interrupts; Interrupts sperren	1	4	–
SIM	Set Interrupt Mask; Interruptmaske setzen	1	4	–
RIM	Read Interrupt Mask; Interruptmaske lesen	1	4	–
Sonstige Befehle				
HLT	Halt	1	5	–
NOP	No Operation	1	4	–

Tab. 52.33 Befehle und Operationscode (Auswahl)

	A	B	C	D	E	H	L	M
MOV A,	7F	78	79	7a	7B	7C	7D	7E
MOV B,	47	40	41	42	43	44	45	46
MOV C,	4F	48	49	4A	4B	4C	4D	7E
MOV D;	57	50	51	52	53	54	55	56
MOV E	5F	58	59	5A	5B	5C	5D	5E
MOV H,	67	60	61	62	63	64	65	66
MOV L,	6F	68	69	6A	6B	6C	6D	6E
MOV M,	77	70	71	72	73	74	75	
MVI A,K	3E XX		INR A	3C		DCR A	3D	
MVI B,K	06 XX		INR B	04		DCR B	05	
MVI C,K	0E XX		INR C	0C		DCR C	0D	
MVI D,K	16 XX		INR D	15		DCR D	15	
MVI E,K	1E XX		INR E	1C		DCR E	1D	
MVI H,K	26 XX		INR H	24		DCR H	25	
MVI L,K	2E XX		INR L	2C		DCR L	2D	
STA ADR	32 XX XX		LDA ADR	3A XX XX				
JMP ADR	C3 XX XX		CALL ADR	CD XX XX				
JZ ADR	CA XX XX		RET	C9				
JNZ ADR	C2 XX XX							
JM ADR	FA XX XX		PUSH PSW	F5		POP PSW	F1	
			PUSH B	C5		POP B	C1	
			PUSH D	D5		POP D	D1	
			PUSH H	E5		POP H	E1	

Beispiel 2

Löschen eines Speicherbereiches

Die Anzahl der zu löschenden Speicherstellen steht im Register B (Anzahl) und die erste zu löschende Speicherstelle im HL-Register (Zeiger). Erhöht sich das HL-Register um 1, steht automatisch die nächste Speicherstelle im Registerpaar. Damit zeigt das HL-Register immer auf die gewünschte Speicherstelle. Nach jedem Löschvorgang wird der Zähler um 1 vermindert und ein erneuter Löschvorgang durchgeführt, bis der Zähler auf Null steht.

Marke	Assembler	Bemerkung
	PUSH PSW	(AF) → Stack
	PUSH H	(HL) → Stack
	SUB A	A = 0
CLEAR:	MOV M,A	1. Stelle löschen
	INX H	Zeiger + 1
	DCR B	Zähler − 1
	JNZ CLEAR	Ende?
	POP H	(Stack) → HL
	POP PSW	(Stack) → AF
	HLT	

◀

Computertechnik

53

Heribert Gierens



53.1 Komponenten eines Computers 730
53.2 Massenspeicher 730
53.3 Eingabegeräte 736
53.4 Ausgabegeräte 736

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_53

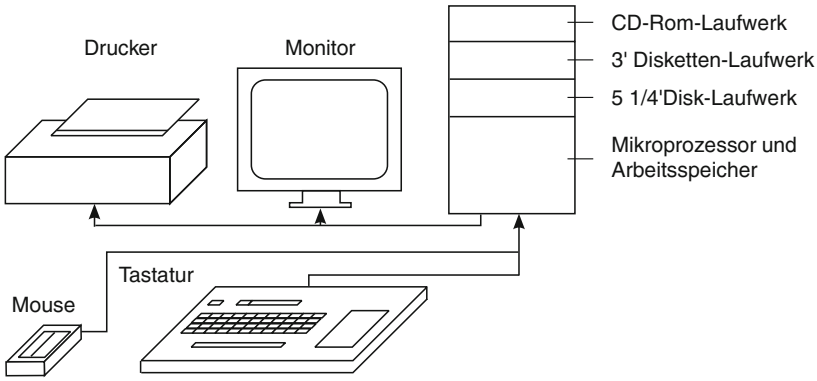


Abb. 53.1 Hardwarekomponenten eines Computers

53.1 Komponenten eines Computers

Ein Computer besteht immer aus der *Hard-* und der *Software*. Die Software lässt sich aufteilen in *Programme*, die die Zentraleinheit ausführt, und *Daten*, mit denen sie arbeitet. Zur Hardware gehört die *Zentraleinheit* und eine Anzahl *peripherer Geräte*. Von ihrer Leistungsfähigkeit in Verbindung mit dem Prozessortyp und der Größe des Arbeitsspeichers hängt die Leistungsfähigkeit des gesamten Computersystems ab. Periphere Geräte können nach ihrer Aufgabe eingeteilt werden in *Ein-* und *Ausgabegeräte* wie Tastatur, Monitor, Drucker und Plotter und in *periphere Speicher* wie Magnetplattenspeicher und optische Speicher. Die Anzahl und Art der Peripheriegeräte hängt von der Aufgabe und vom Einsatzgebiet des Computers ab.

In einem Personalcomputer (PC) (Abb. 53.1) ist die Zentraleinheit mit der Speichereinheit in einem Tisch- oder Towergehäuse untergebracht. Als Speicher für das Betriebssystem und für Anwenderprogramme dient eine *Festplatte* oder ein *CD-ROM-Laufwerk*. Zur Datensicherung und Datenverwaltung werden 3,5″- und 5,25″-Laufwerke eingesetzt. Der angeschlossene *Drucker* wird für die Text- oder Grafikausgaben benutzt. *Tastatur* und *Monitor* ermöglichen die Kommunikation zwischen Mensch und Maschine. Die heute überwiegend grafisch orientierten Anwenderprogramme vereinfachen ihre Bedienung mit einer *Maus*.

Prozessrechner besitzen eine zusätzliche Prozessperipherie für digitale und analoge Eingangs- und Ausgangssignale.

53.2 Massenspeicher

Massenspeicher lassen sich nach ihrer Funktion in *Magnetplatten*, *optische Speichermedien* und *Halbleiterspeicher* einordnen.

53.2.1 Magnetplatten

53.2.1.1 Diskette und Diskettenlaufwerk

Die Diskette (Floppy-Disk) ist ein Speichermedium, auf der Daten gespeichert, gelöscht und wieder überschrieben werden können. Sie besteht aus einer etwa 2 µm dicken Kunststoffträgerfolie, auf der beidseitig eine magnetisierbare Schicht (z. B. Eisenoxidschicht) aufgebracht ist. Man unterscheidet Maxi- (8″), Mini-(5,25″) und Mikro-Disketten (3,5″), wobei die beiden letzten sich bei PC's durchgesetzt haben. Die Größenangabe bezieht sich auf den Durchmesser der Diskettenscheibe, die fest in einer rechteckförmigen Kunststoffhülle mit ausgekleideten Vliesinnenseiten eingeschlossen ist. Dort ist sie gegen kleinste Verschmutzungen (Staub, Fingerabdrücke), die zum Datenverlust führen können, geschützt. Da Disketten unter dem Einfluss magnetischer Felder ebenfalls ihre Daten verlieren, dürfen sie solchen Feldern nicht ausgesetzt werden. Einmal beschriebene Disketten können vor einem erneuten Beschreiben, und damit vor einem eventuellen Datenverlust, dadurch geschützt werden, dass die Schreibschutzkerbe bei den 5,25″-Disketten mit einem lichtundurchlässigen Klebestreifen versehen wird. Die 3,5″-Disketten besitzen einen Schieber, mit dem man die Schreibschutzöffnung schließen kann (Abb. 53.2).

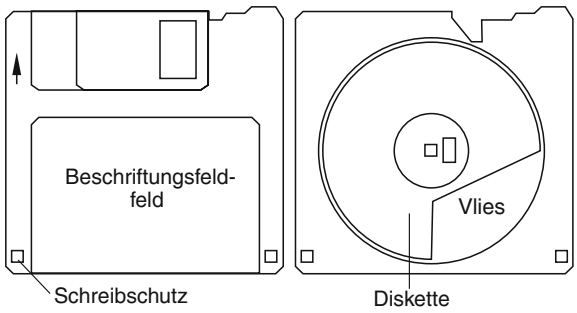


Abb. 53.2 3,5″-Diskette

Disketten erhalten erst ihre Funktion, wenn sie vom Rechner durch ein Programm formatiert wurden. Durch die *Formatierung* wird die Diskette in konzentrische Spuren (Tracks) und radiale Sektoren aufgeteilt (Abb. 53.3). Beispielsweise kann eine Diskettenseite 80 Spuren mit 18 Sektoren besitzen.

Auf den Spuren der einzelnen Sektoren befinden sich nach einem Speichervorgang die entsprechenden Daten, die dort in Datenblöcken zusammengefasst werden. Außer dem Datenblock enthält ein Sektor eine Sektorerkennung, eine Erkennungs- und

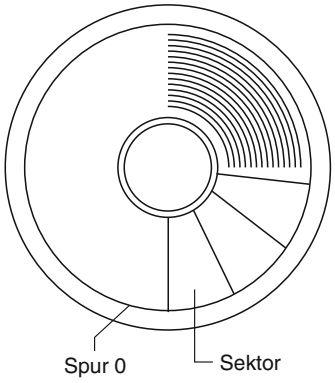


Abb. 53.3 Diskettenscheibe

eine Datenblocklücke. In der Sektorerkennung befinden sich Sektor- und Spurnummer (Abb. 53.4). Liegt in Spur 0 ein Verzeichnis der Datennamen mit Zuordnung von Sektoradresse und Spurnummer, kann die gewünschte Information schnell ausgelesen werden. Die Speicherkapazität einer Diskette ist von der Formatierung abhängig. Beispielsweise ergibt sich bei der oben genannten Diskette bei 512 Byte/Sektor und zwei Aufzeichnungsseiten eine Kapazität von 1440 kB.

Zur Kennzeichnung des Sektors 0 dient ein *Indexloch* oder ein Erkennungsbyte. Die sich ergebenen Datenblöcke sind fortlaufend nummeriert. Auf jeder Spur innerhalb eines Sektors kann eine bestimmte Anzahl Daten, z. B. 256 oder 512 Byte gespeichert werden. Dies ist jeweils abhängig von der Spurdichte. Je mehr Spuren sich auf der Diskette befinden, desto mehr Daten kann sie aufnehmen. Die Spurdichte wird mit TPI (Tracks Per Inch = Spuren pro Zoll) angegeben und die *Aufzeichnungsdichte* in BPI (Bits per Inch). Man spricht auch von FCI (Flux Changes per Inch = Flusswechsel pro Zoll). Da die inneren Spuren einen geringeren Umfang besitzen, sind sie **die Basis** für die Anzahl der Daten. Die mittleren Zugriffszeiten bei Disketten betragen 100 bis 200 ms. Die übliche Speicherkapazität beträgt heute 1,44 MB. Standardlaufwerke verwenden beide Seiten der Diskette und besitzen daher zwei Schreib- und Leseköpfe (Kombiköpfe). Ist ein PC mit mehreren Laufwerken ausgestattet, hat das Erstlaufwerk den Namen A und das Zweitlaufwerk den Namen B. Der Name C ist in der Regel der Festplatte vorbehalten. Das Prinzip der Aufzeichnung und das Lesen von Daten zeigt Abb. 53.5. Beim Speichern der Daten erzeugt ein Schreibstrom im Schreib- und Lesemagnetkopf ein Magnetfeld, das im Luftspaltbereich des Kopfes auf der Diskette eine Magnetisierung bewirkt. Beim Lesen entsteht in der Kopfwicklung durch magnetische Induktion eine Spannung, die von der Magnetisierung (Ausrichtung der Molekularmagnete) der Diskette abhängt.

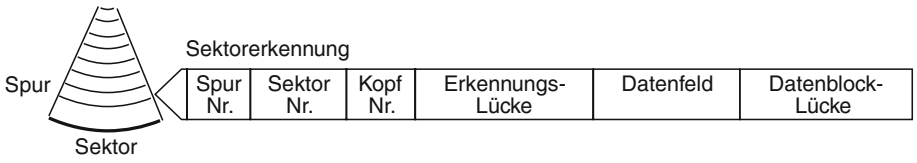


Abb. 53.4 Spurformat

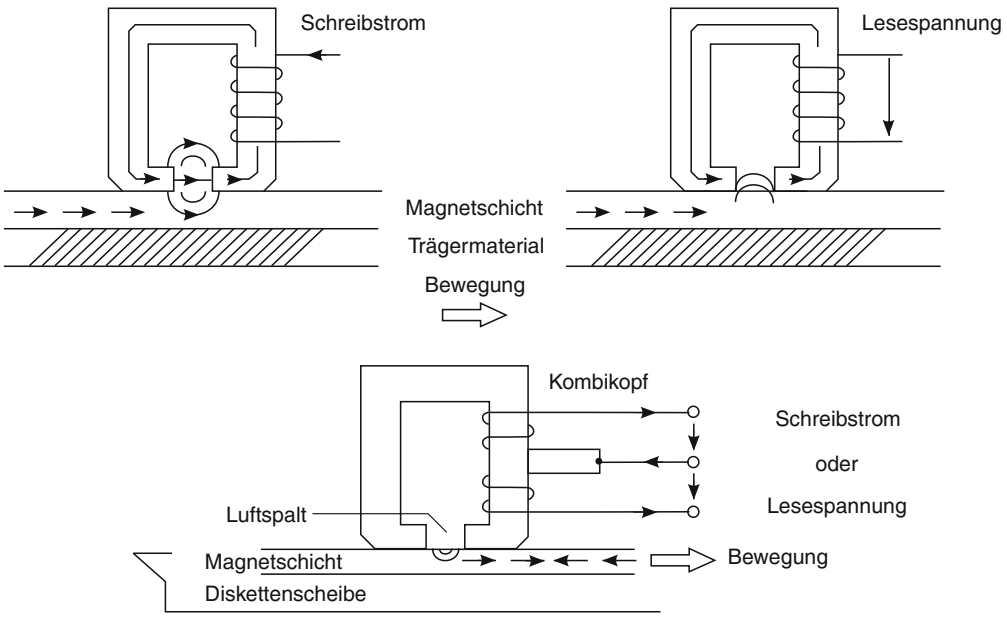


Abb. 53.5 Schreib-Lesekopf

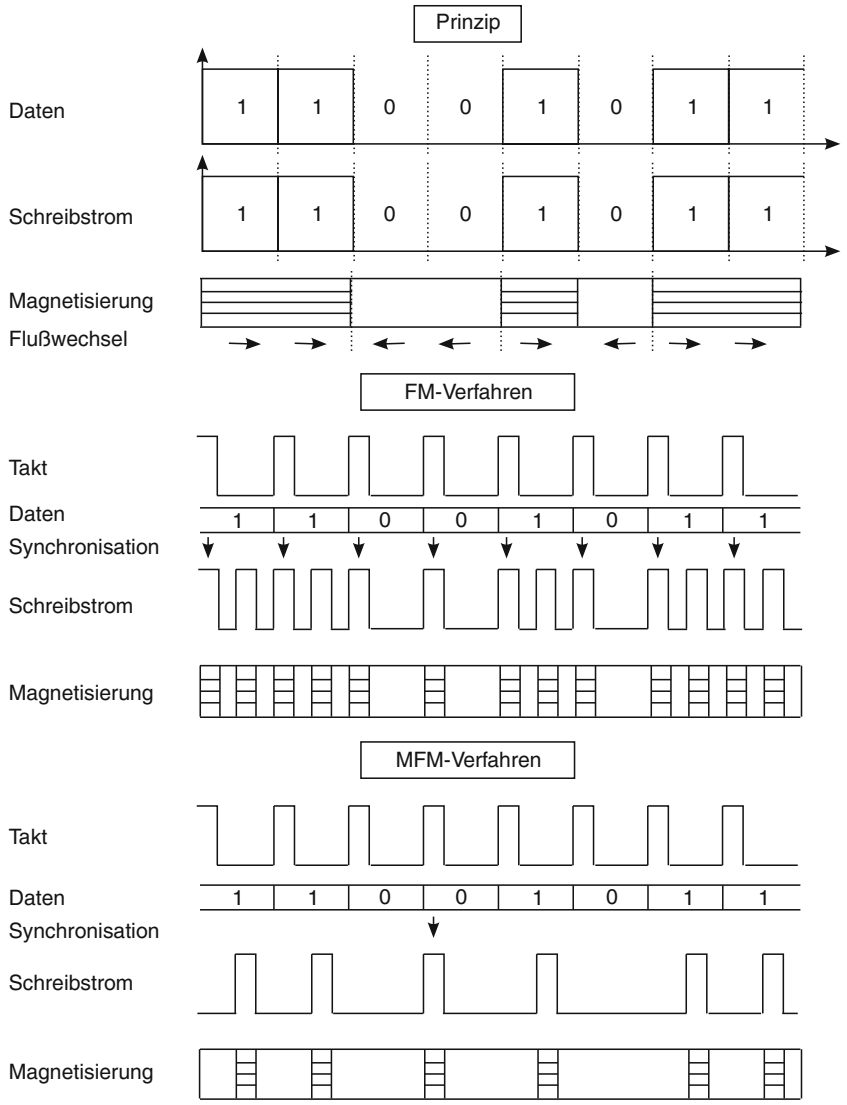


Abb. 53.6 Codierungsverfahren

Abb. 53.6 verdeutlicht den Zusammenhang zwischen Daten, Schreibstrom und Magnetisierung und zeigt mögliche Codierungsverfahren auf. Beim *FM-Verfahren* (Frequency Modulation = Frequenzmodulationsverfahren) werden den Datenbits Synchronisationsbits zugefügt. Ein Datenbit „1“ erhält hierbei die Codierung „1.1“ und ein Datenbit „0“ die Codierung „1.0“. Bei *MFM-Verfahren* (Modified FM) wird nur ein Synchronisationsbit auf die Diskette gebracht, wenn eine „0“ als Datenbit vorausging. Weil dieses Verfahren weniger Flusswechsel benötigt, erhöht sich die mögliche Speicherkapazität.

Das Diskettenlaufwerk wird über den Floppy-Disk-Controller von der CPU angesteuert. Der Antriebsmotor lässt die Diskette 300-mal in der Sekunde rotieren und die Schrittmotorsteuerung bringt die Köpfe in die richtige Position auf der Diskette. Ferner erhält der FDC Signale für die Spur 0- und Schreibschutzerkennung.

Beim Schreibvorgang wandelt der Controller die eingehenden zu speichernden parallelen Daten in einen seriellen Datenstrom um und steuert über einen Schreibverstärker den Schreib-Lesemagnetkopf an. Abb. 53.7 zeigt das Prinzip eines Diskettenlaufwerks mit Controller.

Beim Lesen von der Diskette wird jeder Flusswechsel vom Lesekopf aufgenommen und als Wechselspannung einem nachfolgenden Schreibverstärker zugeführt. Hochfrequente Störspannungen werden vorher über ein Filter (Tiefpass) ausgesiebt. Über eine Differenzierstufe und nachfolgenden Komparator wird das Lesesignal in eine gewünschte digitale Form gebracht, um hier je nach Codierungsverfahren dem FDC die decodierten Daten zur Verfügung zu stellen (Abb. 53.8). Dort wird der serielle Datenstrom dann wieder in das parallele Format gebracht. Die entsprechenden Signale der einzelnen Blöcke sind mit dargestellt.

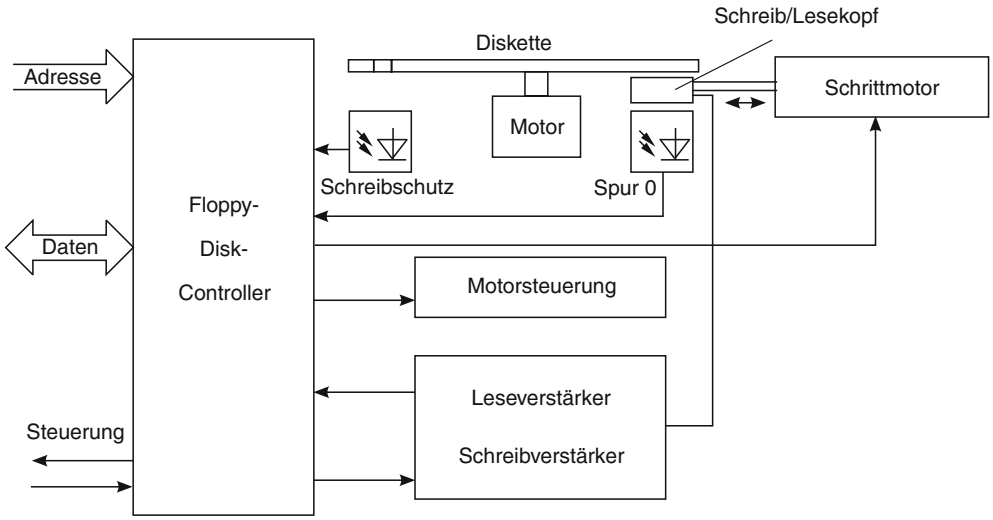


Abb. 53.7 Diskettenlaufwerk mit Cotroller

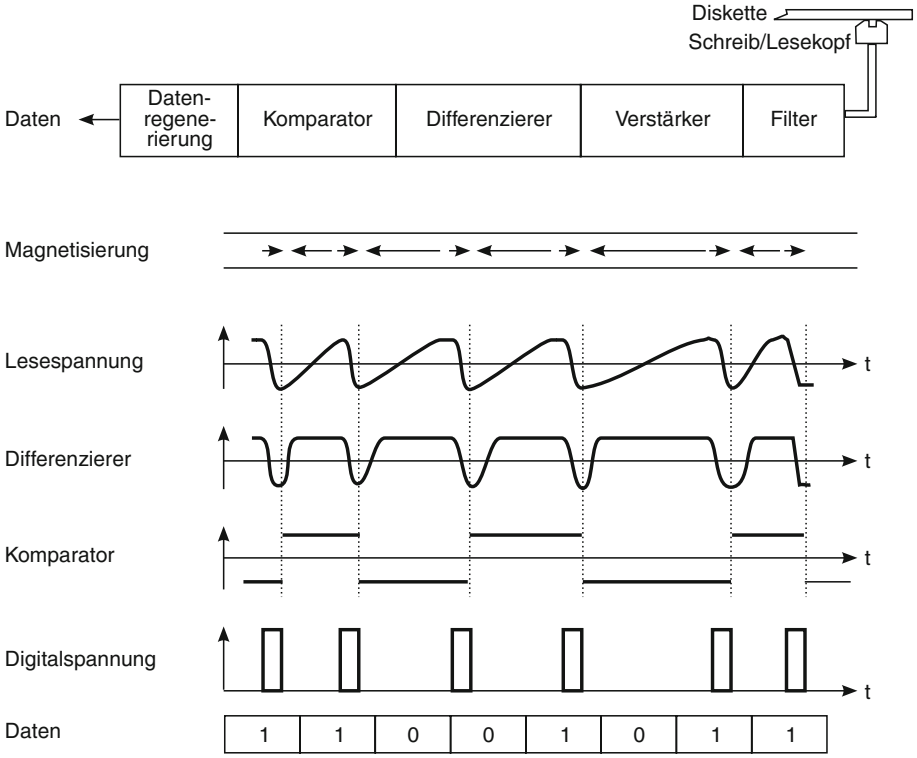


Abb. 53.8 Aufbereitung von Diskettendaten

Abb. 53.9 zeigt beispielhaft einen *Floppy-Disk-Controller*, in dem sich alle notwendigen Funktionseinheiten befinden, mit seiner 34poligen Steckerleiste und dem Anschluss an das Bussystem.

Der Baustein selbst wird wie ein E/A-Baustein über eine Adresse angesprochen. Vorgänger des 82077AA sind die Controller 8272 und NEC765A.

53.2.1.2 Festplatte und Festplattenlaufwerk

Die Funktion der Festplatte beruht wie bei den Disketten auf der Magnetisierung einer Oberfläche. Das Trägermaterial von Festplatten ist Aluminium, auf der eine sehr dünne magnetisierbare Schicht (z. B. Eisenoxid) aufgebracht ist. Um die Speicherkapazität zu erhöhen, sind mehrere Platten übereinander (Abb. 53.10) auf einer gemeinsamen Achse angebracht. Die

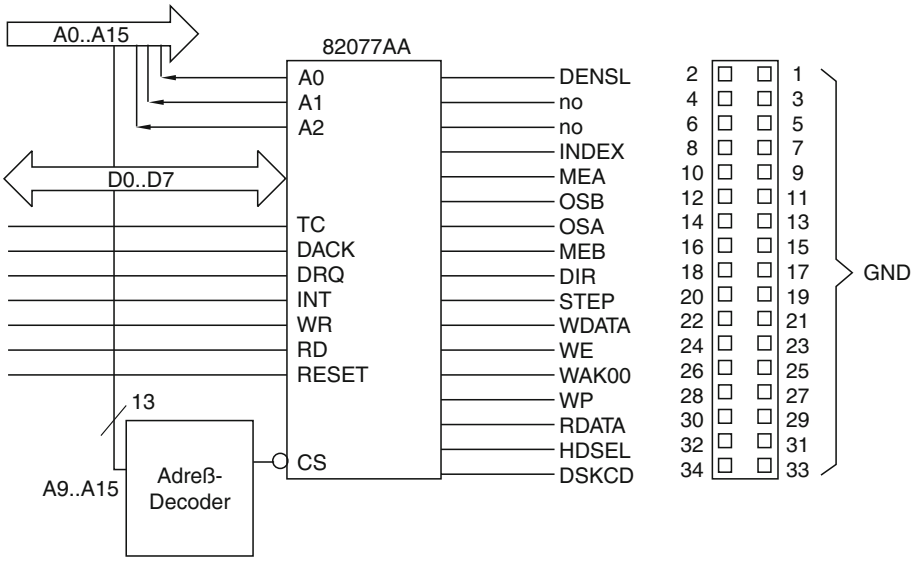


Abb. 53.9 Floppy-Disk-Controller mit Anschlüssen

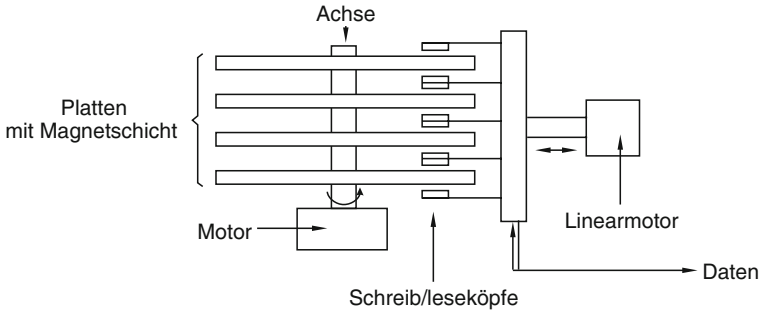


Abb. 53.10 Prinzip des Festplattenlaufwerks

Platte selbst ist – wie bei den Disketten auch – in konzentrische Spuren und radiale Sektoren aufgeteilt.

Die Platten rotieren im Gegensatz zu den Disketten ständig mit Geschwindigkeiten ab 5200 Umdrehungen pro Minute. Da sich bei den hohen Rotationsgeschwindigkeiten zwischen Plattenoberfläche und Kopf ein Luftpolster bildet, können sich Kopf und Platte nie berühren. Der Kopf „fliegt“ somit über die Platte. Zum Starten und Landen des Kopfes dient eine nicht zum Speicherbereich gehörende Randzone, die sogenannte *landing zone*. Beim Ausschalten des Rechners erfolgt das korrekte Landen („Parken“) automatisch.

Festplatten sind stoßempfindlich. Ein Berühren der Platte durch die Köpfe (Headcrash) führt zu deren Zerstörung.

Der Vorteil der Festplatten gegenüber Diskettenlaufwerken liegt in der Geschwindigkeit (mittlere Zugriffszeit 10 ms) und in der wesentlich höheren Speicherkapazität (2 GB und mehr). Unter Zugriffszeit versteht man die Zeit, die der Schreib-Lesekopf benötigt, um zur gewünschten Spur (Einstellzeit) und dem gewünschten Sektor (Latenzzeit) zu kommen.

Durch die Formatierung werden Datenspuren in konzentrischen Kreisen eingerichtet und festgelegt, wo Sektoren und Zylinder auf der Platte liegen. Je dichter die Spuren nebeneinander verlaufen, desto größer ist die mögliche Speicherkapazität. Die genau übereinander liegenden Spuren eines Plattenstapels nennt man Zylinder. Die Zylinder sind jeweils wieder in kleinere Speichereinheiten (Sektoren) aufgeteilt. Die Anzahl der Sektoren ist auf allen Spuren gleich. Die größte Datendichte liegt somit auf der innersten Spur. Da auf der äußeren Spur die gleiche Dateninformation liegt, ist die Datendichte hier geringer. Ausnahmen bilden Platten mit *Zone-Bit-Recording*, wo die inneren Spuren weniger Sektoren als die äußeren enthalten und somit die Platzverschwendung auf den äußeren Sektoren entfällt. Der Controller speichert zusätzlich Kontrolldaten, um die Position der Magnetköpfe zu bestimmen.

Bei den Vorgängertypen war die Geschwindigkeit der Festplatten größer als der mögliche Rechnerzugriff. Die Spur eines Sektors konnte nicht bei jeder Umdrehung der Festplatte, sondern erst bei der zweiten oder dritten Umdrehung gelesen werden. Ein *Interleave-Faktor* passte die Umdrehungsgeschwindigkeit der Festplatte an die Geschwindigkeit des Prozessors bzw. an die Geschwindigkeit an, mit welcher der Controller Daten von oder zur Festplatte übertragen kann. Abb. 53.11 vergleicht die Sektorenfolge der Festplatte mit einem Interleave-Faktor 3 und eine mit dem Interleave-Faktor 1.

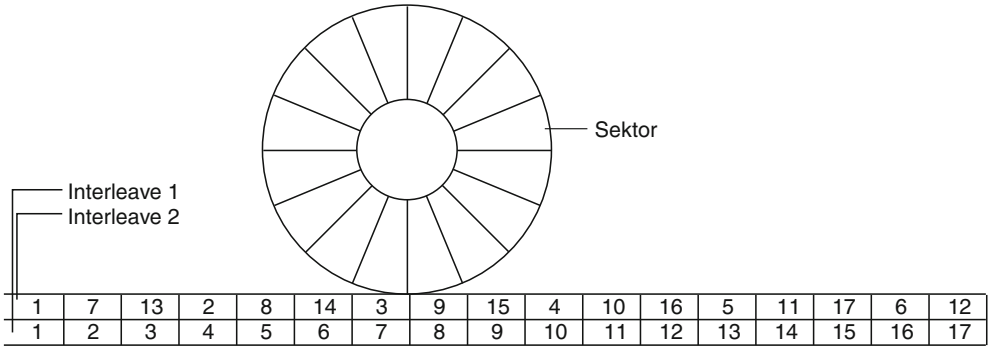


Abb. 53.11 Interleave-Faktor

Tab. 53.1 Standard CD

Durchmesser	120 mm
Drehzahl	200–530 min^{-1}
Abtastrichtung	Von innen nach außen
Abtastgeschwindigkeit	Konstant
Spurbreite	0,6 µm
Abstand benachbarter Spurabschnitte	1 µm

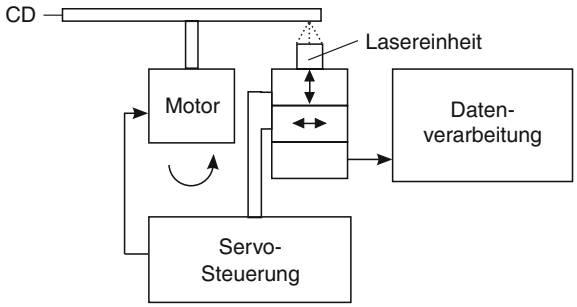


Abb. 53.12 Prinzip eines CD-Laufwerkes

Die Festplatte selbst kann in mehrere logische Laufwerke unterteilt (partioniert) werden, die dann vom Betriebssystem jeweils als eigenständige physikalische Laufwerke behandelt werden. *Wechselplatten* sind nicht fest installiert. Sie können problemlos ausgewechselt und transportiert werden.

53.2.1.3 Magnetbandgeräte

Magnetbandgeräte (Streamer) dienen zur Aufnahme und Sicherung großer Datenmengen. Die Aufzeichnung der Daten erfolgt durch die Magnetisierung einer dünnen Eisenoxidschicht, die sich auf einer Trägefolie befindet. Ein unmittelbarer freier Zugriff auf bestimmte Daten ist nicht möglich.

53.2.2 CD-ROM- und CD-Laufwerk

Das CD-Laufwerk (CD = Compact Disc) gehört heute zur Grundausstattung eines PC-Systems. Die CD zählt zu den optischen Speichern, bei denen zwischen CD-WROM (Write Once Read Multiple, kann einmal beschrieben werden) und CD-ROM unterschieden wird. Ihr Vorteil gegenüber Festplatten liegt in der sehr hohen Speicherkapazität (über 500 MB) und den relativ kurzen Zugriffszeiten (unter 1 ms). Die Datenspur liegt *spiralförmig* auf der CD vor und wird von innen nach außen gelesen. Die Daten der Standard CD sind in Tab. 53.1 zusammengefasst.

Abb. 53.12 zeigt den prinzipiellen Aufbau eines CD-Laufwerkes. Die Motordrehzahl ist zwischen der Innenabtastung und der Außenabtastung variabel, weil beim Lesen der CD ein konstanter Datenstrom erzeugt werden muss. Gegenüber Single-Speed-Laufwerken mit Datenübertragungsraten von 150 kB pro Sekunde ist der Faktor heutiger Laufwerke 20 und mehr. Die

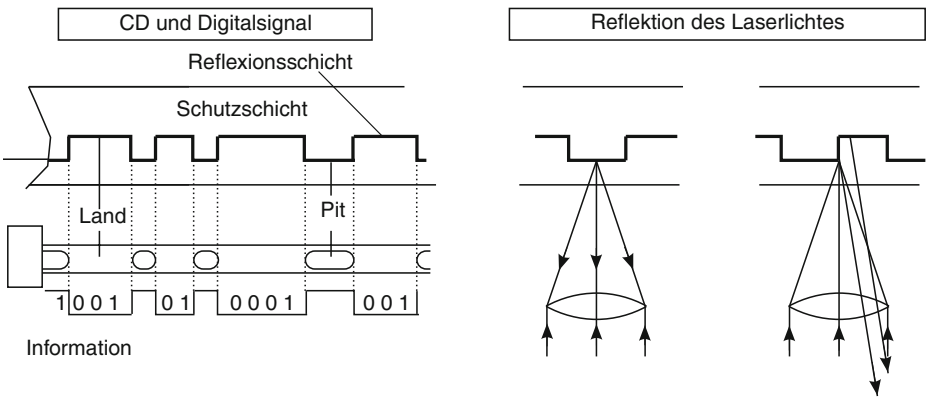


Abb. 53.13 Signale und Reflexion des Laserlichtes

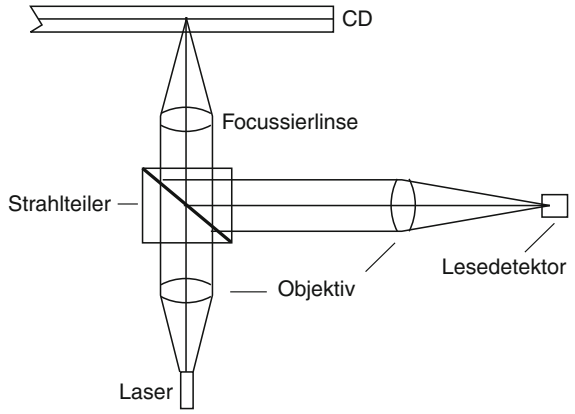


Abb. 53.14 Leseprinzip

Servosteuerung sorgt für eine gezielte Führung der Abtasteinheit (Focus und Spurhaltung) über die Datenspur.

Die *digitalen* Informationen auf der CD sind durch Vertiefungen auf der spiralförmig angeordneten Spur eingeprägt. Die Vertiefungen heißen *Pits* und die Stege dazwischen bezeichnet man mit *Land*. Der Wechsel zwischen Land und Pit wird beim Lesen der Daten von innen nach außen als „1“ interpretiert (Abb. 53.13). Hierbei wird die Spur durch den gebündelten Laserstrahl abgetastet. Trifft der Laserstrahl auf ein Pit, erfolgt eine Reflektion, die im Idealfall zur Absorbtion des Lichtes führt. Trifft er auf keine Vertiefung, dann wird der Strahl voll reflektiert.

Abb. 53.14 verdeutlicht das Leseprinzip. Eine Laserdiode sendet das Laserlicht durch einen halbdurchlässigen Spiegel, einen Strahlteiler und eine Focussierlinse auf die Reflexionsschicht der CD. Das reflektierte Licht gelangt über den Strahlteiler auf den Lesedetektor. Der Lesedetektor besteht aus Fotodioden, die aus dem gewonnenen elektrischen Signal die Daten zur Verarbeitung weitergeben und Signale für die Servosteuerung gewinnen.

53.3 Eingabegeräte

53.3.1 Tastatur

Zur Eingabe von Zeichen und Befehlen wird in PC-Systemen die Tastatur (Keyboard) verwendet. Bei den Tastenelementen selbst sind unterschiedliche Tastenarten und Anordnungen gebräuchlich. Die deutsche Tastatur hat die sogenannte *QWERTZ-Anordnung* mit Funktions- und Cursortasten und Nummernblock. Bei der internationalen Anordnung sind Z mit Y vertauscht (QWERTY-Anordnung).

Die Tastatur besteht aus einer Tastaturmatrix, einem Mikrokontroller und einem Puffer-Speicher für die seriellen Daten, die der Tastaturschnittstelle (Tastatur-Interface) auf dem Motherboard zur Verfügung gestellt werden. Abb. 53.15 zeigt das Blockschaltbild.

Die Funktion der Tastenmatrix ist aus Abb. 53.16 ersichtlich. Die Tastatur erzeugt aus einem Tastendruck ein Bitmuster (Scan-Code = Zuordnung zur Tastennummer).

Der Tastatur-Controller (Beispiel 8048) wertet hierbei die Eingabe der Tastatur aus. Unterschieden wird zwischen dem Drücken der Taste („Make“-Codes) und dem Loslassen der Taste („Break“-Codes). Die beiden Codes unterscheiden sich nur durch das höchstwertigte Bit. Wird eine Taste gedrückt und nach bestimmter Zeit (z. B. 0,5 s) nicht losgelassen, sendet der Tastatur-Controller fortlaufend den gleichen Tastencode (Autorepeat-Funktion). Mit Hilfe des erzeugten Taktes durch den Mikrocontroller werden die Daten *seriell* zum Tastatur-Interface geschrieben. Dort werden die seriellen Daten von einem weiteren Mikrocontroller (z. B. 8042) empfangen und in parallele Daten umgesetzt, die dann der CPU zur Verfügung gestellt werden. Die Zuordnung der Scan-Codes zu den ASCII-Werten kann mit Hilfe eines Tastaturtreibers, z. B. KEYBGR.COM, selbst definiert werden. Die Übertragung zur CPU wird durch INT (Interrupt 1) ausgelöst.

53.3.2 Maus

Text- und Grafik-Software verwenden einen *Cursor* (z. B. kurze blinkende Linie oder Lichtmarke verschiedener Form), um an die Stelle des Bildschirms zu zeigen, an der das nachfolgende Zeichen entsteht. Der Cursor kann sowohl mit Hilfe der Cursortasten als auch mit der Maus bewegt werden. Die Maus-Software bewegt den Cursor auf dem Bildschirm genau in die Richtung, in der die Maus bewegt wird. Hierbei wird die Maus zu einer wichtigen Eingabeeinheit. Installiert wird sie durch die entsprechenden Maustreiber. Auf dem Markt existieren unterschiedliche Maustypen, z. B. mit zwei oder drei Tasten. Bei beiden ist wesentliches Element die Rollkugel. *X*- und *Y*-Richtung und Status der Maustasten wird an die serielle Schnittstelle weitergegeben.

53.4 Ausgabegeräte

53.4.1 Datensichtgeräte

Datensichtgeräte dienen zur Wiedergabe alphanumerischer und grafischer Informationen.

53.4.1.1 Monitor

Nach Abb. 53.17 besteht der Monitor aus einer Bildröhre und aus Schaltungen für die Hochspannungserzeugung, Strahlstromablenkung (Ablenkeinheit), Synchronisierung und Signalverstärkung (Videoverstärker). Angesteuert wird der Monitor vom Videointerface (Grafikkarte) mit dem *Videosignal* und der *Synchronisation*. Das Videosignal ist das eigentliche Bildsignal und die Synchronisationssignale dienen dazu, ein stehendes Bild auf dem Monitor zu erzeugen.

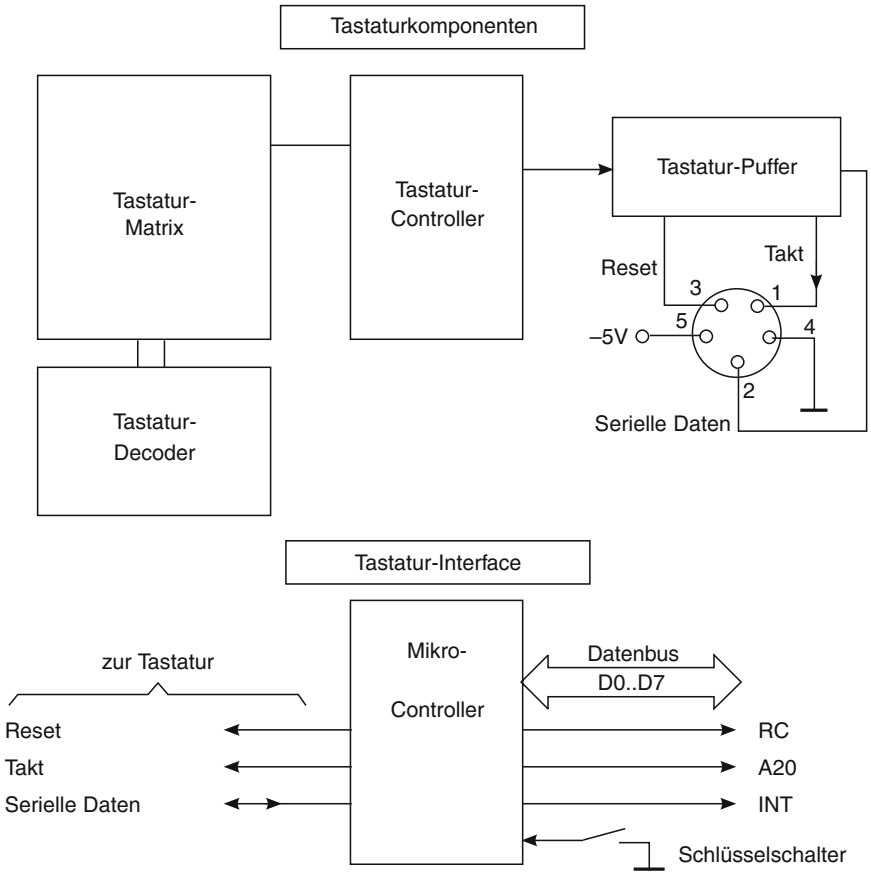


Abb. 53.15 Tastaturinterface und -komponenten

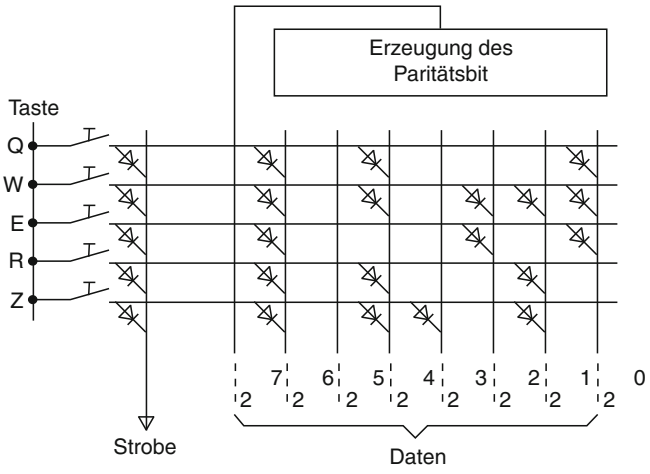


Abb. 53.16 Prinzip: Tastaturmatrix

Würde der Einsetzpunkt des Bildes nicht immer an derselben Stelle liegen, hat man den Eindruck, das Bild würde „durchlaufen“. Die zur Monitoransteuerung notwendigen Daten befinden sich im *Bildschirmspeicher*, der Teil des Arbeitsspeichers ist oder im *Video-RAM* auf der *Grafikkarte*. Die Bildröhre selbst (Kathodenstrahlröhre) besteht aus einem Strahlerzeugersystem (Katode) und dem Bildschirm. Die von der Kathode austretenden Elektronen treffen durch Hochspannung beschleunigt auf dem Bildschirm auf und erzeugen hier einen sichtbaren Punkt, dessen Farbe bei *monochromen* (einfarbigen) Bildschirmen von der Leuchtschicht (weiß, grün, bernstein) des Bildschirmes abhängt. Zur Helligkeitseinstellung dient ein Einsteller, der von außen zugänglich ist und den Strahlstrom durch Potentialänderung zwischen Kathode und Steuergitter (Wehneltzylinder) beeinflusst. Das Steuergitter ist die Elektrode in der Bildröhre, die sich unmittelbar hinter der Katode befindet. Um ein einwandfrei scharfes Bild zu erhalten, muss der Elektronenstrahl gebündelt den Bildschirm erreichen. Dies wird durch ein elektrostatisches Feld erreicht, das wie eine Elektronenoptik wirkt. Eine Einstellmöglichkeit bietet der FOCUS-(Schärfe-) Einsteller.

Da der Elektronenstrahl das Bild zeilenweise erstellt, wird eine Ablenkschaltung benötigt, die den Elektronenstrahl durch Magnetfelder entsprechend ablenkt. Hierzu dienen die Horizontal- und Vertikalablenkspulenpaare (Ablenkeinheit).

Um eine gleichmäßige Ablenkung von links nach rechts zu erhalten, wird den Horizontalablenkspulen ein sägezahnförmiger Strom zugeführt. Während des Rücklaufs (von rechts nach links) wird der Elektronenstrahl dunkel getastet. Die Anzahl der horizontalen Abtastungen pro Sekunde ergibt die *Zeilen-* oder *Horizontalfrequenz*, die bei den einzelnen Herstellern un-

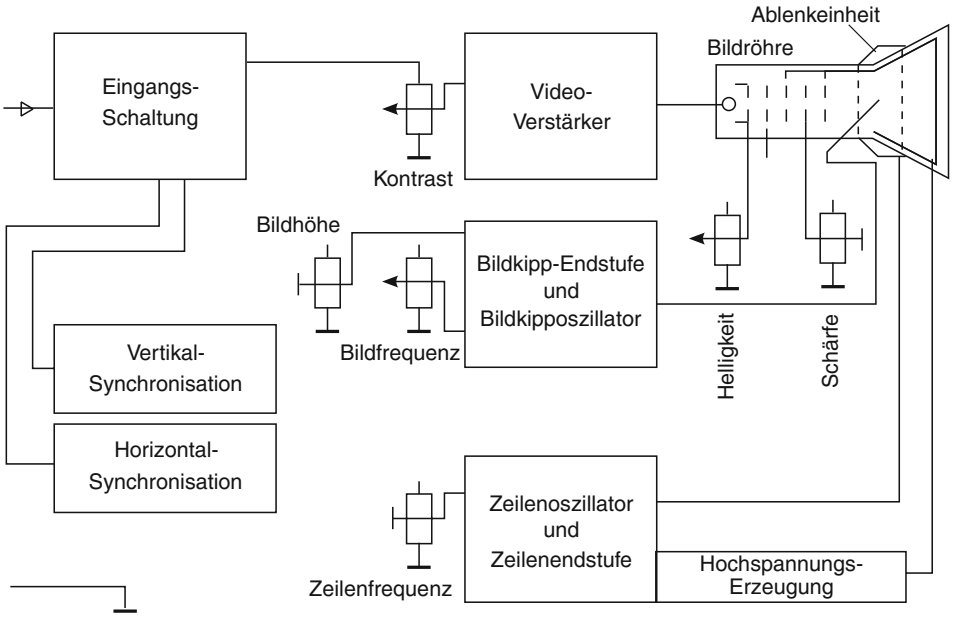


Abb. 53.17 Blockbild eines Monitors

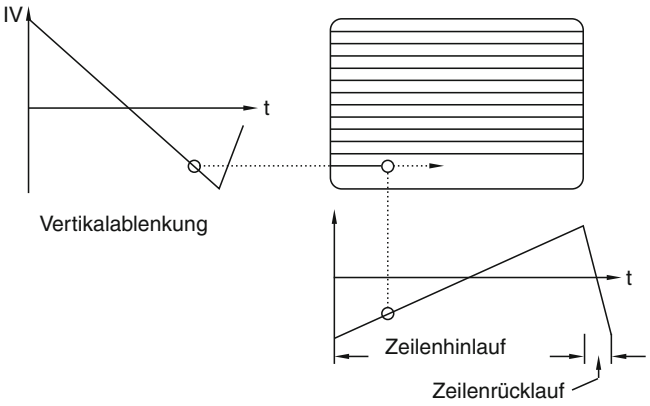


Abb. 53.18 Vertikale und horizontale Ablenkung

terschiedlich ist (z. B. 30 … 80 kHz bei 14″-Monitoren). Die lineare Ablenkung des Elektronenstrahls von oben nach unten bewirken die Vertikalablenkspulen. Abb. 53.18 zeigt den Zusammenhang zwischen vertikaler und horizontaler Ablenkung. Die Amplitude der Bildablenkspannung ist für die Bildhöhe verantwortlich. Häufig ist hierfür ein Einsteller von außen zugänglich.

Zur Darstellung eines flimmerfreien Bildes müssen bei der *Negativdarstellung* (helle Zeichen auf dunklem Grund) mindestens 50 Bilder pro Sekunde geschrieben werden. Dies ergibt eine Bildwiederholfrequenz von 50 Hz. Die Positivdarstellung (dunkle Zeichen auf hellem Grund) erfordert mindestens 70 Hz. Die meisten Monitore arbeiten heute mit Bildfrequenzen zwischen 60 und 120 Hz. Eine Möglichkeit, Flimmern zu reduzieren, ist das Verschachteln zweier Halbbilder nach dem Zeilensprungverfahren (interlaced), wie das bei Fernsehgeräten

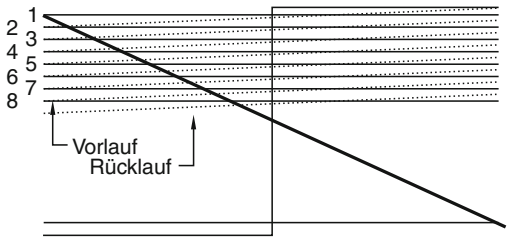


Abb. 53.19 Zeilensprungverfahren

der Fall ist. Hierbei werden zunächst die ungeraden und dann die geraden Zeilen geschrieben (Abb. 53.19). Der Rücklauf ist immer unsichtbar. Bei 25 Halbbildern in der Sekunde wird beispielsweise eine Bildfrequenz von 50 Hz vorgetäuscht (Trägheit der Augen).

Der Farbmonitor besitzt eine *in-line*-Röhre mit drei Elektronenerzeugungssystemen jeweils für die Grundfarben Rot (R), Grün (G) und Blau (B), mit der die entsprechenden Farbstreifen angeregt werden (Abb. 53.20).

Um eine gute Farbwiedergabe zu erhalten, müssen die Elektronenstrahlen ihre zugeordneten Farbstreifen durch den gemeinsamen Schnittpunkt der Schlitzmaske treffen (Konvergenz). Mit Hilfe von Dauermagneten, die um den Röhrenhals angeordnet sind, kann eine eventuell notwendige Nachjustierung erfolgen.

Bei der *Trinitron*-Farbbildröhre besteht die Gittermaske aus senkrecht angeordneten Metallstreifen. Zur Stabilität können zusätzlich zwei Streifen waagerecht eingezogen sein. Das Elektronenerzeugersystem besteht aus drei Kathoden. Im Gegensatz zur In-line-Röhre sind alle anderen Elektroden nur einmal vorhanden.

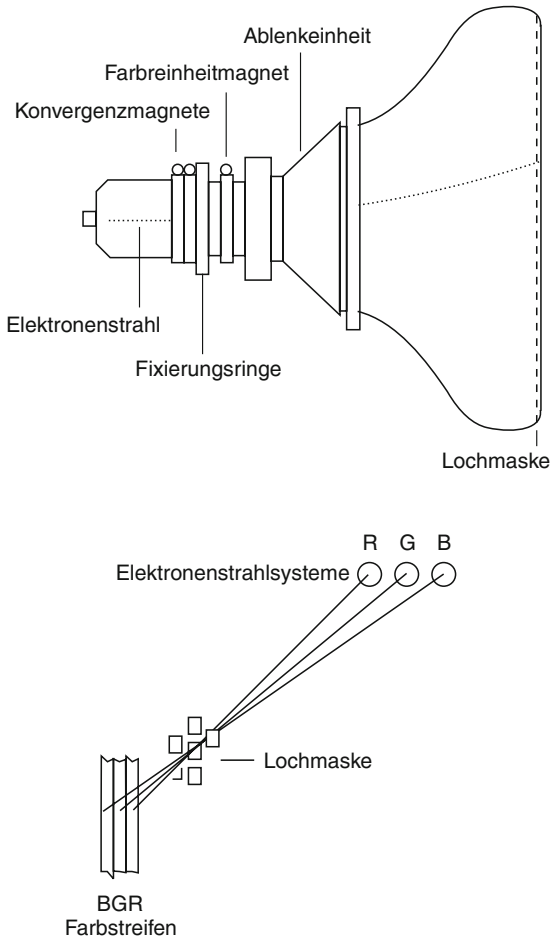


Abb. 53.20 Farbbildröhre und Inlineprinzip

Das Verhältnis von Breite zur Höhe ist bei allen Standardmonitoren 4 : 3. Damit reicht die Bildschirmdiagonale zur Angabe über die Monitorgröße aus.

Die Zahl der vertikal und horizontal ansprechbaren Bildpunkte (Pixel) gibt die *Auflösung* an. Eine Angabe von 1024 × 768 bedeutet, dass horizontal 1024 und vertikal 786 Pixel dargestellt werden. Den Abstand benachbarter Bildmittelpunkte bezeichnet man als *Dot Pitch*.

Tab. 53.2 Grafikkarten (Entwicklung)

Karte	Format/Bildpunkte	Anzahl der Farben	Bild-/Zeilenfrequenz
Hercules	768 × 320	–	50/18430
CGA	320 × 200	4	60/15750
EGA	640 × 350	16	60/21850
VGA	640 × 480 (800 × 600)	256	70/18750

Die vom Videointerface gelieferten Signale (wenige Volt) müssen noch verstärkt werden, um die Bildröhre aussteuern zu können. Die Amplitude des Signals ist mit dem Kontrasteinsteller veränderbar.

Monitore erzeugen Röntgenstrahlen und elektromagnetische Felder. Während für Röntgenstrahlen eine Röntgenverordnung existiert, gibt es für elektrostatische und magnetische Felder nur die allgemeinen Vorschriften VDE 0848. Die heutigen Monitore genügen alle der schwedischen *MPR-Norm* (Statens mät-och prorad), die jeweils Höchstwerte an der Bildschirmoberfläche vorschreiben. Sie können als strahlungsarm eingestuft werden.

Monitor und Grafikarte müssen aufeinander abgestimmt sein, da die Karten mit unterschiedlichen Zeilen- und Bildfrequenzen angeboten werden. Der *Multiscan-Monitor* dagegen ist in der Lage, die Signale aller gängigen Grafikkarten zu verarbeiten, weil er das Bild automatisch nach den angebotenen Frequenzen einstellt (Bildfrequenzen 50 . . . 90 Hz, Zeilenfrequenzen 15 . . . 45 kHz).

Das Videointerface (Abb. 53.21) überträgt Daten zum Monitor. Die Grafikkarte besteht im Prinzip aus einem *Zeichengenerator*, der alle Zeichen, die auf dem Bildschirm dargestellt werden können, enthält, und aus einem *Bildspeicher*, in dem der gesamte Bildschirminhalt abgelegt ist. Der Bildspeicher wird dann z. B. 60-mal in der Sekunde abgefragt und zur Anzeige gebracht. In kürzester Zeit haben sich immer bessere Grafikstandards herausgebildet. Tab. 53.2 gibt einen Überblick über die Entwicklung von Grafikkarten. EGA-Karten können beispielsweise maximal 16 aus 64 Farben darstellen, während im VGA-Standard 256 Farben dargestellt werden können. Weitere

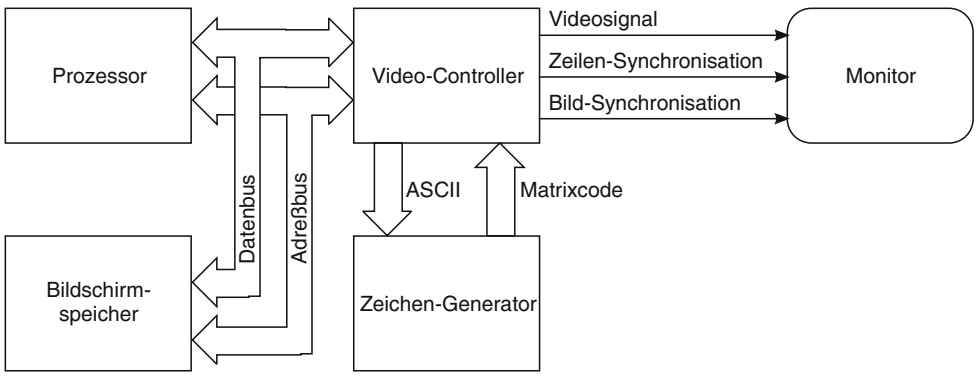


Abb. 53.21 Videointerface

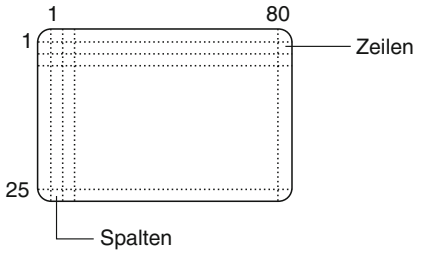


Abb. 53.22 Darstellung im Textmodus

Unterschiede liegen in der Signalübergabe zum Monitor. Während EGA mit digitalen Signalen arbeitet, werden bei den heute üblichen VGA- oder SVGA-Karten analoge Signale übertragen.

Bei Grafikkarten wird nach zwei grundsätzlichen Betriebsarten, dem *Text-* und dem *Grafikmodus*, unterschieden. Im Text- oder Alphamodus liegt die Bildinformation zeichenorientiert im ASCII-Code mit einem zusätzlichen Attributbyte vor. Das Attribut bestimmt z. B. die Farbe des ASCII-Zeichens. Aufgeteilt wird der Bildschirm nach Textzeilen und Textspalten (Abb. 53.22), z. B. in 80 Zeichen pro Zeile und 25 Zeilen. Hierzu ist eine relativ große Videobandbreite erforderlich.

53.4.1.2 LCD-Bildschirm

Neben dem Monitor, der mit einer Bildröhre bestückt ist, sind LCD-Bildschirme im Einsatz. Sie sind als Flachbildschirme aufgebaut und benötigen weniger Platz. Ein weiterer Vorteil ist der geringere Stromverbrauch. Als Nachteile sind Kontrastarmut, vor allem bei Fremdlichteinwirkung, Temperaturabhängigkeit und Trägheit zu nennen.

53.4.2 Drucker

Drucker (Printer) ermöglichen die Ausgabe alphanumerischer Zeichen für die Textverarbeitung und sind in der Lage, Diagramme und Zeichnungen zu erstellen. Im Wesentlichen unterscheidet man heute *Typenrad-*, *Nadel-*, *Tintenstrahl-* und *Laserdrucker*.

53.4.2.1 Typenraddrucker

Typenraddrucker waren die ersten eingesetzten Drucker in Verbindung mit Computersystemen. Sie arbeiten nach dem Prinzip der Typenradschreibmaschine und erzeugen ein ausgezeichnetes Schriftbild. Nachteile sind die geringe Druckergeschwindigkeit und die Unfähigkeit, Grafiken zu drucken. Ein weiterer Nachteil ist die Notwendigkeit, das Typenrad zu wechseln, wenn unterschiedliche Schriftarten in einem Schreiben verwendet werden sollen.

53.4.2.2 Matrixdrucker

Sehr weite Verbreitung haben die Nadel- oder Matrixdrucker gefunden, weil sie relativ preiswert sind. Die Zeichen werden durch ein Punktraster dargestellt. Die Steuerung erfolgt über Nadeln des Druckkopfes. Üblich sind 9- bzw. 24-Nadeldrucker. Matrixdrucker sind relativ schnell, grafik- und farbfähig. Als Nachteil kann das etwas höhere Arbeitsgeräusch empfunden werden

53.4.2.3 Tintenstrahlrucker

Durch sehr kleine Düsen wird Tinte auf das Blatt gebracht. Bei den Vorgängertypen trockneten häufig diese Düsen ein, wenn der Drucker wenig im Gebrauch war. Tintenstrahldrucker wurden in letzter Zeit durch unterschiedliche technische Verfahren wesentlich verbessert. Sie arbeiten sehr leise und sind farb- und grafikfähig.

53.4.2.4 Laserdrucker

Das Drucksystem eines Laserdruckers kann mit dem Prinzip eines Fotokopierers verglichen werden. Die Druckgeschwindigkeit ist sehr groß und das Druckbild hervorragend.

Programmiertechnik

Heribert Gierens



54.1 Programmiersprachen 742
54.2 Grundlagen der Programmierung 744

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_54

Aufgabe der Programmiertechnik ist es, Verfahren bereitzustellen oder zu entwickeln, mit deren Hilfe die Eigenschaften programmierbarer Geräte verändert werden können. Voraussetzung dafür ist, dass diese Geräte über einen Befehlsvorrat sowie über die Freiheit verfügen, die Auswahl und die Reihenfolge der auszuführenden Befehle zu variieren. Die genannten „Geräte„ beinhalten in der Regel mindestens einen Prozessor oder Controller.

Während eine SPS eindeutig programmierbar ist, erscheint eine Waschmaschine, obwohl sie einen Controller enthält, nur auf den ersten Blick programmierbar: über den Programmwahlschalter wählt der Benutzer aus fertigen, von ihm nicht veränderbaren Abläufen einen aus. Seine „Programmierung“ bezieht sich auf die Auswahl der berücksichtigten Daten (Art der Wäsche, Temperatur, . . .). Da diese Daten (oder Parameter) dem Gerät keine grundsätzlich neuen Eigenschaften verleihen, nennt man diese Vorgang nicht Programmieren sondern *Parametrieren*.

54.1 Programmiersprachen

Der Begriff *Programmiersprache* entstammt der Analogie mit der menschlichen Sprache. Hier wie dort gibt es die Vokabeln (Befehle) und die Grammatik (Regeln).

Dem Programmierstil der Pionierzeit lag keine „Sprache“ in diesem Sinne zugrunde, denn es fehlten die „Vokabeln“, die Begriffe, die synonym für bestimmte Inhalte (Befehle) standen. So war die Programmierung dadurch gekennzeichnet, dass jedes einzelne Bit im Speicher per Hand (Schalter) gesetzt bzw. gelöscht werden musste. Unter einer höheren Programmiersprache ist im Folgenden zu verstehen, dass *ein Befehl* in eben dieser Sprache *mehrere* Maschinenbefehle zu seiner Umsetzung benötigt.

In PASCAL lautet z.B. die Anweisung zur Addition zweier Variablen-Werte :

```
U_Gesamt := U_R1 + U_R2 ;
```

In Z80-Assembler sind für den (prinzipiell) gleichen Vorgang die folgenden Anweisungen nötig:

Vorplanung der Registerbelegung:

$$\begin{aligned} \text{HL} &= \text{Adresse des 1. Operanden} \\ (\text{HL}+1) &= \text{Adresse des 2. Operanden} \\ (\text{HL}+3) &= \text{Adresse des Ergebnisses} \\ \text{A} &= \text{Rechenregister} \\ \text{B} &= \text{Hilfsregister} \end{aligned}$$

```
LD HL, 8000 ; Adresse des 1. Operanden in das HL-Register
LD A , (HL)  ; Lade das A-Register mit dem 1. Operanden
INC HL       ; erhöhe den Inhalt des HL-Registers um 1
LD B, (HL)   ; Lade das B-Register mit dem 2. Operanden
ADD B        ; Addiere den Inhalt des B-Registers zum Inhalt
             ; des A- Registers (das Ergebnis verbleibt im A-Register)
INC HL       ; erhöhe den Inhalt des HL-Registers um 1
LD (HL), A   ; lade die Speicheradresse, auf die das HL-Register zeigt,
             ; mit dem Ergebnis
```

Aus der Verwendung einer höheren Programmiersprache resultiert daher eine bessere Übertragbarkeit (Portierung) auf andere Rechnersysteme sowie eine effizientere Programmierung; allerdings steigt der Speicherbedarf teilweise drastisch an.

Guter Programmierstil ist nur sehr bedingt von der verwendeten Programmiersprache abhängig. Mit der nötigen Disziplin und Sorgfalt lässt sich auch mit Sprachen, die nicht modular angelegt sind, sauber arbeiten. Es folgt daher eine unvollständige, alphabetisch sortierte Liste einiger Programmiersprachen, um die babylonische Vielfalt wenigstens zu benennen. Aufgrund der enormen Innovationsgeschwindigkeit können viele der bei einer Sprache genannten Eigenschaften zurzeit bereits in eine andere Sprache übernommen worden sein.

54.1.1 Assembler

Der Begriff Assembler ist doppeldeutig. Zum einen wird damit die (hier gemeinte) Assemblersprache und zum anderen das Programm, das aus einem Quelltext (siehe Beispiel) die Maschinenbefehle übersetzt, bezeichnet. Diese Dualität der Bezeichnungen findet sich auch bei anderen Programmiersprachen.

„Assembler“ ist eine Programmier*sprache*, weil der einzelne, vom Zielprozessor verstehbare Befehl durch ein sinnvolles Synonym dargestellt wird. Diese Synonyme werden Mnemonics (Mneme: Gedächtnis, Erinnerung) genannt, da sie die Aktion des betreffenden Befehls in kürzester Form beschreiben.

Beispiel

LD B,3E ; Lade das B-Register mit dem Wert $3E_{16}$

Hier steht die Abkürzung (Mnemonic) LD für „lade“!

Obwohl die Programmierung in Assembler hohe Anforderungen an den Programmierer stellt, weil er den Zielprozessor sehr gut kennen muss, gehört diese Sprache zu den niederen: *jeder* benutzte Maschinenbefehl wird durch *eine* Assembler-Anweisung dargestellt. Daher liefert diese Sprache keine Strukturierungshilfen (Wiederhol-Schleifen, Datentypen etc.) mit. Wer diese Sprache beherrscht, wird durch keine Sprachkonventionen reglementiert; man muss allerdings die volle Verantwortung selbst für einfachste Programmelemente (z.B. Ende einer Zählschleife) tragen. Der Vorteil dieser Sprache liegt in dem sehr schnellen Ablauf der erzeugten Programme sowie dem geringen Speicherplatzbedarf. ◄

54.1.2 ADA

Sowohl für den kommerziellen, den industriellen und den militärischen Bereich geeignet, ist ADA (Augusta *Ada* Byron, 1816–1862) eine universelle Sprache. Sie wurde mit den

Schwerpunkten leichte Wartung/Prüfung sowie Portierbarkeit entwickelt. Ähnlich PASCAL dokumentiert sie sich weitgehend selbstständig und erlaubt aufgrund des modularen Konzeptes das arbeitsteilige Vorgehen bei großen Softwareprojekten.

Eine geringe Innovationsgeschwindigkeit dieser Sprache ist der Preis für ein stabiles, an feste Spezifikationen gebundenes Konzept.

54.1.3 ALGOL

Die technisch-wissenschaftlich orientierte Sprache ALGOL (*Algo*rithmic *L*anguage) ist seit ca. 1960 verfügbar. Sie ist der „Urahn" der modernen, blockorientierten und prozeduralen Sprachen (Klammerung durch BEGIN ... END). Hier wird erstmalig zwischen der Zuweisung := und dem Vergleich = schon bei den verwendeten Operatoren unterschieden. Dazu wird die Prozedur als Statement und die Funktion als Variable einsetzbar. Die geschriebenen Prozeduren und Funktionen können sich selber aufrufen (echte Rekursion) und die Bool'sche Algebra wird im Befehlsvorrat berücksichtigt.

54.1.4 BASIC

Die leichte Erlernbarkeit stand im Vordergrund, als die Sprache BASIC (*B*eginner's *a*ll pupose *s*ymbolic *i*nstruction *c*ode) anfang der sechziger Jahre am Dartmouth College/USA aus dem Vorbild FORTRAN als „verdünnter Aufguss" entstand. Da BASIC ein Interpreter ist, ergeben sich kurze Testzeiten für die Programme. Allerdings werden diese recht langsam, wenn das Programm lang ist und viele Erklärungen (REM) sowie umfangreiche Kontrollstrukturen enthält.

Die Sprache BASIC ist einfach, beansprucht wenig Speicherplatz und kann auf (fast) jedem Mikroprozessorsystem installiert werden.

Da jeder Hersteller „sein" BASIC weiterentwickelte und/oder an sein Mikroprozessorsystem anpasste, entstanden im Laufe der Zeit unzählige, nichtkompatible aber immer leistungsfähigere BASIC-Dialekte. Die modernen BASIC-Versionen sind zu vollwertigen Konkurrenten der Sprachen C oder PASCAL geworden.

Der große Bekanntheitsgrad sowie die Leistungsfähigkeit haben dazu geführt, dass professionelle Anwenderprogramme (Textverarbeitung, Tabellenkalkulation, Datenbanken) BASIC-Dialekte als Makrosprache zur Parametrierung benutzen.

54.1.5 C

Die Eigenschaften von C stellen eine Mittlerfunktion zwischen den Assembler-Sprachen und höheren Sprachen dar. So besitzt C von Assembler die extrem hohe Abarbeitungsgeschwindigkeit mit einem sehr kompakten Code und von den (blockorientierten) höheren Sprachen deren Modularität und damit die gute Dokumentation und Wartungsfähigkeit.

Mit der größeren Nähe zur Prozessorebene obliegt dem Programmierer, ähnlich wie bei den Asembler-Sprachen, eine größere Verantwortung bei der Kontrolle des Programmes. Beispielsweise müssen Typüberprüfungen von Objekten veranlasst werden, während diese in PASCAL automatisch durchgeführt werden.

Aus diesen Gründen wird C gern als Programmiersprache für systemnahe Probleme verwendet. Beschränken sich die Programmierer auf den Kernighan-Ritchie- oder den ANSI-Standard, so ist eine leichte Portierung auf die verschiedensten Rechnersysteme möglich.

54.1.6 FORTRAN

Seit ca. 1955 verfügbar, ist FORTRAN (*For*mular *Tran*slator) eine der ersten Hochsprachen für den technisch-wissenschaftlichen Bereich. Ermöglicht wurde die (teilweise) Ablösung von den Assemblersprachen zu diesem Zeitpunkt durch größere, zur Verfügung stehende Arbeitsspeicher (4 bis 16 kB).

Durch die Hochsprachen-Eigenschaften sind die mit FORTRAN geschriebenen Programme austauschbar (portabel). So entstanden rasch große Programmsammlungen, die dieser Programmiersprache auch heute noch ihr Dasein sichern, obwohl es zu dieser Programmiersprache mittlerweile Alternativen gibt.

54.1.7 PASCAL

Die modernen PASCAL-Sprachen lehnen sich alle an das von Niklaus Wirth Anfang der 70er-Jahre in Zürich entwickelte PASCAL (Blaise Pascal, 1623–1662) an.

PASCAL, das ebenso wie ALGOL eine blockorientierte, prozedurale Sprache ist, wurde von Wirth ursprünglich für seine Studenten entwickelt, damit diese sich einen sauberen, strukturierten Programmierstil erwerben konnten. Hält sich der Programmierer an wenige Regeln, entsteht ein sich selbst dokumentierender Quelltext, so dass Wartung und Pflege der Programme vereinfacht werden. Zudem ermöglicht die Modularisierung sowie die Unterscheidung in lokale und globale Objekte eine Entwicklung großer Projekte in arbeitsteiligen Gruppen.

Diese Eigenschaften, sowie wesentlich verbesserte PASCAL-Dialekte, tragen dazu bei, dass diese Sprache zunehmend als (semi-)professionelle Programmiersprache gilt. Die Verbesserungen der einzelnen Dialekte ohne Normung bzw. Standard führt allerdings dazu, dass die erzeugten Programme kaum portabel sind.

54.1.8 PL/M

Für die Mikroprozessoren 8080/85 und 8088/8086 sowie für die Controller-Familie 8051 wurde die Sprache PL/M (*P*rogramming *L*anguage for *M*icroprozessors) aus der älteren Sprache PL/1 entwickelt. Diese erlaubt eine strukturierte Programmierung in Modulen, die separat übersetzt und später gebunden werden können. Für zeitkritische Aufgaben können auch Programmteile in Assembler eingefügt werden.

Die Nutzung dieser Sprache setzt allerdings gute Kenntnisse des Zielprozessor-Systems voraus, so dass die Anwender im professionellen Bereich zu finden sind.

54.2 Grundlagen der Programmierung

Die Aufgabe beim Programmieren besteht darin, eine Befehlsfolge (Programm) für ein bestimmtes Prozessorsystem in dessen Speicher zu schreiben, so dass es vom Prozessor ausgeführt werden kann.

Da der Prozessor nur binäre Befehle verstehen und verarbeiten kann, der Mensch wiederum lieber mit einer ihm vertrauten Sprache umgeht, wird das Programm möglichst in einer höheren Programmiersprache verfasst und dann in die maschinenübliche, binäre Form übersetzt. Die lästige, syntaktische Strenge der meisten Programmiersprachen wird nun zu einem Vorteil, weil diese Arbeit des Übersetzens durch einen Computer mit einem entsprechenden Übersetzungsprogramm ausgeführt werden kann.

54.2.1 Interpreter

Wie der Name ausdrückt, deutet (übersetzt) ein Interpreter das ihm zur Verfügung stehende Quellprogramm Anweisung für Anweisung, *während* das zu übersetzende Programm abläuft. Dieses Verfahren bedingt, dass sich der Interpreter *und* das zu übersetzende Programm immer gleichzeitig im Speicher befinden müssen, wenn das Programm ablaufen soll. Weiterhin wird jede Anweisung bei jedem Programmdurchlauf erneut gedeutet (übersetzt), wodurch sich zu den Programmablaufzeiten auch noch die Übersetzungszeiten addieren.

Der größte Nachteil liegt somit in der großen Ausführungszeit der damit laufenden Programme. Da auch Leerstellen, lange Bezeichner und (teilweise) erklärende Kommentare die Laufzeit verlängern, verzichten viele Programmierer auf eine übersichtliche Programmierung zugunsten der Geschwindigkeit. Die Folge sind komplizierte, weil optimierte Programme, die kaum gewartet werden können.

Ein Vorteil liegt in der Möglichkeit, relativ kleine und damit begrenzt leistungsfähige Interpreter einzusetzen. Diese Interpreter liefern aufgrund ihrer engen Zweckbestimmung (z. B. Mess-, Steuer- und Regelaufgaben) einen beachtlichen Befehlsvorrat und sind so klein, dass sie im Masken-ROM eines Controllers Platz finden.

54.2.2 Compiler

Ein Compiler (Übersetzer) ist ein Programm, das ein Quellprogramm komplett übersetzt und dann *anschließend* bei Bedarf ablaufen lässt. Bei der Übersetzung werden die Kommentare zwar gelesen, aber nicht bearbeitet; ebenso erzeugt ein langer Variablenname (oder ein beliebiger Bezeichner) genau den gleichen Verweis auf einen Speicherplatz wie ein kurzer Name, so dass die Übersetzungszeit zwar ansteigt, nicht aber die Laufzeit.

Als weiterer Vorteil verbleibt die gute Selbstdokumentation eines solchen, mit aussagefähigen Variablennamen geschriebenen Quellprogrammes, das auch durch die reichliche Verwendung von Leerzeilen und Zeichen zur Verbesserung der Übersicht nicht langsamer ausgeführt wird.

Das so erzeugte, lauffähige Programm kann später ohne Anwesenheit des Compilers gestartet werden. Diese Eigenschaft macht die erzeugten Programme kommerziell nutzbar, weil das erzeugte Programm (im Gegensatz zum Compiler oder Interpreter) geistiges Eigentum des Programmierers ist und somit nicht unter den Urheberrechtsschutz des Compiler-Lieferanten fällt.

54.2.3 Editor

Der Editor ist ein Programm, das dem Schreiben des Quellprogrammes dient. Im Grunde ist dazu jedes Textverarbeitungsprogramm, das den geschriebenen Text (Quellprogramm oder Quelltext) in unformatierter Form abspeichern kann, geeignet. Unformatiert bedeutet hier ohne Textmerkmale wie fett, unterstrichen oder ähnlichen, z. B. für die Korrespondenz wichtigen Eigenschaften.

Der Quelltext muss im reinen ASCII-Format vorliegen, damit der Interpreter, Compiler oder Assembler ihn verarbeiten können. Gleichzeitig bieten die Editoren großer, kommerzieller Textverarbeitungen vielfältige Annehmlichkeiten:

- Blockbefehle (suchen, verschieben, kopieren, löschen, . . .)
- komfortable Dateiverwaltung
- Arbeit mit mehreren Dateien gleichzeitig (Fenster-Technik)

54.2.4 Integrierte Entwicklungsumgebung

Unter einer integrierten Entwicklungsumgebung versteht man ein Programmpaket, das alle Hilfsprogramme enthält, die zur Programmerstellung gebraucht werden: Editor, Interpreter oder Compiler, Linker und Debugger. Diese Hilfsprogramme werden dann je nach geforderter Aktivität menügesteuert aufgerufen.

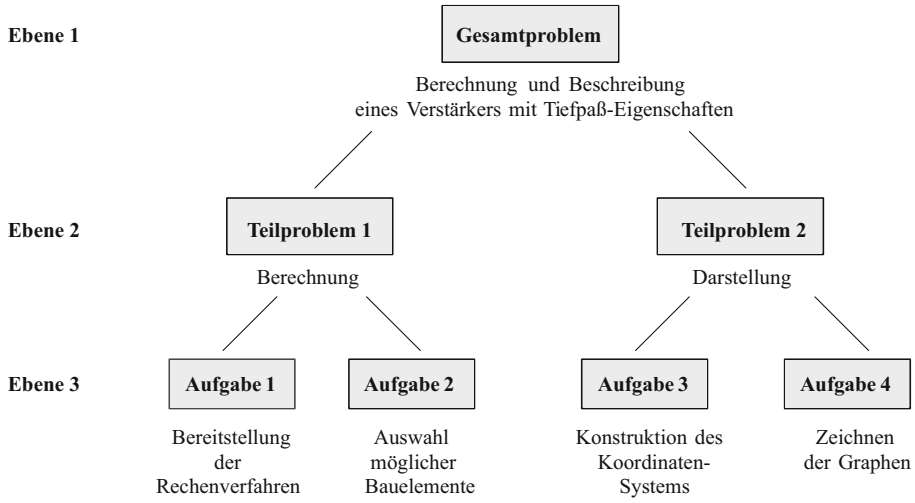


Abb. 54.1 Top-Down-Methode

Häufig gehört auch ein umfassendes Hilfesystem dazu, das Syntaxfehler erkennt und Abhilfe vorschlägt. Diese Entwicklungsumgebungen, die für die meisten höheren Programmiersprachen verfügbar sind, beschleunigen die Programmerstellung enorm, weil die Zeit für einen Durchlauf – Quelltext erstellen und Programm testen – (Turn-Around-Time) drastisch verkürzt wird.

54.2.5 Methoden der Programmentwicklung

Obwohl es theoretisch unendlich viele Wege zum Programm gibt, haben sich die Top-Down- und die Bottom-Up-Methoden etabliert, die es erlauben, den Weg nach ausgesuchten Kriterien zu optimieren. Eine exakte *Problembeschreibung* setzen allerdings beide Methoden voraus. Die Bearbeitung sehr großer Software-Projekte setzt weitergehende Verfahren voraus (Software-Engineering). Dazu gehört vor allem die Planung der zeitlichen Abhängigkeiten verschiedener Arbeitsteams (Netzplantechnik).

54.2.6 Problembeschreibung

Der Abnehmer eines Programmes sollte vom Programmierer (-team) dazu gedrängt werden, eine schriftliche Beschreibung der erwarteten Leistungen des Programmes anzufertigen. Diese Beschreibung kann oder sollte gemeinsam mit dem Fachmann (Programmierer) erstellt werden, um überzogene, vom System nicht erfüllbare Anforderungen sofort auszuschließen oder ein anderes Zielsystem einzuplanen. Diese Beschreibung, auch Pflichtenheft genannt, sollte u. a. enthalten:

- erwartete Ausgangsinformationen des Systems (eventuell grafisch; Farbdarstellung/Monochrom?)
- benötigte Eingangsinformationen (Datei-Formate; Tastatur; Maus; Sensoren; . . .)
- Vorkenntnisse der Benutzergruppe (Online-Hilfe-System; Handbuch für Benutzer?)
- gemeinsamer Datenzugriff mehrerer Benutzer? (Konsistenz des Datenbestandes, Vermeidung unvollständiger Transaktionen)

Eine große Sorgfalt im frühen Planungsstadium zahlt sich durch eine geringe Änderungswahrscheinlichkeit im Laufe der Arbeit aus.

54.2.7 Top-Down-Methode

Unter Anwendung der Top-Down-Methode wird das Problem zu Beginn global, also in seiner Gesamtheit beschrieben. Dieses Gesamtproblem wird dann hierarchisch immer weiter aufgeteilt in Teilprobleme, bis diese zu lösbaren Aufgaben werden.

Nach Bearbeitung der ersten, globalen Ebene stehen die beiden Teilprobleme exakt beschrieben fest. Je nach Umfang sollte auch hier die Problembeschreibung, jetzt für die zweite Ebene, detailliert und schriftlich erfolgen. Wird die Arbeit an den Teilproblemen verschiedenen Arbeitsgruppen übertragen, sind diese Beschreibungen verbindliche Abgrenzungen zu den anderen Teilproblemen.

Auf der niedrigsten Stufe der Hierarchisierung werden sämtliche Aufgaben bearbeitet und die Ergebnisse entsprechend zu den höheren Ebenen zurückgereicht. Bis auf die unvermeidbare Fehlersuche ist das Programm damit fertig, denn der obige Hierarchisierungsplan liefert gleichzeitig das Rahmen- oder Hauptprogramm nebst Dokumentation.

54.2.8 Bottom-Up-Methode

Auch bei der Bottom-Up-Methode wird das Gesamtproblem in einem Pflichtenheft beschrieben. Der daran anschließende Lösungsweg beginnt nun aber auf der niedrigsten, der Aufga-

benebene. Zu dem Beispiel in Abbildung würde das bedeuten, je nach Interessenlage oder Vorkenntnissen des Programmierers stehen etliche, mögliche Aufgaben zur Bearbeitung an. Er könnte mit einer Sammlung in Frage kommender Formeln der komplexen Rechnung beginnen bzw. bei mangelnden Kenntnissen, sich in diese einarbeiten. Nachdem alle für notwendig erachteten Aufgaben erledigt sind, werden sie in einem übergeordneten Rahmen (-Programm) zusammengefasst, bis das Gesamtproblem gelöst ist.

54.2.9 Bewertung der Methoden

Die Top-Down-Methode entspricht in ihrer Struktur am ehesten der Struktur effizienter, menschlicher Handlungsweise und verspricht daher ein zielgerichtetes, von unnötigen Seitenwegen freies Vorgehen.

Damit wird gleichzeitig eine geringe Fehlerwahrscheinlichkeit sowie eine gute Selbstdokumentation erzielt.

Die Gefahr dieser Methode liegt darin, „das Rad ständig neu“ zu erfinden: auf der letzten, der Aufgaben-Ebene werden Programmteile (Module) geschrieben, die die geforderte Aufgabe möglichst genau erledigen. Das führt zwangsläufig zu einer geringen Wiederverwertbarkeit der Module und schließt zudem die Verwendung bereits vorhandener, ähnlicher Module aus.

Die Bottom-Up-Methode empfiehlt sich ausschließlich für kleinste Programmiervorhaben, denn die Gefahr, sich in Detaillösungen zu verirren, ist sehr groß. Zusätzlich besteht die Wahrscheinlichkeit, dass ergänzende Anforderungen einer der höheren Hierarchisierungsebenen an die (fertigen) Module von diesen dann nicht erfüllt werden können; dies führt häufig zur völligen Neukonzeption einiger Module.

Erschwerend kommt die nachträgliche Dokumentation des Vorhabens hinzu. Da diese erst am Ende des dann fertigen Programmes erfolgen kann, steht die Dokumentation zwangsläufig nicht als Strukturierungshilfe während der Arbeit zur Verfügung; strukturelle Schwächen des Programmes sind so unvermeidbar.

Als „Königsweg“ bietet sich die Top-Down-Methode unter Beachtung folgender Bedingungen an:

1) frühzeitig, auf höchstmöglicher Ebene, wird bereits geprüft, ob es für die zu lösenden Teilprobleme/Aufgaben Standard-Lösungen (Module) gibt, bzw. ob es sich lohnt ein vorhandenes anzupassen
2) muss ein Modul vollständig neu geschrieben werden, ist auf gute Wiederverwertbarkeit zu achten; daraus folgt, dass Module, die nicht wiederverwertbar sind, so klein wie möglich sein sollten (Verweis auf Modularisierung/Bibliotheken)

54.2.10 Programm-Test

Hiermit ist die Prüfung eines Programmes auf logische Fehler gemeint, denn die syntaktischen Fehler (Fehler in der Schreibweise, Verletzung von Konventionen) werden bereits vom Interpreter oder Compiler erkannt.

Fehlerfreie Programme gibt es nicht! Denn Fehlerfreiheit bedeutet nicht nur, dass aus der Anweisung A := 12/B das Ergebnis „3“ resultiert, wenn die Variable B augenblicklich den Wert 4 besitzt. Darin eingeschlossen ist die Erwartung, dass das Programm nicht einfach die Bearbeitung abbricht, falls „B“ den Wert Null annehmen sollte, sondern den Benutzer darüber informiert, dass dieser Fall eingetreten ist. Noch besser wäre es, wenn der Benutzer dann unter Alternativen wählen könnte, wie das Programm damit umgehen soll.

Ein Instrument zur Fehlererkennung ist der Bleistifttest: alle vorkommenden Variablen werden in einer Tabelle aufgeführt und das Modul mit extremen Werten für die Eingangswerte „auf dem Papier“ gestartet. Alle Änderungen der Variablen werden in der Tabelle festgehalten und geben so ein Bild über die Zulässigkeit und den Erfolg der einzelnen Befehle des Programmes.

Weiterhin existiert bei der Top-Down-Methode bereits ein Rahmenprogramm (Hauptprogramm), wenn die Module getestet werden sollen. Diesem Rahmenprogramm werden die oben genannten extremen Eingangswerte für die zu testenden Module übergeben und das Hauptprogramm mit den Modulen gestartet. Noch nicht fertige Module können in ihrem Ein-und Ausgabeverhalten durch kleine Simulationsmodule ersetzt werden, um das Hauptprogramm lauffähig zu machen.

Zeigen sich dabei Unregelmäßigkeiten, bieten die Entwicklungsumgebungen die Möglichkeit, den vorliegenden Quelltext des Modules schrittweise abzuarbeiten und dabei die aktuellen Werte der Variablen in einer Tabelle auf dem Bildschirm zu verfolgen.

Die Simulation von Programmen ist ein Verfahren, das bei der Software-Entwicklung für Microprozessoren und Controllern angewandt wird. Hier wird das Programm ja nicht auf dem Zielsystem (z. B. Controller), sondern auf einem PC entwickelt. Um die so erstellten Programme dennoch auf dem PC testen zu können, bieten viele derartige Entwicklungsumgebungen eine Simulation auf dem Entwicklungsrechner an.

Ungefähr die Hälfte der vorhandenen Fehler können so vom Programmierer selbst entdeckt und behoben werden. Weitere Fehler können durch andere (objektivere) Mitglieder des Entwicklungsteams gefunden werden und ca. 10 % der Fehler bleiben unerkannt.

Datenkommunikation

55

Ulrich Lindemann



55.1 Einführung 748
55.2 Computernetzwerke 749
55.3 Einteilung von Netzwerken 750
55.4 Netzwerk-Topologien 752
55.5 Das OSI-Referenzmodell 754
55.6 Industrial Ethernet 769
Literatur 769

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_55

55.1 Einführung

Eine Datenverarbeitungsanlage (DVA) arbeitet nach dem Prinzip *Eingabe – Verarbeitung – Ausgabe*. Eine isoliert arbeitende DVA, ein „System“, kann nur die Daten verarbeiten, auf die sie lokal zugreifen kann. Viele Aufgabenstellungen machen es wünschenswert oder sogar zwingend erforderlich, Daten zwischen Systemen auszutauschen; oder aber die Vernetzung der Systeme erlaubt neue Funktionalitäten, mehr Effektivität, mehr Produktivität – oft bei gleichzeitiger Kostensenkung. Mit *System* ist hier nicht nur ein Großrechner oder ein PC gemeint, auch Produktionsmaschinen oder Roboter in der Industrie, ein Mikrokontroller zur Steuerung eines Außenspiegels am Pkw ist in diesem Sinn ein System.

55.1.1 Warum Datenkommunikation?

Ohne triftigen Grund wird keine neue Technologie eingeführt. Erst wenn deutliche Vorteile sichtbar sind, erfolgt der Schwenk in die neue Richtung. Die Vernetzung von eigenständigen Systemen bietet diese Vorteile:

- **Kostensenkung:**
  - durch höhere Effektivität:
    schnelle und preiswerte Kommunikation über E-Mail und Voice-over-IP (VoIP); gemeinsamer Zugriff auf Projektdaten, . . .
  - durch Automatisierung von Produktionsabläufen:
    Vernetzung von der Unternehmensleitebene bis zum Produktionsprozess; Fernwartung und Fernsteuerung von Produktionsanlagen; . . .
  - durch Rationalisierungen:
    E-Banking; E-Shopping; B2B, Business-to-Business-Systeme („Elektronischer Einkauf“); . . .
- **Datenverbund:**
  - Zugriff auf räumlich verteilte Daten einmal erfasste Daten sind netzweit verfügbar
  - Sicherung der Datenkonsistenz und Aktualität:
    aktuelle Projektdateien sind überall verfügbar; Flugreservierungen sind weltweit möglich; . . .
- **Funktionsverbund:**
  - Zugriff auf spezielle Funktionen im Netz:
    Großrechner; Datenbanken; . . .
  - keine räumlichen Einschränkungen für den Arbeitsplatz:
    Teleworking; . . .
- **Verfügbarkeitsverbund:**
  - redundante und fehlertolerante Systeme: hohe Zuverlässigkeit für sicherheitskritische oder unternehmenskritische Anwendungen
    Ampelsteuerung; Autopilot; redundante Firmen-Server; . . .
- **Lastverbund:**
  - jede Maschine kann jede Aufgabe bearbeiten
  - kürzere mittlere Antwortzeiten durch gleichmäßige Verteilung der Rechenlast auf alle Maschinen
  - Bereitstellen skalierbarer (leicht erweiterbarer) Rechnerleistung für viele Nutzer:
    „Server-Farmen“ für WWW-Server; Datenbankserver (z. B. bei Google); . . .
- **Leistungsverbund:**
  - Parallelrechner: die Aufgabe wird aufgeteilt auf verschiedene Maschinen und dann parallel (zeitgleich) von allen Maschinen bearbeitet (falls die Aufgabe auch parallelisierbar ist!):
    Wettervorhersage; Erstellen von Computeranimationsfilmen

Diesen Vorteilen stehen leider auch einige Nachteile gegenüber:

- **Abhängigkeit vom Netzwerk:**
  - bei Abhängigkeit von der Netzwerkverbindung geht ohne das Netz nichts mehr . . .
- **Gefahr durch Angriffe über das Netz:**
  - mögliche Angriffe: Datendiebstahl, Spionage, Sabotage:
    durch Computer-Viren, Würmer, Trojaner, Hacker; „Denial-of-Service-Angriffe“: können das System zum Stillstand oder zum Absturz bringen (PC, Firmen-Server, Produktionsanlagen, Infrastruktur, z. B. Stromnetze; . . .)

Bezieht man das Internet und Menschen mit ein in die Datenkommunikation, so können gerade hier noch weitere gravierende Nachteile auftreten:

- **Rationalisierungen:**
  - Personalabbau besonders in kundennahen Bereichen wie E-Banking, E-Shopping
- **Datenabgleich und Überwachung:**
  - einfache elektronische Überwachung
  - Tendenz zur Aufhebung der Trennung von Privatsphäre und betrieblichen Interessen
    beobachten der Internet-Nutzer durch Cookies; Personen-Recherche im Internet vor Einstellung (Achtung: das „Netz“ vergisst nichts, seien Sie vorsichtig mit dem, was Sie dort von sich eintragen. . .)
- **soziale Auswirkungen:**
  - Heimarbeit
  - Outsourcing
  - „Internet-Sucht“

55.1.2 Geschichte der Datenkommunikation

Batch-Betrieb

In den 60er-Jahren war der Batch-Betrieb mit Operateuren für die Rechnerbedienung Standard in der Datenverarbeitung.

Die Programmeingabe erfolgte typischer Weise über Lochkarten, die Ergebnisse kamen auf einem Zeilendrucker aus der Maschine. Vereinzelt gab es Prozessrechner, z. B. in der Chemie. Eine Datenkommunikation zwischen den DV-Anlagen fand nicht statt.

Zentralrechner

In den 70er-Jahren ging die Entwicklung hin zu Zentralrechnern („Mainframe“, Großrechner) mit Time-Sharing-Betrieb und Dialogterminals. Die Terminals standen immer weiter entfernt vom „Rechnerraum“ in den „Terminalräumen“ oder direkt am Arbeitsplatz des Sachbearbeiters, z. B. in Banken und Reisebüros.

Die erforderliche Datenkommunikation zwischen Rechner und Terminal erfolgte meist über analoge Leitungen, wobei jeder Hersteller seine eigene Technik einsetzte. Die Terminalnetze verschiedener Hersteller waren inkompatibel zueinander, manchmal sogar die neuen Varianten zu den älteren desselben Herstellers. Bekannte Standards waren:

- SNA (System Network Architecture, IBM)
- DECNET (DEC)
- TRANSDATA (Siemens)

Der PC

1971 war mit dem Intel 4004 der erste Mikroprozessor verfügbar; 1978 erschienen dann der Intel 8086 und gleichzeitig der IBM PC, der auf der Basis der 8086-Architektur das Konzept eines kompakten, *Persönlichen Computers* („PC“) realisierte. Der PC hatte genügend Rechenleistung, um sowohl autark zu arbeiten als auch Datenkommunikation betreiben zu können.

Vernetzung

Die Leistungsfähigkeit der PCs wuchs mit jedem neuen Mikroprozessor extrem schnell. Anfangs wurden PCs nur isoliert als Einzelplatzrechner betrieben. Aber bald erkannte man, dass vernetzte PCs eine viel preiswertere Alternative zu Zentralrechner sind – solange die zu verarbeitenden Datenmengen klein genug sind. Hieraus ergab sich ein wirtschaftlicher Druck zur standardisierten Vernetzung aller Rechner.

Internet

Das Internet entstand ursprünglich aus einer Initiative der US-Regierung über ein Forschungsprojekt zur Realisierung eines ausfallsicheren kontinentalen Netzwerkes. 1969 startete es mit vier Rechnerknoten. In der Folgezeit schlossen sich erst immer mehr Universitäten, dann auch Forschungseinrichtungen und andere Einrichtungen an. Heute besteht das Internet aus einem Zusammenschluss von vielen Netzwerken verschiedener Institutionen und Provider, die einen Internetanschluss auch für Privatkunden bereitstellen.

Grundlage des Internets ist seit 1983 das Protokoll TCP/IP Version 4 (IPv4), das nur sehr zögerlich von der verbesserten Version 6 (IPv6) abgelöst wird. TCP/IP ist zurzeit der De-facto-Marktstandard: es steht kostengünstig für alle wichtigen Rechnertypen zur Verfügung!

55.2 Computernetzwerke

Ein Computernetzwerk besteht aus mehreren unabhängigen Computern, die über ein Netzwerk verbunden sind. Man unterscheidet Server und Clients:

- ein **Server** stellt im Netzwerk einen Dienst bereit, den
- ein **Client** über das Netzwerk in Anspruch nehmen kann.

Je nach Organisation der Stationen im Netz unterscheidet man Peer-to-Peer-Netze („P2P“) und Client/Server-Netze („C/S“).

55.2.1 Peer-to-Peer-Netz

In einem Peer-to-Peer-Netz (P2P-Netz) sind alle Stationen gleichberechtigt: jede Station kann gleichzeitig Client und Server sein. Jede Station dient primär als Arbeitsplatzrechner. Der Aufbau des Netzes ist einfach: es müssen nur alle Stationen miteinander verbunden werden; es ist kein spezieller Server notwendig. Nachteilig ist der hohe Wartungsaufwand durch die **dezentrale Verwaltung**, da in jeder Station alle Rechte für alle anderen Stationen verwalten werden müssen. Daher sind P2P-Netze nur in sehr kleinen Netzen bis etwa zehn Stationen sinnvoll, z. B. in kleinen Büros. Ein typischer Vertreter der P2P-Netze war Windows für Workgroups.

55.2.2 Client/Server-Netz

In einem Client/Server-Netz sind die Aufgaben klar getrennt:

- der/die Server stellen zentrale Dienste bereit
- die Clients rufen die Dienste ab.

Mit „Server“ bezeichnet man meist einen besonders leistungsfähigen und betriebssicheren Rechner; oder aber das Programm, das einen Server-Dienst realisiert. Es gibt eine Vielzahl von möglichen Server-Diensten, z. B.

- WWW-Server: stellt Webseiten ins Netz
- Datenbank-Server
- File-Server: zentraler Datenspeicher
- Druck-Server: zentrales Drucken
- DNS-Server: Domain Name Service
- Mail-Server: für E-Mails
- Authentifikations-Server (Anmelde-Server)
- ...

Wenn ein Server hinreichend leistungsstark ist und die Datenverarbeitung für die Clients mit übernimmt, muss der Client nur noch die Ergebnisse anzeigen. Dafür reicht dann auch ein „Thin Client“, ein leistungsschwächerer und entsprechend preiswerter Rechner. Hiermit nähert man sich wieder der Struktur mit Zentralrechner von früher.

55.3 Einteilung von Netzwerken

Netzwerke lassen sich nach räumlicher Ausdehnung oder Dienstart in Kategorien einteilen:

55.3.1 Ausdehnung

GAN

Weltweite Netze bezeichnet man als *GAN*. Typische Verbindungen erfolgen über Satellit mit z. B. $n \times 36\,\text{MBit/s}$ (n = Anzahl der Kanäle) oder über Glasfasern mit typisch $n \times 10\,\text{GBit/s}$.

WAN

Im Bereich um 1000 km spricht man vom *WAN*. Der Datentransport erfolgt per Telefon, Datex-P, ISDN (9,6 bis 64 kBit/s) oder über Glasfasern (100 MBit/s ... 10 GBit/s).

MAN

Ein *MAN*, umfasst etwa den Bereich einer großen Stadt (100 km). Typische Verbindungen sind als Glasfaserstrecke mit Datenraten von 100 ... 10.000 MBit/s ausgelegt (FDDI, Fibre Distributed Data Interface, 100 MBit/s; ATM, Asynchronous Transfer Mode, 622 MBit/s, GE, Gigabit Ethernet, 1000 MBit/s).

LAN

Netzwerke bis etwa 10 km Ausdehnung bezeichnet man als *LAN*. Auf kurzen Entfernungen dominieren verdrillte Kupferkabel (Twisted-Pair-Kabel, TP-Kabel) in Haus und Büro bis 100 m (CAT-5-Kabel: 100 MBit/s bis 100 m; CAT-7 S/FTP-Kabel: 10 GBit/s bis 100 m). Auch größere Entfernungen sind mit TP-Kabeln möglich (DSL: 3 km mit 1 MBit/s); interessanter sind aber die bereits verlegten Kabel-TV-Netze, die technisch Bandbreiten von 800 MBit/s und mehr bieten können. Für hohe Datenraten setzt man Glasfasern oder Kunststofffasern ein; diese bieten gleichzeitig eine galvanische Trennung.

55.3.2 Dienstarten und Verbindungen

Man unterscheidet verbindungsorientierte und verbindungslose Dienste, Punkt-zu-Punkt-Kanäle und Rundsendekanäle.

55.3.2.1 Verbindungsorientierter Dienst

Ein verbindungsorientierter Dienst liefert einen zuverlässigen Nachrichtenstrom (z. B. *Dokumente zum Drucker)*, einen zuverlässigen Bytestrom *(z. B. Sitzung am Remote Terminal*) oder auch nur eine unzuverlässige Verbindung bei minimaler Verzögerung (*z. B. VoIP, „Internet-Telefonie“).*

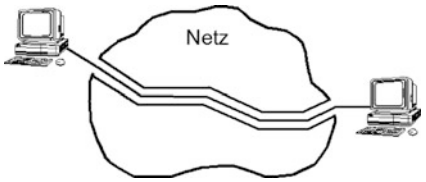


Abb. 55.1 Kanalvermittlung

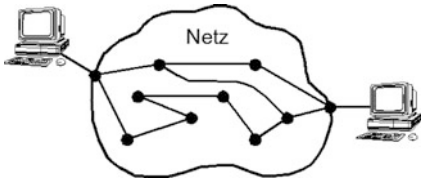


Abb. 55.2 Verbindungsloser Dienst

55.3.2.2 Kanalvermittlung

Die Kanalvermittlung (Abb. 55.1) erstellt eine feste Leitung für die Dauer der Vermittlung, einen „Tunnel“ durch das Netz. Dazu ist ein Auf- und Abbau der Verbindung erforderlich.

Die Daten fließen nach dem **FIFO-Prinzip** (First In, First Out) durch das Netz. So bleibt die Reihenfolge der Pakete erhalten.

Beispiele:

- Telefon: feste physikalische Verbindung (früher) oder virtuelle Verbindung (heute)
- ATM, TCP: feste virtuelle Verbindung

55.3.2.3 Verbindungsloser Dienst

Ein verbindungsloser Dienst (Abb. 55.2) benutzt eine Paketvermittlung: Pakete werden an das Netz übergeben und im Netz „irgendwie“ von Knoten zu Knoten weitergereicht; jeder Knoten ist bei der Weiterleitung autark. Im Internet sind die Knoten als IP-Router realisiert.

Es ist kein Auf- oder Abbau der Verbindung erforderlich, der Datendurchsatz ist maximal; allerdings können Pakete sich überholen und in falscher Reihenfolge am Ziel ankommen.

Beispiele:

- Briefpost
- IP, Internet Protokoll
- UDP, User Data Protocol

55.3.2.4 Punkt-zu-Punkt-Kanäle

Je zwei Netzknoten sind mit einer Leitung verbunden, Abb. 55.3. Die Netzknoten reichen die Nachricht weiter, bis sie am Ziel ist.

Da alle Netzknoten gleichzeitig senden können, ist der Gesamtdurchsatz des Netzes maximal. Die Verbindungen werden als TP-Kabel oder als Glasfaser ausgeführt; seltener auch als Funkstrecke.

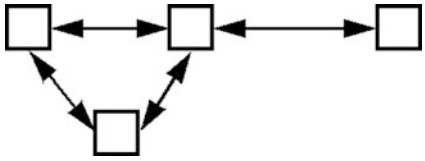


Abb. 55.3 Punkt-zu-Punkt-Kanäle

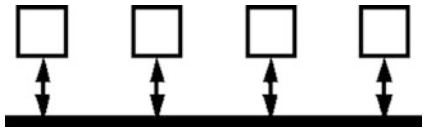


Abb. 55.4 Rundsendekanäle

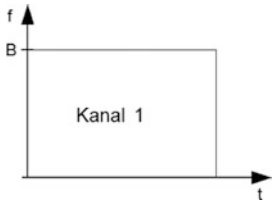


Abb. 55.5 Basisbandübertragung

55.3.2.5 Rundsendekanäle

Alle Netzknoten hören auf einem gemeinsamen Medium alle Nachrichten mit und ignorieren die nicht an sie selbst gerichteten, Abb. 55.4.

Die Verzögerungszeit je Nachricht ist minimal; für Rundsendungen (**Broadcasts**, ein Sender und viele Empfänger) ist diese Struktur ideal; bei hoher Datenlast wird der Kanal jedoch zum Flaschenhals. Ethernet (speziell mit Koax-Kabeln: 10Base5, 10Base2; veraltet) arbeitete nach diesem Prinzip; ebenso aktuell viele Funknetze.

55.3.3 Basisband – Breitband

Bei der *Basisbandübertragung* (Abb. 55.5) gehen die Digitalsignale direkt ins Kabel und belegen den Frequenzbereich von der Frequenz null bis zur System-Bandbreite *B*.

Die Schnittstellen sind preiswert, das Netz ist einfach und billig zu installieren und zu warten. Das Medium ist bidirektional verwendbar, die Datenraten reichen bis 10 GBit/s bei 100 m oder 10 MBit/s bei 1 km (Kupfer-Kabel). Die Basisbandübertragung ist sehr verbreitet, viele Standards bauen darauf auf: V.24, USB, Ethernet, ...

Bei der *Breitbandübertragung* (Abb. 55.6) ist die gesamte Bandbreite des Mediums aufgeteilt in mehrere Frequenzbänder mit einer Kanalbandbreite von z. B. 8 MHz; dies entspricht einem Fernsehkanal. Die Datenrate beträgt typisch 50 MBit/s (je 8-MHz-Kanal).

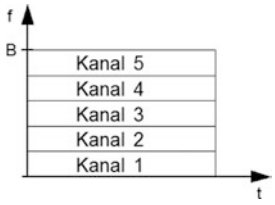


Abb. 55.6 Breitbandübertragung

Typische Satellitenkommunikationskanäle haben eine Bandbreite von 36 MBit/s und können damit ca. 500 PCM-Kanäle mit 64 kBit/s gleichzeitig übertragen.

In Kabelfernsehnetzen (Standard DOCSIS 3, Data over Cable Service Interface Specification) erreicht man bei Zusammenlegen von z. B. 16 Kanälen eine Bandbreite von 800 MBit/s (Downstream), die sich alle Teilnehmer in einem Netzwerksegment teilen müssen. Die Übertragung erfolgt über analoge Verstärker mit der vorhandenen Kabelfernsehnetz-Infrastruktur. Da das Kabelfernsehnetz ursprünglich nur unidirektionalen Datentransport zulässt, müssen für den Rückkanal (Upstream) entweder ein zweites Medium (z. B. Telefon; selten) oder ein zweiter Frequenzbereich bereitgestellt werden (DOCSIS 3: VHF-Kanäle 2, 3 und 4); dies setzt „Rückkanalfähige" Komponenten voraus (z. B. Hausanschlussverstärker).

55.3.4 FDM – DWDM – TDM

Das Frequenzmultiplexverfahren *FDM* (Frequency Division Multiplexing) teilt jeder Station ein Frequenzband zu. Bei wenigen Stationen erfolgt eine statische Vergabe der Frequenzbänder. Es handelt sich um die analoge Technik der Breitbandübertragung, siehe oben.

Das Wellenlängenmultiplexverfahren *DWDM* (Dense Wavelength Division Multiplexing) ist die optische Variante von FDM. Da die Bandbreite der elektrischen Sender und Empfänger begrenzt ist, koppelt man mehrere Strahlen mit leicht unterschiedlichen Wellenlängen über ein Gitter oder ein Prisma in die Glasfaser ein, Abb. 55.7.

DWDM-Systeme mit 160 Wellenlängen à 10 GBit pro Sekunde sind bereits verfügbar; die Übertragungskapazität einer einzelnen (!) Glasfaser liegt damit im Bereich Terabit pro Sekunde (TBit/s). Systeme mit mehr als der zehnfachen Kapazität sind

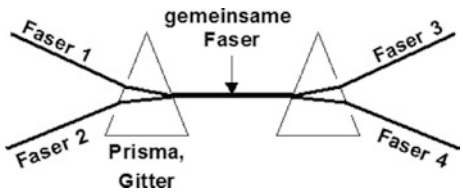


Abb. 55.7 Wellenlängenmultiplexverfahren

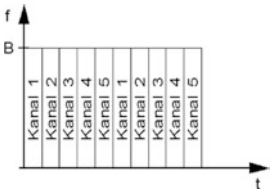


Abb. 55.8 Zeitmultiplexverfahren

ebenfalls im Labor machbar, werden zurzeit aber (noch?) nicht benötigt.

Das Zeitmultiplexverfahren *TDM* (Time Division Multiplexing) verschachtelt die verschiedenen Kanäle zeitlich hintereinander, Abb. 55.8. Jede Station erhält für einen Zeitintervall T die gesamte Bandbreite B des Übertragungskanals. Kritisch ist die Steuerung der Vergabe der Zeitfenster, hierin unterscheiden sich die verschiedenen realisierten Verfahren ganz erheblich:

- ATM ist deterministisch, echtzeitfähig und kann QoS (Quality of Service) bieten
- Token-Ring-Netze sind ebenfalls berechenbar und gut geeignet auch für „harte“ Echtzeit
- CSMA/CD ist sehr einfach, bei wenig Netzlast auch schnell, aber weder deterministisch noch zuverlässig und nicht „hart“ echtzeitfähig
- EtherCAT teilt die Daten auf in Prozessdaten, die extrem zeitgenau versendet werden, und restliche Daten, die in der verbleibenden Zeit über das Netz geschickt werden

55.4 Netzwerk-Topologien

Zur Verkabelung des Netzwerkes können die verschiedensten Topologien angewandt werden: Stern, Bus, Ring, Baum, Vollverbindung oder Kombinationen davon.

55.4.1 Sternnetz

Im Sternnetz (Abb. 55.9) hat die Zentrale Punkt-zu-Punkt-Verbindung zu allen Stationen. Der Datenaustausch erfolgt nur über die Zentrale.

Die Vorteile sind:

- sehr einfache Verkabelung
- der Ausfall einer Station ist unkritisch

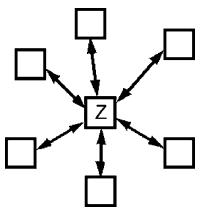


Abb. 55.9 Sternnetz

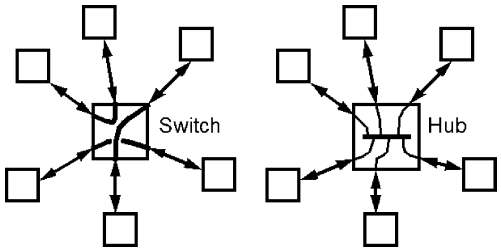


Abb. 55.10 Hub und Switch im Sternnetz

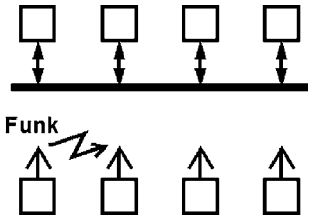


Abb. 55.11 Busnetz

Die Nachteile sind:

- der Ausfall der Zentrale legt das Netz lahm
- der Verkabelungs*aufwand* ist mittel bis hoch.

Beispiele: Ethernet-LAN mit TP-Kabeln oder Glasfasern (10 . . . 10.000 MBit/s)

Die Zentrale wurde früher als „*HUB*“ (*„Multiport-Repeater“, Signalverstärker mit mehreren Ein- und Ausgängen*) ausgeführt, Abb. 55.10. Dies ist zwar sehr preiswert, hat aber den Nachteil, dass der Hub immer nur eine einzige Verbindung herstellen kann; was schnell zum Flaschenhals werden kann.

Verbesserung bringt ein „*Switch*“, der als „Schalter“ nur genau die Leitungen durchschaltet, die für den Datentransport erforderlich sind und gegebenenfalls auch Datenpakete zwischenspeichern kann, falls eine Zielleitung gerade blockiert sein sollte („Store-and-Forward“). Durch das Schalten mehrerer paralleler Datenpfade wird der Flaschenhals im zentralen Knotenpunkt entschärft.

55.4.2 Busnetz

Im Busnetz (Abb. 55.11) sind alle Stationen parallel an den Bus angeschlossen. Jede Station kann Daten direkt zu jeder anderen Station senden („*Rundsendungen*“). Das Medium für den Bus ist ein Koax-Kabel, TP-Kabel mit einem Hub oder Funk.

Die Vorteile sind:

- es können alle gleichzeitig hören
- der Ausfall einer Station ist unkritisch
- der Verkabelungsaufwand ist gering
- kein Aufwand für Vermittlung

Die Nachteile sind:

- es ist nur eine Verbindung auf dem Bus möglich
- bei hoher Netzlast wird der Bus zum Flaschenhals

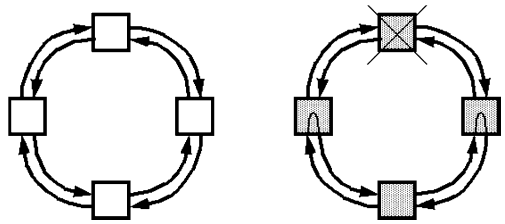


Abb. 55.12 Ringnetz

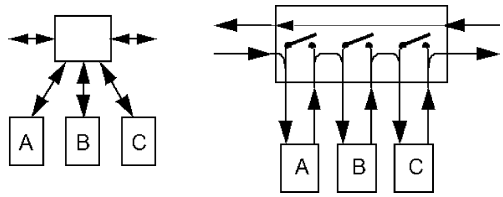


Abb. 55.13 Kabelkonzentratoren

- die Fehlersuche im Bus ist schwierig
- der Bus ist leicht abzuhören
- ein Ausfall des Mediums legt das Netz lahm

55.4.3 Ringnetz

Alle Stationen sind ringförmig verbunden, Abb. 55.12. Die Daten werden mit genau 1 Bit Verzögerung von Station zu Station weitergereicht. Die Vorteile sind:

- geeignet für Echtzeitanwendungen
- der Verkabelungsaufwand gering

Die Nachteile sind:

- die Übertragungsdauer steigt mit der Zahl der Stationen
- der Ausfall einer Station legt das Netz lahm.

Im doppelten Ring wird der Ausfall einer Station durch Umkonfigurieren aufgefangen.

Beispiele:

- IBM Token Ring: 4/16 MBit/s
- FDDI: Fibre Distributed Data Interface 100 MBit/s
- HSTR: High Speed Token Ring 100/1000 MBit/s
- SERCOS III, EtherCAT: Ringnetzwerke zur schnellen Steuerung von Maschinen und Anlagen („Industrial Ethernet“)

Zur Überbrückung nicht aktiver Stationen hat die Firma IBM Kabelkonzentratoren eingeführt (Abb. 55.13), die als Verteilerstationen am Ring hängen und extern eine einfache sternförmige Verkabelung erlauben. Der Kabelkonzentrator stellt dann intern die logische Ringstruktur her; nicht aktive Stationen kann er erkennen und überbrücken. Neue Entwicklungen (z. B. EtherCAT) besitzen diese Funktionalität in jedem Netzwerkanschlussknoten.

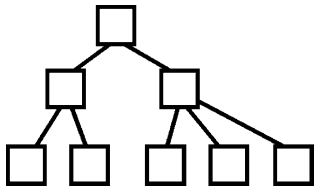


Abb. 55.14 Baumnetz

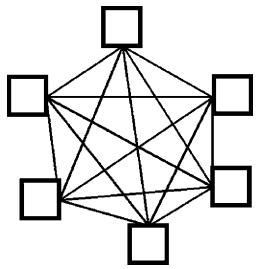


Abb. 55.15 Vollverbindung

55.4.4 Baumnetz

Im Baumnetz senden alle Stationen zur Kopfstation, diese verteilt die Nachrichten wieder im Netz (Abb. 55.14). Dies Verfahren ist ideal zur Verwendung auf bestehenden Kabelfernsehnetzen.

Beispiele: Kabelnetze; USB *(Universal Serial Bus)*.

55.4.5 Vollverbindung

Bei einer Vollverbindung (Abb. 55.15) ist jede Station direkt mit jeder anderen verbunden. Dies ergibt minimale Transportverzögerung und viele redundante Wege im Netz; jedoch auf Kosten eines maximalen Verkabelungsaufwandes.

Beispiele: die Anfänge der Telefonnetze in den USA; redundante Verbindung einiger (z. B. $2 + 2$) Router zur Verbesserung der Ausfallsicherheit.

55.4.6 Mischformen

55.4.6.1 Stern-Stern-Netz

- ein Stern als Hauptverteiler (Abb. 55.16)
- dazu ein Stern je Arbeitsgruppe
- oder ein Stern je Etage
- wenn der Hauptverteiler ausfällt, ist das Netz in der Etage noch funktionsfähig.

55.4.6.2 Stern-Bus-Netz

- Bus als Hauptverteiler („Backbone“, Abb. 55.17)
- sonst wie Stern-Stern-Netz.

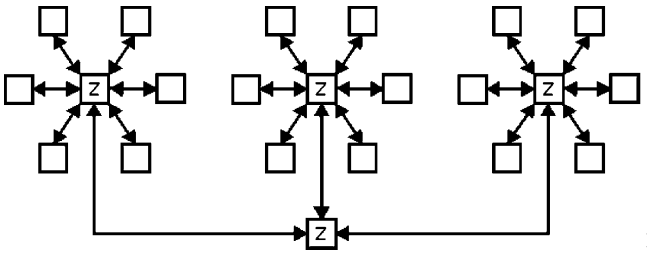


Abb. 55.16 Stern-Stern-Netz

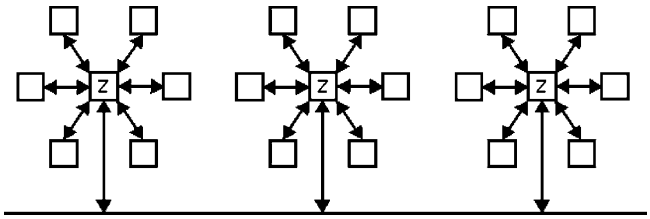


Abb. 55.17 Stern-Bus-Netz

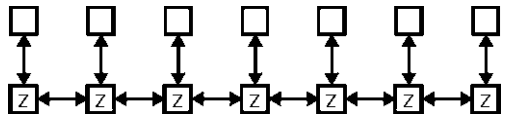


Abb. 55.18 Linien-Netz

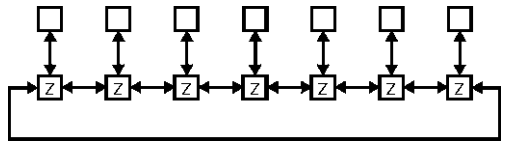


Abb. 55.19 Ring-Netz

55.4.7 Sonderformen

55.4.7.1 Linien-Netz

- Sonderform eines Sternnetzes (Abb. 55.18)
- Switch im Gerät eingebaut
- die Verbindung wird von Gerät zu Gerät weitergezogen
- Einsatz in der Automatisierungstechnik.

55.4.7.2 Ring-Netz

- Liniennetz mit einfacher Redundanz (Abb. 55.19): Ausfall eines Kabels oder Switches ist unkritisch
- Switche müssen die Schleife erkennen, z. B. mit:
  - MRP: Media Redundancy Protocol
  - RSTP: Rapid Spanning Tree Protocol
- Einsatz in der Automatisierungstechnik.

55.5 Das OSI-Referenzmodell

Das OSI-Referenzmodell (Abb. 55.20) besteht aus sieben aufeinander aufbauenden Ebenen sowie der Schicht 0, dem Übertragungsmedium. Es unterscheidet Dienste, Schnittstellen und Protokolle und ist genormt in ISO 7498.

- Ein **Protokoll** ist eine logische Verbindung in einer Schicht: z. B. http, das *Hyper Text Transfer Protocol*, verbindet einen WWW-Client mit einem WWW-Server
- Ein **Dienst** stellt in einer Schicht eine definierte Menge von Funktionen bereit und bietet diese der nächst höheren Schicht an: IP bietet TCP an, Pakete durchs Netz zu transportieren
- Eine **Schnittstelle** liegt zwischen zwei Schichten und definiert Zahl und Art der Parameter und Ergebnisse für den Zugriff auf Dienste der tieferen Schicht

Im OSI-Modell erfüllt jede Schicht genau eine Funktion und kann dabei auf Dienste der unter ihr liegenden Schicht zugreifen; als Ergebnis stellt die Schicht Dienste für die nächst höhere Schicht bereit. Die Grenzen zwischen den Schichten sind so gewählt, dass der Informationsfluss über die Schnittstellen zwischen den Schichten möglichst gering ist. Nur auf derselben Schicht fließen Informationen vom Sender zum Empfänger und zurück. Art, Umfang und Inhalt der Informationen sind im Protokoll dieser Schicht festgelegt: jede Schicht hat also ihr eigenes Protokoll.

Das Protokoll stellt damit eine logische Verbindung zur gleichen Schicht in verschiedenen Stationen her: die Bitübertragungsschicht überträgt **Bits** entsprechend dem Bitübertragungsprotokoll; die Datensicherungsschicht überträgt **Datenrahmen** entsprechend dem Datensicherungsprotokoll, die Vermittlungsschicht überträgt **Datenpakete** entsprechend dem Vermittlungsprotokoll usw.

Sind zwei Stationen nicht direkt über das Übertragungsmedium verbunden, sondern über eine Vermittlungsstelle, die auf Schicht 3 arbeitet (**Router**), so werden hier die Daten bis auf Schicht 3 hoch „ausgepackt“, entsprechend dem Vermittlungsprotokoll von Schicht 3 analysiert, wieder „eingepackt“ und zur Schicht 0 geschickt, wo sie dann zur Zielstation gelangen (Abb. 55.21). Liegen mehrere Vermittlungsstationen zwischen Sender und Empfänger, so wiederholt sich der Vorgang entsprechend oft.

55.5.1 Aufteilen langer Nachrichten

Die mit einer Nachricht maximal übertragbaren Datenmengen sind je nach Schicht begrenzt. Zu lange Nachrichten werden in mehrere kleine aufgeteilt (Abb. 55.22).

Um die ursprüngliche Nachricht wieder korrekt zusammensetzen zu können, erhalten alle Teilnachrichten (Fragmente)

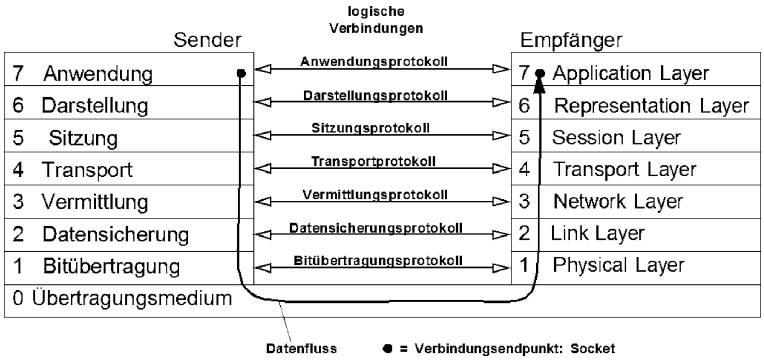


Abb. 55.20 OSI-Referenzmodell

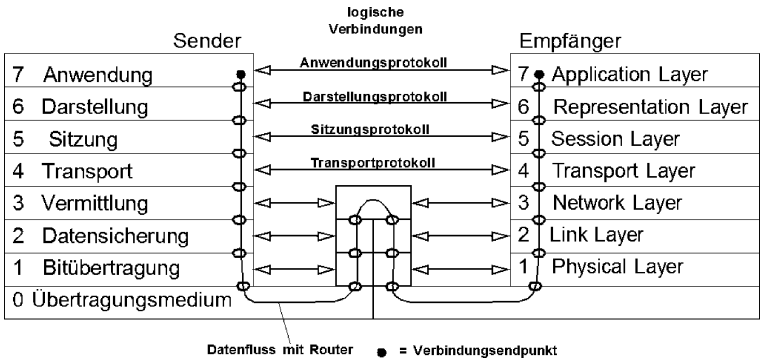


Abb. 55.21 Datenfluss mit Router

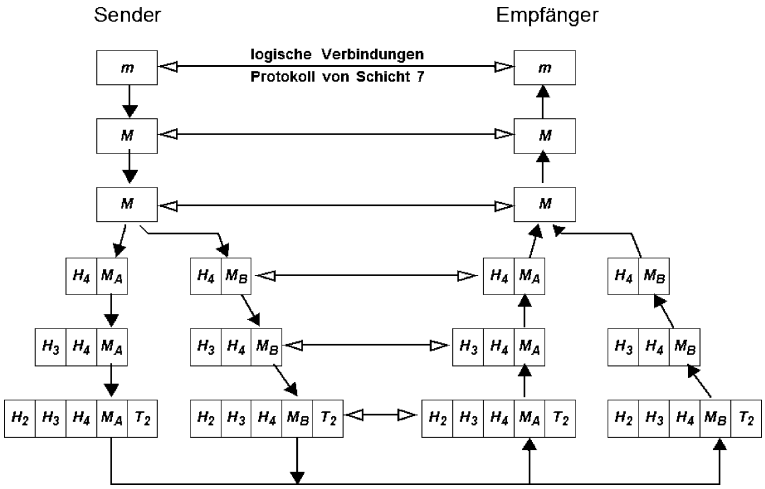


Abb. 55.22 Aufteilen langer Nachrichten

Zusatzinformationen in einem Nachrichtenkopf („Header“), der direkt vor der ursprünglichen Nachricht gesendet wird. Bei Bedarf kann jede Schicht einen eigenen Protokoll-Header zur Nachricht hinzufügen; dieser wird aber nur innerhalb derselben Schicht verwendet und im Empfänger ausgewertet. Bevor die Nachricht im Empfänger von einer Schicht „nach oben“ weitergereicht wird, wird der Header der Schicht entfernt. Die Datensicherungsschicht fügt neben dem Header auch noch eine Endanweisung („Trailer“) zur Nachricht hinzu; in ihr steht bei Ethernet die 32-Bit-CRC-Prüfsumme zur Fehlererkennung.

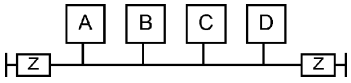


Abb. 55.23 Koaxialkabel

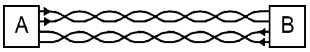


Abb. 55.24 Twisted-Pair-Kabel

55.5.2 Schicht 0: Übertragungsmedien

55.5.2.1 Koaxialkabel

Koaxialkabel eignen sich ideal zum Aufbau von Busnetzen (Abb. 55.23). Auch die ersten Ethernet-Netzwerke benutzten Koaxialkabel:

- 10base2: „Thinwire Ethernet“ oder „Cheapernet“, 10 MBit/s bis 185 m (1985), 5 mm Außendurchmesser
- 10base5: „Yellow Cable“, 10 MBit/s bis 500 m (1986), 12 mm Außendurchmesser

Das „Yellow Cable“ hat einen 2,6 mm starken Innenleiter aus massivem Kupfer. Das „Cheapernet“-Kabel ist wesentlich dünner, flexibler und preiswerter; wegen seiner höheren Dämpfung ist die maximale Länge eines Segmentes aber auch stark reduziert. Sehr nachteilig ist die maximale Datenrate von nur 10 MBit/s, die das Busnetz zum Flaschenhals macht. Koaxialkabel sind heute nur noch in Altinstallationen zu finden, sie werden beim Netzausbau gegen Twisted-Pair-Kabel ausgetauscht. Einzig in Kabelfernsehverteilernetzen sind Koaxialkabel nach wie vor Stand der Technik.

55.5.2.2 Twisted-Pair-Kabel

TP-Kabel sind seit langer Zeit im rauen Einsatz erprobt und bewährt. Durch die Verdrillung der zwei zusammengehörenden Adern und die Übertragung differenzieller Signalpegel ist eine sehr störarme Übertragung auch hoher Frequenzen möglich (Abb. 55.24), zudem sind die Kabel preiswert. Neuverkabelungen erfolgt mindestens mit CAT5-Kabel („Kategorie 5“, für Fast-Ethernet, 100 MBit/s bis 100 m) oder besser gleich mit CAT7-Kabel (für 10-Gigabit-Ethernet bis 100 m).

Die Verbindung zur Netzwerkkarte erfolgt über preiswerte RJ45-Stecker. Beim Anschluss eines Rechners an einen Hub oder Switch wird ein **Patch-Kabel** benötigt, in dem alle Adern 1 : 1 durchgeführt sind. Bei der direkten Verbindung zweier Rechner muss ein Kabel mit gekreuzten Leitungen („**Crossover-Kabel**“) benutzt werden, damit die Sendeleitungen der Station A auf den Eingangsleitungen der Station B und umgekehrt ankommen.

Neuere Geräte für Gigabit-Ethernet (1000Base-T) und 10GBase-T beherrschen mittlerweile eine automatische Erkennung der Sende- und Empfangsleitungen (Auto-MDIX); damit sind keine Crossover-Kabel mehr erforderlich.

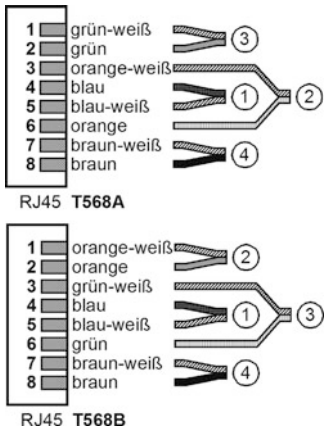


Abb. 55.25 RJ45-Stecker, Kontakt- und Farbschema

55.5.2.3 RJ45-Stecker

Der RJ45-Stecker hat acht Kontakte für vier verdrillte Aderpaare. Folgende Aderpaare werden benutzt (Abb. 55.25):

- Ethernet 10/100: Paare 2 + 3
- Gigabit-Ethernet: Paare 1 + 2 + 3 + 4
- ISDN: Paare 1 + 3

Ein Kabel mit vier Aderpaaren kann bei passender (Spar-) Verdrahtung also zwei Ethernet 10/100 Anschlüsse oder einen Ethernet-10/100 und einen ISDN-Anschluss versorgen; ein späterer Umstieg auf Gigabit-Ethernet ist dann aber ausgeschlossen, weil dieses alle vier Aderpaare benötigt. Daher sollte bei Neuinstallationen jede Anschlussdose mit einem kompletten Kabel versorgt werden.

Für die Adern im RJ45-Stecker gibt es die Farbschemata T568A und T568B. Ein Crossover-Kabel hat zwei verschiedene Stecker, einen mit dem Farbschema T568A und einen mit T568B. Sind beide Enden gleich, so handelt es sich um ein Patch-Kabel.

55.5.2.4 Lichtwellenleiter (LWL)

Bei hohen Datenraten und großen Entfernungen kommen Glasfasern als Lichtwellenleiter zur Anwendung. Unterseekabel mit Glasfasern verbinden zuverlässiger und preiswerter als Satelliten die Kontinente. FLAG: Fiber-Optic Link Around the Globe: 28.000 km, 10 GBit/s, von England bis Japan; SEA-ME-WE3: South-East Asia–Middle East–Western Europe 3, 39.000 km, 48 × 10 GBit/s, von Europa bis Australien;

TAT-14: Transatlantisches Telefonkabel Nr. 14, Deutschland–USA, 64 × 10 GBit/s. Unterseekabel sind besonders gefährdet durch Seebeben und Schiffsanker.

In der Automatisierungstechnik werden zur galvanischen Trennung Kunststofflichtwellenleiter eingesetzt, die robust und preiswert sind, aber auch eine sehr hohe Dämpfung aufweisen; dies ist bei den hier typischen kurzen Entfernungen von unter 500 m aber kein Problem.

Anhand ihres physikalischen Aufbaus unterscheidet man drei Arten von Lichtwellenleitern:

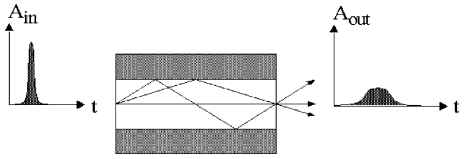


Abb. 55.26 Stufenindexfaser

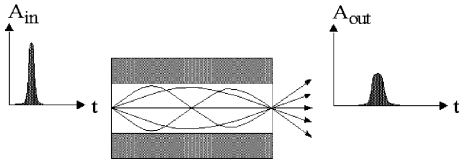


Abb. 55.27 Gradientenindexfaser

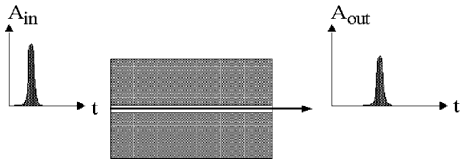


Abb. 55.28 Mono-Mode-Faser

Stufenindexfaser

Die maximale Länge der Stufenindexfaser (Abb. 55.26) ist begrenzt durch die starke Moden-Dispersion in der Faser. Mit Moden-Dispersion bezeichnet man die Impulsaufweitung durch unterschiedliche Weglängen der verschiedenen Ausbreitungswege der Lichtstrahlen in der Faser.

Der Durchmesser des Kerns einer Glasfaser beträgt 50 µm oder 62,5 µm; der des Mantels 125 µm (50/125 oder 62,5/125). Die Dämpfung beträgt $D < 5\,\mathrm{dB/km}$ bei Glasfasern.

Kunststofffasern gibt es als HCS-Faser (Hard Clad Silica Optical Fiber, 200/230 µm, bis 2 km; Kern Quarzglas, Mantel Kunststoff) und als POF-Faser (Polymeric Optical Fiber, 980/1000 µm, bis 100 m, Kern PMMA, Polymethylmethacrylat, Mantel Kunststoff).

Gradientenindexfaser

Der Durchmesser ist identisch zur Stufenindexfaser (50/125 oder 62,5/125). Ein optimierter gradueller Verlauf des Brechungsindexes von der Mitte der Faser zum Rand hin verringert die Moden- Dispersion, Abb. 55.27. Die Dämpfung beträgt $D < 3{,}5\,\mathrm{dB/km}$ bei 850 nm und $D < 1{,}0\,\mathrm{dB/km}$ bei 1300 nm Wellenlänge.

Mono-Mode-Fasern

Der Außendurchmesser ist auch hier 125 µm; der Kerndurchmesser ist mit 9 µm so klein, dass die Moden-Dispersion nicht mehr auftreten kann (Abb. 55.28). Die Dämpfung beträgt $D < 1{,}0\,\mathrm{dB/km}$ bei 1310 nm Wellenlänge. Mono-Mode-Fasern eignen sich daher am besten für hohe Datenraten und hohe Entfernungen, sind aber teurer und mit dem sehr kleinen Kerndurchmesser schwieriger zu verarbeiten.

Weitere Einzelheiten zu Lichtwellenleitern siehe Kapitel Nachrichtentechnik XII Nachrichtenübertragung über Lichtwellenleiter (LWL).

55.5.2.5 Ethernet-Medien

Ethernet bezeichnet einen Firmenstandard, der ursprünglich von Xerox, DEC und Intel für ein 10-MBit/s-Netz erarbeitet wurde und der später als Vorlage für die Norm IEEE 802.3 diente. Heute bezeichnet man mit „Ethernet“ meist alle Netze, die dem Standard IEEE 802.3 entsprechen. Alle Ethernet-Netze haben eine logische Bus-Struktur, aber nur mit Koax-Kabeln hat ein Ethernet-Netz auch eine physikalische Bus-Struktur: TP-Kabel und Glasfasern sind physikalisch immer als Stern-Struktur verlegt. Derzeit sind u. a. diese Kabeltypen für Ethernet genormt:

- (Standard-)Ethernet, 10 MBit/s:

10Base2	185 m	dünnes Koax-Kabel
10Base5	500 m	dickes Koax-Kabel, das *„Yellow Cable“*
10BaseT	150 m	Twisted Pair Kabel
10BaseF	2 km	Glasfaserkabel

- Fast-Ethernet, 100 MBit/s:

100BaseT	100 m	TP-Kabel CAT 5
100BaseF	2 km	Glasfaserkabel

- Gigabit-Ethernet, 1 GBit/s:

1000Base-T	100 m	TP-Kabel CAT 5e
1000Base-SX	500 m	Multi-Mode-Faser
1000Base-LX	5 km	Mono-Mode-Faser

- 10 Gigabit-Ethernet mit 10 GBit/s:

10GBase-T	100 m	TP-Kabel CAT 6a
10GBase-LX	300 m	Multi-Mode-Faser
10GBase-LX	10 km	Mono-Mode-Faser
10GBase-EX	40 km	Mono-Mode-Faser

55.5.2.6 Strukturierte Verkabelung

Die Europäische Norm EN 50173-1 definiert Grundsätze für eine dienst- und anwendungsneutrale strukturierte Verkabelung. In einem systematischen „Top-Down-Ansatz“ ist jeder Anschlusspunkt nach Ort und Leistungsfähigkeit spezifiziert. Dies reduziert die Kosten bei Betrieb und Fehlersuche im Netz. Die Norm unterscheidet Primärbereich, Sekundärbereich und Tertiärbereich:

Primärbereich

Verbindung verschiedener Standortverteiler oder Gebäudeverteiler; meist mit Glasfasern (galvanische Trennung!); eventuell auch DSL.

Große Entfernungen, hohe Datenraten, wenig Anschlusspunkte.

Sekundärbereich

Anschluss der Etagen- oder Gruppenverteiler (Switches) an die Standort- oder Gebäudeverteiler; mit Glasfasern; bis ca. 500 m.

Tertiärbereich

Anschluss der Endgeräte an die Etagen- oder Gruppenverteiler. Typisch mit max. 100 m Twisted-Pair- (TP-) Kupferkabel, davon 90 m fest verlegt und zweimal fünf Meter Patchkabel; bei besonderen Anforderungen auch Glas- oder Kunststofffasern.

55.5.3 Schicht 1: Bitübertragung

Das Protokoll der Bitübertragungsschicht legt fest, wie die Bits im Übertragungsmedium physikalisch kodiert sind hinsichtlich Spannung, Strom, Frequenz, Phasenwinkel, Amplitude oder Intensität und Zeitdauer. Hier sind die Elektrotechnik oder auch die Optik gefragt. Aktive Komponenten in Schicht 1 sorgen nur für eine Signalverstärkung im Medium (Repeater, Multiport-Repeater = Hub) und sind völlig unabhängig von allen Protokollen höherer Schichten.

55.5.3.1 Bit/s und Baud

Die Übertragungsrate einer Verbindung wird angegeben in Bit/s, das ist die Zahl der übertragenen Bits je Sekunde. Mit Baud („Schritt“) bezeichnet man die Zahl der Signaländerungen je Sekunde; also ist 1 Baud gleich 1 Signaländerung je Sekunde.

Ist also 1 Bit/s gleich 1 Baud?

- **Ja**, falls 1 Signaländerung genau 1 Bit überträgt. Dies ist der Fall bei einer digitalen Übertragung, in der nur die Zustände 0 oder 1 vorkommen. Diese Art der Übertragung ist sehr störsicher und daher auch sehr weit verbreitet. Einige Beispiele:
  - serielle V.24-Schnittstellen (PC: COM1)
  - Ethernet 10BaseX
  - USB
- **Nein**, falls 1 Signaländerung mehrere Bits gleichzeitig überträgt. Kann ein mehrstufiges Signal 8 Zustände annehmen, dann kann nach jeder Signaländerung einer der Zustände $0 = 000_2$ bis $7 = 111_2$ angenommen werden. Damit hat 1 Schritt 3 Bit Informationsgehalt; in diesem Beispiel gilt: 1 Baud = 3 Bit/s. Beispiele:
  - 10GBase-T (10 GBit/s Ethernet), Modulationsverfahren: 128-DSQ 128 Punkte, 7 Bit je Symbol: drückt die Signalfrequenz auf 400 MHz; 800 MBaud: 1 Schritt = 1,25 ns
  - DVB-T: max. 64-QAM (6 Bit je Symbol)
  - DVB-C: 256-QAM (8 Bit je Symbol)
  - DVB-C2: 4096-QAM (12 Bit je Symbol)

55.5.3.2 Maximale Bitrate

Für einen idealen Übertragungskanal mit der Bandbreite B, das heißt, er überträgt nur Frequenzanteile im Bereich von 0 bis B (Basisbandübertragung), ergibt sich bei einer binären Übertragung eine maximale Datenrate von

$$\mathrm{BR}_{\max} = 2 \cdot B \quad [\mathrm{Bit/s}]$$

Bei einer mehrstufigen Übertragung mit N Stufen je Schritt vergrößert sich die Datenrate um die Zahl der Bits je Schritt $\mathrm{ld}(N)$, wobei ld der Logarithmus Dualis, also der Logarithmus zur Basis 2, ist und sich über den natürlichen Logarithmus berechnen lässt:

$$\mathrm{BR}_{\max} = 2 \cdot B \cdot \mathrm{ld}(N) \quad [\mathrm{Bit/s}]$$
$$\mathrm{ld}(N) = \ln(N)/\ln(2)$$

Hat ein Telefonkanal eine Bandbreite von $B = 2400$ Hz, so ergibt die bei binärer Übertragung ($N = 2$) eine maximale Datenrate von:

$$\begin{aligned}\mathrm{BR}_{\max} &= 2 \cdot B \cdot \mathrm{ld}(N) \\ &= 2 \cdot 2400 \cdot 1 = 4800\,\mathrm{Bit/s}\end{aligned}$$

Das ist deutlich weniger als Standard-Ethernet mit 10.000.000 Bit/s.

Das Rauschen begrenzt die maximale Zahl der fehlerfrei erkennbaren Stufen; binäre Codierungen mit $N = 2$ sind ja gerade wegen ihrer Stör**un**empfindlichkeit so weit verbreitet. Die Größe des Rauschen wird angegeben als Rauschabstand, der aus dem Quotienten von Signalleistung und Rauschleistung berechnet wird (S/N, Signal-to-Noise), meist in dB:

$$S/N = 10\ \log(S/N) \quad [\mathrm{dB}]$$

Im Telefonnetz ist typisch $S : N = 1000 : 1$ oder $S/N = 30$ dB. Die maximale Bitrate eines verrauschten Datenkanals berechnet sich zu

$$\mathrm{BR}_{\max} = B \cdot \mathrm{ld}(1 + S/N)$$

Für den Telefonkanal mit $B = 2400$ Hz und $S/N = 30$ dB ergibt dies als maximale Datenrate von:

$$\begin{aligned}\mathrm{BR}_{\max} &= B \cdot \mathrm{ld}(1 + S/N) \\ &= 2400 \cdot \mathrm{ld}(1 + 1000/1) \\ &= 24.000\,\mathrm{Bit/s}\end{aligned}$$

Hätte der Telefonkanal CD-Qualität ($B = 40.000$ Hz, $S/N = 90$ dB), dann ergäbe das auch nur ca. 1,2 MBit/s.

Hohe Datenraten kann man auf einfache Weise nur mit einem Medium erreichen, das eine möglichst hohe Bandbreite hat. Reicht die Bandbreite nicht aus, muss man auf mehrstufige Codierungen ausweichen.

55.5.3.3 Bitkodierungen

55.5.3.3.1 V.24

Die Norm CCITT V.24, gleich mit RS 232 C und DIN 66020, beschreibt eine weit verbreitete, sehr einfache serielle, asynchrone binäre Datenübertragung, die auch vom PC her als

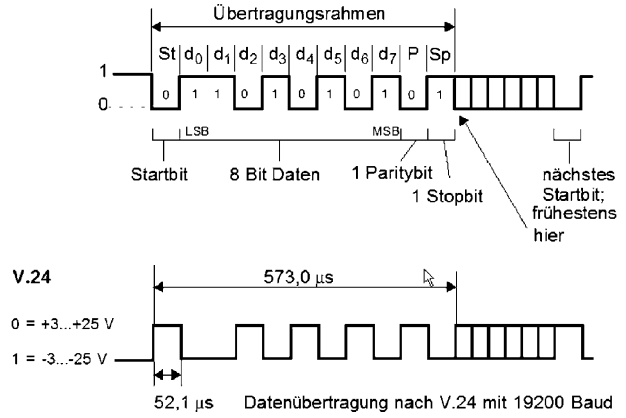


Abb. 55.29 V.24-Schnittstelle: Datenübertragung

Schnittstelle COM1 bekannt ist. Die Baudrate – hier gleich mit der Bitrate – ist in einem weiten Bereich von z. B. 110 bis 115.200 Baud einstellbar. Die maximale Reichweite bei kleinen Datenraten liegt bei ca. 100 m, bei 115.200 Baud eher im Meterbereich. Trotz der hohen – aber auf Masse bezogenen – Signalpegel sind längere Kabel doch sehr störempfindlich.

Zu Beginn der Datenübertragung synchronisiert das Startbit den Empfänger, Abb. 55.29. Es folgen die Datenbits, beginnend mit dem niederwertigstem Bit (LSB, Least Significant Bit). Die Zahl der Datenbits ist meist einstellbar, z. B. von fünf bis acht. Nach den Daten kann optional ein Parity-Bit (Prüfbit) folgen; mit diesem sind alle 1-Bit-Fehler erkennbar. Den Abschluss des Übertragungsrahmens bildet immer ein Stopbit. Stehen keine weiteren Daten zur Übertragung bereit, so wird das Stopbit beliebig verlängert.

Die Redundanz bei der Übertragung eines Bytes (8 Bit) beträgt mindestens 20 % (ein Startbit plus ein Stopbit zu jedem Byte). Je nach Dateninhalt kann ein Gleichanteil im Signal enthalten sein; damit ist nur eine Basisbandübertragung möglich. Eine direkte Übertragung über Telefon oder eine transformatorische galvanische Trennung sind nicht möglich, da der Gleichanteil im Signal nicht übertragen werden kann.

55.5.3.3.2 Manchester-Kodierung

Die Manchester-Kodierung (Abb. 55.30) ist frei von Gleichanteilen im Signal und lässt sich daher sehr gut übertragen. Weitere Vorteile sind der im Signalstrom enthaltene Takt zur Synchronisierung der Datenbits sowie die Erkennbarkeit von Kurzschluss und offener Leitung. Den Vorteilen steht die recht hohe Redundanz von 50 % gegenüber; daher wird die Manchester-Kodierung nur bei kleinen Datenraten eingesetzt (10Base-T).

Die Codierung erfolgt als 1B/2B-Codierung, d. h. aus einem Datenbit werden zwei Bits im Datenkanal:

- aus 1 wird 10
- aus 0 wird 01

In der Bitmitte ist also immer ein Pegelwechsel, der zur Synchronisation der PLL im Empfänger genutzt wird. Ethernet mit

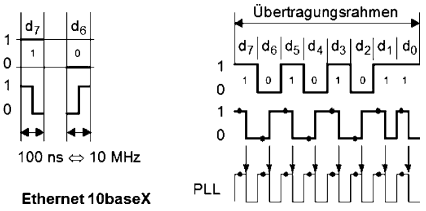


Abb. 55.30 Manchester-Codierung

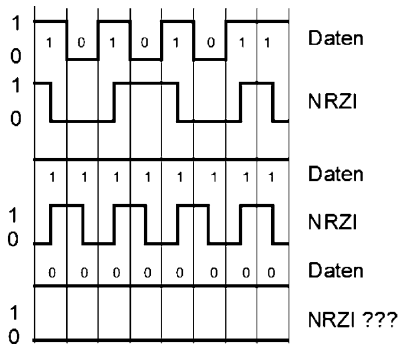


Abb. 55.31 NRZI/Invert-Codierung

10 MBit/s (10BaseX) arbeitet mit der Manchester-Kodierung; die maximale Frequenz im 10BaseX-Kabel beträgt 10 MHz, die minimale 5 MHz.

55.5.3.3.3 NRZI – Non Return to Zero/Invert

Die Codierung (Abb. 55.31) verringert die Frequenzen im Kabel:

- bei 0 bleibt das Signal konstant
- bei 1 erfolgt eine Pegelumkehr in der Mitte des Bitintervalls

Damit entspricht die maximale Frequenz im Kabel nur noch der halben Datenrate. Auch hier ist der Takt im Signalstrom enthalten. Jedoch ist je nach Kodierung etwas Gleichanteil im Signal: eine lange Folge von Nullen erzeugt einen konstanten Signalpegel und muss daher vermieden werden. Dafür sorgt ein RLL-Code (Run Length Limited Code), z. B. mit einer 4B/5B-Kodierung bei Ethernet 100BaseTx oder mit einer 8B/10B-Kodierung wie bei 1000BaseX. Die Redundanz beträgt 20 %.

55.5.3.3.4 4B/5B-Code

Abb. 55.32 zeigt die Codetabelle.

55.5.3.3.5 MLT3-Kodierung

Zur Übertragung von Ethernet mit 100 MBit/s benötigt man bei binärer Übertragung ein Kabel mit mindestens 50 MHz Bandbreite. Mit einer mehrstufigen Übertragung kann man Bandbreite sparen, also auch preiswertere Kabel nutzen. MLT-3 (Multilevel Threshold-3, Abb. 55.33) verwendet eine dreistufige (ternäre) Übertragung und kommt so bei einer Datenrate von

4 Bit	4B/5B-Code
0	11110
1	01001
2	10100
3	10101
4	01010
5	01011
6	01110
7	01111
8	10010
9	10011
A	10110
B	10111
C	11010
D	11011
E	11100
F	11101

Abb. 55.32 4B/5B-Code

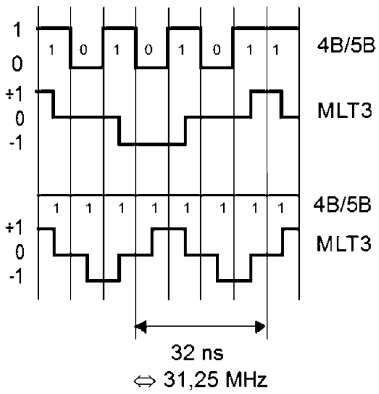


Abb. 55.33 MLT-3-Codierung

100 MBit/s mit einer Bandbreite von 31,25 MHz (statt 50 MHz) aus.

Bei der MLT-3-Codierung führt eine 0 zu einem konstanten Signal, eine 1 bewirkt eine Pegeländerung. Daher ist auch hier zur Minimierung des Gleichanteils ein RLL-Code notwendig. Das Signal kann drei Pegel annehmen: +1, 0, −1. Der Takt ist im Signalstrom enthalten; je nach RLL-Code verbleibt ein minimaler Gleichanteil im Signal. Mit dem 4B/5B-Code beträgt die Redundanz 20 %.

55.5.3.3.6 PAM-5-Kodierung

Gigabit-Ethernet (1000Base-X) nutzt einen 8B/10B-Code mit einer Symbolrate von 125 MBaud. Auf Kupferkabeln (1000Base-T) wird über alle vier Adernpaare gleichzeitig gesendet und empfangen, Abb. 55.34. Jedes Adernpaar überträgt zwei Bit; alle vier Adernpaare parallel zusammen ein Byte. Die zwei Bit auf einem Adernpaar sind über Trellis-Code-Modulation (TCM) auf drei Bit erweitert; damit erhält man die notwendige Redundanz (33,3 %) für eine Vorwärtsfehlerkorrektur bei einer Bruttoübertragungsrate von 1,5 GBit/s. Je drei Bit der TCM werden dynamisch auf eine der fünf Stufen der PAM-5 (Puls-Amplituden-Modulation mit fünf Stufen) abgebildet; mit

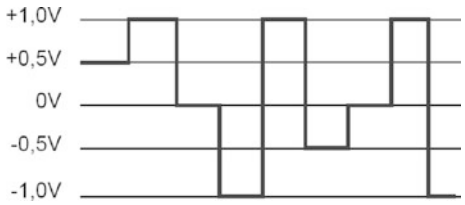


Abb. 55.34 PAM-5-Codierung

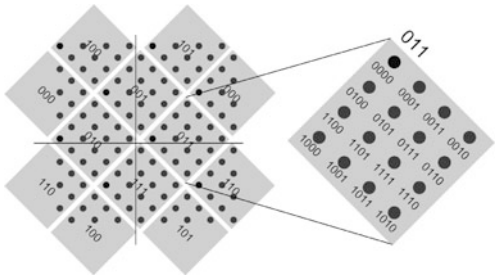


Abb. 55.35 128 DSQ-Codierung

vier Adernpaaren und fünf Zuständen je Paar ergeben sich 625 mögliche Symbole für ein Byte.

55.5.3.3.7 128 DSQ-Kodierung

Für die Datenübertragung mit 10 GBit/s über 100 m auf TP-Kabeln bei 10 Gigabit-Ethernet (10GBase-T) nutzt man die 128 DSQ-Kodierung (DSQ = Double Square), Abb. 55.35. 128 Punkte je Symbol ergeben sieben Bit je Symbol, drei Bit als Feld-Index („Coset“) und vier Bit für die Position im Feld. Alle vier Adernpaare werden bidirektional benutzt; die Symbolrate beträgt 800 MBaud, die Leitungskodierung erfolgt über D/A-Wandler (DAC) mit einer PAM-16 (16 Stufen). Die Signalverarbeitung übernimmt ein digitaler Signalprozessor (DSP), der in einem Trainingsmodus – vor dem normalen Betrieb – einige Leitungsparameter bestimmt und so bekannte Eigenstörungen aus dem Empfangssignal heraus rechnet (Echo-Anteile, NEXT, Near End Cross Talk, Nahnebensprechen, FEXT, Far End Cross Talk, Fernnebensprechen).

55.5.4 Schicht 2: Datensicherung

Schicht 2 legt fest, wie der Sender Datenübertragungsrahmen (*Data Frames*) erstellt, ob und wie der Empfänger Quittungsrahmen erstellt und wie die üblichen Probleme – beschädigte, verloren gegangene oder duplizierte Datenrahmen – behandelt werden. Teilweise erfolgt auch eine Datenflusssteuerung um einen Überlauf im Empfänger bei zu schnellem Sender zu vermeiden. Aktive Komponenten in Schicht 2 sind die *Bridges* (Brücken), die verschiedene LANs nach 802.3, 802.4 und 802.5 verbinden können. Ein „Ethernet-Switch“ arbeitet ebenfalls auf Schicht 2 und sorgt als Zentrale im Sternnetz für maximalen Datendurchsatz.

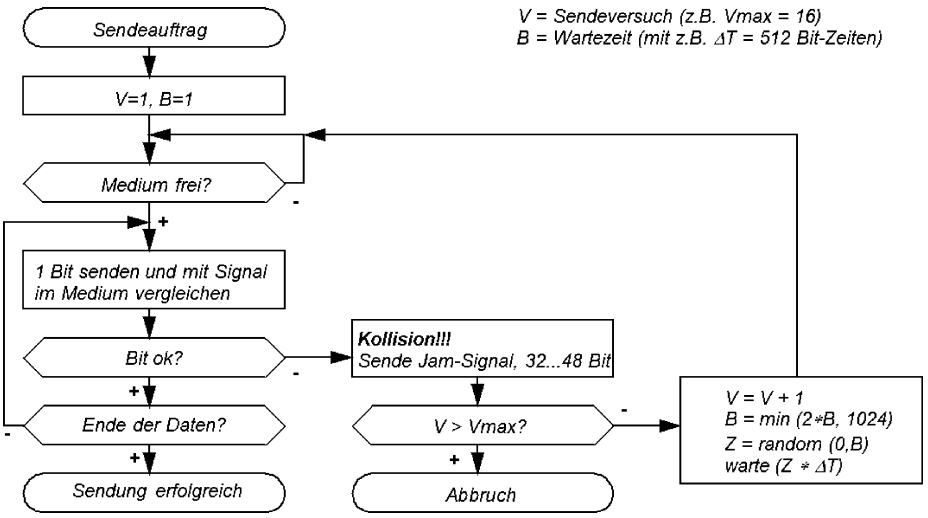


Abb. 55.36 Buszuteilungsverfahren nach IEEE 802.3: CSMA/CD

Nach Festlegung der sieben Schichten des OSI-Modells stellte man fest, dass Schicht 2 etwas überladen war und so teilte man sie auf in Schicht 2a und 2b, wobei 2b noch weiter unterteilt wurde:

- 2a: *LLC, Logical Link Control*, IEEE 802.2
- 2b: *MAC, Media Access Control*
- IEEE 802.3, *CSMA/CD*
- IEEE 802.4, *Token Bus*
- IEEE 802.5, *Token Ring*
- IEEE 802.11, *Wireless*

55.5.4.1 IEEE 802.3: CSMA/CD

CSMA/CD (Carrier Sense Multiple Access with Collision Detection) ist ein Buszuteilungsverfahren mit sehr geringer logischer Komplexität (Abb. 55.36). Es lässt sich sehr flexibel mit beliebig vielen Stationen einsetzen; allerdings steigt mit der Zahl der Stationen auch die Wahrscheinlichkeit von Kollisionen. Die Effizienz ist extrem abhängig vom Gesamtverkehrsaufkommen im Netz, bis 40 % Auslastung ist es problemlos; bei deutlich höherer Netzlast kann das Netz zusammenbrechen. Das Zugangsverfahren ist nicht determiniert: weder Erfolg noch Wartezeit lassen sich vorhersagen. Zudem neigt CSMA/CD dazu, „unfair“ zu sein: eine Station kann durch viele andere beliebig lange am Senden gehindert werden. Damit ist CSMA/CD nicht geeignet für „harte“ Echtzeitanwendungen.

55.5.4.1.1 Funktionsweise

Laufende Übertragungen werden nicht unterbrochen. Eine sendewillige Station hört daher zunächst das Medium ab, bis dieses frei ist *(Carrier Sense)*. Ist das Medium frei, so dürfen *alle* sendewillige Stationen gleichberechtigt mit ihren Übertragungen beginnen *(Multiple Access)*. Sollten jetzt zwei oder mehr Stationen gleichzeitig senden, so kommt es zur Kollision der Daten im Medium. Alle beteiligten Stationen erkennen die Kollision *(Collision Detection)* und bestätigen sie mit einem Jam-Signal. Zur Auflösung der Kollision „würfelt“ jede Station über eine Zufallszahl eine Wartezeit bis zu ihrem nächsten Sendeversuch. Solange die Zufallszahlen unterschiedlich sind, wird die Station mit der kleinsten Wartezeit als nächste allein das (dann hoffentlich freie) Medium beanspruchen und ihre Nachricht senden können. Sollte es nach einer Wartezeit erneut zu einer Kollision kommen, so wird die maximale Wartezeit solange verdoppelt, bis der Sendeversuchszähler sein Maximum überschritten hat und der Sendeversuch mit einer Fehlermeldung abgebrochen wird.

55.5.4.1.2 Late Collisions

CSMA/CD setzt voraus, dass Kollisionen erkannt werden. Daher müssen alle Stationen eine Mindestnachrichtenlänge einhalten, die so groß ist, dass auch in großen Netzen mit endlicher Signalausbreitung im Medium alle Kollisionen sicher erkannt werden. Ist das Netz zu groß – oder die Nachricht zu kurz – so ist es möglich, dass die sendende Station die Nachricht zwar fehlerfrei auf die Reise schickt, aber in einiger Entfernung im Medium die Nachricht doch noch zerstört wird – ohne das die sendende Station dies bemerken kann. Diese *Late Collisions* können den Netzbetrieb empfindlich stören.

55.5.4.1.3 Autonegotiation

Nerzwerkkarten nach dem Standard 10baseT prüften in Übertragungspausen alle 16 ms mit einem *Normal Link Pulse (NLP)* die Verbindung zur Gegenseite, Abb. 55.37. Wird die Verbindung durch Einstecken des Kabels hergestellt und werden zwei bis zehn NLPs empfangen, gilt die Leitung als aktiv und an der Netzwerkkarte leuchtet die „Link LED“ auf. Fehlt der NLP länger als 100 ms, so gilt die Leitung als unterbrochen und die „Link LED“ erlischt.

Der Standard 100baseT definiert einen *Fast Link Pulse Burst (FLP)*, der eine kompatible Erweiterung des NLPs ist. Zum Verbindungsaufbau oder nach Leitungsunterbrechung sendet die

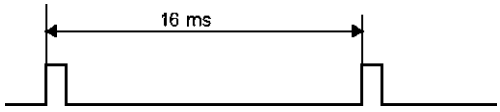


Abb. 55.37 Autonegotiation

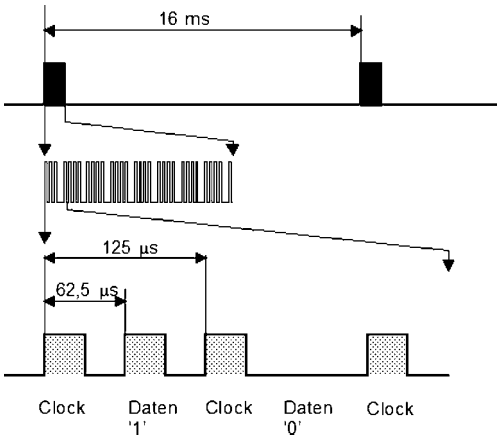


Abb. 55.38 100baseT-Standard: Fast Link Pulse Burst

Station alle 16 ms eine Impulsserie mit 17 Takt- und 16 Datenimpulsen, Abb. 55.38. Diese bilden zusammen das 16 Bit-Link-Code-Wort, das den höchsten möglichen Übertragungsmodus der Station anzeigt:

1. 100Base-TX Vollduplex
2. 100Base-T4
3. 100Base-TX Halbduplex
4. 10Base-TX Vollduplex
5. 10Base-TX Halbduplex

Beide Stationen wählen den höchsten gemeinsamen Modus.

55.5.4.2 IEEE 802.5: Token Ring

Ein Ringnetz nach IEEE 802.5 besteht aus einer Summe von Punkt-zu-Punkt-Verbindungen in Form eines Rings. Es kommt felderprobte Verbindungstechnik (Koax-Kabel, TP-Kabel, Lwl) zum Einsatz. Die Daten sind im differentiellen Manchester-Code verschlüsselt; die ungültigen Codes dienen als *Frame Start* und *End Delimiter*.

Das gesamte Netz ist in rein digitaler Technik realisiert; jegliche analogen Komponenten wie die Kollisionserkennung bei IEEE 802.3 fehlen. Jede Station verzögert die Nachricht um genau 1 Bit; dazu kommen noch die Laufzeiten im Kabel.

55.5.4.2.1 Funktionsweise

Es kreist immer genau 1 Token im Ring (Abb. 55.39). Das Token ist 3 Bytes gleich 24 Bit lang. Ist der Ring nicht lang genug, um das Token aufzunehmen, dann verlängert die Überwachungsstation den Ring künstlich. Wer das Token besitzt, darf senden; aber spätestens nach der Token-Haltezeit (10 ms) muss das Token weitergegeben werden.

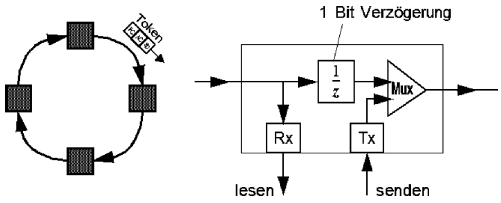


Abb. 55.39 Token Ring

Eine Station kann ein Token reservieren, falls die eigene Priorität höher ist als die des Frei-Tokens. Die Station mit der höchsten Priorität erhält das Senderecht; Stationen mit zu kleiner Priorität können unter Umständen bei ausgelastetem Netz nie senden. Die zu sendenden Daten werden an das reservierte freie Token angehängt. Die Zielstation bestätigt den Empfang; Daten und Bestätigung kommen am Sender wieder an; der Sender kann dann die Integrität der Daten prüfen. Nach Abschluss der Sendung muss der Sender ein Frei-Token generieren. Aus der Token-Haltezeit und der Zahl der Stationen im Ring (plus Laufzeit im Kabel) kann eine maximale Wartezeit bis zur nächsten Sendung berechnet werden – sofern alle Stationen mit gleicher Priorität arbeiten.

55.5.5 Schicht 3: Vermittlung

Die Vermittlungsschicht bestimmt die Paketleitwege, z. B. über Tabellen (Routing-Tabellen). Optionale Funktionen sind die Abrechnung der Kosten, die Behandlung von Überlastsituationen (z. B. zu viele Pakete im Netz) und – immer wichtiger werdend – die Abwehr unzulässiger Zugriffe und Angriffe durch Paket-Filter-Firewalls.

55.5.5.1 Internet Protokoll, Version 4 (IPv4)

Das Internet Protokolls, Version 4 ist genormt in RFC 791 und MIL-STD 1777. IP fragmentiert Pakete bei Bedarf; ein Paket *kann* maximal 64 kB (inkl. Kopf) groß sein; mindestens 576-Byte-Pakete *muss* jeder Knoten verarbeiten können. Es gibt eine Kopfprüfsumme, aber keine Datenprüfsumme; Prüfsummen für die Daten müssen bei Bedarf die höheren Schichten berechnen. Im LAN rechnet Ethernet auch eine 32-Bit-CRC-Prüfsumme. Die Lebensdauer eines Paketes ist begrenzt.

IP-Protokollkopf

Der Protokollkopf von IPv4 besteht aus mindestens fünf 32-Bit-Worten und kann durch Angabe von Optionen noch länger werden, Abb. 55.40. Die wichtigsten Felder sind die 32-Bit Internet-Adressen von Sender und Empfänger. Die Felder sind im Einzelnen:

- *Version*: 4 (→ IPv4)
- *Länge*: Zahl der 32-Bit-Worte für den IP-Header (variabel in IPv4!)
- *Servicetypen*: Priorität, Wartezeit, Durchsatz und Zuverlässigkeit (wird aber derzeit von fast allen Router ignoriert)

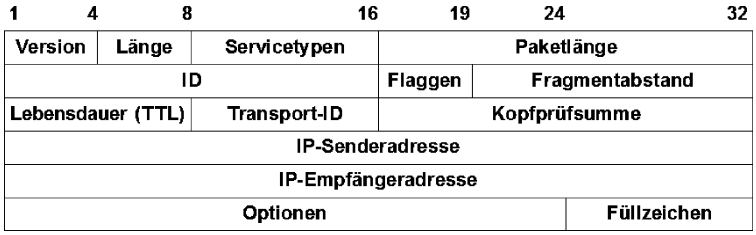


Abb. 55.40 IPv4: Protokollkopf

- *Paketlänge*: max. 65.535 Bytes, Header + Daten
- *ID*: eindeutige Identifikation für Fragmente, um diese wieder zusammensetzen zu können
- *Flaggen*:
  - *DF*: „don't fragment" Paket wird nie fragmentiert (sondern ggf. gelöscht)
  - *MF*: „more fragments": es folgen weitere Fragmente
- *Fragmentabstand*: bei MF = 1: Startadresse (geteilt durch 8) der Daten dieses Fragmentes im Gesamtpaket
- *Lebensdauer: TTL*, „time to live", jeder Router zählt TTL um eins herunter; bei einer TTL von null wird das Paket gelöscht
- *Transport-ID:* ID des Transportprotokolls, z. B. 6 = TCP, 17 = UDP, ...
- *Kopfprüfsumme*: Prüfsumme nur für den Header; muss wegen kleiner werdender TTL auf jedem Knoten neu berechnet werden
- *Senderadresse*,
- *Empfängeradresse*: 4-Byte Internet-Adressen
- *Optionen*: z. B. für Netzwerkmanagement oder Sicherheit
- *Füllzeichen*: zum Auffüllen auf 32-Bit-Worte

55.5.5.2 IP-Adressen

55.5.5.2.1 Schreibweisen

Mit 32 Bit Länge könnten die IP-Adressen theoretisch vier Milliarden Stationen eindeutig adressieren. Zur einfacheren Handhabung der 32-Bit-Zahlen sind diese bei Eingabe und Anzeige stets unterteilt in vier Gruppen zu je 8 Bit, also vier Bytes, die dann getrennt durch Punkte als einzelne Dezimalzahlen von 0 bis 255 dargestellt werden *(Dotted Decimal Format)*. Eine Darstellung als Hexadezimalzahl hat sich bei IPv4 nicht durchsetzen können; erst bei IPv6 kam man darauf zurück. Zwei Beispiele:

IP-Adresse:

- dezimal: 10.1.42.2
- hexadezimal: 0A011A02
- binär: 00001010.00000001.00101010.00000010

Netzmaske:

- dezimal: 255.255.255.192
- hexadezimal: FFFFFFC0
- binär: 11111111.11111111.11111111.11000000

55.5.5.2.2 Netzmaske

Netzmasken oder Subnetzmasken (*Subnet Mask*) unterteilen den IPv4-Adressraums in Teilnetze (Subnetze). IP-Adresse und Netzmaske legen das Subnetz einer Station fest. Eine Station kann nur andere Stationen im „eigenen" Subnetz direkt erreichen; für alle anderen Stationen muss ein Router vermitteln.

Als Binärzahl betrachtet besteht eine Netzmaske immer aus einer links beginnenden, lückenlosen Folge von Einsen. Hiervon ist eine andere Schreibweise abgeleitet: anstelle der Dezimal-Punkt-Schreibweise für die Netzmaske schreibt man die Zahl ihrer Einsen hinter einen Schrägstrich. Das für die Konfiguration einer Station notwendige Paar IP-Adresse/ Netzmaske kann dann so aussehen:

IP-Adresse/Netzmaske
10.1.42.2/255.255.255.192
10.1.42.2/11111111.11111111.11111111.11000000
10.1.42.2/26

Zwei Stationen sind genau dann in demselben Subnetz, wenn das Ergebnis der Bit-weisen UND-Verknüpfung der binären IP-Adresse mit der binären Subnetzmaske bei beiden gleich ist. Ein Beispiel:

IP-Adresse/NM	Netzmaske	Netzadresse
10.1.42.10/24	255.255.255.0	10.1.42.0
10.1.42.50/24	255.255.255.0	10.1.42.0
10.1.43.50/24	255.255.255.0	10.1.43.0

Die beiden ersten Stationen liegen im selben Netzwerk und können sich gegenseitig erreichen; die letzte Station liegt isoliert in einem anderen Netzwerk und wäre nur über einen zusätzlichen Router zu erreichen. Falls möglich, kann man aber auch die Netzmasken so weit verkürzen, dass alle Stationen in einem Netz liegen:

IP-Adresse/NM	Netzmaske	Netzadresse
10.1.42.10/23	255.255.254.0	10.1.42.0
10.1.42.50/23	255.255.254.0	10.1.42.0
10.1.43.50/23	255.255.254.0	10.1.42.0

Jetzt liegen alle Stationen im selben Netz und jede kann jede erreichen. Bei der Konfiguration eines Netzes sollte dringend beachtet werden, dass alle Stationen dieselbe Netzmaske haben!

55.5.5.2.3 Host-Adressen

Der Teil der IP-Adresse, der nicht zur Netzadresse gehört, ist die Host-ID. Je kürzer die Netzmaske ist, desto mehr Host-Adressen passen in ein Netz. Die kleinste (binär alles Nullen) und die größte (binär alles Einsen) Host-ID sind reserviert als Netzwerkadresse und als Broadcast-Adresse (BC-Adresse) und dürfen nicht für Stationen benutzt werden:

- IP-Adresse/NM: 10.1.42.10/16
- Netzwerkadresse: 10.1.0.0
- kleinste Host-Adresse: 10.1.0.1
- Host-IDs: 0.1 … 255.254
- größte Host-Adresse: 10.1.255.254
- Broadcast-Adresse: 10.1.255.255

Damit berechnet sich die maximale Zahl der Stationen in einem Netz zu zwei hoch (Anzahl der Nullen in der Netzmaske) minus zwei (keine Station auf Netz- oder BC-Adresse):

Netzmaske	max. Stationen im Netz
/8	$2^{24} - 2 = 16.777.214$
/16	$2^{16} - 2 = 65.534$
/24	$2^{8} - 2 = 254$
/23	$2^{9} - 2 = 510$
/30	$2^{2} - 2 = 2$

Große Netze mit sehr vielen Stationen können zur effizienteren Organisation und Verwaltung des Netzes durch Wahl einer längeren Netzmaske in mehrere kleine Netze aufgeteilt werden; dies erfordert dann aber den Einsatz von zusätzlichen Routern.

55.5.5.2.4 IP-Adressklassen

Die IP-Adressen sind in Adressklassen unterteilt, die eine einfache und bedarfsgerechte Verteilung für große und kleine Netze möglich machen sollten:

- Klasse A: 0.0.0.0 bis 127.255.255.255/255.0.0.0
- Klasse B: 128.0.0.0 bis 191.255.255.255/255.255.0.0
- Klasse C: 192.0.0.0 bis 223.255.255.255/255.255.255.0

Klasse A umfasst 128 Netze für je bis zu 16 Millionen Stationen. Diese Netze belegen den halben Adressraum; sie wurden in der Anfangszeit schnell vergeben, sind aber praktisch leer. Klasse B umfasst 16.384 Netze mit je 65.534 nutzbaren Adressen; diese sind gut geeignet für große Firmen. Es bleiben noch ca. zwei Millionen Klasse-C-Netze mit je 254 nutzbaren Adressen.

Die Adressen ab 224.0.0.0 sind nicht im Internet nutzbar; sie werden teilweise für spezielle Anwendungen in lokalen Netzen benutzt.

Bei der Vergabe von IP-Adressen sind einige Einschränkungen zu beachten:

- die IP-Adresse muss der Klasse A, B, oder C angehören
- sie darf mit der Netzmaske der Station keine Netzwerk- oder Broadcast-Adresse ergeben
- falls die Station ans Internet angeschlossen werden soll, muss die IP-Adresse weltweit eindeutig sein und es darf keine private IP-Adresse sein (siehe unten)
- 0.0.0.0 ist reserviert für „keine Adresse“
- das Netz 127.0.0.0/8 ist reserviert für interne Loopback-Tests (*localhost*)
- 255.255.255.255 ist reserviert für Broadcast

55.5.5.3 Private IP-Adressen

In RFC 1597 und RFC 1918 *Address Allocation for Private Internets* (RFC, Request for Comment; „Internet-Normen“) sind drei Adressbereiche zum Aufbau privater Intranets beschrieben:

- Klasse A: 10.0.0.0 bis 10.255.255.255
- Klasse B: 172.16.0.0 bis 172.31.255.255
- Klasse C: 192.168.0.0 bis 192.168.255.255
- APIPA: 169.254.1.1 bis 169.254.254.254

In Intranets sind dieselben Adressen weltweit beliebig oft verwendbar, solange die Netze voneinander isoliert sind. Kein Internet-Router leitet ein Paket mit einer privaten IP-Adresse weiter. Daher ist keine direkte Verbindung zum Internet möglich; aber auch kein direkter Angriff aus dem Internet auf eine private Adresse.

APIPA (Automatic Private IP Addressing, „Auto-IP“, RFC 3927, Dynamic Configuration of IPv4 Link-Local Addresses) oder Zeroconf (Zero Configuration Networking) ermöglicht eine konfigurationsfreie Vernetzung von Geräten im lokalen Netzwerk: jede Station berechnet für sich eine pseudozufällige Netzwerkadresse aus diesem Bereich. Sollte es damit im Netz zu einer Adresskollision kommen, so berechnet sich die Station eine neue pseudozufällige Adresse.

55.5.5.4 NAT: Network Address Translation

Eine Verbindung aus einem Netz mit privaten IP-Adressen in das Internet erfordert ein Gateway oder einen Router mit *NAT* (*Network Address Translation oder IP-Masquerading*), Abb. 55.41. Erste Entwürfe für NAT gab es ab Anfang der 90er Jahre. Die Ziele waren:

- den Engpass bei den IP-Adressen zu beseitigen
- das private Netz nach außen zu verstecken
- das private Netz vor Angriffen von außen zu schützen

Der NAT-Router hat eine offizielle Internet-Adresse und eine im privaten Netz. Er nimmt ein Paket aus dem privaten Netz an, merkt sich den Absender im privaten Netz in einer Tabelle, setzt als Absender seine eigene offizielle Internet-Adresse ein und leitet das Paket ins Internet weiter. Die Antwort aus dem Internet kommt zurück an die Adresse des Routers und er leitet das Paket weiter an die richtige Station im privaten Netz.

Das gesamte private LAN 192.168.1.0/24 besitzt hier nur genau eine öffentliche IP-Adresse, die des Routers (141.71.24.151). Übliche DSL-Router arbeiten immer mit NAT, ihre offizielle

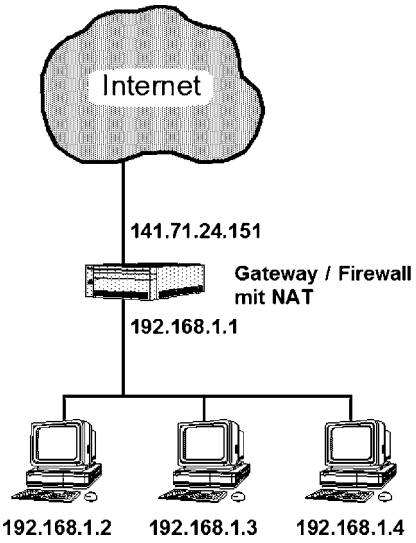


Abb. 55.41 Network Address Translation

IP-Adresse beziehen sie bei der Einwahl dynamisch per DHCP (siehe unten) vom jeweiligen Provider.

Für größere Netzwerke gibt es auch NAT-Gateways, die mehrere Internet-Adressen besitzen.

55.5.5.5 Router

Router verbinden verschiedene Netze indem sie Pakete annehmen und bestmöglich weiterleiten. Per Definition kann eine Station nur andere Stationen im *selben* Netz erreichen; Stationen in *fremden* Netzen sind nur über einen Router erreichbar.

IP	NM	Gateway	Schnittstelle
0.0.0.0	0.0.0.0	141.71.24.1	141.71.24.2
10.0.2.2	255.255.255.0	10.0.2.2	10.0.2.2
10.0.2.2	255.255.255.255	127.0.0.1	127.0.0.1
141.71.24.0	255.255.255.0	141.71.24.2	141.71.24.2
141.71.24.2	255.255.255.255	127.0.0.1	127.0.0.1
192.168.1.0	255.255.255.0	10.0.2.3	10.0.2.2

Eine Routing-Tabelle enthält u. a. diese Einträge: Zieladresse (IP), Netzmaske (NM), nächster Router zum Ziel (Gateway, GW) und Schnittstelle des Routers, an die das Gateway angeschlossen ist. Der Router vergleicht die Ziel-IP-Adresse eines ankommenden Paketes, UND-verknüpft mit der NM aus der Tabelle, mit der IP aus derselben Zeile der Routing-Tabelle. Bei einem Treffer schickt er das Paket entsprechend dem am besten passenden Eintrag weiter:

Ziel-IP	via Schnittstelle	an Gateway
10.0.2.50	10.0.2.2	10.0.2.2
10.0.2.2	127.0.0.1	127.0.0.1
10.1.2.50	141.71.24.2	141.71.24.1

Passt kein Eintrag oder ist die Lebensdauer (TTL) des Paketes abgelaufen, dann löscht der Router das Paket.

Der Eintrag 0.0.0.0/0.0.0.0 in der Tabelle passt auf alle Adressen; dies ist der *Standard-Gateway*-Eintrag. Das Standard-Gateway ist der Default-Router, der immer dann benutzt wird, wenn kein anderer Eintrag passt. Bei mehreren Standard-Gateway-Einträgen wird zuerst das erste Standard-Gateway benutzt, fällt dieses aus, dann wird automatisch auf das nächste umgeschaltet.

55.5.5.6 Routing-Protokolle

Router tauschen ihre Informationen über vorhandene Paketleitwege mit Routing-Protokollen untereinander aus (RIP: Router Information Protocol; OSPF: Open Shortest Path First; BGP: Border Gateway Protocol).

Wenn ein Router Anschluss an ein neues Netzwerk erhält oder ein benachbarter Router ausfällt, wird die Änderung der Netzkonfiguration von Router zu Router weitergegeben. In sehr großen Netzen kann es eine Weile dauern, bis alle Router auf dem aktuellen Stand sind; unmittelbar nach dem Ausfall eines Routers können daher noch einige Pakete verloren gehen.

55.5.5.7 Paketfilter-Firewall

Ein Router packt die Datenpakete bis auf Schicht drei aus, um die Ziel-IP-Adresse auswerten zu können. Damit kann er auf die Absender-IP-Adresse zugreifen und es besteht die Möglichkeit, den Datenverkehr an Hand der IP-Adressen im Paket zu filtern und unerwünschte Verbindungen zu unterbinden. Eine solche Paketfilter-Firewall ist sehr schnell, aber auch sehr grob: Angriffe von zugelassenen IP-Adressen kann sie nicht verhindern; Würmer und Trojaner, die in den Datenpaketen z. B. via E-Mail kommen, sind für eine Paketfilter-Firewall absolut unsichtbar.

55.5.5.8 Kostenabrechnung

Der Router kann an jedem Paket an den IP-Adressen erkennen, woher es kommt und wohin es geht. Damit kann er auch für eine Kostenabrechnung die Datenmenge je IP-Adresse (oder je Gruppe/Abteilung) erfassen.

55.5.5.9 IPv6

IPv6 ist wesentlich leistungsfähiger, bringt einige Sicherheitsoptionen mit, benötigt aber auch wesentlich mehr Rechenleistung, weshalb man gerne bei IPv4 bleibt. Aber spätestens wenn keine offiziellen IPv4-Adressen mehr verfügbar sind, müssen neue System mit IPv6 arbeiten. So schätzt die Internet Assigned Numbers Authority (IANA), dass dies um 2011/2012 der Fall sein könnte. IPv4 hat einige Schwachstellen:

- Die IP-Adressen werden knapp, da die Verteilung mit den Klasse-A/B/C-Adressen wenig effektiv war.
- Jeder Router muss für jedes der max. 2 Mio. Klasse-C-Netze einen Tabelleneintrag im Speicher halten.
- In absehbarer Zeit sollen etwa eine Milliarde Handys und „Wireless Devices“ angeschlossen werden.

- Das Routing benötigt sehr viel Prozessorleistung für das Fragmentieren von Paketen im Router und zum Berechnen der Prüfsummen im IP-Kopf, die sich wegen der TTL ständig ändern.
- Es gibt kein QoS (Quality of Service), etwa reservierte Bandbreite für Echtzeitdaten wie Telefon (VoIP, Voice over IP) oder Video.
- Es gibt keinen Standard für Verschlüsselung oder Authentifizierung.

IPv6 (Internet Protokoll, Version 6) soll diese Schwächen beseitigen. Klassische Protokolle werden weiter unterstützt (TCP, UDP, ICMP; RIP, OSPF, BGP). Die IPv6-Adressen sind auf 128 Bit erweitert, die alten IPv4-Adressen sind darin enthalten. Das Routen kostet weniger Prozessorleistung und wird schneller: weil Pakete nicht mehr fragmentiert werden (das muss bei Bedarf der Sender machen). Der Basis-Header hat eine feste Länge, weniger Felder und enthält keine Checksumme mehr; bei Bedarf folgen dem Basis-Header noch Erweiterungs-Header.

55.5.6 Schicht 4: Transport

Schicht 4 ist zuständig für den Aufbau und Abbau der Verbindung. Sie kann die Art der Verbindung festlegen, etwa hinsichtlich Dienstgüte wie Bandbreite, Zuverlässigkeit, usw. Abhängig von den Fähigkeiten der Schicht 3 muss sie eventuell die Nachrichten im Sender in kleinere Pakete zerlegen und im Empfänger wieder in der richtigen Reihenfolge zusammensetzen. Allgemein bietet die Transportschicht eine genormte Schnittstelle für die Sitzungsschicht durch Kapselung der tieferen Schichten.

55.5.6.1 TCP/IP

Die Entwicklung von TCP/IP (Transport Control Protocol/Internet Protocol) begann 1969 im ARPANET mit 4 Knoten: UCLA (University of California at Los Angeles), UCSB (University of California at Berkley), SRI (Stanford Research Institute) und die University of Utah koppelten zum ersten Mal vier Rechner.

1983 erfolgte die Umstellung des ARPANETs von NCP (*Network Control Program*) auf TCP/IP Version 4 (**IPv4**); diese Version ist auch heute noch aktuell(!). 1990 bestand das Internet aus einem Zusammenschluss von 3000 Netzen und 200.000 Rechnern auf der Basis von TCP/IP; 2000 waren es ca. 20 Mio. Rechner. Laut EITO nutzen Anfang 2008 1,23 Milliarden Menschen das Internet.

TCP/IP ist zurzeit der De-facto-Marktstandard: es steht kostengünstig für alle wichtigen Rechnertypen zur Verfügung!

55.5.6.2 Ziele von TCP/IP

Die Architektur von TCP/IP verfolgte diese Ziele:

- robust bei Ausfall von Teilen des Netzes
- unabhängig von Netzwerktechnologien und Hostrechnern
- universelle Verbindungsmöglichkeiten im Netz

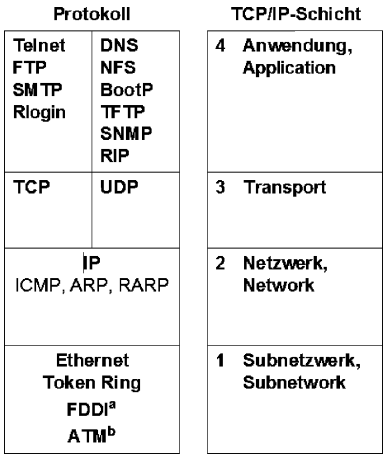


Abb. 55.42 TCP/IP-Referenzmodell

- zuverlässige Übertragung
- standardisierte Anwendungsprotokolle

TCP/IP realisiert diese Anforderungen, indem es Netzknoten als Paketvermittlungsrechner *(Router)* einsetzt und über Routing-Protokolle dynamisches Routing ermöglicht: die Router informieren sich gegenseitig über neue und ausgefallene Teilnetze. TCP nutzt den Paket-Datenübertragungsdienst von IP. IP ist ein verbindungsloses, unzuverlässiges Protokoll, das ggf. auch die Nachricht in kleinere Pakete zerlegt und wieder zusammensetzt; und bei Überlast notfalls auch mal Pakete löschen darf. Die Robustheit gegen Ausfälle von Teilen des Netzes erreicht man mit IP.

TCP sorgt als zuverlässiges Transportprotokoll mit Sicherungsfunktionen für eine sichere Ende-zu-Ende-Datenübertragung. Insbesondere muss TCP die Unzuverlässigkeit von IP wieder ausgleichen.

55.5.6.3 Das TCP/IP-Referenzmodell

Das TCP/IP-Referenzmodell (Abb. 55.42) besteht aus vier Schichten:

1. *Subnetzwerk*: überträgt die Daten; der Dienst wird gebraucht, aber nur die Schnittstelle zu ihm definiert. Hier findet man Ethernet, Token Ring usw.
2. *Netzwerk*: hier sorgt IP für die optimale Zustellung der Pakete durch das Netz.
3. *Transport*: hier kümmert sich TCP um die zuverlässige Zustellung der Nachrichten von Programm zu Programm oder von Prozess zu Prozess; alternativ kann UDP die Daten auch schneller, aber unzuverlässig transportieren
4. *Anwendung*: hier arbeiten die Programme und Prozesse, die über das Netzwerk Daten austauschen wollen

55.5.6.4 TCP/IP-Socket

Zur Adressierung von Anwendungen und Diensten benutzt TCP/IP-*Sockets*. Ein Socket ist ein Kommunikationsendpunkt

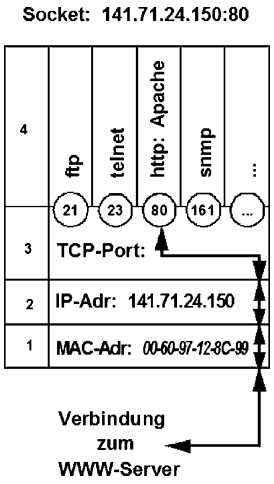


Abb. 55.43 TCP/IP-Socket, Ports

und besteht aus einer 32-Bit-IPv4-Adresse sowie einer 16-Bit-TCP-Portnummer (Abb. 55.43). Mit der IP-Adresse ist eine Station eindeutig identifiziert; mit der Portnummer ist in der Maschine ein Socket eindeutig festgelegt. Jedes Programm, das Daten über das Netz austauschen will, muss dazu vorher einen Socket öffnen.

Mit der 16-Bit-Portnummer sind 65.536 Ports in einer Station möglich; jeder Port ermöglicht einen Datenübertragungskanal. Der Bereich der Ports ist aufgeteilt in diese Bereiche:

0 ... 1023 „*Well-Known Ports*:“ für Serverprozesse
1024 ... 49151 „*Registered Ports*:“ offiziell bei der IANA (Internet Assigned Numbers Authority) beantragte Ports
49152 ... 65535 „*Private Ports*:“ frei verfügbare Ports

Beispiele für „*Registered Ports*“:

Port: Dienst/Anwendung:

21 FTP, File Transfer Protocol
22 ssh, Secure Shell (sicher)
23 Telnet, virtuelles Terminal (unsicher)
53 DNS, Domain Name Service
80 WWW, World Wide Web, HTTP
110 POP3, Post Office Protocol Version 3

55.5.6.5 TCP: Transport Control Protocol

TCP benutzt für seine Aufgaben eine voll duplex-fähige, bidirektionale, sichere, virtuelle Verbindung mit geregeltem Verbindungsauf- und -abbau (*TCP-3-Wege-Handshake*), Abb. 55.44. Der Benutzer sieht nur einen bidirektionalen Datenstrom von Bytes, aber keine Blöcke oder Pakete. Die Transportbenutzer-Adressierung erfolgt über 16-Bit-Portnummern. 32-Bit-Sequenznummern (SEQ) und 32-Bit-Quittungsnummern (ACK) mit Zeitüberwachung, sowie Prüfsummen mit Empfangsquittungen garantieren die Datensicherheit.

Der Verbindungsaufbau findet über das *TCP-3-Wege-Handshake* statt.

1. Eine Station sendet einen Verbindungsaufbauwunsch mit der Flagge SYN = 1 und der Sequenznummer X.
2. Die Gegenstation sendet ihrerseits einen Verbindungsaufbauwunsch für den Rückkanal mit SYN = 1 und ihrer Sequenznummer Y; gleichzeitig bestätigt sie dem Empfang mit über die Sequenznummer ACK = $X + 1$.
3. Die erste Station bestätigt den Aufbau des Rückkanals über die Sequenznummer ACK = $Y + 1$. Die Verbindung ist aufgebaut, die Daten können ausgetauscht werden.

Die Datenflusssteuerung übernimmt das *Sliding Window*: jede Station übermittelt eine Fenstergröße, die die Zahl der unbestätigt, also im Voraus gesendeten Bytes festlegt. Ist diese Zahl von Bytes gesendet, muss der Sender auf eine Bestätigung warten. Kommt eine Bestätigung, so darf der Sender um die Zahl der bestätigten Bytes weiter senden. Ist die Fenstergröße null, so muss der Sender anhalten.

55.5.6.6 UDP: User Data Protocol

UDP ist eine vereinfachte Alternative zu TCP. UDP benutzt wie TCP Portnummern zur Adressierung. Aber im Gegensatz zu TCP ist UDP verbindungslos, hat keine Quittungen, ist unzuverlässig – aber schnell! Datensicherung und Flusssteuerung übernehmen bei Bedarf die Anwendung oder ein zuverlässiges Netzwerk (z. B. Ethernet). UDP eignet sich damit sehr gut im LAN für eine unzuverlässige, aber sehr schnelle Datenübermittlung wie Sprach- und Videoübertragung sowie für lokale Services wie NFS (Network File System) oder DNS (Domain Name Service).



Abb. 55.44 Transport Control Protocol

55.5.7 Aufgaben der Schichten 5–7

Im TCP/IP-Referenzmodell sind die OSI-Schichten fünf bis sieben zu einer Schicht „Anwendung“ zusammengefasst, was in den meisten Fällen auch kein Problem macht.

Im OSI-Modell regelt Schicht 5 *(Sitzungsschicht)* Datentransport und gehobene Dienste, z. B. den Zugang zu Time-Sharing-Systemen, die Dialogsteuerung oder ggf. ein Token-Management (die Operation darf nur ausführen, wer das Token hat). Die Synchronisierung großer Dateien mit Fixpunkten (check points).

Schicht 6 *(Darstellung)* stellt häufig gebrauchte Funktionen für den Anwender bereit, wie Datenkompression, Kryptographie (SSL, Secure Socket Layer) und analysiert ggf. Syntax und Semantik der Daten zur Konvertierung von ASCII in EBCDIC, 1er- in 2er-Komplement oder Little- in Big-Endian.

In Schicht 7 *(Anwendung)* laufen die typischen Server- und Client-Anwendungen: WWW-Server und WWW-Browser; Terminal-Emulationen, Dateitransfer, E-Mail-Server und -Clients, Application Gateway Firewalls und sonstige Anwenderprogramme.

55.5.7.1 DNS: Domain Name Service

Zum Aufbau einer Verbindung benötigt man die IP-Adresse der Gegenstation. In der Anfangszeit benutzte man eine Datei HOSTS.TXT, die den wenig lesbaren IP-Adressen gut lesbare Name zuordnete. Im ARPANET lag eine zentrale Datei HOSTS.TXT mit den Adressen *aller* Stationen im Netz im NIC (Network Information Center), die per FTP von allen Stationen lesbar war und als lokale Kopie im Verzeichnis /etc/hosts auf jeder Unix-Maschine verfügbar war. Mit der stark steigenden Zahl der Stationen wurde diese Praxis zu aufwändig.

Abhilfe brachte der DNS *(Domain Name Service; RFCs 1034, 1035, 1101, 1183, 1032)* als „Internet Directory Service“, der die Zuordnung von symbolischen Namen zu Internet-Adressen über ein hierarchisches, auf Domänen basierendes Benennungsschema einführte. Der DNS beruht auf einer verteilten Datenbank, die DNS-Server verwalten.

55.5.7.1.1 Funktionsweise

Ein Client will eine Verbindung zu einem Ziel-Host aufbauen, hat dessen DNS-Namen und benötigt nun für den Verbindungsaufbau die IP-Adresse (Abb. 55.45). Zur Namensauflösung schickt der Client den Namen über den Resolver (einen DNS-Client) an seinen lokalen DNS-Server, den er aus seiner Netzwerkkonfiguration kennt. Kann der lokale DNS-Server den Namen auflösen, so schickt den Namen mit zugehöriger IP-Adresse an den Client zurück. Kann er den Namen nicht auflösen, so befragt der lokale DNS-Server den in der Namenshierarchie nächst höheren Server. Dieser Vorgang wiederholt sich, bis der DNS-Server gefunden ist, der zu diesem Namen verbindlich Auskunft geben kann. Das Ergebnis ist dann die gesuchte IP-Adresse oder die Auskunft, dass der Name ungültig ist.

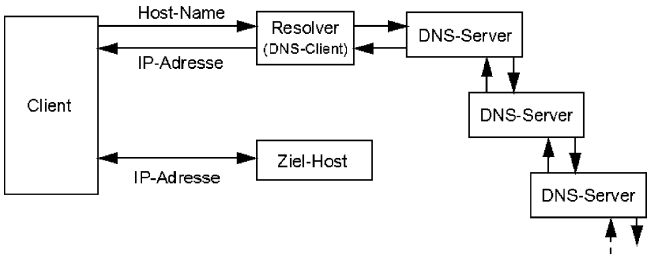


Abb. 55.45 Domain Name Service

DNS-Einträge waren ursprünglich alle statisch, d. h. vom Administrator der Domäne von Hand eingepflegt. Eine Verbesserung ist der DDNS *(Dynamic Domain Name Service; RFC 2136)*. Active Directory von Windows 2000/2003 benutzt ebenfalls einen DDNS als Name Service.

Die DNS-Server im Internet müssen besonders geschützt werden, da verfälschte Adresse/Namenzuordnungen in der DNS-Datenbank die betroffenen Adressen praktisch unerreichbar machen.

55.5.7.1.2 DNS-Namen

Ein vollständiger DNS-Name *(FQDN, Fully Qualified Domain Name)* kann so aussehen:

hostname.subdomain.domain.top-level-domain

Ein Beispiel:

FQDN:	www.etech.fh-hannover.de
top-level-domain:	.de
domain:	.fh-hannover
subdomain:	etech
hostname:	www

Die höchste Hierarchiestufe der Internet-Domänen steht am Ende des Namens; die Namen der einzelnen Hierarchiestufen sind durch Punkte getrennt. Groß- und Kleinschreibung wird nicht beachtet. Zusammengesetzte Namen können bis 63 Zeichen lang sein; der gesamte Pfadname kann maximal 255 Zeichen lang sein. Seit kurzem sind auch deutsche Umlaute (ä, ö, ü) im Namen erlaubt (*aber noch nicht in allen Programmen problemlos benutzbar*).

Einige allgemeine Top-Level Domänen sind:

.int	internationale Organisationen
.mil	Amerikanische Streitkräfte
.com	Kommerzielle Unternehmen
.net	Netzbetreiber
.edu	education: Universitäten u. ä.
.org	nicht-kommerzielle Organisationen
.gov	US-Bundesregierung

Ebenso hat jedes Land eine eigene Top-Level Domäne:

.de	Deutschland
.fr	Frankreich
.jp	Japan
.tv	Tuvalu (Inselstaat im Stillen Ozean)
	usw.

55.5.7.2 DHCP: Dynamic Host Configuration Protocol

Jede Station muss eine netzwerkweit eindeutige IP-Adresse haben. In großen Netzwerken kann dies schnell unübersichtlich werden.

Abhilfe schafft ein DHCP-Server, der dynamisch IP-Adressen und Netzmasken für den Client aus einem Bereich von freien IP-Adressen an die DHCP-Clients verteilt. Neben den Adressen verteilt der DHCP-Server auch die IP-Adressen von DNS-Server, Standard-Gateway, den DNS-Domänennamen u. a. m. Der DHCP-Client erhält eine *Lease* mit allen IP-Parametern, deren Nutzungsdauer (*Lease Time*) allerdings zeitlich begrenzt ist (z. B. auf 24 Stunden).

Der DHCP-Client erhält beim Booten vom DHCP-Server seine Konfiguration. Bieten mehrere DHCP-Server eine Lease an, kann der Client sich eine Lease aussuchen.

Der Vorteil des DHCP-Dienstes liegt in der zentralen Verwaltung aller Clients über den DHCP-Server. Der DHCP-Client erhält nach dem Starten immer eine aktuelle IP-Konfiguration. Damit gibt es weniger Konfigurationsfehler und der Aufwand zur Netzadministration sinkt; Änderungen in der Netzwerkkonfiguration werden einfach via DHCP verteilt. Neu in Netz aufgenommene Stationen werden automatisch mit den passenden Konfigurationsdaten versorgt.

Bei Bedarf können feste Adressen von der dynamischen Vergabe ausgeschlossen werden. Für Router, Server oder Drucker können auch feste Adressen reserviert werden; die Erkennung der Geräte erfolgt über die MAC-Adressen. In Netzen mit hohen Sicherheitsanforderungen können auch nur Clients mit bekannten MAC-Adressen Zugang zum Netz erhalten. Damit kann sich ein Angreifer nicht ohne weiteres ins Netz einschleichen.

Die Erneuerung der DHCP-Lease erfolgt immer beim Neustart des Clients innerhalb der Gültigkeitsdauer. Ist die Lease-Time zu 50 % abgelaufen, dann versucht der DHCP-Client „prophylaktisch“, seine Lease beim aktuellen DHCP-Server zu verlängern um einen unterbrechungsfreien Netzwerkbetrieb zu gewährleisten. Ist der DHCP-Server nicht erreichbar und die Lease-Time ist zu 87,5 % abgelaufen, dann der DHCP-Client sucht mit einem Broadcast nach beliebigen DHCP-Servern im Netz. Gelingt auch dies nicht und ist die Lease-Time zu 100 % abgelaufen, dann muss der Client seine Lease freigeben; und es ist für ihn kein weiterer Zugriff auf das Netz mehr möglich.

Eine Anzeige der aktuellen IP-Konfiguration erhält man unter Windows 9x mit dem Befehl *winipcfg* (ausgeführt in einer DOS-Konsole); unter Windows NT/2000/2003 mit *ipconfig/all* und unter Linux mit *ifconfig*.

55.6 Industrial Ethernet

Industrial Ethernet beschreibt die Anwendungen von Ethernet in industriellen Produktionsprozessen. Der Einsatz von Ethernet-Standards in der industriellen Automatisierungstechnik hat diese Vorteile:

- die Ethernet-Technologie ist bekannt, bewährt und preiswert
- die Fertigung und die Produkte lassen sich verbessern und Kosten reduzieren
- Fernwartung, Ferndiagnose und Fernzugriff über WAN sind möglich
- nur eine Netzwerk-Technologie vom Büro bis zur Produktionsmaschine ist zu pflegen

Die transparente Kommunikation über alle Geschäftsbereiche, von der Unternehmensleitebene (ERP-Systeme, Enterprise Resource Planning) über die Nutzung etablierter IT-Techniken zur Datenvisualisierung und Kommunikation bis hin zur Datenerfassung und Steuerung von Maschinen in der Produktionszelle kann mit dem Ethernet-Standard abgedeckt werden.

Gegenüber dem Einsatz im Büro ergeben sich je nach Einsatzort deutlich höhere Anforderungen an die Robustheit der Netzwerkkabel und Netzwerkgeräte, speziell hinsichtlich der Umgebungsbedingungen, z. B. bei mechanischen Beanspruchungen, Vibrationen, Schmutz und Feuchtigkeit, Temperaturbereich, UV-Strahlung und Resistenz gegenüber elektromagnetischen Störungen.

Echtzeit-Betrieb

Industrielle Steuerungen müssen oft schnell und zuverlässig reagieren: soll die Bewegung eines mehrachsigen Roboters in Echtzeit über ein Netzwerk gesteuert werden, so müssen alle Steuerdaten nicht nur hinreichend *rechtzeitig* (nicht zu spät), sondern auch *gleichzeitig* (nicht zu früh *und* nicht spät) eintreffen. Die Abweichung von der Sollzeit heißt „Jitter“; dieser muss für die Anwendung klein genug sein. Meist ist hierfür eine Synchronisation der Uhren in allen Stationen notwendig, z. B. über das Precision Time Protocol (PTP, IEEE 1588).

Die benutzten Industrial-Ethernet-Protokolle unterscheiden sich stark in ihrer Echtzeitfähigkeit und dem Grad der dazu notwendigen Modifikation von Standard-Ethernet-Software und -Hardware:

Software-Modifikationen:

- Modbus-RTPS: Zykluszeit 5 . . . 10 ms

zusätzlich mit IEEE 1588-Hardware:

- EtherNet/IP mit CIPsync: Zykluszeit 300 µs bei 30 Antrieben

Spezieller Netzwerk-Controller (ASIC/FPGA):

- PROFInet Echtzeit-Subnetz V3: Zykluszeit 1 ms bei 100 Antrieben
- SERCOS III: Zykluszeit 250 µs bei 100 Antrieben
- EtherCAT: Zykluszeit 100 µs bei 100 Antrieben

Literatur

1. Tanenbaum, A. S.: Computernetzwerke, 4., aktualisierte Auflage, Pearson, 2003
2. c’t extra 03/2009 Netzwerke, Heise
3. c’t magazin für computertechnik, Heise
4. http://www.heise.de/netze/
5. http://de.wikipedia.org/wiki/Rechnernetz

Automatisierungs-technik

VIII



56 Einführung 773
57 Automatisierungsgeräte 775
58 Grundzüge der SPS-Norm IEC 61131-3 779
59 Programmstrukturen 787
60 Eingabe- und Ausgabesignale 789
61 Eingabe-/Ausgabebaugruppen 791
62 Verknüpfungssteuerungen 793
63 Aufruf und Wertübergaben zwischen Bausteinen nach IEC 61131-3 803
64 Ablaufsteuerung 807
65 Kommunikation in Automatisierungssystemen 813
66 Steuerungssicherheit 817
67 Regelungstechnische Grundbegriffe der Automatisierungstechnik 821
68 Regelstrecken 825
69 Regler 831

Einführung

56

Günter Wellenreuther und Dieter Zastrow



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_56

Automatisierung Reale Aufgaben der Automatisierungstechnik sind im Allgemeinen sehr komplex. Als umfassender Ausdruck für Steuerungs-, Regelungs- und Visualisierungs-Vorgänge hat sich der Begriff der Automatisierung durchgesetzt. Er beinhaltet, dass Automatisierungsgeräte selbsttätig Programme befolgen und dabei Entscheidungen auf Grund vorgegebener Führungsgrößen und rückgeführter Prozessgrößen aus der Anlage sowie erforderlicher Daten aus internen Speichern des Systems treffen, um daraus notwendige Ausgangsgrößen für den Betriebsprozess zu bilden.

Steuerung Steuern oder Steuerung wird als Ablauf in einem System definiert, bei dem eine oder mehrere Eingangsgrößen andere Größen als Ausgangsgrößen aufgrund der dem System eigentümlichen Gesetzmäßigkeiten beeinflussen. Kennzeichen für das Steuern ist der *offene Wirkungsablauf über der Steuerstrecke*. Eine Steuerung liegt also vor, wenn Eingangsgrößen nach einer festgelegten Gesetzmäßigkeit Ausgangsgrößen beeinflussen. Die Auswirkung einer nicht vorhersehbaren Störgröße wird nicht ausgeglichen.

Regelung Immer dann, wenn Störgrößenänderungen das System nicht hinnehmbar beeinflussen können, werden Regelungen erforderlich. Die Regelgröße (Aufgabengröße) muss sich messtechnisch erfassen lassen, denn eine Regelung ist ein Vorgang, bei dem die Regelgröße fortlaufend erfasst, mit der Führungsgröße verglichen und abhängig vom Ergebnis dieses Vergleichs im Sinne einer Angleichung an die Führungsgröße beeinflusst wird. Der sich dabei ergebende Wirkungsablauf findet in einem geschlossenen Kreis, dem *Regelkreis*, statt. Programmierbare Automatisierungsgeräte können die Aufgaben des Steuerns und Regelns ausführen, denn beide Funktionen beruhen auf Programmen unter Verwendung desselben Operationsvorrates.

Automatisierungsgeräte

57

Günter Wellenreuther und Dieter Zastrow



Teil VIII

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_57

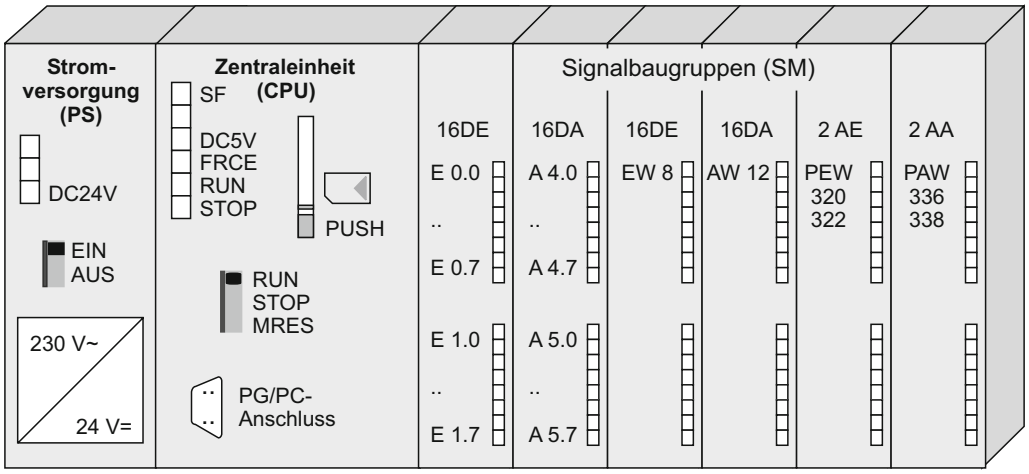


Abb. 57.1 Aufbau einer Speicherprogrammierbaren Steuerung ohne Feldbusanschluss. DE = Digitale Eingänge, DA = Digitale Ausgänge, AE = Analoge Eingänge, AA = Analoge Ausgänge

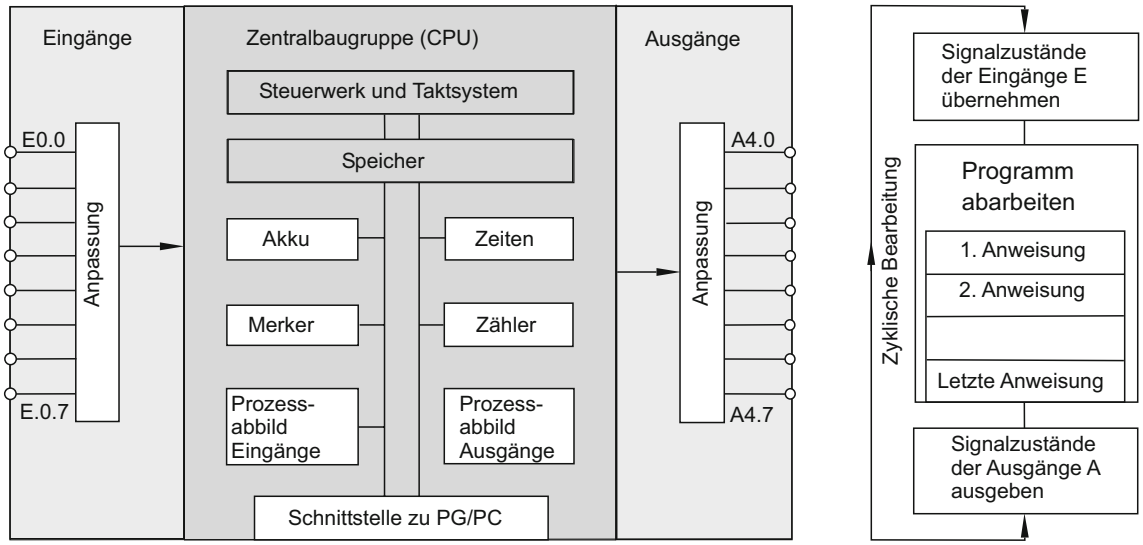
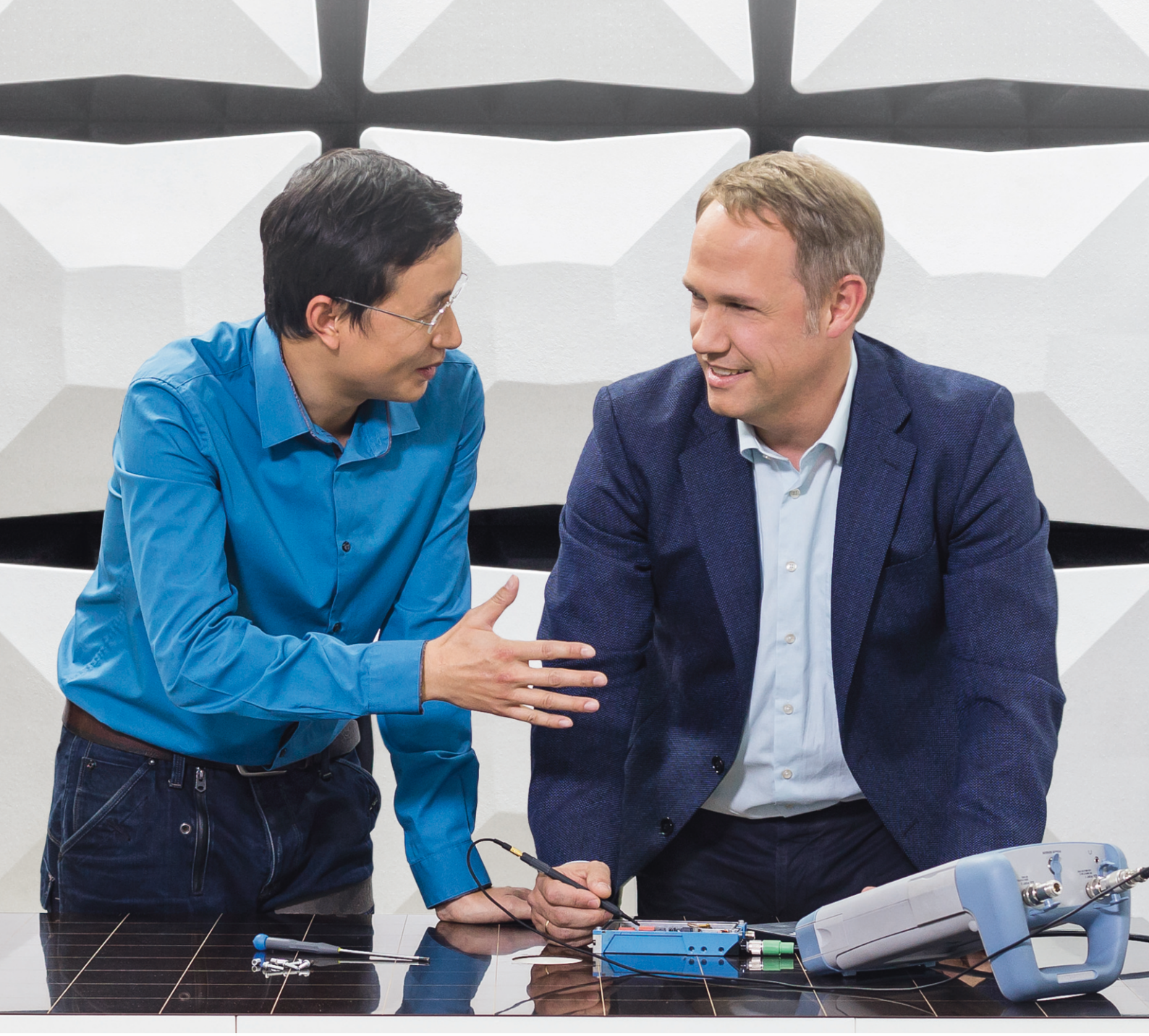


Abb. 57.2 SPS-Struktur und zyklische Programmbearbeitung

Die derzeit am weitesten verbreitete Hardware-Plattform der Steuerungstechnik ist die *Speicherprogrammierbare Steuerung SPS*, wie sie in Abb. 57.1 abgebildet ist, dort jedoch ohne den heute schon üblichen Anschluss an ein Feldbussystem zur Vernetzung mit anderen Steuerungskomponenten.

Eine Speicherprogrammierbare Steuerung hat die Struktur eines Rechners, deren Funktion als Programm gespeichert ist. Sie besteht im einfachsten Fall aus einer *Stromversorgung PS*, einem *Steuerungsprozessor CPU*, einigen zentralen *digitalen Eingabe- und Ausgabebaugruppen* sowie einem internen Bussystem. Bei Bedarf können auch Baugruppen zur *Analogwertverarbeitung* oder für besondere Funktionen wie *Regler*, schnelle *Zähler* und *Positionierungen* hinzukommen. Die Peripheriebaugruppen und die Programmiersprachen sind auf die Belange der Steuerungstechnik ausgerichtet. Speicherprogrammierbare Steuerungen gibt es als modulare und kompakte Systeme für unterschiedliche Anforderungsniveaus. Ein typisches Merkmal von SPS-Steuerungen ist die zyklische Programmbearbeitung, wie in Abb. 57.2 angedeutet.

In der Praxis werden nicht mehr ausschließlich SPS-basierte Automatisierungssysteme eingesetzt, sondern vernetzte Steuerungssysteme bestehend aus SPSen und PCs mit unterschiedlichen Aufgaben, wobei die SPS mehr prozessnah und der PC mehr übergeordnet und datenverarbeitend genutzt wird. Soll die Steuerungsfunktionalität nicht mehr von einer modernen SPS, sondern von einem PC ausgeführt werden, so gibt es diese Lösung als Industrie-PC mit eingebauter SPS-Karte und zusätzlicher Feldbus-Schnittstelle. In Sonderfällen kommt auch eine sog. Soft-SPS-Lösung in Frage, bei der in einem PC die Funktion einer SPS softwaremäßig nachgebildet und Steuerungsaufgaben „nebenher“ miterledigt werden. Für die Programmerstellung macht es keinen Unterschied, ob die Hardware-Plattform ein PC oder eine SPS ist.



JE HELLER DER KOPF, UMSO BRILLANTER DIE IDEE.

THIS IS SICK

Sensor Intelligence.

GREAT PLACE TO WORK® DEUTSCHLANDS BESTE ARBEITGEBER 2016

Visionäre Denkerinnen und Denker mit Erfindergeist gesucht. Entwickeln Sie gemeinsam mit uns Sensorlösungen, die weltweit Standards setzen und die nächste industrielle Revolution mitgestalten. Ihre Karriere: anspruchsvoll, abwechslungsreich und mit besten persönlichen Entwicklungschancen. Ihr Umfeld: hochprofessionell, international und inspirierend. Ihr neuer Arbeitgeber: ein Hightech-Unternehmen mit weltweit mehr als 7.400 Mitarbeiterinnen und Mitarbeitern. Ihre Zukunftsadresse: www.sick.de/karriere

Grundzüge der SPS-Norm IEC 61131-3

58

Günter Wellenreuther und Dieter Zastrow



58.1 Programmorganisationskonzept 780
58.2 Deklaration von FB- und FC-Bausteinen 780
58.3 Variablen 782
58.4 Programmiersprachen 784

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_58

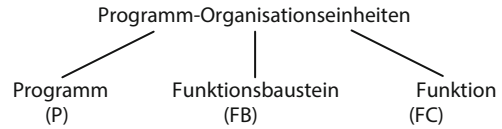
Auch wenn das Programmieren von Steuerungen und Regelungen noch immer ein sehr individueller Prozess ist, muss seit Jahren der Standard der SPS-Programmier-Norm IEC 61131-3 zu Grunde gelegt werden, der eine rationellere Programmerstellung und größere Herstellerunabhängigkeit zum Ziel hat. Den Anwendungsprogrammierern müssen wirksame Mittel an die Hand gegeben werden, um wiederverwendbare und damit kostensparende Programme entwickeln zu können. Die Norm IEC 61131-3 bzw. DIN EN 61131-3 richtet sich daher in erster Linie an SPS-Hersteller. Für Anwendungsprogrammierer ist die SPS-Norm eher ein Dokument im Hintergrund, denn sie sind für ihre Arbeit auf ein reales IEC 61131-3 kompatibles Programmier- und SPS-System angewiesen.

Die Grundzüge dieser SPS-Norm werden in den folgenden Abschnitten dargelegt und bilden die Voraussetzungen dafür, mit einem der SPS-Norm entsprechenden Programmiersystem richtig umgehen zu können.

58.1 Programmorganisationskonzept

Es ist ein hierarchisch gegliedertes System von so genannten Programmorganisationseinheiten (POE) eingeführt worden, bestehend aus einem Hauptprogrammtyp (Schlüsselwort: PROGRAM) mit Zugriffsmöglichkeit auf SPS-Eingänge/-Ausgänge und zwei Unterprogrammtypen davon einen Typ mit Gedächtnisfunktion (Schlüsselwort: FUNCTION_BLOCK) und einen Typ ohne Gedächtnisfunktion (Schlüsselwort: FUNCTION). Durch dieses Programmorganisationskonzept in Verbindung mit dem Datentyp- und Variablenkonzept zur Entwicklung strukturierter Programme wird die Wiederverwendbarkeit von Bausteinen bei späteren Anwendungen ermöglicht. Dadurch ist die neue Anwendungsprogrammierung sehr viel anspruchsvoller als die herkömmliche SPS-Programmierung geworden.

Ein Steuerungsprogramm (Anwenderprogramm) ist eine in Programm-Organisations-Einheiten (POE) gegliederte Einheit. Anstelle der unhandlichen Bezeichnung Programmorganisationseinheit wird vereinfacht auch nur von Bausteinen gesprochen.



Jeder Baustein besteht aus einem

- Deklarationsteil für die Definition aller lokal verwendeten Variablen und einem
- Rumpf für die Anweisungen (Befehle) des ausführbaren Programms.

Funktion Dieser POE-Typ ist geeignet, wenn ein Funktionsergebnis ausschließlich aus den Eingangsvariablen des Bausteins zu ermitteln ist und unter dem Funktionsnamen des Bausteins zur Verfügung gestellt werden soll. Der Aufruf einer Funktion mit denselben Werten der Eingangsvariablen liefert deshalb immer denselben Ausgangswert zurück. Die SPS-Norm enthält einen Katalog von Standardfunktionen, die in SPS-Systemen zur Verfügung stehen sollten. Falls eine spezielle Funktion benötigt wird, kann diese vom Anwender selbst erzeugt werden, dabei muss jedoch beachtet werden, dass keine internen Zustandsvariablen deklarierbar sind, da der Bausteintyp Funktion dafür keine Speicherfähigkeit (Gedächtnis) besitzt. Eine Funktion stellt das Funktionsergebnis unter dem deklarierten Funktionsnamen zur Verfügung, sodass keine Ausgangsvariable deklariert werden muss. Es ist jedoch zulässig, Funktionen mit mehreren Ausgangsvariablen zu bilden. Funktionen können innerhalb eines Programmzyklus mehrfach aufgerufen werden, um mit unterschiedlichen Werten der Eingangsvariablen entsprechende Funktionsergebnisse zu ermitteln.

Funktionsbaustein Dieser POE-Typ ist geeignet, wenn aus den Werten von Eingangs- und Ausgangsvariablen sowie bausteininterner Zustandsvariablen neue Ergebnisse für eine oder mehrere Ausgangsvariablen ermittelt werden sollen. Alle Werte der Ausgangs- und Zustandsvariablen bleiben von einer Bearbeitung des Funktionsbausteins bis zur folgenden erhalten. Das bedeutet, dass es bei einer erneuten Bearbeitung des Funktionsbausteins mit den gleichen Werten der Eingangsvariablen zu anderen Ausgangsergebnissen kommen kann. Anschaulich spricht man hier von einem Bausteintyp mit Gedächtnis. Um die Fähigkeiten eines Funktionsbausteins in einem Programm auch mehrfach nutzen zu können, ist die sog. *Instanziierung* der Funktionsbausteine erforderlich, worunter man das Erzeugen einer Kopie (Instanz) des Bausteines versteht. Jede Instanz muss mit einem eigenen Namen versehen werden. Unter dem Instanznamen werden die jeweils letztgültigen Variablenwerte auf entsprechenden Speicherplätzen verwaltet, während das Bausteinprogramm im Original verbleibt und dort den Instanzen bei deren Ausführung zur Verfügung steht. Die SPS-Norm schlägt viele Standardfunktionsbausteine vor, die in SPS-Systemen verfügbar sein sollten. Spezielle Funktionsbausteine können vom Anwender selbst erzeugt werden. Auch sie unterliegen bei ihrer Anwendung der Instanzenbildung.

Programm Dieser POE-Typ bildet die oberste Hierarchieebene der Programmorganisationseinheiten. Die SPS-Norm sieht vor, dass Programme (P) in Ressourcen (SPS-Systeme) auch instanziiert werden können. Einige SPS-Systeme verwenden den Bausteintyp Programm (P) als alleiniges Hauptprogramm zur Organisation des Anwenderprogramms, dass aus Funktionen (FC) und Funktionsbausteinen (FB) besteht. Der Programminhalt eines solchen (Haupt-)Programms besteht dann nur aus Aufrufen der Funktionen und Funktionsbausteine und um deren Eingangs-/Ausgangs-Variablen mit realen SPS-Ein-/Ausgängen zu verbinden.

58.2 Deklaration von FB- und FC-Bausteinen

Anwenderprogramme bestehen immer auch aus sog. abgeleiteten Funktionen (FC) und/oder Funktionsbausteinen (FB), die erst durch Deklaration und Programmierung erzeugt werden

müssen. Dabei sind Vorschriften zu beachten. Die Deklaration bezieht sich auf die Festlegung des Bausteintyps und auf die Bildung der Außenschnittstelle mit ihren Eingangs- und Ausgangsvariablen (-parametern) sowie der außen nicht erkennbaren bausteininternen Zustandsvariablen bei Funktionsbausteinen. Die Programmierung bezieht sich auf den Bausteinrumpf, der die Steuerungslogik enthalten muss. Deklaration und Programmierung kann in Textform oder in Grafik erfolgen.

Erstellen einer Funktion mit dem Funktionsnamen FC 1

Bei der Deklaration in Textform sind folgende Elemente zu verwenden:

- das einleitende Schlüsselwort **FUNCTION** gefolgt vom Funktionsnamen, einem Doppelpunkt und dem Datentyp des Funktionswertes,
- das Konstrukt **VAR_INPUT...END VAR**, mit dem die Namen und Datentypen der Eingangsvariablen der Funktion festlegt werden,
- das Konstrukt **VAR_OUT...END VAR** für die Namen und Datentypen von Ausgangsvariablen,
- das Konstrukt **VAR_IN_OUT...END VAR**, das die Namen und Datentypen von Durchgangsvariablen der Funktion festlegt,
- falls erforderlich das Konstrukt **VAR...END VAR**, mit dem die Namen und Datentypen von internen temporären Hilfsvariablen festlegt werden können, deren Daten jedoch bei Beendigung der Funktion verloren gehen,
- einen Funktionsrumpf mit dem auszuführenden Programm,
- das abschließende Schlüsselwort **END_FUNCTION**.

Beispiel 1

Deklaration einer Funktion FC

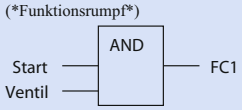
Allgemein	Ausführung in Textform
FUNCTION FC1 : BOOL (*Außenschnittstelle*) **VAR_INPUT** Bezeichner1:Datentyp; Bezeichner2:Datentyp; **END_VAR** (*Funktionsrumpf*) Programm **END_FUNCTION**	**FUNCTION FC1 : BOOL** (*Außenschnittstelle*) **VAR_INPUT** Start:BOOL; Ventil:BOOL; **END_VAR** (*Funktionsrumpf*) AND: Start, Ventil → FC1 **END_FUNCTION**

Ausführung in Grafik

FUNCTION FC1 : BOOL
(*Außenschnittstelle*)



(*Funktionsrumpf*)



END_FUNCTION

◄

Erstellen eines Funktionsbausteins mit dem Namen FB 1

Die Deklaration eines Anwender-Funktionsbausteins in Textform erfolgt in ähnlicher Weise wie bei der Anwender-Funktion, dabei sind folgende Elemente zu verwenden:

- das einleitende Schlüsselwort **FUNCTION_ BLOCK** gefolgt vom Funktionsbausteinnamen ohne einen Datentyp,
- das Konstrukt **VAR_INPUT...END VAR**, mit dem die Namen und Datentypen der Eingangsvariablen des Funktionsbausteins festlegt werden,
- das Konstrukt **VAR_OUTPUT...END VAR**, mit dem die Namen und Datentypen der Ausgangsvariablen des Funktionsbausteins deklariert werden (mehrere Ausgangsvariablen sind zulässig),
- das Konstrukt **VAR_IN_OUT...END VAR**, das die Namen und Datentypen von Durchgangsvariablen des Funktionsbausteins festlegt, deren Werte innerhalb des Bausteins durch das Programm verändert werden dürfen,
- das Konstrukt **VAR...END VAR**, das die Namen und Datentypen der bausteininternen Zustandsvariablen des Funktionsbausteins festlegt. Dieses Konstrukt wird auch verwendet, um im Funktionsbaustein eine Instanz eines Standardfunktionsbausteins zu erzeugen,
- einen Funktionsbausteinrumpf mit dem auszuführenden Programm,
- das abschließende Schlüsselwort **END_FUNCTION_BLOCK**.

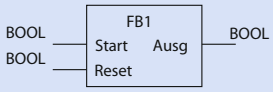
Beispiel 2

Deklaration eines Funktionsbausteins FB

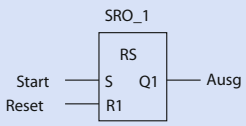
Allgemein	Ausführung in Textform
FUNCTION_BLOCK FB 1 (*Außenschnittstelle*) **VAR_INPUT** Bezeichner_1:Datentyp; Bezeichner_2:Datentyp; **END_VAR** **VAR_OUTPUT** Bezeichner_3:Datentyp; **END_VAR** **VAR** Bezeichner_4:Datentyp; **END_VAR** (*Funktionsbausteinrumpf*) Programm **END_FUNCTION_BLOCK**	**FUNCTION_BLOCK FB 1** (*Außenschnittstelle*) **VAR_INPUT** Start:BOOL; Reset:BOOL; **END_VAR** **VAR_OUTPUT** Ausg:BOOL; **END_VAR** **VAR** SRO_1:RS; **END_VAR** (*Funktionsbausteinrumpf*) SRO_1: RS (Start, Reset → S, R1; Q1 → Ausg) **END_FUNCTION_BLOCK**

Ausführung in Grafik

FUNCTION_BLOCK FB 1
(*Außenschnittstelle*)



(*Funktionsbausteinrumpf*)



END_FUNCTION_BLOCK

◄

Tab. 58.1 Elementare Datentypen

Schlüsselwort	Datentyp	Größe	Schreibweise und Wertebereiche
Bit-Datentypen			
BOOL	Boolesche Variablen	1 Bit	FALSE, TRUE
BYTE	Bit-Folge und 8 Bit-Hex-Zahlen	8 Bit	16# 0...FF
WORD	Bit-Folge und 16 Bit-Hex-Zahlen	16 Bit	16# 0000...FFFF
DWORD	Bit-Folge und 32 Bit-Hex-Zahlen	32 Bit	16# 0000_0000...FFFF_FFFF
CHAR	ASCII-Zeichen	8 Bit	'A'
Arithmetiktypen			
INT	Ganze Zahlen (Festpunktzahlen)	16 Bit	−32.768 bis +32.767
DINT	Ganze Zahlen (Festpunktzahlen)	32 Bit	L# −2.147.483.648 bis +2.147.483.647
REAL	Reelle Zahlen (Gleitpunktzahlen)	32 Bit	Dezimalzahl mit Punkt: 341,7 oder Exponentialdarstellung: 3,417E+02
Zeittypen			
TIME	Zeitdauer (IEC-Format)	32 Bit	TIME# −24d20h31m bis +24d20h31m
TIME OF DAY	Uhrzeit (Tageszeit)	32 Bit	TIME_OF_DAY#23:59:59,9
DATE	Datum	16 Bit	DATE#1990-01-01

Hinweis: Es existiert kein besonderer Datentyp für BCD-Zahlen (Binär Codierte Dezimalzahlen), diese sind eine Teilmenge der Hexadezimalzahlen, für die es die Datentypen BYTE, WORD und DWORD gibt.

58.3 Variablen

In Funktionen (FC) und Funktionsbausteinen (FB) sollte nur mit symbolischen Variablen programmiert werden, um damit bibliotheksfähige Programme zu erhalten, die keine Festlegungen bezüglich der Verwendung realer SPS-Eingänge/Ausgänge oder Merker enthalten. Erst auf der Ebene der Programme (P) sollte die Zuordnung der SPS-Eingänge/Ausgänge/ Zähler/Zeitglieder zu den direkten (realen) Eingangs- und Ausgangsvariablen erfolgen.

Eine *Variable* ist ein mit einem Namen (Bezeichner) versehener Speicherplatz, der im Anwenderprogramm als Platzhalter für Daten fungiert, die sich zur Laufzeit des Programms ändern können. Diese Variablen können symbolisch oder direkt adressiert sein.

Unter *Daten* sollen hier Informationen aus technischen Anlagen verstanden werden, wie z. B. Messdaten über Temperaturen, Füllstände, Durchflussmengen, die verarbeitet und gespeichert werden müssen. Die Variablen sind die Mittel, um die Daten zu erfassen. Dabei wird für die Variablen ein bestimmter Datentyp festgelegt. Dieser Datentyp hängt direkt zusammen mit den auf ihn zulässigen Operationen.

Einzelelement-Variablen enthalten einzelne Datenelemente mit einfachem Datentyp, die in der Tab. 58.1 auszugsweise dargestellt und bei der Deklaration der Variablen zu berücksichtigen sind.

Die Deklaration symbolischer Variablen erfolgt im dafür vorgesehenen Deklarationsteil der Bausteine unter einem Namen (Bezeichner) durch Verwendung von Schlüsselwörtern wie VAR, VAR_INPUT, VAR_OUTPUT, VAR_IN_OUT und Angabe eines Datentyps wie bei der Deklaration der Bausteine FC und FB bereits gezeigt. Der Speicherort wird durch das Betriebssystem automatisch festgelegt. Der Gültigkeitsbereich einer symbolischen Variablen ist lokal auf den Baustein beschränkt, in dem sie deklariert wurde, d. h. die Variable ist nur in diesem Baustein bekannt. Eine Ausnahme von dieser Regel bilden symbolisch adressierte Globalvariablen, die in allen Bausteinen der SPS bekannt sind aber nur sparsam verwendet werden sollten. Bei der Deklaration von Variablen in Textform sind die in Tab. 58.2 angegebenen Schlüsselwörter zu verwenden. Bei grafischer Deklaration führt das Programmiersystem die Anwendung dieser Regeln im Hintergrund aus.

Tab. 58.2 Schlüsselwörter zur Deklaration von Variablen

Schlüsselwort	Bezeichnung und Gebrauch der Variablen
VAR_INPUT ... END_VAR	Eingangsvariable: von außerhalb kommend, nicht innerhalb des Bausteins änderbar
VAR_OUTPUT ... END_VAR	Ausgangsvariable: nach außen lieferbar
VAR_IN_OUT ... END_VAR	Durchgangsvariable: von außen kommend, innerhalb des Bausteins änderbar, nach außen lieferbar
VAR ... END_VAR	Zustandsvariable: Gebrauch nur innerhalb des Bausteines
VAR_GLOBAL ... END_VAR	Globalvariable: Globaler Geltungsbereich in SPS
VAR RETAIN	Zustandsvariable gepuffert (Remanenzverhalten)
CONSTANT	„Konstante Variable", nicht veränderbar
AT	Zuweisung eines direkten Speicherortes

Multielement-Variablen enthalten mehrere Datenelemente die in Feldern (Arrays) oder Strukturen zusammengefasst sind.

Feld (Array)

Ein Feld ist eine Sammlung von Datenelementen des gleichen Datentyps, die durch in eckigen Klammern [ ] angegebene Feldindizes angesprochen werden. Als Datentyp für Indizes sind

Tab. 58.3 Präfix für Speicherort und Größe der Operanden

Präfix	Bedeutung	Beispiele für direkte Variablen	
I	Speicherort Eingang	Einzel-Eingänge	%IX0,7 ... %IX0,0
Q	Speicherort Ausgang	Einzel-Ausgänge	%QX0,7 ... %QX0,0
M	Speicherort Merker	Eingangsbyte	%IB0 = %IX0,7 ... %IX0,0
X	(Einzel)-Bit-Größe	Ausgangsbyte	%QB0 = %QX0,7 ... %QX0,0
B	Byte-(8 Bit) Größe	Eingangswort	%IW0 = %IB0 + %IB1
W	Wort-(16 Bit) Größe	Ausgangswort	%QW0 = %QB0 + %QB1
D	Doppelwort-(32 Bit)	Eingangsdoppelwort	%ID0 = %IW0 + %IW1

z. B. INT, WORD, BYTE, BOOL zulässig. Es gibt eindimensionale und mehrdimensionale Felder.

Beispiel 3

Eindimensionales Feld

```
(*Deklaration*)
VAR
  Tabelle : ARRAY [0..3] OF BYTE := 16#00, 16#0F, 16#80, 16#FF ;
END_VAR
VAR_OUTPUT
  Wert : BYTE ;
END_VAR
(*Abfrage*)
  Wert := Tabelle [2] ;
```

◀

Struktur

Eine Struktur ist eine Sammlung von Datenelementen unterschiedlicher Datentypen. Die Datenelemente sind in einer Hierarchie geordnet, z. B. Produkt, Version, Seriennummer, Datum.

Direkte Variablen

In der obersten Hierarchieebene der Programmorganisationseinheiten steht der Bausteintyp Programm (P). In diesem Baustein müssen nicht nur die zum Anwenderprogramm gehörenden Funktionsbausteine (FB) und Funktionen (FC) aufgerufen, sondern auch die Verbindungen zu den SPS-Eingängen/Ausgängen hergestellt werden. Nach den Vorschriften der Norm IEC 61131-3 sind diese Adressen dem Programm (P) jedoch nicht automatisch bekannt, d. h. sie müssen erst durch Deklaration bekannt gemacht werden, dazu dienen die sog. direkten Variablen.

Bei der direkten Adressierung von Variablen wird als Variablenname der physikalische Speicherort des Datenelements verwendet, also ein SPS-Eingang/Ausgang oder auch Merker. Zur Unterscheidung der direkten Variablen von ihrem Speicherort wird ein vorgesetztes Prozentzeichen (%) gefolgt von einem Präfix zur Kennzeichnung des Speicherortes und ein Präfix für die Speichergröße verwendet, wie in der Tab. 58.3 angegeben.

Beispiel 4

Deklaration einer direkten Variablen

Allgemein

```
VAR
  AT %Operand : Datentyp ;
END_VAR
```

Ausführung in Textform

```
VAR
  AT %IX4,7 : BOOL ;
END_VAR
```

◀

In einer zweiten Variante können zur Erzielung einer besseren Programmlesbarkeit Variablennamen eingeführt werden, die jedoch im Unterschied zu den richtigen symbolischen Variablen direkt mit dem physikalischen Speicherort (SPS-Eingang/Ausgang, Merker) verbunden sind (Beispiel 5).

Beispiel 5

Deklaration einer direkten Variablen mit symbolischen Namen

Allgemein

```
VAR
  Bezeichner AT %Operand : Datentyp ;
END_VAR
```

Beispiel

```
VAR
  Endschalter AT %IX4,7 : BOOL ;
END_VAR
```

◀

Anm.: Direkte Variablen sollten nicht in den Bausteintypen FB und FC verwendet werden, weil diese Programmteile dadurch hardwarespezifische Festlegungen enthalten würden wie beispielsweise, an welchen SPS-Eingang ein bestimmter Sensor angeschlossen wird. Das aber widerspricht der Forderung nach einer bibliotheksfähigen Gestaltung von FB- und FC-Bausteinen.

Tab. 58.4 Operatoren der Anweisungsliste (AWL)

Operator	Modifizierer	Operand/Typ	Bedeutung
LD (Load)	N	Variable/Bool	Setzt das aktuelle Ergebnis (AE) dem Operanden gleich
ST (Store)	N		Speichert das aktuelle Ergebnis (boolesche 1 oder 0) auf die Operandenadresse
S (Set)			Setzt booleschen Operator auf 1
R (Reset)			Setzt booleschen Operator auf 0 zurück
AND	N, (		Boolesches UND
OR	N, (		Boolesches ODER
XOR	N, (		Boolesches Exklusiv-ODER
)			Bearbeitet die eingeklammerte Operation
CAL (Call)	C, N	Instanzname	Aufruf eines Funktionsbausteins
JMP	C, N	Marke	Sprung zur Marke
GT	(		Vergleich auf größer als, $>$
GE	(		Vergleich auf größer gleich, $\geq$
EQ	(		Vergleich auf ist gleich, $=$
NE	(		Vergleich auf ungleich, $\neq$
LE	(		Vergleich auf kleiner gleich, $\leq$
LT	(		Vergleich auf kleiner als, $<$
ADD	(		Addition
SUB	(		Subtraktion
MUL	(		Multiplikation
DIV	(		Division

Globale Variablen

Globale Variablen sind solche, die durch eine entsprechende Deklaration in allen Bausteintypen innerhalb einer SPS bekannt gemacht wurden, also einen globalen Geltungsbereich haben. Der Deklarationsort ist der Baustein Programm (P).

Globale Variablen werden verwendet, um auf einfache Art einen bausteinübergreifenden Datenaustausch zu erreichen. Das kann jedoch unerwünschte Nebenwirkungen haben und die Verwendbarkeit eines Bausteins einschränken. Eine bessere Lösung besteht in der Deklaration von Übergabevariablen im Baustein Programm (P), um die entsprechenden Eingänge und Ausgänge der dort aufgerufenen Bausteine miteinander zu verbinden.

Beispiel 6

Deklaration einer Globalvariablen

```
PROGRAM PRG
(*Deklaration*)
VAR_GLOBAL
  Bezeichner : Datentyp ;
END_VAR
(*Programmrumpf*)
  Programm
END_PROGRAM
```

◀

58.4 Programmiersprachen

Zur Erstellung der Steuerungsprogramme mit Hilfe einer Programmiersoftware stehen gemäß DIN EN 61131-3 fünf Programmiersprachen zur Verfügung stehen: Zwei textuelle Fachsprachen (AWL, ST) und zwei grafische Fachsprachen (KOP, FBS) sowie die übergeordnete Ablaufsprache (AS) mit grafischen und textuellen Elementen.

Der Programmierer kann auswählen, ob er das Bausteinprogramm in einer grafischen Sprache oder in einer Textsprache programmieren will.

- Für die grafischen Sprachen stehen in Standardbibliotheken grafische Symbole für Standardfunktionen (FC) und Standardfunktionsbausteine (FB) zur Verfügung. Diese Bausteinsymbole müssen in das Anwenderprogramm je nach Bedarf eingefügt werden, um die bezeichneten Aktionen dort ausführen zu lassen.
- Die Textsprachen benutzen anstelle von Standardbausteinen sog. Standardoperatoren, die der Programmiersprache implizit bekannt sind. Ein Operator ist ein textliches Symbol, das eine Aktion darstellt, die an einem Operanden, d. h. einer symbolischen oder direkten Variable, ausgeführt wird.

Bei der Textsprache AWL stehen die einzelnen Anweisungen in Listenform untereinander. Jede Anweisung besteht aus einem Operator, der ggf. mit einem Modifizierer ergänzt wird [N für eine Negation, C für eine Bedingung, (für Klammer auf,) für Klammer zu], gefolgt von einem oder mehreren durch Kommas getrennten Operanden. Eine Anweisung kann auch durch eine Sprungmarke gekennzeichnet sein. Die Verarbeitung der Anweisungen erfolgt in einem Ergebnisregister, dieses enthält das sog. Aktuelle Ergebnis (AE), das auf folgende Weise für das Beispiel in Abb. 58.1 ermittelt wird:

LD E0	Lade den Wert des 1. Operanden (E0) in das Ergebnisregister.
ANDN E1	Verknüpfe das Aktuelle Ergebnis mit dem negierten Wert des 2. Operanden (E1).
ST A4	Speichere das veränderte Aktuelle Ergebnis im Zieloperanden (A4).

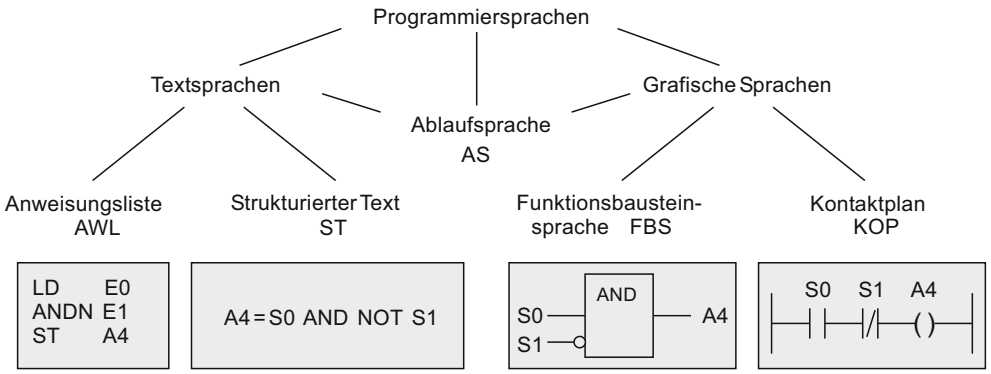


Abb. 58.1 Übersicht zu genormten SPS-Programmiersprachen

Die Tab. 58.4 zeigt die Standardoperatoren der Anweisungsliste (AWL) mit kurzen Erläuterungen, auf die nicht mehr weiter eingegangen wird. Die den Standardoperatoren entsprechenden Standardbausteine werden weiter unten jedoch noch näher dargestellt.

Programmstrukturen

59

Günter Wellenreuther und Dieter Zastrow



59.1 Lineares Programm 788
59.2 Gegliedertes Programm 788
59.3 Parametrierbares Programm 788

Teil VIII

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_59

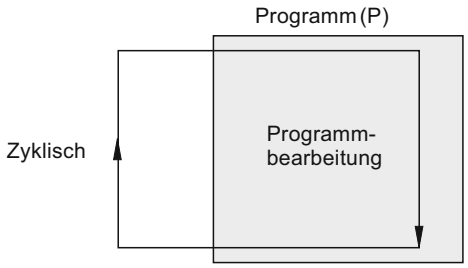


Abb. 59.1 Lineares Programm

Unter einer Programmstruktur versteht man den Aufbau eines Anwenderprogramms aus Codebausteinen. Es können drei Strukturen unterschieden werden.

59.1 Lineares Programm

Das gesamte Programm befindet sich in dem zyklisch bearbeiteten Programm (P). Die CPU arbeitet die Anweisungen der Reihe nach ab und beginnt dann wieder von vorne, wie in Abb. 59.1 gezeigt.

59.2 Gegliedertes Programm

Das Programm ist in mehrere Bausteine aufgeteilt, wobei jeder Baustein nur das Programm einer Teilaufgabe enthält. Die Programmorganisationseinheit Programm (P) enthält die Aufruf-Anweisungen, nach deren Reihenfolge die einzelnen Bausteine bearbeitet werden. Gegenüber dem linearen Programm besteht der Vorteil in den besser überschaubaren kleinen Einheiten. Abb. 59.2 lässt die Gliederungsstruktur und den Aufruf der beiden Funktionen mit der erforderlichen Parameterübergabe erkennen.

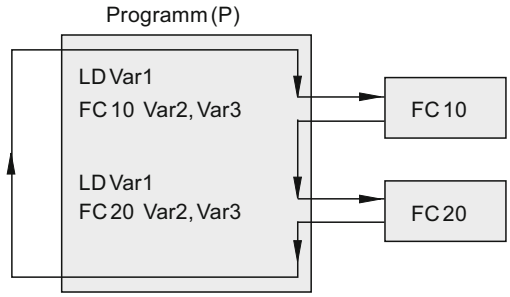


Abb. 59.2 Das Steuerungsprogramm besteht aus zwei unterschiedlichen Programmteilen

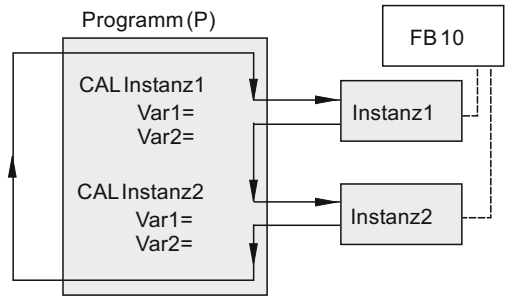


Abb. 59.3 Das Steuerungsprogramm besteht aus zwei Instanzen eines Programmteils

59.3 Parametrierbares Programm

Ein Unterprogramm ist parametrierbar, wenn es seine Eingangs- und Ausgangsvariablen dem Hauptprogramm bei jedem Aufruf erneut zur Verfügung stellt und sein Programm für jeden Aufruf erneut ausführt, um so aufrufspezifische Ergebnisse zu erzielen. Im Abb. 59.3 ist die Parametrierbarkeit am Beispiel eines Funktionsbausteins (FB) dargestellt. Grundsätzlich ist dies auch bei Funktion (FC) möglich, jedoch mit der Einschränkung der nicht gegebenen Gedächtnisfähigkeit dieses Bausteintyps.

Eingabe- und Ausgabesignale

60

Günter Wellenreuther und Dieter Zastrow



60.1 Binäre Signale 790
60.2 Digitale Signale 790
60.3 Analoge Signale 790

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_60

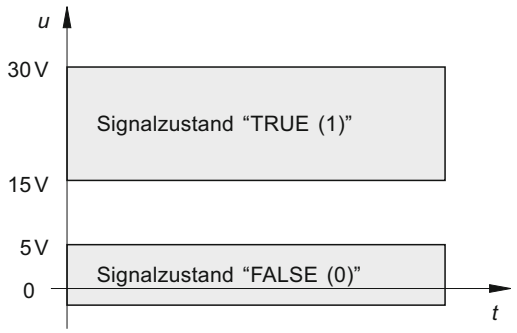


Abb. 60.1 Signalzustände und Spannungspegel

In den technischen Prozessen treten physikalische Größen wie Temperaturen, Drucke, Durchflüsse etc. auf. Automatisierungsgeräte können in der Regel nur elektrische Signale erkennen und ausgeben. Wo erforderlich, muss also eine Signalumwandlung erfolgen. Man unterscheidet verschiedene Signalarten.

60.1 Binäre Signale

Ein binäres Signal ist ein 1 Bit-Signal, das nur einen von zwei möglichen Signalzuständen annehmen kann. Ein typischer Binärsignal-Geber ist ein Schalter.

Ein Signal heißt binär, wenn es nur zweier Werte fähig ist: (TRUE = 1, FALSE = 0). Die SPS-Hersteller haben für ihre Steuerungskomponenten ein Toleranzschema festgelegt, das den Wertebereich konkreter Spannungen den binären Signalzuständen zuordnet, die von den Geräten verarbeitet werden.

Die Automatisierungsgeräte können nicht den Schaltzustand von angeschlossenen Schaltern, sondern nur anliegende Signale erkennen, d. h. die unterschiedliche Wirkung von Öffner- und Schließerkontakten in Anlagen muss bei der Programmerstellung bedacht werden.

Offene (unbeschaltete) Steuerungseingänge erzeugen Signalzustand „0“.

60.2 Digitale Signale

Ein digitales Signal ist eine mehrstellige Bitkette, die durch Codierung eine festgelegte Bedeutung erhält, z. B. als Zahlenwert. Ein typischer Digitalsignal-Geber ist ein Zifferneinsteller. Um z. B. die Zahlen 0 bis 9 darstellen zu können, sind vier Binärstellen erforderlich.

	Dualzahl				
Dezimalzahl	8	4	2	1	Wert
0	0	0	0	0	
1	0	0	0	1	
2	0	0	1	0	
3	0	0	1	1	
4	0	1	0	0	
5	0	1	0	1	
6	0	1	1	0	
7	0	1	1	1	
8	1	0	0	0	
9	1	0	0	1	

1 Binärstelle = 1 Bit
1 Byte = 8 Bit
1 Wort = 2 Byte = 16 Bit
1 Doppelwort = 32 Bit

60.3 Analoge Signale

Für ein analoges Signal ist charakteristisch, dass der Signalparameter (z. B. die Spannung) innerhalb bestimmter Grenzen jeden beliebigen Wert annehmen kann. Automatisierungsgeräte können intern keine analogen Signale verarbeiten. So genannte Analogbaugruppen nehmen eine Signalumsetzung vor und wandeln ein analoges Signal in ein digitales Signal um bzw. auch umgekehrt.

Eingabe-/Ausgabebaugruppen 61

Günter Wellenreuther und Dieter Zastrow



Teil VIII

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_61

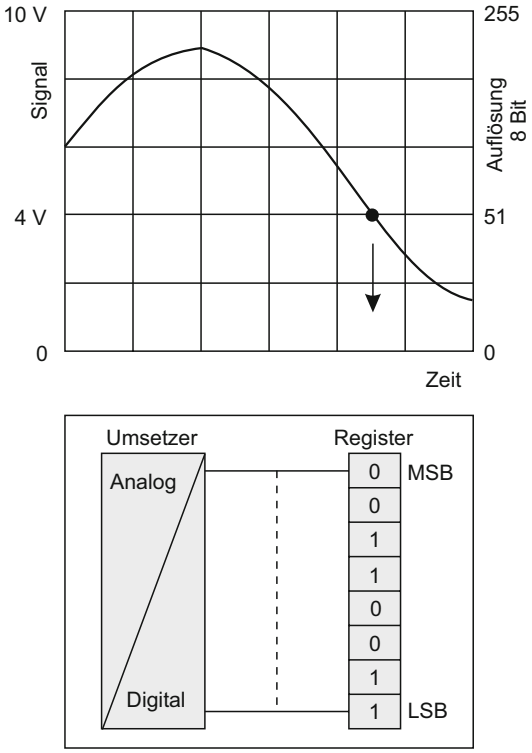


Abb. 61.1 Ein Spannungswert wird in eine Zahl umgesetzt (vereinfacht)

Die in Abb. 57.1 angedeuteten Eingabe- und Ausgabebaugruppen der SPS werden üblicherweise als zentrale digitale Eingabe- und Ausgabebaugruppen bezeichnet im Gegensatz zu den dezentralen Baugruppen (Slaves), die über ein Feldbussystem angeschlossen sind. Die digitalen Eingabe- und Ausgabebaugruppen umfassen meistens 1 Byte = 8 Bit; 2 Byte = 16 Bit; 4 Byte = 32 Bit Eingänge bzw. Ausgänge. Im Steuerungsprogramm können Bits, Bytes, Worte oder Doppelworte abgefragt oder angesteuert werden. Bei Analogbaugruppen sind entsprechend die Anzahl der Eingänge bzw. Ausgänge angegeben.

Digitaleingabebaugruppen gibt es für DC 24 V und AC 120/230 V mit Potenzialtrennung über Optokoppler sowie Anzeige des aktuellen Signalzustandes durch Leuchtdioden. Aufgrund von Filtermaßnahmen gegen Störsignale liegt die Frequenzobergrenze für Eingangssignale bei etwa 50 Hz. Die Digitaleingabebaugruppen formen die Pegel der externen digitalen Signale aus dem Prozess in den internen Signalpegel des SPS-Systems um. Die Baugruppen sind z. B. geeignet für den Anschluss von Schaltern und 2-Draht-Näherungsschaltern (BERO).

Digitalausgabebaugruppen gibt es für Lastspannungen DC 24 V oder AC 120/230 V bei spezifizierter Strombelastbarkeit und Potenzialtrennung mittels Optokoppler. Die Schaltfrequenz der Ausgänge wird nach ohmscher Last, induktiver Last und Lampenlast unterschieden und liegt im Bereich bis 100 Hz. Die Digitalausgabebaugruppen formen den internen Signalpegel des SPS-Systems in die externen, für den Prozess benötigten Signalpegel um. Die Baugruppen sind z. B. geeignet für den Anschluss von Magnetventilen, Schützen, Kleinmotoren, Lampen und Motorstartern.

Analogeingabebaugruppen wandeln analoge Signale aus dem Prozess in digitale Signale für die interne Verarbeitung innerhalb der SPS um. Es können Spannungs- und Stromgeber, Thermoelemente, Widerstände und Widerstandsthermometer angeschlossen werden:

Spannung	z. B. ±10 V
Strom	z. B. 4 bis 20 mA
Widerstand	z. B. 0 … 300 Ω
Thermoelement	z. B. Typ E, N, K mit Kennlinien-Linearisierung
Widerstandsthermometer	z. B. Pt 100-Standard mit Kennlinien-Linearisierung

Die Baugruppen verfügen über eine parametrierbare Auflösung von z. B. 9 bis 15 Bit + Vorzeichen, unterschiedliche Messbereiche (einstellbar durch Messbereichsmodule und Software) sowie Alarmfähigkeit (Diagnose und Grenzwertalarme an die CPU).

Analogausgabebaugruppen wandeln digitale Signale aus der SPS in analoge Signale für den Prozess um und sind für den Anschluss analoger Aktoren geeignet. Als Ausgangsbereiche werden angeboten:

Spannungsausgang	z. B. ±10 V
Stromausgang	z. B. 0 bis 20 mA

Die Baugruppen haben eine Auflösung von 12 bis 15 Bit. Es sind unterschiedliche Messbereiche je Kanal einstellbar.

Verknüpfungssteuerungen

62

Günter Wellenreuther und Dieter Zastrow



62.1 Logische Grundverknüpfungen in verschiedenen Darstellungen . . 794
62.2 Zusammengesetzte logische Grundverknüpfungen 794
62.3 Schließer- und Öffnerkontakte, Drahtbruchsicherheit, Erdschlussgefahr 794
62.4 Speicherfunktionen 795
62.5 Flankenauswertung 797
62.6 Darstellung und Eigenschaften elektropneumatischer Stellglieder . 797
62.7 Regeln für das Umsetzen von Schützschaltungen in SPS-Programme 798
62.8 Zeitfunktionen 799
62.9 Zählerfunktionen 799
62.10 Vergleichsfunktionen 800
62.11 MOVE-Funktion 801
62.12 EN/ENO-Mechanismus 801

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_62

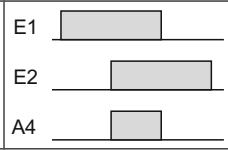
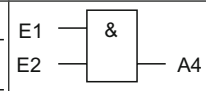
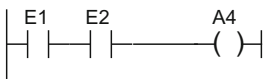
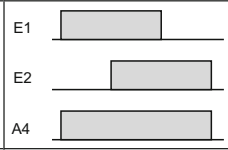
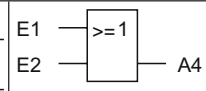
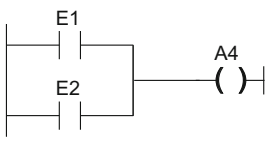
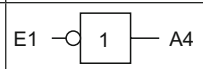
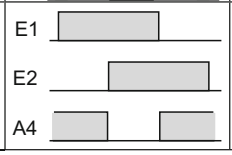
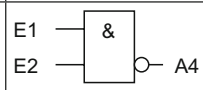
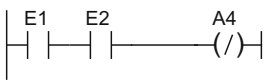
Funktion	Zeitdiagramm	FBS	KOP	AWL
UND A4 = E1 ∧ E2 A4 = E1 & E2 A4 = E1 E2				LD E1 AND E2 ST A4
ODER A4 = E1 ∨ E2				LD E1 OR E2 ST A4
NICHT A4 = $\overline{E1}$				LDN E1 ST A4
Ausgangs-NEGATION A4 = $\overline{E1 \wedge E2}$				LD E1 AND E2 STN A4

Abb. 62.1 Logische Grundverknüpfungen nach IEC 61131-3

SPS-Programme werden mit einem IEC 61131-3 orientierten Projektierungssystem (z. B. CoDeSys) als Projekte angelegt, in Bausteine programmiert und unter einem Dateinamen abgespeichert.

Als Verknüpfungssteuerungen bezeichnet man solche Programme, die Ausgangssignale überwiegend unter Verwendung einfacher Logikbeziehungen, Speicherfunktionen, Zeitglieder, Zähler u. a. in zyklischer Bearbeitungsweise erzeugen. Damit stellen Verknüpfungssteuerungen den größten Teil der SPS-Programme dar. Neben den Verknüpfungssteuerungen gibt es noch die Programmart der Ablaufsteuerungen.

62.1 Logische Grundverknüpfungen in verschiedenen Darstellungen

Nachfolgend wird eine Auswahl wichtiger Programmierfunktionen zur Realisierung von Verknüpfungssteuerungen gezeigt. In Abb. 62.1 werden die logischen Grundverknüpfungen UND, ODER, NEGATION gezeigt. Die in diesen Beispielen vorkommenden Eingangs-/Ausgangsbezeichnungen wie E1, E2, A4 oder auch S1 usw. sind als kurze Namen von deklarierter Variablen mit zutreffenden Datentyp in entsprechenden Bausteinen zu betrachten und nicht zu verwechseln mit bekannten SPS-Operanden.

62.2 Zusammengesetzte logische Grundverknüpfungen

Zu den logischen Grundverknüpfungen zählen auch die beiden häufig vorkommenden Strukturen UND-vor-ODER sowie ODER-vor-UND, die in Abb. 62.2 in grafischer Form (FBS) und Textform (AWL) dargestellt sind. Die zugehörigen Schaltfunktionen heißen Disjunktive Normalform für die UND-vor-ODER-Struktur und Konjunktive Normalform für die ODER-vor-UND-Struktur.

62.3 Schließer- und Öffnerkontakte, Drahtbruchsicherheit, Erdschlussgefahr

Befehlsgeber können auf einer Schließer- oder Öffner-Funktion beruhen, d. h. bei Betätigung ein 1-Signal oder ein 0-Signal an den Steuerungseingang liefern.

Ein Startbefehl für eine Steuerung ist mit einem 1-Signal, ein Haltbefehl mit einem 0-Signal auszuführen. Bei Gleichzeitigkeit muss der Haltbefehl Vorrang haben.

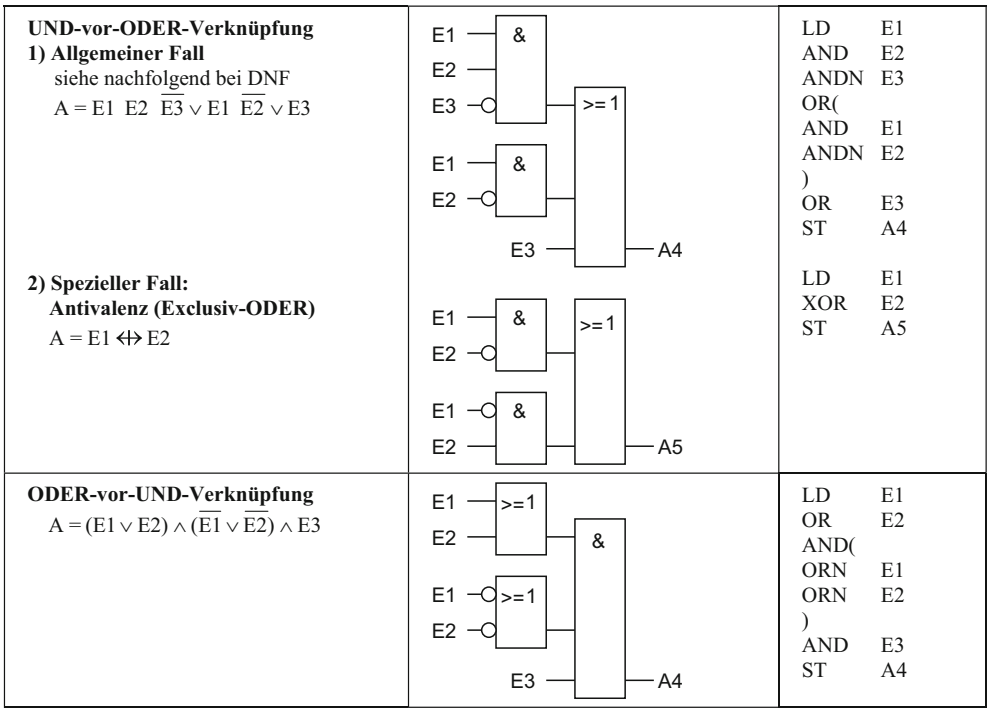


Abb. 62.2 Zusammengesetzte logische Grundverknüpfungen nach IEC 61131-3

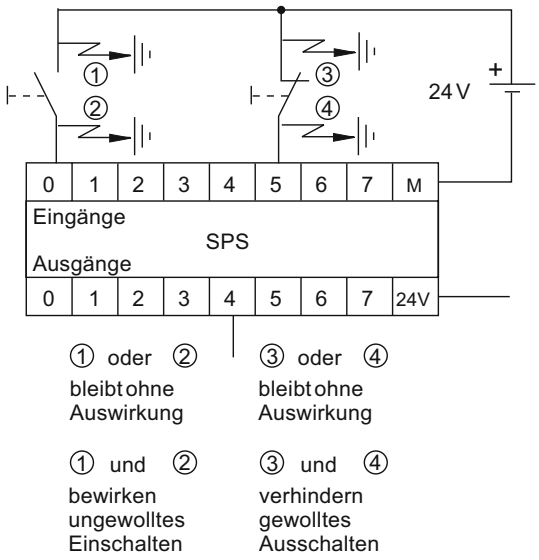


Abb. 62.3 Auswirkungen von Erdschlüssen im nicht geerdeten Steuerkreis

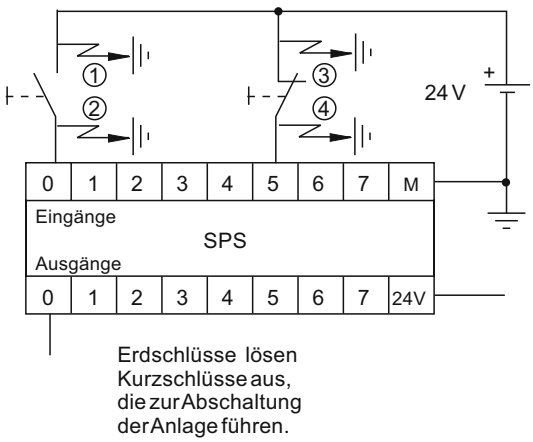


Abb. 62.4 Auswirkung von Erdschlüssen im geerdeten Steuerkreis

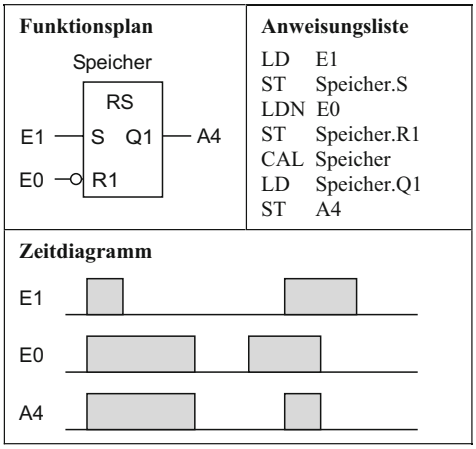
62.4 Speicherfunktionen

Das Einschalten einer Steuerung durch einen Schließerkontakt (Arbeitsstromprinzip) und das Ausschalten mit einem Öffnerkontakt (Ruhestromprinzip) macht die Steuerung drahtbruchsicher. Bei Auftreten eines Drahtbruchs erfolgt kein unbeabsichtigtes Einschalten der Steuerung, jedoch wird eine eingeschaltete Steuerung abgeschaltet. Die Erdschlussgefahr erfordert geerdete Steuerkreise oder Isolationsüberwachung (Abb. 62.3 und 62.4).

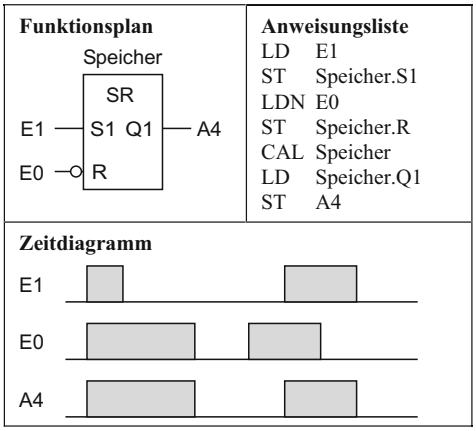
Viele Steuerungsaufgabe erfordern die Verwendung von Speicherfunktionen. Eine Speicherfunktion liegt dann vor, wenn ein kurzzeitig auftretender Signalzustand über den Programmzyklus hinaus festgehalten, d. h. gespeichert und erst zu einem späteren Zeitpunkt wieder gelöscht werden muss. Die Ausführung einer Speicherfunktion umfasst das Setzen und Rücksetzen des Speichers. Für die Speicherfunktionen stehen bistabile Standardfunktionsbausteine zur Verfügung, aus denen sich Instanzen zur Verwendung im Anwenderprogramm bilden lassen. Zu un-

terscheiden sind zwei Speicherbausteintypen, die unterschiedlich auf gleichzeitiges Eintreffen von Setz- und Rücksetzsignalen reagieren.

Speichern mit vorrangigem Rücksetzen



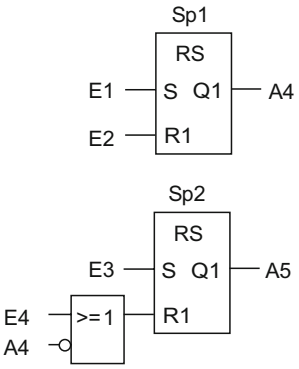
Speichern mit vorrangigem Setzen



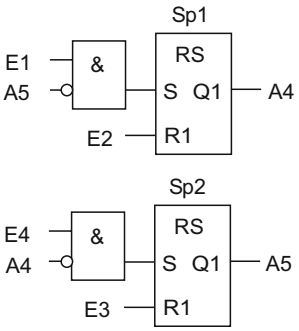
Gegenseitiges Verriegeln

Zwei Arten von Verriegelungen lassen sich unterscheiden, Bei dem gegenseitigen Verriegeln dürfen die Speicher nicht gleichzeitig gesetzt sein, z. B. für Motor-Rechtslauf und Linkslauf. Für die Ausführung von Speicherverriegelungen gibt es zwei Varianten.

Verriegeln über die Rücksetzeingänge



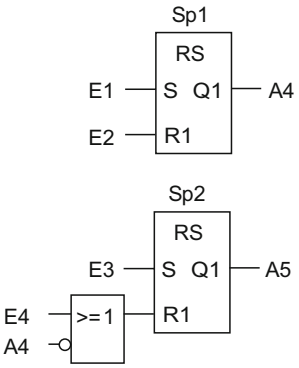
Verriegeln über die Setzeingänge



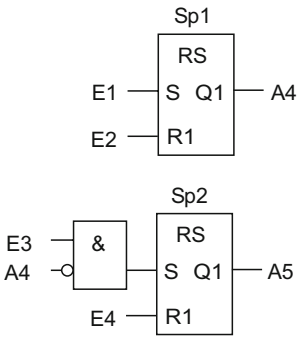
Reihenfolgeverriegelung

Bei einer Reihenfolgenverriegelung darf ein Speicher nur gesetzt werden, wenn zuvor ein oder mehrere andere Speicher gesetzt sind. Ein Beispiel sind hintereinander geschaltete Förderbänder. Auch bei der Reihenfolgeverriegelung gibt es zwei Ausführungsvarianten.

Verriegeln über den Rücksetzeingang



Verriegeln über den Setzeingang



62.5 Flankenauswertung

Die Flankenauswertung eines Signals ist dann erforderlich, wenn aus einem Signalwechsel von FALSE nach TRUE oder umgekehrt ein kurzer Impuls der Zeitdauer von einer Programmzykluszeit herzuleiten ist, unabhängig von der Betätigungsdauer des Signalgebers. Das ist z. B. wichtig, wenn bei einer Sicherheitsschaltung überprüft werden muss, ob ein Taster wirklich betätigt wurde oder nur dauerhaft durch Manipulation niedergehalten wird.

Im Abb. 62.5 werden Beispiele gezeigt, wie die Impulsgewinnung unter Verwendung von Standardfunktionsbausteinen für Flankenerkennung gelöst werden kann. P-Flanke und N-Flanke sind die Namen der Instanzen der Funktionsbausteine. Der Zykluszeitimpuls steht als Impulsoperand IO zur Verfügung.

62.6 Darstellung und Eigenschaften elektropneumatischer Stellglieder

Elektropneumatische Stellglieder können mit Ausgangssignalen einer SPS angesteuert werden. Ihre Hilfsenergie ist Druckluft, mit der pneumatische Zylinder betätigt werden. Abb. 62.6 zeigt drei häufig verwendete elektropneumatische Stellgliedtypen.

FUP	AWL	Zeitdiagramm
Positive (steigende) Flanke 0 → 1 P_Flanke R_TRIG S1 — CLK Q — IO	LD S1 ST P_Flanke.CLK CAL P_Flanke LD P_Flanke.Q ST IO	S1 P_Flanke IO Zyklen 1 2 3 .. n n+1
Negative (fallende) Flanke 1 → 0 N_Flanke F_TRIG S1 — CLK Q — IO	LD S1 ST F_Flanke.CLK CAL F_Flanke LD F_Flanke.Q ST IO	S1 F_Flanke IO Zyklen 1 2 3 .. n n+1
P_Flanke: R_TRIG, E1 — CLK Q Speicher: RS, S, R1 Q — A4 E0	LD E1 ST P_Flanke.CLK CAL P_Flanke LD P_Flanke.Q ST Speicher.S LDN E0 ST Speicher.R1	Der Impulsoperand IO kann entfallen, wenn die Flankenauswertung nur an einer Stelle des Programms benötigt wird, im Beispiel zum Setzen des RS-Speichers.

Abb. 62.5 Flankenauswertung nach IEC 61131-3

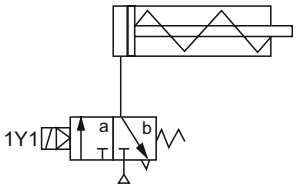
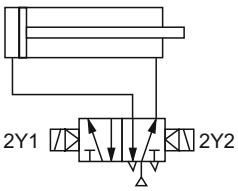
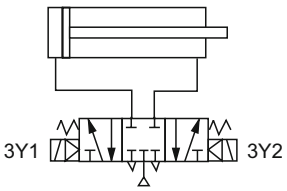
Einfachwirkender Zylinder mit Ansteuerung durch ein 3/2-Wegeventil	Doppeltwirkender Zylinder mit Ansteuerung durch ein 5/2-Wegeventil	Doppeltwirkender Zylinder mit Ansteuerung durch ein 5/3-Wegeventil
		

Abb. 62.6 Beispiele elektropneumatischer Stellglieder

3/2-Wegeventil Elektromagnetisches Ventil mit Rückstellfeder hat nur einen elektrischen Steuereingang Y1 und kann durch ein Stellsignal aus der Schaltstellung b in die Schaltstellung a geschaltet werden. Nach Beendigung des Stellsignals erfolgt eine federmechanische Rückstellung des Ventils, dessen Vorteil die definierte Schaltstellung im unbetätigten Zustand ist.

5/2-Wegeventil Elektromagnetische Impulsventile haben zwei elektrische Steuereingänge Y1 und Y2, sie können durch kurze Ansteuerimpulse aus einer Schaltstellung in die andere umgeschaltet werden. Die Ventile übernehmen die RS-Speicherfunktion der Steuerung. Nachteil des Speicherverhaltens ist die nicht definierte Schaltstellung im unbetätigten Zustand, daher erfolgt meistens eine Ansteuerung von Y1 und Y2 mit inversen Signalen.

5/3-Wegeventil Bei elektromagnetischen Impulsventilen mit Federzentrierung geht das Ventil im unbetätigten Zustand in die Mittelstellung. Daher kann neben der Vorwärts- und Rückwärtsbewegung des Zylinderkolbens auch eine Halteposition veranlasst werden.

Funktionsdiagramme zeigen den Bewegungsverlauf von Zylindern. In der Ordinate wird der zurückgelegte Weg und in der Abszisse werden Schritte oder Zeiten aufgetragen. Zusätzlich können die Zustände von Magnetspulen der Ventile dargestellt werden.

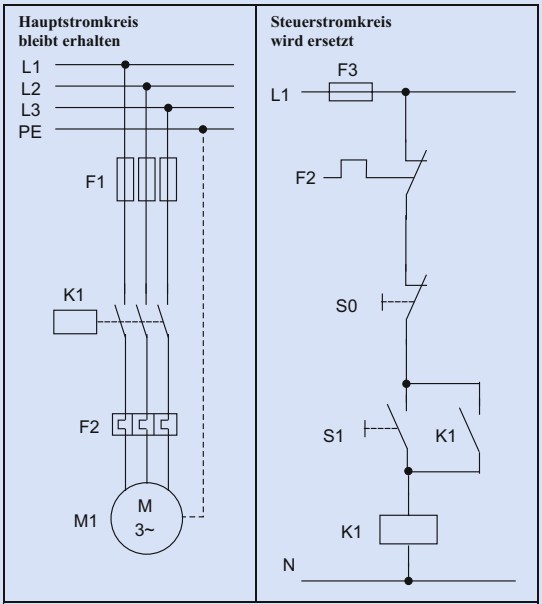
62.7 Regeln für das Umsetzen von Schützschaltungen in SPS-Programme

Eine gegebene Schützsteuerung, die z. B. zur Ansteuerung von Elektromotoren oder Elektropneumatik eingesetzt wurde, kann durch eine SPS-Steuerung unter sinngemäßer Anwendung der nachfolgenden Umsetzungsregeln ersetzt werden.

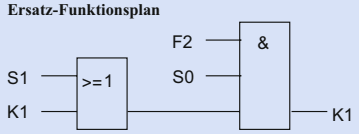
1. Der Hauptstromkreis wird unverändert außerhalb der SPS beibehalten.
2. Hauptschütze werden von den SPS-Ausgängen angesteuert und werden als OUT- oder IN-OUT-Variable deklariert. Der Schaltzustand des Hauptschützes wird somit außerhalb des Bausteins in einem SPS-Ausgang gespeichert. Wird ein Kontakt K1 des Hauptschützes K1 am Bausteineingang abgefragt, ist K1 als Durchgangsvariable (IN_OUT) zu deklarieren.
3. Hilfsschütze werden durch temporäre oder statische Variablen *innerhalb* des Bausteins ersetzt. Wenn Hilfsschütze interne Speicherfunktionen ausüben, muss für jedes Hilfsschütz eine statische Speichervariable (VAR) deklariert werden.
4. Parallelschaltungen von Schützkontakten werden durch ODER-Verknüpfungen und Reihenschaltungen durch UND-Verknüpfungen der Variablen ersetzt.
5. Öffner von Schützkontakten werden negiert und Schließer bejaht abgefragt.
6. Öffner- und Schließerkontakte von Signalgebern wie Taster und Schalter werden im Programm bejaht abgefragt und als Eingabevariablen deklariert, wenn derselbe Kontakttyp beibehalten wird.
7. Die bejahte Abfrage von Signalgebern in Regel 6 gilt nicht bei Verwendung der Speicherfunktion anstelle der Selbsthaltung.
8. Die Umsetzungsregeln 1 bis 6 verändern nicht die vorgegebene Steuerungsstruktur, wenn die Schütze keine Zeitverzögerungen oder Wischerkontakte enthalten. Einschalt- und Ausschaltverzögerungen müssen mit Zeitgliedern und Wischerkontakten mit Flankenauswertung nachgebildet werden. Impulse von Wischerkontakten beim Einschalten (Ausschalten) entsprechen steigenden (fallenden) Flanken.

Beispiel 1

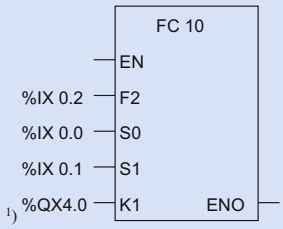
Umsetzung einer Schützschaltung in ein SPS-Programm



Ersatz-Funktionsplan

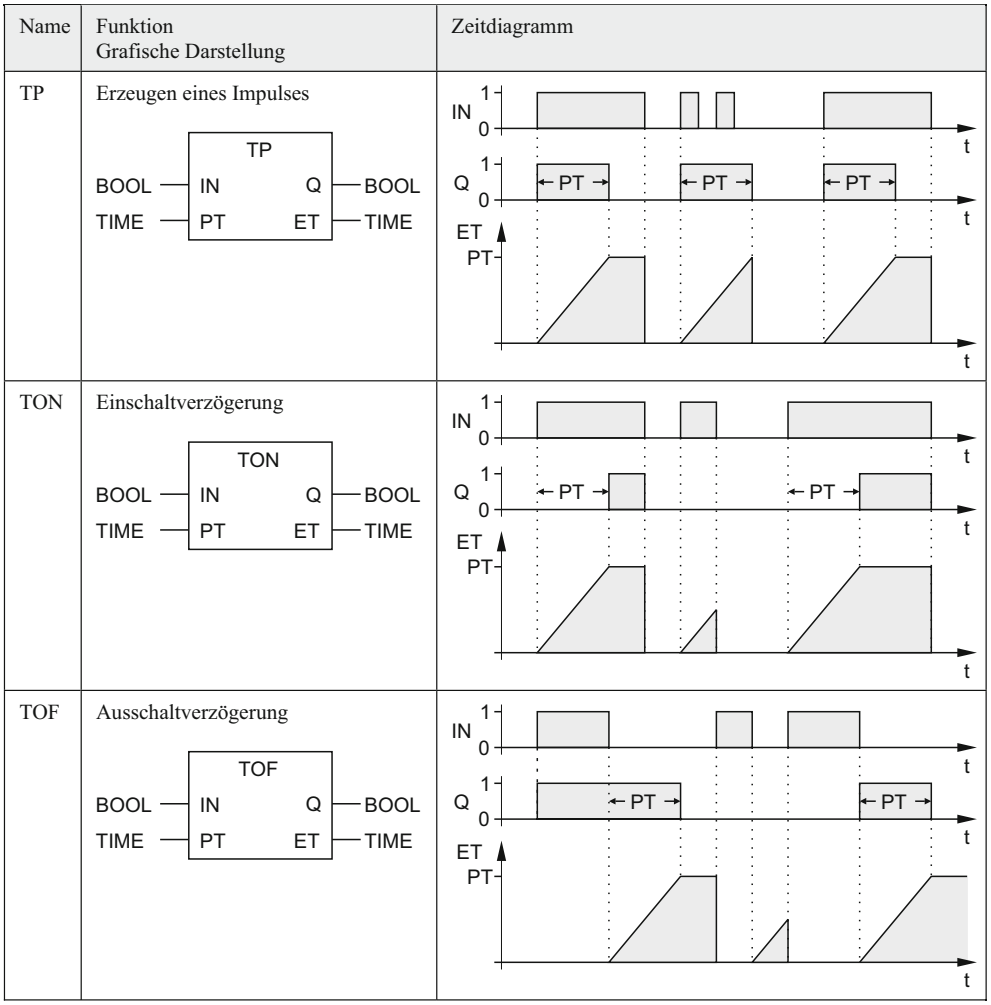


Aufruf FC 10 im Programm (P)
Die Signalspeicherung für das Hauptschütz K1 erfolgt im SPS-Ausgang A 4.0, daher genügt der Bausteintyp Funktion. Die direkten Variablen müssen in Baustein (P) deklariert werden.



[1]) K1 ist als Durchgangsvariable (IN_OUT) deklariert. Auf die SPS-Ausgangsadresse %QX4.0 kann lesend und schreibend zugegriffen werden.

◄



Operandenbedeutung:
IN = Startbedingung; PT = Zeitvorgabe; Q = Status der Zeit; ET = Aktueller Zeitwert

Abb. 62.7 Zeitglieder nach IEC 61131-3

62.8 Zeitfunktionen

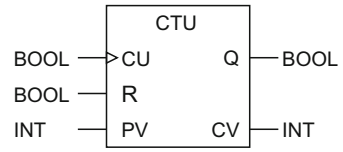
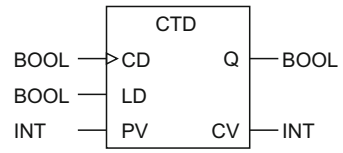
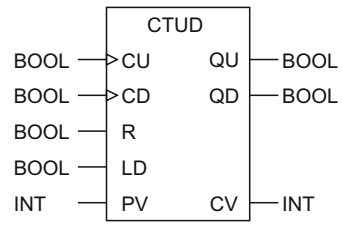
Die Zeitbildung ist eine Grundfunktion der Steuerungstechnik. Es können zeitliche Abläufe, wie Warte- und Überwachungszeiten, Zeitmessungen oder Taktimpulse programmiert werden. In der Norm DIN EN 61131-3 sind neben einer Echtzeituhr drei Standardfunktionsbausteine für die Zeitbildung aufgeführt, die in Abb. 62.7 gezeigt werden.

62.9 Zählerfunktionen

Zählerfunktionen werden in Steuerungsaufgaben benötigt, um bestimmte Mengen oder Positionen durch Aufzählen von Impulsen zu erfassen, Frequenzen zu ermitteln oder die Funktion von Steuerwerken zu übernehmen.

Grundsätzlich lassen sich Zähler in Automatisierungssystemen auf drei verschiedene Arten realisieren.

- Zähler werden als Funktionsbausteine oder Funktionen im Steuerungsprogramm aufgerufen und parametriert. Diese Zähler können je Zykluszeit nur einen Vorwärts- und einen Rückwärtszählimpuls verarbeiten. Die Verarbeitung von externen Zählimpulsen ist von der Zykluszeit und der Schaltfrequenz der Signaleingänge abhängig. Innerhalb dieses Kapitels wird ausschließlich diese Zählerart verwendet.
- Zähler werden durch eine Variable vom Datentyp Integer oder Doppelinteger realisiert. Das Auf- bzw. Abwärtszählen erfolgt bei diesen Zählern mit Additions- bzw. Subtraktionsbefehlen. Diese Zähler können je Zykluszeit mehrere interne Vorwärts- bzw. Rückwärtszählimpulse verarbeiten. Die Verarbeitung von externen Zählimpulsen ist jedoch wieder von der Zykluszeit abhängig.
- Zähler sind auf einer speziellen Baugruppe untergebracht oder sind als „Schnelle Zähler“ mit separaten Signaleingängen Teil des Betriebssystems der Zentralbaugruppe CPU.

Name	Grafische Darstellung	Arbeitsweise in ST-Sprache
Aufwärts-Zähler		IF R THEN CV := 0; ELSIF CU AND (CV < Pvmax) Then CV := CV+1; END_IF; Q := (CV >= PV)
Abwärts-Zähler		IF LD THEN CV := PV; ELSIF CD AND (CV < Pvmin) Then CV := CV-1; END_IF; Q := (CV <= 0)
Auf-Abwärts-Zähler		IF R THEN CV := 0; ELSIF LD THEN CV := PV; ELSIF CU AND (CV < Pvmax) THEN CV := CV+1; ELSIF CD AND (CV < Pvmin) THEN CV := CV-1; END_IF; QU := (CV >= PV) QD := (CV <= 0)

Bedeutung der Eingänge und Ausgänge: CU = Vorwärtszähleingang steigende Flanke; CD = Rückwärtszähleingang steigende Flanke; R = Rücksetzeingang; LD = Ladeeingang; PV = Ladewert; QU = Status Zählerstand (CV größer gleich Ladewert); QD = Status Zählerstand (CV kleiner gleich 0); CV = Zählerstand.
Hinweis: Die numerischen Werte der Grenzvariablen Pvmax und Pvmin sind implementierungsabhängig. Mögliche Werte sind Pvmax = 32767 und Pvmin = –32768.

Abb. 62.8 Zählfunktionen nach IEC 61131-3

Mit diesen Zählern ist es möglich, externe Zählimpulse zu erfassen, die schneller als die Zykluszeit sind. Innerhalb des Steuerungsprogramms können Zählerstände mit Übergabevariablen abgerufen werden.

Bei der Realisierung von Zählern unterscheidet man Vorwärts- bzw. Aufwärts-Zähler, Rückwärts- bzw. Abwärts-Zähler und Vor-Rückwärts- bzw. Auf-Abwärts-Zähler. Der Zählerstand ergibt sich aus der Anzahl der positiven Signalflanken an den Eingängen für Vorwärts- und Rückwärtszählen.

In der Norm DIN EN 61131-3 sind für die drei Zähler die grafische Darstellung und die Arbeitsweise der Standardfunktionsbausteine festgelegt wie in Abb. 62.8 abgebildet.

62.10 Vergleichsfunktionen

Mit Vergleichsfunktionen werden die Werte zweier digitaler Operanden des gleichen Datentyps verglichen. Das sind in den meisten Fällen Variablen mit Datentypen für Zahlenwerte wie REAL (Dezimalzahlen mit Komma), INT (Ganzzahlen mit Vorzeichen) und WORD (vorzeichenlose Zahlen).

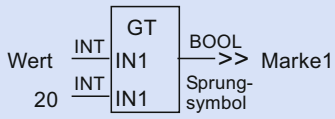
Zur Durchführung von Vergleichen stehen Standardfunktionen (FC) zur Verfügung, die beliebig oft mit verschiedenen Eingangsparametern aufgerufen werden können. Das Vergleichsergebnis steht als binärer Signalzustand im aktuellen Ergebnis (AE) zur Verfügung, und zwar mit TRUE, wenn der Vergleich zutrifft und mit FALSE, wenn er nicht zutrifft. Die Tab. 62.1 zeigt die Vergleichsmöglichkeiten für den einfachsten Fall von zwei Vergleichseingängen IN1 und IN2.

Tab. 62.1 Vergleichsfunktionen

Name	Symbol	Beschreibung
GT	>	IN1 größer IN2
GE	>=	IN1 größer gleich IN2
EQ	=	IN1 gleich IN2
LE	<=	IN1 kleiner gleich IN2
LT	<	IN1 kleiner IN2
NE	<>	IN1 ungleich IN2

Beispiel 2

Vergleich auf „größer“



Die Eingangsvariable „Wert“ vom Datentyp Integer wird mit der Integerkonstanten „20“ auf „größer als“ verglichen, siehe Abbildung. Am Ausgang des Vergleichers ist

das Sprungsymbol durch eine Signallinie mit Doppelpfeil am Ende angegeben. Der Sprung ist bedingt, weil er vom Ausgang des Vergleichs abhängig ist. Ist die Vergleichsbedingung erfüllt, wird das Programm an der Stelle von Marke1 fortgeführt. Ist die Vergleichsbedingung nicht erfüllt, läuft das Programm linear weiter. Im Vergleichsbaustein ist der Name der Vergleichsfunktion angegeben (GT). ◄

62.11 MOVE-Funktion

Die MOVE-Funktion zählt zu den arithmetischen Standardfunktionen. Sie hat genau einen Eingang und einen Ausgang und bewirkt das „Durchreichen“ von Daten der Eingangsvariablen an die Ausgangsvariable in Abhängigkeit von einem booleschen Freigabesignal EN, dabei darf der Datentyp auf der Ausgangs- und Eingangsseite sogar verschieden sein. Die Nutzung der MOVE-Funktion zur Datentypumwandlung hat enge Grenzen. Sinnvoll ist z. B. die Umwandlung von BYTE in WORD. Mit der Datentypangabe ANY in Abb. 62.9 ist kein neuer Datentyp gemeint. ANY ist der ranghöchste der sog. allgemeinen Datentypen der SPS-Norm, die sich nur an Hersteller wendet und mitteilt, dass jeder nachgeordnete Datentyp verwendet werden darf. Die Ausgestaltung der MOVE-Funktion ist herstellerabhängig.

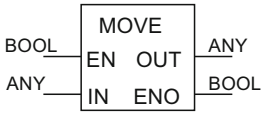


Abb. 62.9 MOVE-Funktion mit Ausführungssteuerung EN/ENO

62.12 EN/ENO-Mechanismus

EN steht für Enable IN (boolescher Freigabe-Eingang) und ENO für Enable OUT (boolescher Freigabeausgang). Die SPS-Norm sieht vor, dass Funktionen (FC) in den Programmiersprachen FBS und KOP über diese Möglichkeit der Ausführungssteuerung verfügen müssen. Der Freigabe-Eingang EN kann mit einer Logikverknüpfung beschaltet sein (z. B. UND-Glied, Vergleicher). Der boolesche Ausgang ENO kann zur Ansteuerung nachfolgender Funktionen benutzt werden, weil er anzeigt, ob die Bausteinfunktion ordnungsgemäß ausgeführt wurde oder nicht.

Die Variablen EN/ENO müssen nicht benutzt werden, wenn die Ausführung der Funktion in jedem Fall erfolgen soll.

Ist bei Funktionsaufruf EN = 0, wird die Funktion nicht ausgeführt und ENO = 0 ausgegeben.

Ist bei Funktionsaufruf EN = 1, wird die Funktion ausgeführt und der Ausgang meldet mit ENO = 1 die Fehlerfreiheit der Ausführung oder mit ENO = 0 einen Ausführungsfehler.

Aufruf und Wertübergaben zwischen Bausteinen nach IEC 61131-3

63

Günter Wellenreuther und Dieter Zastrow



63.1 Aufrufhierarchie der Bausteine P, FB und FC 804
63.2 Aufruf von Funktionsbausteinen in FBS 804
63.3 Aufruf von Funktionsbausteinen in AWL 804
63.4 Aufruf von Funktionen in AWL 805

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_63

63.1 Aufrufhierarchie der Bausteine P, FB und FC

Das Anwenderprogramm einer SPS hat einen hierarchischen Aufbau

An oberster Stelle steht ein Baustein des Typs Programm (P), dessen Deklaration und Gebrauch identisch ist mit denen der bereits beschriebenen Funktionsbausteine, jedoch mit den begrenzenden Schlüsselwörtern **PROGRAM ... END_PROGRAM**. In diesem Baustein können Instanzen von Funktionsbausteine (FB) sowie Funktionen (FC) aufgerufen werden.

An mittlerer Stelle stehen die Bausteine des Typs Funktionsbaustein (FB). Innerhalb eines Funktionsbausteins können Instanzen anderer Funktionsbausteine oder auch Funktionen (FC) aufgerufen werden

An unterster Stelle stehen die Bausteine vom Typ Funktionen (FC). Innerhalb einer Funktion können andere Funktionen (FC) aufgerufen werden.

63.2 Aufruf von Funktionsbausteinen in FBS

Am einfachsten gestalten sich Bausteinaufrufe und die damit verbundenen Werteübergaben in grafischer Programmiersprache, z. B. FBS. Die Einbindung eines Funktionsbausteins (FB) in ein Programm (P) geschieht in der logischen Reihenfolge: Deklarieren einer Instanz des Funktionsbausteins, Einfügen des Bausteinsymbols aus der Bausteinbibliothek und Zuordnung des Instanznamens, Anbindung der Instanz-Eingänge/Ausgänge an Variablen des aufrufenden Bausteins, z. B. auch an direkte Variablen, welche die Verbindung zu den realen SPS-Eingängen/Ausgängen halten.

Instanzen werden deklariert wie Variablen durch Angabe eines Namens und eines Typs. Als Typ wird der Name des Funktionsbausteins verwendet.

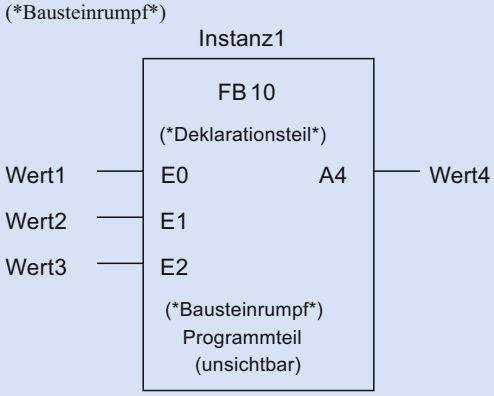
63.3 Aufruf von Funktionsbausteinen in AWL

Gleicher Deklarationsteil wie in der Grafiksprache FBS. Etwas schwieriger gestaltet sich der Bausteinaufruf und die damit verbundenen Werteübergaben in Textsprache, z. B. AWL. Es stehen mehrere Methoden zur Verfügung, deren einfachste ist die mit der formalen Argumentenliste, die auch vom Hersteller fest vorgegeben sein kann. Bei dieser Methode muss ein Funktionsbaustein mit dem Befehl CAL (unbedingt) oder CALC (bedingt) aufgerufen werden, gefolgt vom Instanznamen und der offenen Klammer. Darunter erscheint die auszufüllende Liste mit den Parametern des aufgerufenen Instanzbausteins, die mit der geschlossenen Klammer beendet wird.

Beispiel 1

Programm P ruft Funktionsbaustein FB auf

```
PROGRAM PRG
(*Deklaration*)
VAR_INPUT
   Wert1 AT %IX0.0 : BOOL ;
   Wert2 AT %IX1.0 : BOOL ;
   Wert3 AT %IX1.7 : BOOL ;
END_VAR
VAR
   Instanz1 : FB 10 ;
END_VAR
VAR_OUTPUT
   Wert4 AT %QX4.0 : BOOL ;
END_VAR
(*Bausteinrumpf*)
```




```
END_PROGRAM
```

◄

Beispiel 2

Programm P ruft Funktionsbaustein FB auf

```
PROGRAM PRG
(*Deklaration*)
VAR_INPUT
   Wert1 AT %IX0.0 : BOOL ;
   Wert2 AT %IX1.0 : BOOL ;
   Wert3 AT %IX1.7 : BOOL ;
END_VAR
VAR
   Instanz1 : FB 10 ;
END_VAR
VAR_OUTPUT
   Wert4 AT %QX4.0 : BOOL ;
END_VAR
(*Bausteinrumpf*)
CAL Instanz1 (
   E0 := Wert1,
   E1 := Wert2,
   E2 := Wert3,
   A4 => Wert4 )
END_PROGRAM
```

◄

63.4 Aufruf von Funktionen in AWL

Gegenüber dem Aufruf von Funktionsbausteinen entfällt die Instanzbildung. Der Aufruf in der Grafiksprache FBS ist sonst der gleiche wie bei Funktionsbausteinen. Unterschiede bestehen beim Aufruf in der Textsprache AWL, da nicht der Operator CAL verwendet werden darf. Der Funktionsname ist als Operator in der Anweisung anzugeben. Es sind zwei Aufrufmethoden zu unterscheiden.

Beim „*formalen Aufruf*“ wird als erste Anweisung einfach der Name der aufzurufenden Funktion geschrieben, gefolgt von der offenen Klammer. Danach wird je Anweisungszeile ein Eingangsparameter übergeben, wie bei der obigen formalen Argumentenliste, jedoch nur für die Funktionseingänge. Dann folgt die geschlossene Klammer, mit der die Funktion bearbeitet und das Ergebnis im Ergebnisregister gespeichert wird. Die letzte Anweisung speichert das Aktuelle Ergebnis (AE) des Ergebnisregisters in der Zielvariablen.

Beispiel 3

Formaler Funktionsaufruf einer Funktion

```
PROGRAM PRG
(*Deklaration*)
VAR_INPUT
   Wert1 AT %IX0.0 : BOOL ;
   Wert2 AT %IX1.0 : BOOL ;
   Wert3 AT %IX1.7 : BOOL ;
END_VAR
VAR_OUTPUT
   Wert4 AT %QX4.0 : BOOL ;
END_VAR
(*Bausteinrumpf*)
FC 10 (
   E0 := Wert1,
   E1 := Wert2,
   E2 := Wert3,
   );
   ST    Wert4
END_PROGRAM
```

◀

Beim „*nichtformalen Aufruf*“ muss zuerst der erste Eingangsparameter in das Ergebnisregister geladen werden. Dann kommt die Anweisung mit dem Namen der Funktion als Operator, gefolgt von den restlichen Eingangsparametern in richtiger Reihenfolge und durch Kommas getrennt im Operandenteil. Mit dem Abschluss dieser Anweisung berechnet die Funktion das Ergebnis und stellt es in das Ergebnisregister. In der letzten Anweisung wird das Aktuelle Ergebnis (AE) des Ergebnisregisters in der Zielvariablen gespeichert.

Beispiel 4

Nichtformaler Aufruf einer Funktion

```
PROGRAM PRG
(*Deklaration*)
VAR_INPUT
   Wert1 AT %IX0.0 : BOOL ;
   Wert2 AT %IX1.0 : BOOL ;
   Wert3 AT %IX1.7 : BOOL ;
END_VAR
VAR_OUTPUT
   Wert4 AT %QX4.0 : BOOL ;
END_VAR
(*Bausteinrumpf*)
LD  Wert1
FC 10  Wert 2, Wert 3
ST  Wert4
END_PROGRAM
```

◀

Ablaufsteuerung

64

Günter Wellenreuther und Dieter Zastrow



64.1 Ablauf-Funktionsplan 808
64.2 Grafische Darstellung von Ablaufsteuerungsfunktionen 808
64.3 Betriebsartenteil und Bedienfeld 811

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_64

Steuerungsprogramme, die einen schrittweisen Prozessablauf nach den Vorgaben von *Ablauf-Funktionsplänen* ausführen, werden *Ablaufsteuerung*en genannt.

64.1 Ablauf-Funktionsplan

Der Ablauf-Funktionsplan ist eine eigenständige Planart zur prozessorientierten Darstellung von Steuerungsaufgaben. Eine verbale Aufgabenstellung soll aus Gründen der Klarheit und Vollständigkeit durch eine grafische Darstellung ersetzt werden, die bei der Planung, Inbetriebnahme und Störungssuche hilfreich sein soll. Ein richtig entworfener Ablauf-Funktionsplan muss bereits die Lösung einer entsprechenden Steuerungsaufgabe darstellen.

Geeignete Steuerungsaufgaben sind solche, bei denen unterscheidbare Aktionen in einer ereignis- oder zeitgesteuerten Reihen- oder auch Parallelfolge ablaufen und die auf Wiederholung gerichtet sind. Anschauliche Beispiele dafür sind Produktionsanlagen.

Der Ablauf-Funktionsplan stellt nur die grafischen Elemente für eine Ablaufbeschreibung zur Verfügung. Die Entwurfsmethode besteht darin, dass Steuerungszustände eingeführt und mit Aktionen verknüpft sowie Übergangsstellen zur Berücksichtigung von Steuersignalen vorgesehen werden.

Die Darstellung von Ablauf-Funktionsplänen kann auf zwei verschiedenen Normen beruhen:

- DIN EN 60848 GRAFCET, Spezifikationssprache für Funktionspläne der Ablaufsteuerung. (Nachfolger der DIN 40719-6). Diese Norm definiert eine grafische Entwurfssprache für die funktionale Beschreibung des Verhaltens des Ablaufteils eines Steuerungssystems. Die Entwurfssprache wird „GRAFCET“ genannt.
- DIN EN 61131, Speicherprogrammierbare Steuerungen, Teil 3: Programmiersprachen, hierin Elemente der Ablaufsprache (AS). Der Zweck der Ablaufsprache ist die Darstellung von Ablauffunktionen in SPS-Programm-Organisationseinheiten des Typs Funktionsbaustein oder Programm. Dazu gibt die Norm zwei Darstellungsvarianten für ihre Elemente an, eine ausführlich behandelte grafische Variante und eine textuelle Variante.

DIN EN 60848 und DIN EN 61131-3 haben jeweils ihren eigenen spezifischen Anwendungsbereich. Während die Entwurfssprache GRAFCET für die Beschreibung des Verhaltens unabhängig von einer speziellen Realisierung (elektronisch, elektromechanisch, pneumatisch oder gemischt) ist, legt IEC 61131-3 die Beschreibungsmittel der Ablaufsprache AS zwecks Programmrealisierung fest. Bei der Darstellung der Aktionsblöcke wird auf die Unterschiede eingegangen.

64.2 Grafische Darstellung von Ablaufsteuerungsfunktionen

Zur Darstellung von Ablaufstrukturen werden Ablauf-Funktionspläne verwendet, die auch als SFC (Sequential Function Chart) bezeichnet werden.

Darstellung von Schritten und Transitionen

Jeder mögliche Zustand einer Steuerung wird im Ablauf-Funktionsplan durch einen *Schritt* dargestellt. Ein Schritt ist entweder aktiv oder inaktiv und stellt einen Zustand der Steuerung dar. Ein Schritt muss grafisch durch einen Block dargestellt werden, der einen Schrittnamen in Form eines Bezeichners enthält, siehe Abb. 64.1.

Eine Sonderstellung nimmt der Anfangsschritt ein, der durch eine doppelte Umrahmung gekennzeichnet ist. Der Anfangsschritt muss beim Start des Ablaufsals einziger Schritt aktiv sein. Die Schritte sind durch *gerichtete Verbindungen* in Form von vertikalen Linien miteinander verbunden. Die Eigenschaft eines Schrittes, aktiv oder inaktiv sein zu können, setzt ein Speicherverhalten im Steuerungsprogramm voraus. Jedem Schritt ist normalerweise eine Aktion zugeordnet, die in einem Aktionsblock angegeben und mit dem Schritt verknüpft werden kann. Die Aktion ist nicht Bestandteil des Schrittes. Die Ablaufkette besteht aus einer Folge von *Schritten* und *Transitionen*, die einfach oder verzweigt sein kann.

Einfache Ablaufkette

Der Wechsel von Schritt und Transition wird als Folge wiederholt. Es wird eine Kettenschleife gebildet, um wieder zum Anfang zurückzukehren. Die zurückführende Wirkungslinie kann durch eine Pfeildarstellungen mit Angabe von Schrittnummern ersetzt werden (Abb. 64.2).

Verzweigte Ablaufkette

Man unterscheidet je nach Art der Verzweigung zwischen der Alternativ-Verzweigung und der Simultan-Verzweigung (Abb. 64.3).

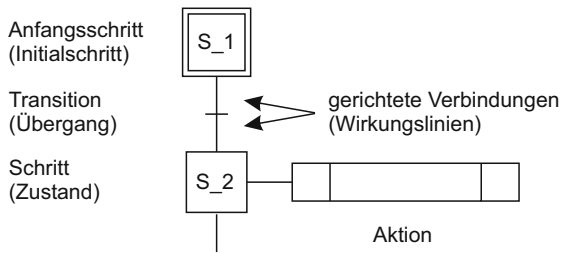


Abb. 64.1 Darstellung von Ablaufschritten

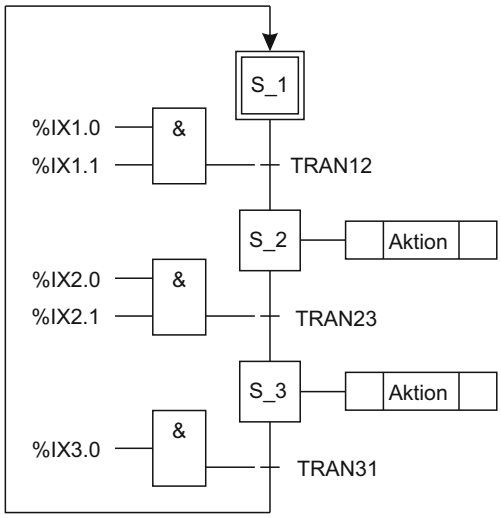


Abb. 64.2 Einfache (lineare) Ablaufkette

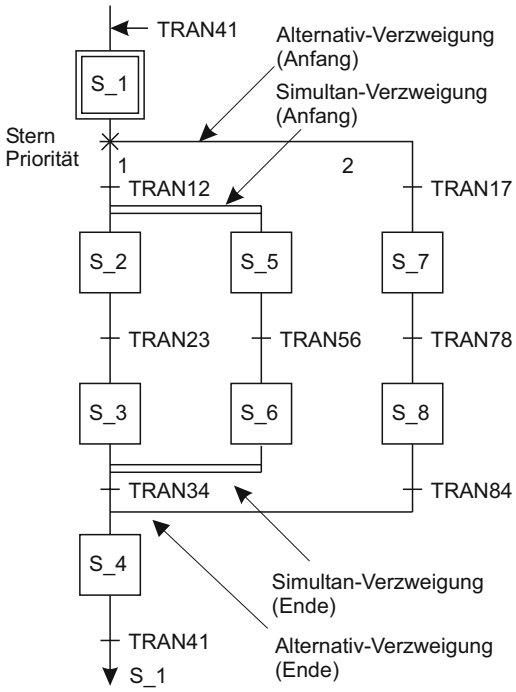


Abb. 64.3 Verzweigte Ablaufkette

- *Alternativ-Verzweigung* (1- aus *n*-Verzweigung, ODER-Verzweigung); es erfolgt die Auswahl und Bearbeitung nur eines Kettenstranges aus mehreren Kettensträngen. Am Verzweigungsanfang darf zur gleichen Zeit nur eine Transitionsbedingung wahr sein (Verriegelung) oder es muss eine Priorität vorgegeben werden, indem der Strang mit der niedersten Nummer die höchste Priorität hat. Zusätzlich wird mit einem Stern (*) angegeben, dass die Transitionen von links nach rechts bearbeitet werden. Jedes Strang-Ende muss eine eigene Transitionsbedingungen zum Verlassen des Kettenstranges aufweisen. Anfang und Ende von Alternativ-Verzweigungen werden durch waagerechte Einfachlinien dargestellt.
- *Simultan-Verzweigung* (Parallelbearbeitung mehrerer Kettenstränge, UND-Verzweigung); es erfolgt die gleichzeitige Aktivierung der Anfangsschritte mehrerer Kettenstränge, die dann aber unabhängig voneinander bearbeitet werden. Alle Kettenstränge unterliegen auf der Anfangsseite nur einer vorgelagerten gemeinsamen Transitionsbedingung. Bei der Zusammenführung der Kettenstränge (Endeseite) darf nur eine gemeinsame Transitionsbedingung vorhanden sein. Anfang und Ende von Simultan-Verzweigungen werden durch waagerechte Doppellinien dargestellt.

Wenn zum Beispiel der Anfangsschritt S_1 aktiv ist und die Transitionsbedingung von TRAN_17 den booleschen Wert TRUE liefert während die Transitionsbedingung von TRAN_12 den booleschen Wert FALSE hat, wird der Schritt S_7 aktiviert und Schritt 1 deaktiviert. Damit kann es in dieser Phase des Ablaufs nicht mehr zu einer Aktivierung der Schritte S_2 und S_5 kommen. Der Übergang von Schritt S_8 zu Schritt S_4 erfolgt erst, wenn die Transitionsbedingung von TRAN_84 den booleschen Wert TRUE annimmt. Schritt S_8 wird dann von S_4 zurückgesetzt.

Für den Fall, dass der Anfangsschritt S_1 aktiv ist und die Transitionsbedingungen von TRAN_12 und TRAN_17 den booleschen Wert TRUE haben, wird wegen der festgelegten Priorität die Simultan-Verzweigung bearbeitet. Es werden die Schritte S_2 und S_5 aktiviert und der Vorgängerschritt S_1 deaktiviert. Der Schritt S_4 wird erst erreicht, wenn die Vorgängerschritte S_3 und S_6 aktiv sind und die Transitionsbedingung von TRAN_34 den booleschen Wert TRUE annimmt. Die Schritte S_3 und S_6 werden dann von S_4 zurückgesetzt.

Schleifen und Sprünge

Ein *Sprung* führt unter Steuerung durch eine Transitionsbedingung von einem Schritt zu einem entfernten anderen Schritt, wobei der durch den Sprung gebildete Zweig keine Schritte enthält (Abb. 64.4).

Eine *Schleife* kann die Folge eines Sprunges sein, indem unter Steuerung durch eine Transitionsbedingung auf einen Vorgängerschritt zurückgesprungen wird. Dabei muss verhindert werden, dass sich das Programm in einer Endlosschleife verfängt.

Aktionen, Aktionsblock

Mit einem Schritt ist in der Regel eine Aktion verbunden. Ein Schritt ohne zugehörige Aktion übt eine Warte-Funktion aus bis die nachfolgende Transitionsbedingung erfüllt ist. Da der Ablauf-Funktionsplan beschreiben soll, was in der gesteuerten Anlage zu geschehen hat, wird in der Norm der Begriff *Aktion* anstelle von Befehl verwendet.

Der *Aktionsblock* ist ein grafisches Element zur Darstellung von Aktionen. Der Aktionsblock ist nicht Teil eines Schrittes und damit auch nicht Teil der Ablaufkette. Der Aktionsblock kann mit einem Schritt verknüpft oder als grafisches Element in einer Kontaktplan- bzw. Funktionsplandarstellung verwendet werden.

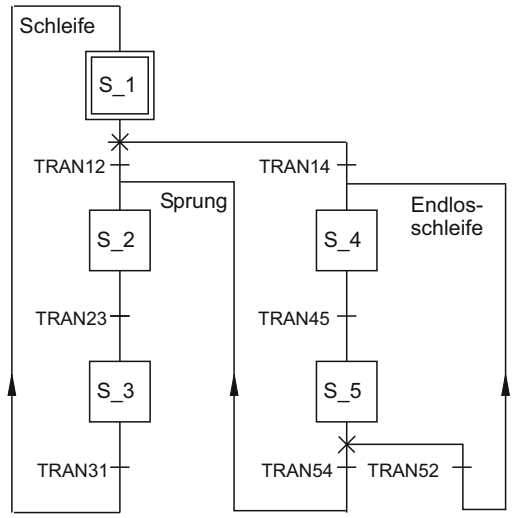
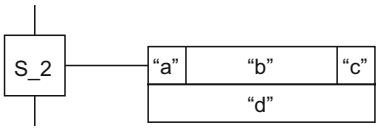


Abb. 64.4 Sprung und Schleife

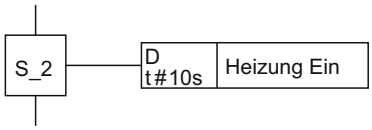
Nachfolgend wird jedoch die Verknüpfung von Schritt und Aktionsblock bevorzugt.

In vollständiger Darstellung besteht der Aktionsblock aus vier Teilflächen, die nicht alle genutzt werden müssen:



- Feld „a“: Bestimmungszeichen
- Feld „b“: Aktionsname
- Feld „c“: Anzeigevariable
- Feld „d“: Beschreibung der Aktion in AWL; ST; KOP, FBS

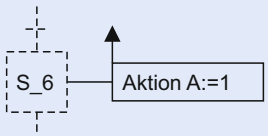
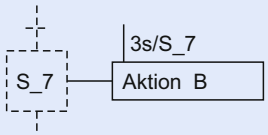
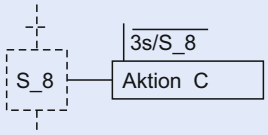
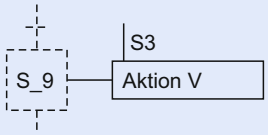
Im einfachsten Fall wird in Feld „b“ der Name einer booleschen Variablen und in Feld „a“ ein zutreffendes Bestimmungszeichen eingetragen.

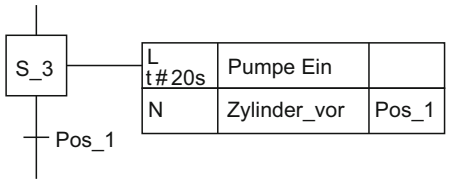


Die Aktion wird ausgeführt, wenn der zugehörige Schritt gesetzt ist und eine Aktionssteuerung die Freigabe erteilt. Die Aktionssteuerung sorgt für die richtige Umsetzung der im Feld „a“ eingetragenen Bestimmungszeichen. Im Beispiel lautet die Aktion „Heizung EIN“ und es ist das Bestimmungszeichen D zusammen mit einer Zeitangabe im Feld „a“ eingetragen. Ergebnis: Die Heizung wird verzögert eingeschaltet.

Es ist zulässig, mehrere Aktionen mit einem Schritt zu verbinden. Dies wird grafisch dargestellt durch aneinander gereihte Aktionsblöcke.

Tab. 64.1 Symbole für Aktionen nach GRAFCET

	**Speichernde Aktion** Aktion wird bei Aktivierung des Schrittes gespeichert Beispiel: Setze den Wert der booleschen Variablen A auf TRUE
	**Zeitverzögerte Aktion** Die verzögerte Aktion ist eine kontinuierlich wirkende Aktion, bei der die Zuweisungsbedingung erst nach 3 s erfüllt ist Beispiel: Ausgang B nimmt den Wert TRUE an, nachdem drei Sekunden seit der Aktivierung von Schritt S_7 vergangen sind
	**Zeitbegrenzte Aktion** Die begrenzte Aktion ist eine kontinuierlich wirkende Aktion, bei der die Zuweisungsbedingung während der Dauer von 3 s erfüllt ist Beispiel: Ausgang C erhält den Wert TRUE für drei Sekunden nach Aktivierung von Schritt S_8
	**Bedingte Aktion** Die Zuweisungsbedingung S3 beeinflusst die kontinuierlich wirkende Aktion Beispiel: Ausgang V erhält den Wert TRUE, wenn in Schritt 9 Bedingung S3 erfüllt ist Als boolesche Gleichung: V = S_9 & S3



Im Aktionsfeld „c“ kann eine boolesche Anzeige-Variable eingetragen werden, die durch die Aktion gesetzt werden kann, um die Erledigung oder einen Fehlerfall (z. B. Zeitüberschreitung) anzuzeigen oder als Weiterschaltbedingung verwendbar ist.

Bestimmungszeichen für Aktionen (IEC 61131-3)

Die Verknüpfung einer Aktion mit einem Schritt erfolgt formal durch ein sog. Bestimmungszeichen. Für jedes Bestimmungszeichen (N, S, R, L, D u. a.) muss in der Ablaufsteuerung eine Programmteil mit einer entsprechenden Ansteuerlogik für die im Feld „b“ angegebene boolesche Ausgangsvariable vorhanden sein. Im Abb. 64.5 ist für einzelne Bestimmungszeichen die Art der Ansteuerlogik mit Liniendiagrammen beschrieben.

Die **DIN EN 60848** unterteilt den Aktionsblock nicht. Die Art der Aktionsausgabe wird dort durch Symbole am Aktionsblock festgelegt, siehe Tab. 64.1.

Welche der Darstellungen verwendet wird, bestimmen in der Praxis die Vorteile für den jeweiligen Anwendungszweck. Ablaufsteuerungen werden mit speziellen Software-Tools der Automatisierungsfirmen entwickelt, die viel Detailarbeit übernehmen.

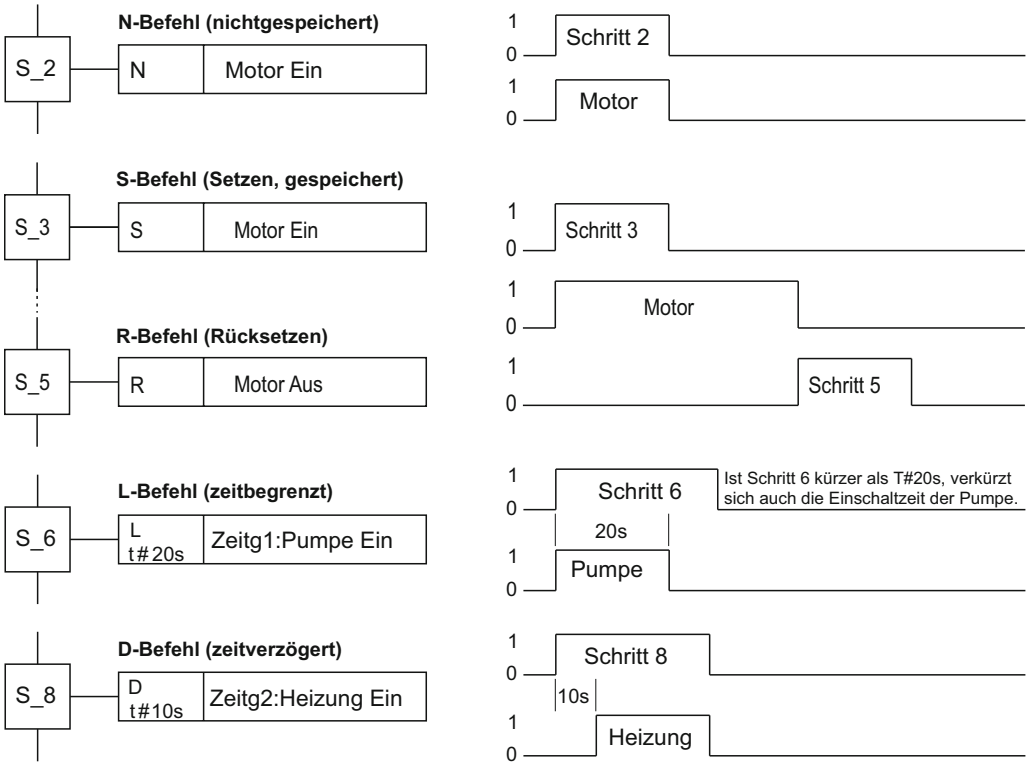


Abb. 64.5 Bestimmungszeichen N, S, R, L und D für Aktionen nach IEC 61131-3

64.3 Betriebsartenteil und Bedienfeld

Ablaufsteuerungen in der Praxis bestehen aus Ablaufketten und verfügen über einen übergeordneten Betriebsartenteil für die Inbetriebnahme, Anlagenbetrieb und Störungsbeseitigung. Ablaufsteuerungen bestehen im Prinzip immer aus folgenden Komponenten:

Betriebsartensignale

Betriebsartensignale sind die vom Betriebsartenprogramm erzeugten Steuersignale für das Weiterschalten und Rücksetzen der Ablaufkette sowie für die Freigabe von Aktionen (Abb. 64.6).

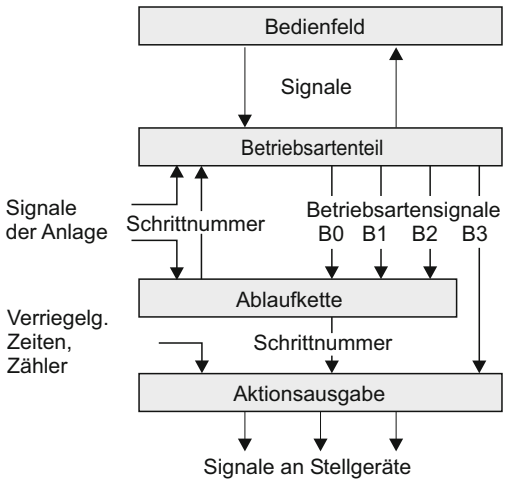


Abb. 64.6 Struktur einer Ablaufsteuerung

Rücksetzen RESET (B0) RESET wirkt durch die Übergabevariable B0 auf einen Eingangsparameter RUECKSETZ des Ablaufkettenbausteins und versetzt die Schrittkette in die Grundstellung oder wirkt über den RESET-Eingang des Aktionsbausteins und setzt dort Speicher, Zeitglieder und Zähler zurück.

Freigabe der Kette mit Bedingungen FREI_K_mB (B1) FREI_K_mB wirkt durch die Übergabevariable B1 auf einen Eingangsparameter WEITER_mB im Ablaufkettenbaustein. Nur wenn B1 = 1 ist, wird bei erfüllten Weiterschaltbedingungen der nächste Schritt gesetzt. Das B1-Signal kann ein Impuls bei Hand-Einzelschrittbetrieb oder ein Dauersignal bei Automatikbetrieb sein.

Freigabe der Kette ohne Bedingungen FREI_K_oB (B2) FREI_K_oB wirkt durch die Übergabevariable B2 auf einen Eingangsparameter WEITER_oB im Ablaufkettenbaustein. Nur wenn B2 = 1 ist, wird im Hand-Einzelschrittbetrieb der nächste Schrittspeicher gesetzt, ohne dass die Weiterschaltbedingung erfüllt sein muss. Das B2-Signal muss immer ein Impulssignal mit der Länge von einer Zykluszeit sein.

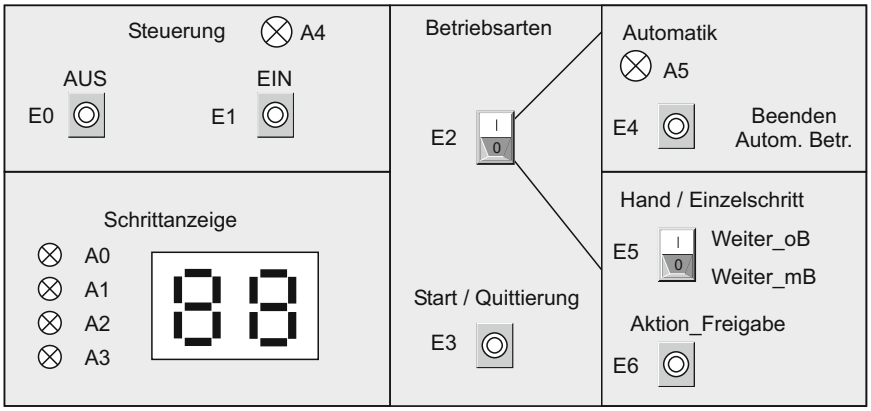


Abb. 64.7 Beispiel eines Bedienfeldes für Ablaufsteuerung

Freigabe Aktion FR_AKTION (B3) FR_AKTION wirkt durch die Übergabevariable B3 auf einen Eingangsparameter FREIGABE des Aktionsbausteins und kann dort einen Aktionsausgang freischalten bzw. sperren.

Funktionen eines Bedienfeldes

Eingriffe in Steuerungen sollen von einem Bedienfeld aus erfolgen, dessen Bedienoberfläche auf die Erfordernisse von Ablaufsteuerungen mit Betriebsartenteil abgestimmt ist.

Die Kenntnisnahme der Einzelfunktionen der Befehlsgeber und Anzeigen des beispielhaften Bedienfeldes verdeutlicht die Komplexität und Praxisnähe von Ablaufsteuerungen (Abb. 64.7).

Taster E0: **Steuerung: AUS**

Taster E1: **Steuerung: EIN**
Hinweis: Statt der beiden Taster kann auch ein EIN-AUS-Schalter bzw. ein Schlüsselschalter verwendet werden.

Schalter E2: **Betriebsart**
Wahl der Betriebsart: E2 = 1 für Automatik, E2 = 0 für Handbetrieb.
Während eines Bearbeitungsprozesses kann zwischen Automatik und Handbetrieb umgeschaltet werden.

Taster E3: **Start/Einzelschritt**
Bei E2 = 1 (Automatik) werden durch Betätigung von E3 der Automatikbetrieb für das Weiterschalten der Ablaufkette mit Bedingungen sowie die Freigabe der Aktionen eingeschaltet und somit der automatische Ablauf der Kette gestartet.
Bei E2 = 0 (Hand-/Einzelschrittbetrieb) erfolgt bei Betätigung von E3 eine Einzelschritt-Weiterschaltung der Schrittkette.
Bei ausgeschalteter Steuerung kann durch Betätigen der Start/Quittierungs-Taste ein RESET zum Rücksetzen der Schrittkette und der gespeicherten Aktionen ausgelöst werden, wenn sich die Anlage in der Grundstellung befindet.

Taster E4: **Beenden Automatikbetrieb**
Vorwahl zum Beenden der Betriebsart Automatik bei Erreichen der Grundstellung der Ablaufkette (vollständig beendeter Bearbeitungsprozess).

Schalter E5: **Weiterschaltbedingung für Hand-/Einzelschrittbetrieb**
Bei E2 = 0 (Hand-/Einzelschrittbetrieb) muss die Art der Ablaufketten-Weiterschaltung festgelegt werden. Zur Auswahl stehen E5 = 1 (Weiterschalten ohne Bedingungen) und E5 = 0 (Weiterschalten mit Bedingungen).

Taster E6: **Aktions-Freigabe**
In der Betriebsart Handbetrieb wird durch Taster E6 die Aktion des aktiven Schrittes solange ausgeführt, wie E6 betätigt ist.

Kommunikation in Automatisierungssystemen

65

Günter Wellenreuther und Dieter Zastrow



65.1 Bussysteme 814
65.2 PROFINET – Offener Industrial Ethernet Standard 815
65.3 OPC-Technologie 816

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_65

Die Automatisierungstechnik befindet sich im Umbruch. Durch Anwendung moderner Kommunikationstechnik erhofft man sich entscheidende Verbesserungen bei Effizienz und Flexibilität der Automatisierungsprozesse. Der Begriff der Automatisierungstechnik umfasst heute mehr als nur das Automatisieren im Sinne von Steuern, Regeln und Visualisieren, eingeschlossen ist auch das Kommunizieren.

Zwischen den Anlagenkomponenten eines Automatisierungssystems müssen in der Regel Informationen ausgetauscht werden. Das sind im einfachsten Fall Signale von Sensoren und Aktoren, die zum übergeordneten Automatisierungsgerät gelangen müssen oder von dort herkommen. In anderen Fällen handelt es sich um Messwerte, Statusmeldungen und Diagnoseinformationen, die schon kompliziertere Daten darstellen. Realität ist auch, dass Auftragsdaten der Produktion zwischen Büro und Fertigungsanlage übertragen werden müssen. Es liegen also umfangreiche Kommunikationsbeziehungen in der Automatisierungstechnik vor, zu deren Bewältigung moderne Kommunikationssysteme verwendet werden.

65.1 Bussysteme

Die klassische Informationsübertragung von Prozesssignalen mittels analoger Spannungs- oder Stromwerte passt nicht zur digitalen Datenverarbeitung in den Automatisierungsgeräten. Es ist deshalb naheliegend, auch die Informationsübertragung auf eine digitale Grundlage zu stellen. Um gleichzeitig den Verkabelungsaufwand so gering wie möglich zu halten, überträgt man Daten über Bussysteme in Form serieller Zweidrahtverbindungen, an die alle Teilnehmer angeschlossen sind. Für den Informationsaustausch werden Telegramme mit entsprechenden Sende- und Empfangsadressen sowie den Nutzdaten gebildet. Ein solches digitales Kommunikationssystem ist leichter erweiterbar durch Anschluss weiterer Teilnehmerstationen und Änderungen erfordern zum großen Teil nur softwaremäßige Eingriffe. Digitale Kommunikationssysteme vereinfachen nicht nur den sonst erforderlichen Verdrahtungsaufwand radikal, sondern bringen für den Anlagenbetreiber auch noch einen Zusatznutzen, indem außer den eigentlichen Messwerten nützliche Anlageninformationen wie z. B. Fehlermeldungen ohne großen Mehraufwand zusätzlich zur Verfügung stehen und Fernparametrierungen intelligenter Sensoren möglich sind.

Je nach Anwendungsbereich werden in der Automatisierungstechnik unterschiedliche Bussysteme eingesetzt, deren wichtigste Merkmale ihre sog. Echtzeitfähigkeit und Störsicherheit sind. Echtzeit bedeutet, dass die neuen Daten immer „rechtzeitig" eintreffen. Zu spät eintreffende Daten können sonst zu gefährlichen Anlagenzuständen führen. Verwendet werden nur noch *standardisierte, offene Bussysteme*. Standardisiert bedeutet international genormt in der IEC 61158 und *Offenheit* gewährt den Zugang zu Spezifikationen und Technologien, damit sich neue Anbieter mit eigenen Produkten am System beteiligen können. Für die Nutzer ergibt sich so der Vorteil einer größeren Unabhängigkeit von Herstellern bei der Auswahl der zu vernetzenden Anlagenkomponenten.

Im prozessnahen Bereich der Anlage, der Feldebene, findet man überwiegend sogenannte **Feldbussysteme** mit Master-Slave-Kommunikation vor. Die Slaves sind die Buskomponenten, über die alle Eingangs- und Ausgangssignale der Anlage erfasst bzw. ausgegeben werden. Die Master-Station ist ein Kommunikationsprozessor, der für die zyklische Bedienung der zugeordneten Slave-Stationen sorgt, indem er die Daten von Eingängen der Slaves abholt bzw. an Ausgänge von Slaves ausliefert. Bekannte Master-Slave-Systeme für den Feldbereich sind der **AS-i-Bus** (Aktor-Sensor-Interface), der **PROFIBUS-DP** (Process Field BUS-Dezentrale Peripherie), der **INTERBUS-S** und andere Systeme. In der Feldebene zählen Bussysteme seit vielen Jahren zum Stand der Technik. In komplexeren Anlagen mit mehreren SPSen kann jedoch auch ein Datenaustausch zwischen Steuerungsstationen (SPSen) erforderlich sein. Das erfordert bereits die höherwertige Master-Master-Kommunikation, die aber auch noch dem Produktionsbereich der Fabrik zuzurechnen ist. Muss ein noch weitergehender Datenaustausch unter Einbeziehung von Bürobereichen, z. B. der Fertigungssteuerung, verwirklicht werden, so wird eine zweite Netzinfrastruktur neben dem vorhandenen Feldbussystem erforderlich. Hierfür bietet sich dann das im Bürobereich bereits etablierte **Ethernet-TCP/IP-Netz** an. Schon lange wurde gefordert, das TCP/IP-Netz in die Fertigungsebene zu verlängern bei gleichzeitiger Erfüllung der Echtzeitbedingung. Ein Beispiel hierfür ist **PROFINET** als neues Kommunikationssystem. Abb. 65.1 gibt eine Übersicht zu Bussystemen in der industriellen Kommunikation.

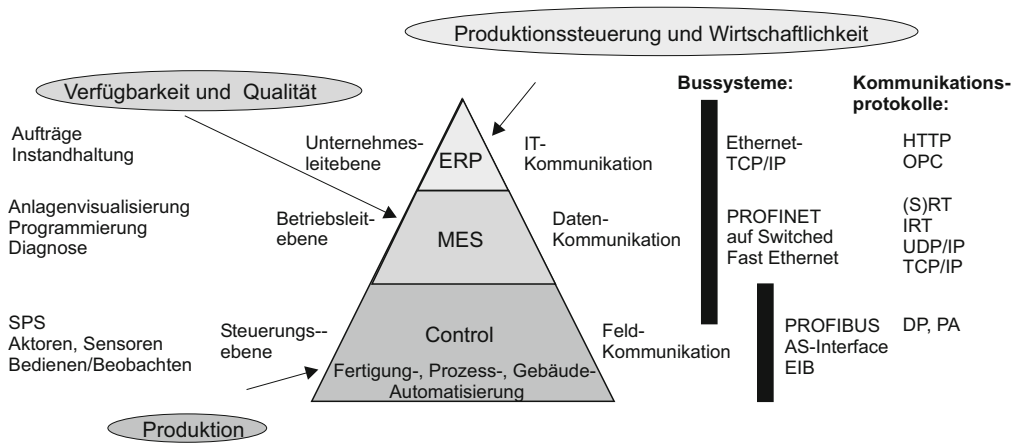


Abb. 65.1 Industrielle Kommunikation. ERP = Enterprise Ressource Planning, MES = Manufacturing Execution Systems

65.2 PROFINET – Offener Industrial Ethernet Standard

PROFINET bietet sich als ein offenes und durchgängiges Konzept für Automatisierungslösungen auf Ethernetbasis an, im Bereich von Einzelmaschinen bis hin zu modular aufgebauten Anlagen mit verteilter Steuerungsintelligenz und gewährleistet dabei durch Einbinden von PROFIBUS DP einen Investitionsschutz.

Das Grundkonzept besteht aus

- PROFINET IO (dezentrale Feldgeräte) und
- PROFINET CBA (verteilte Automatisierung)

und schließt folgenden Leistungsumfang ein:

- Industrial Ethernet-Netzwerke mit aktiven Netzkomponenten (Switches, Router)
- Integration bestehender Feldbussysteme (PROFIBUS DP, INTERBUS, . . .)
- Kommunikationskanäle für anforderungsabhängige Übertragungsleistung
- Herstellerübergreifendes Engineeringkonzept (Projektierung, Programmierung)
- Einsatz von IT-Technologien (Netzwerkadministration, Webserver, E-Mail, OPC)
- Sicherheitsgerichtete Kommunikation

PROFINET IO

PROFINET IO ist die Kurzbezeichnung für das Steuerungskonzept „Dezentrale Feldgeräte“, dass mit einem Industrial Ethernet Netzwerk ein zentrales Steuerungsgerät (IO-Controller) mit dem dezentralen Feldgerätebereich (IO-Devices) verbindet und somit eine Ähnlichkeit mit dem weitverbreiteten Feldbussystem PROFIBUS-DP besitzt. Die Nutzdaten der Feldgeräte werden auch wieder zyklisch in Echtzeit in das Prozessabbild des IO-Controllers übertragen oder in umgekehrter Richtung an die IO-Devices ausgegeben. Das verwendete Kommunikationsmodell heißt Provider-Consumer-Verfahren und nicht mehr Master-Slave-Verfahren, obwohl der Datenverkehr mit den Feldgeräten in beiden Systemen nach dem gleichen Prinzip abläuft. PROFIBUS regelt den Buszugriff über die Token-Weitergabe, von der die DP-Slaves ausgeschlossen sind. Im Ethernet-System ist das nicht möglich, weil alle Teilnehmer am Netz beim Buszugriff gleichberechtigt sind. Der Provider sendet seine Daten ohne Aufforderung des Kommunikationspartners. Den IO-Devices wird deshalb beim System-Hochlauf mitgeteilt, dass sie mit einem Buszyklus von z. B. 10 ms mit aktuellen Daten versorgt werden. Die Eigenschaften der IO-Devices werden durch deren GSD-Datei (General Station Description auf XML-Basis) beschrieben, wie dieses auch von PROFIBUS DP her bekannt ist.

Die Steuerungsintelligenz in Form eines Anwenderprogramms befindet sich bei PROFINET IO oftmals nur im IO-Controller, kann aber auch teilweise in intelligenten IO-Devices (Feldgeräte mit eigener CPU) untergebracht sein, auch wieder vergleichbar mit dem PROFIBUS DP-System.

Abb. 65.2 zeigt zusammen mit der anschließenden Gegenüberstellung der Grundbegriffe die Ähnlichkeit beider Systeme auf.

Ziffer	PROFINET IO	PROFIBUS DP	Bemerkung
1	IO-System	DP-Master-system	Alle Geräte (IO-Controller, IO-Devices) und Kommunikationsverbindungen
2	IO-Controller	DP-Master	Gerät, über das angeschlossene Feldgeräte angesprochen werden
3	IO-Device	DP-Slave	Dezentrale Feldgeräte z. T. mit eigener CPU
4	Industrial Ethernet	Profibus	Netzwerkinfrastruktur mit Switches (in Abb. 65.2 nicht dargestellt)
5	IO-Supervisor	PG/PC DP-Master Klasse 2	Programmieren, Inbetriebnahme/Diagnose
6	HMI = Human Machine Interface	HMI	Gerät zum Bedienen und Beobachten mit Zugriff auch auf IO-Devices über Ethernet!
7	GSD (XML-Datei)	GSD (ASCII-Datei)	Gerätebeschreibungsdatei für die IO-Devices und DP-Slaves

PROFINET-Kommunikationskanäle

Die PROFINET-Kommunikation findet über Industrial Ethernet statt. Dabei werden die folgenden *Übertragungsarten* unterstützt:

- Zyklische Übertragung von zeitkritischen Daten (Nutzdaten)
- Azyklische Übertragung von Engineering-Daten und zeitunkritische Parametrierungs-, Konfigurierungs- und Diagnose-Daten.

Für die genannten Übertragungsarten werden unterschiedliche *Transportprotokolle* verwendet, die man sich vereinfacht als Transportkanäle unterschiedlicher Leistungsstufen vorstellen kann. Abb. 65.3 gibt eine Übersicht.

- **TCP/UDP-IP-Kanal** (Standard-Transportprotokoll der IT-Welt) für die Übertragung zeitunkritscher PROFINET-Daten. Dieser Transportkanal steht auch zur generellen Anbindung der Automatisierungssysteme an die übergeordneten Ethernet-Netze wie den Firmen-eigenen Intranets und dem öffentlichen Internet zur Verfügung.
- **SRT-Kanal** (Soft Real Time) für zeitkritische PROFINET-Daten. Hierbei handelt es sich um ein spezielles Transportprotokoll von PROFINET, um die im Feldbusbereich geforderte Echtzeitkommunikation zu ermöglichen. Werden Aktualisierungszeiten von ca. 10 ms bei zyklischer Datenübertragung gefordert, wird von „weicher“ Echtzeitbedingung gesprochen. Das SRT-Transportprotokoll wird als Software auf Basis vorhandener Controller realisiert.
- **IRT-Kanal** (Isochrone Real Time) für ganz besonders anspruchsvolle Anforderungen an die Übertragung von PROFINET-Daten wie beispielsweise für Antriebssteuerungen. Hier sind „harte“ Echtzeitbedingungen einzuhalten, d. h. Aktualisierungszeiten von ca. 1 ms bei einer garantierten Taktgenauigkeit bis auf 1 µs. Die IRT-Kommunikation ist zeitschlitzgesteuert und setzt eine entsprechende Konfigurierung mit IRT-fähigen Geräten einschließlich der Switches voraus. IRT-fähige Switches schalten die Verbindungen zeitsynchronisiert (nicht adressgesteuert) bereits vor

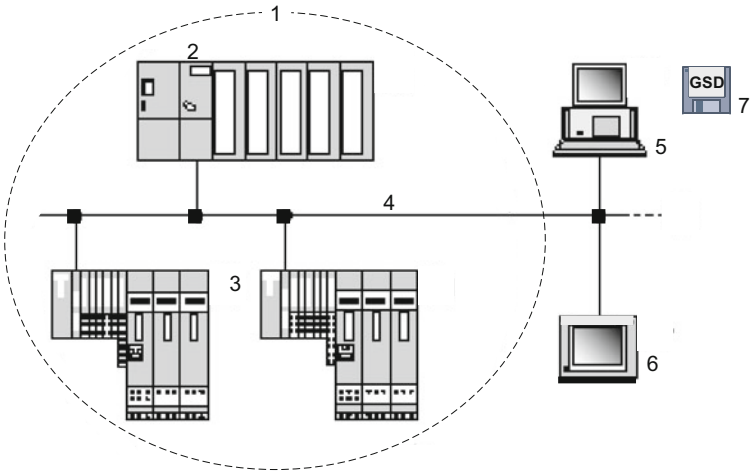


Abb. 65.2 Bussystem PROFINET IO

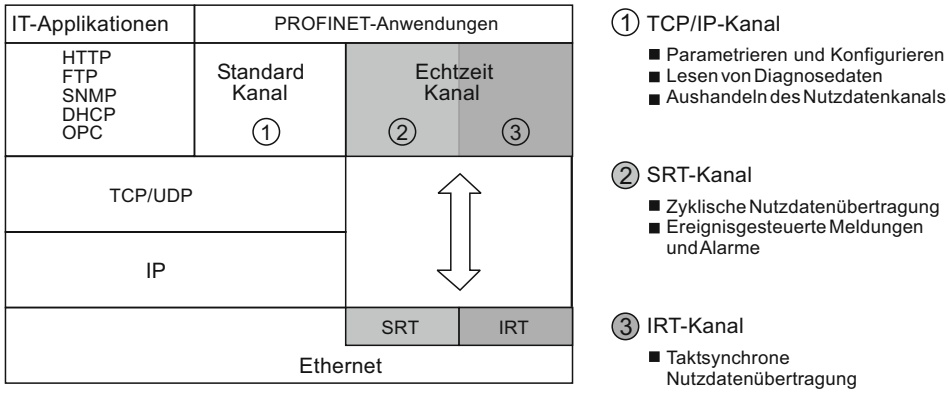


Abb. 65.3 Kommunikationskanäle bei PROFINET

dem Eintreffen der Ethernet-Telegramme durch. Die Realisierung des IRT-Transportprotokolls erfolgt auf Hardware-Basis durch einen ASIC. PROFINET nutzt auch das Prinzip der „Telegramm-Priorisierung“, um die Übertragung der Daten durch das Ethernet-Netzwerk zu verbessern. Vordringlichere Telegramme sollen die weniger eiligen Telegramme überholen können. Netzwerkkomponenten wie Switches können den Datenfluss priorisierter Telegramme steuern, dazu verwenden sie Zwischenspeicher.

65.3 OPC-Technologie

Die OPC-Technologie (OLE for Process Control) bildet eine Datenbrücke zwischen einer Applikation, die Prozessdaten übergeordnet zu verarbeiten hat und einer Steuerungs-Hardware, die als gerätespezifischer Datenlieferant angesehen werden kann. OPC ist ein Client-Server-System.

Ein OPC-Server ist eine Software-Komponente, die der Hersteller einer SPS-Hardware für diese zur Verfügung stellt, damit von übergeordneten Anwenderprogrammen (z. B. MS Excel) aus auf die spezifische Hardware zugegriffen werden kann. Für den OPC-Server, der auf einem PC läuft, muss eine unterlagerte Kommunikationsverbindung zur Hardware des Herstellers eingerichtet werden, z. B. PROFIBUS oder Industrial Ethernet-TCP/IP.

Im Anwenderprogramm muss ein OPC-Client angelegt und konfiguriert werden, um auf einen OPC-Server und somit auf die Prozessdaten beispielsweise einer SPS lokal oder entfernt zugreifen zu können. Die Entwicklung eines OPC-Client-Programms kann beispielsweise mit Excel-VBA oder durch Einbindung eines fertigen ActiveX-Elements gelöst werden.

Steuerungssicherheit

66

Günter Wellenreuther und Dieter Zastrow



66.1 Europäische Richtlinien und Sicherheitsnormen 818
66.2 Sicherheitsbegriff 818

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_66

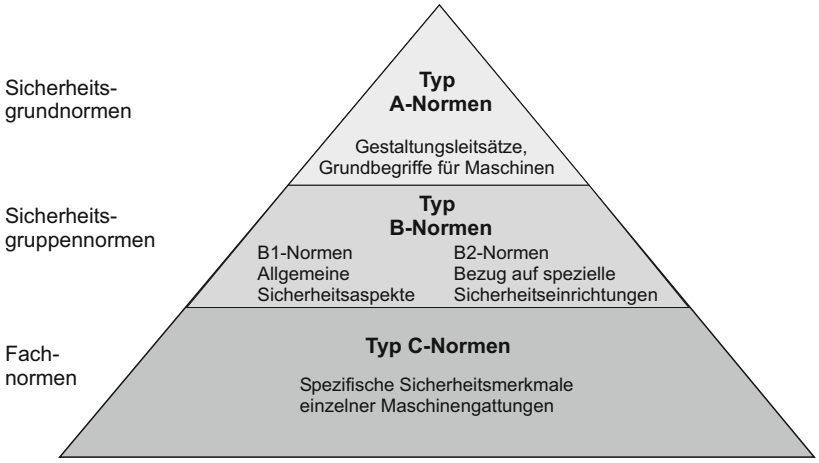


Abb. 66.1 Hierarchie des europäischen Normenwerkes für Sicherheit von Maschinen

Jede Maschinensteuerung ist mit einem Fehlerrisiko behaftet, das sich durch besondere Maßnahmen bei der Entwicklung, Fertigung, Inbetriebnahme und Bedienung verringern aber nicht völlig ausschließen lässt. Fehlerursachen sind technisches oder menschliches Versagen.

Sicherheitsrelevante Maßnahmen sind nicht allein unter dem Aspekt der technischen Funktion (Wie funktioniert die Sicherheitsmaßnahme?) sondern besonders auch unter rechtlichen Gesichtspunkten (Welche Sicherheitsvorschriften gelten?) zu sehen. Hinter allem aber steht eine soziale Verantwortung des Geräteherstellers und Betreibers, der sich bewusst sein sollte, dass der Werker an einer Maschine davon ausgeht, dass diese sicher ist.

66.1 Europäische Richtlinien und Sicherheitsnormen

Drei für die Steuerungstechnik wichtige EG-Richtlinien seien hier besonders erwähnt:

- Niederspannungsrichtlinie 73/23/EWG: Schutz vor Gefahren durch elektrischen Strom bei Niederspannungsgeräten im Spannungsbereich 50 … 1000 VAC, 75 … 1500 VDC, CE-Kennzeichnungspflicht seit 1997.
- Maschinenrichtlinie 89/392/EG, letzte Fassung 98/37 EG: Grundlegende Anforderungen an die Sicherheit der Maschinen zum Schutz der Gesundheit des Betreibers. Inzwischen gilt diese Richtlinie auch für Sicherheitsbauteile, CE-Kennzeichnungspflicht seit 1995.
- EMV-Richtlinie 89/336/EWG (Elektromagnetische Verträglichkeit): Zwei grundlegende Anforderungen an die Geräte sind die sehr allgemein gehaltenen Grenzen für Störaussendung und Störfestigkeit bei Einstrahlung, CE-Kennzeichnungspflicht seit 1996.

Die europäischen Normen zur Sicherheit von Maschinen weisen eine dreigeteilte hierarchische Struktur auf (Abb. 66.1):

- Typ A-Normen: Enthalten Gestaltungsleitsätze, die für alle Maschinen gültig sind.
- Typ B-Normen: Sind Sicherheits-Gruppennormen, die auf unterschiedliche Maschinengruppen anwendbar sind, wie z. B. DIN EN 60204-1: Elektrische Ausrüstung von Maschinen. B2-Normen behandeln spezielle Sicherheitseinrichtungen, wie z. B. DIN EN 418: NOT-AUS-Einrichtungen.
- Typ C-Normen: Sind Sicherheits-Fachnormen und beschreiben konkrete Anforderungen an einzelne Maschinenarten.

Für Normen von Typ C gilt das „Vermutungsprinzip“, d. h. bei Einhaltung dieser Normen darf vermutet werden, dass die Anforderungen der betreffenden EG-Richtlinien erfüllt sind.

66.2 Sicherheitsbegriff

Der Begriff *Sicherheit eines Steuerungssystems* ist auf die möglichen Folgen von auftretenden Fehlern bezogen, die Personen und Sachen betreffen. Davon zu unterscheiden ist der Begriff der *Verfügbarkeit* eines technischen Systems, die zwischen 0 und 100 % liegen kann, unabhängig von der Bedeutung der möglichen Folgen eines Ausfalls.

Um funktionale Sicherheit einer Maschine zu erreichen, ist es erforderlich, dass die sicherheitsbezogenen Teile der Schutz- und Steuereinrichtungen korrekt funktionieren und sich im Fehlerfall so verhalten, dass die Maschine in einem sicheren Zustand bleibt oder dorthin gebracht wird (Tab. 66.1).

Das Maß für die erreichte funktionale Sicherheit wird in den Normen mit unterschiedlichen Wertigkeitsstufen ausgedrückt, und zwar mit:

- Safety Integrity Level in IEC 61508: SIL 1 bis SIL 4 (Höchststufe)

Tab. 66.1 Beschreibung der Anforderungen für Sicherheitskategorien

Kategorie	Kurzfassung der Anforderungen	Systemverhalten	Maßnahmen
B	Sicherheitsbezogene Teile von Steuerungen nach dem Stand der Technik Bauteile müssen den zu erwartenden Einflüssen standhalten	Das Auftreten eines Fehlers kann zum Verlust der Sicherheitsfunktion führen.	z. B. 1-kanaliger Sicherheitskreis, Erdung des Steuerstromkreises
1	Anforderungen von „B“ müssen erfüllt sein Einsatz bewährter Bauteile und Sicherheitsprinzipien	Das Auftreten eines Fehlers kann zum Verlust der Sicherheitsfunktion führen, aber höhere Zuverlässigkeit als in Kategorie B	Zusätzlich z. B.: zwangsöffnende und zwangsgeführte Kontakte
2	Anforderungen von „B“ müssen erfüllt sein. Einsatz bewährter Bauteile und Sicherheitsprinzipien Testung der Sicherheitsfunktion in angemessenen Zeitabständen durch die Steuerung	Das Auftreten eines Fehlers kann zum Verlust der Sicherheitsfunktion zwischen den Prüfungsabständen führen. Der Verlust der Sicherheitsfunktion wird durch die Prüfung erkannt	Zusätzlich z. B.: Funktions-/Anlauftestung
3	Anforderungen von „B“ sind zu erfüllen, Einsatz bewährter Bauteile und Sicherheitsprinzipien Einfehler-Sicherheit: Ein einzelner Fehler führt nicht zum Verlust der Sicherheitsfunktion Der einzelne Fehler wird erkannt mit der Einschränkung „wann immer und in angemessener Weise durchführbar“	Wenn der einzelne Fehler auftritt, bleibt die Sicherheitsfunktion immer erhalten. Einige aber nicht alle Fehler werden erkannt. Eine Anhäufung unerkannter Fehler kann zum Verlust der Sicherheitsfunktion führen	Zusätzlich z. B.: 2-kanalige Ausführung von Sicherheitskreisen
4	Anforderungen von „B“ sind zu erfüllen, Anwendung bewährter Prinzipien 1-Fehlersicherheit ist gewährleistet Erkennung des einzelnen Fehlers vor oder bei nächster Anforderung an die Sicherheitsfunktion (Selbstüberwachung) Falls die Erkennung des einzelnen Fehlers nicht möglich ist, darf eine Anhäufung von Fehlern nicht zum Verlust der Sicherheitsfunktion führen	Wenn Fehler auftreten, bleibt die Sicherheitsfunktion immer erhalten. Die Fehler werden rechtzeitig erkannt, um einen Verlust der Sicherheitsfunktion zu verhindern	Zusätzlich z. B.: Selbstüberwachung der Sicherheitskreise, Querschlusserkennung

- Kategorien in der DIN EN 954-1: KAT 1 bis KAT 4 (Höchststufe)
- Performance Level in ISO 13849-1: PLa bis PLe (Höchststufe)

Startfunktion

- Start-Funktionen müssen durch Erregen des entsprechenden Kreises erfolgen.

Stoppfunktionen

- Stopp-Funktionen müssen durch Entregen des entsprechenden Kreises erfolgen und haben Vorrang vor zugeordneten Start-Funktionen. Das Rücksetzen der Stopp-Funktion darf keinen Gefahr bringenden Zustand einleiten. Bei den Stopp-Funktionen gibt es die Kategorien 0, 1 und 2.

Anforderungen der Stopp-Kategorien:

1. **Kategorie 0** ist ein Stillsetzen durch sofortiges Abschalten der Energiezufuhr zu den Antrieben, d. h. ein *ungesteuertes Stillsetzen*, z. B. durch Betätigung des Motor Leistungsschalters und aller Bremsen (mechanischer Stillsetzeinrichtungen), (Abb. 66.2).
2. **Kategorie 1** ist ein *gesteuertes Stillsetzen*, bei dem die Energiezufuhr zu den Maschinenantrieben beibehalten wird, um das gesteuerte Stillsetzen ausführen zu können. Die Energiezufuhr wird erst dann unterbrochen, wenn der Stillstand erreicht ist, z. B. Gegenstrombremsung von Drehstrommotoren (Abb. 66.3).
3. **Kategorie 2** ist ein *gesteuertes Stillsetzen*, bei dem die Energiezufuhr zu den Maschinenantrieben erhalten bleibt, z. B. Anhalten durch Vorgabe von Sollwert „0“ (Abb. 66.4).

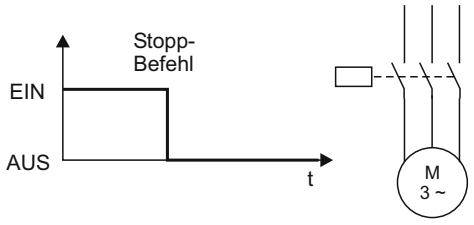


Abb. 66.2 Stopp-Kategorie 0

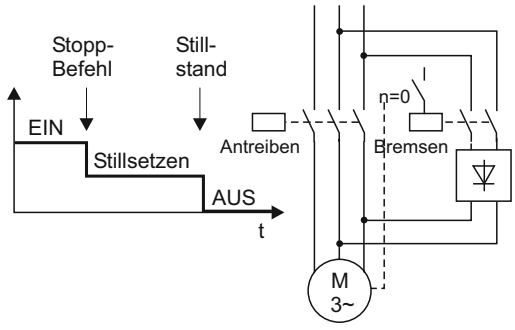


Abb. 66.3 Stopp-Kategorie 1

Handlungen im Notfall (NOT_AUS)

Der neue Begriff *Handlungen im Notfall* ist Teil eines neuen Sicherheitskonzepts und steht für eine differenzierte Betrachtung der Handlungen, die im Notfall auszuführen sind. Im informativen Anhang der Norm DIN EN 60204-1 wird das noch in der

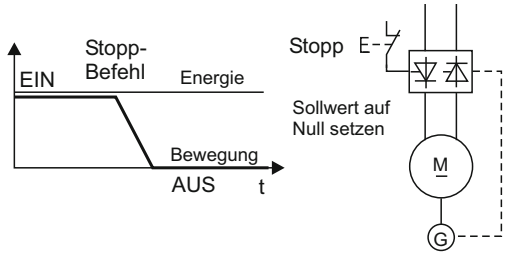


Abb. 66.4 Stopp-Kategorie 2

Beratung befindliche Konzept erläutert. Eine Handlung im Notfall schließt einzeln oder in Kombination ein:

1. Stillsetzen im Notfall (Risiko durch einen Prozessablauf oder eine Bewegung)
2. Ingangsetzen im Notfall!
3. Ausschalten im Notfall (Risiko durch elektrische Gefährdung)
4. Einschalten im Notfall!

Die Norm macht nur Aussagen über *Stillsetzen im Notfall* sowie *Ausschalten im Notfall* und bestimmt, dass beide durch eine einzige menschliche Handlung auszulösen sind.

Es besteht eine enge Anbindung der Notfallsignale an die Stoppkategorien:

- *Stillsetzen im Notfall* muss entweder als Stopp der Kategorie 0 oder 1 wirken und über die Risikoanalyse bestimmt werden.
  Für das Stillsetzen im Notfall der Kategorie 0 dürfen nur festverdrahtete, elektromechanische Betriebsmittel verwendet werden. Zusätzlich darf die Funktion nicht von einer elektronischen Schaltlogik (Hardware oder Software) oder von der Übertragung von Befehlen über ein Kommunikationsnetzwerk oder eine Datenverbindung abhängen.
  Bei der Stopp-Funktion der Kategorie 1 für die Stillsetz-Funktion im Notfall muss die endgültige Abschaltung der Energie der Maschinen-Antriebselemente sichergestellt sein und muss durch Verwendung von elektromechanischen Betriebsmitteln erfolgen.
  Anmerkung der Norm: Die Norm weist in ihrem deutschen Vorwort daraufhin, dass in Fällen, in denen andere Normen andere technische Lösungen zulassen als in DIN EN 60204-1 festgelegt ist, diese anderen technischen Lösungen zur Anwendung gelangen dürfen! Damit ist auch klargestellt, dass elektronische Betriebsmittel doch für NOT-AUS-Einrichtungen unabhängig von der Stopp-Kategorie eingesetzt werden dürfen, wenn diese unter Anwendung der Normen DIN EN 954-1 (Risikoanalyse) und/oder IEC 61508 die gleiche Sicherheit erfüllen, wie nach DIN EN 60204-1 gefordert.
- *Ausschalten im Notfall* sollte vorgesehen werden, wo ein Schutz gegen direktes Berühren aktiver Stromkreisteile (z. B. Schaltgeräte in elektrischen Betriebsräumen) nur durch Abstand oder Hindernisse erreicht wird oder wo es die Möglichkeit einer anderen Gefährdung durch elektrische Energie gibt.
  Ein Ausschalten im Notfall wird durch Abschalten der Maschine von der Versorgung erreicht mit der Folge eines ungesteuerten Stillsetzens (Stopp-0-Kategorie) der Maschine. Ist das nicht zulässig, kann es notwendig sein, einen anderen Schutz z. B. gegen direktes Berühren vorzusehen, sodass ein Ausschalten im Notfall nicht notwendig ist.

Regelungstechnische Grundbegriffe der Automatisierungstechnik

67

Günter Wellenreuther und Dieter Zastrow



67.1 Unterschied zwischen Steuern und Regeln, regelungstechnische Größen 822
67.2 Regler-Technologien 823

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_67

Die folgenden Abschnitte über regelungstechnische Grundbegriffe, Regelstrecken und Regler sind nur eine enge Ausschnittsbetrachtung einer umfassenden Wissenschaft, die sich mit der gezielten Beeinflussung dynamischer Prozesse während des Prozessablaufs beschäftigt und dafür mathematische Methoden zur Systembeschreibung und -untersuchung entwickelt hat. Es geht hier nicht um die mathematische Behandlung von Regelkreisen, sondern im Kern nur darum, für eine bestimmte Regelungsaufgabe einen geeigneten Reglertyp auszuwählen und den Einfluss seiner Regelparameter auf das Verhalten im Regelkreis kennenzulernen.

67.1 Unterschied zwischen Steuern und Regeln, regelungstechnische Größen

Kennzeichen einer reinen Steuerung ist der sog. offene Wirkungsablauf, d. h. die Eingangssignale enthalten keine Rückmeldungen über ihre aktuelle Auswirkung auf den Prozess. Dieser Fall ist z. B. bei einer einfachen Drehzahlsteuerung eines Motors gegeben, wie in Abb. 67.1 dargestellt. Mit dem einstellbaren Widerstand kann die Motordrehzahl gesteuert werden. Eine unterschiedlich schwere Last tritt als Störgröße auf und beeinflusst die tatsächliche Motordrehzahl. Eine solche Wirkungsstruktur bezeichnet man als *Steuerung*, im Sinne von Vorwärtssteuerung.

Um die Darstellung unabhängig von einer bestimmten Anwendung zu machen, führt man allgemeingültige Symbole und steuerungstechnische Größen ein, wie Abb. 67.2 zeigt, die man sich für den jeweiligen Anwendungsfall anschaulich übersetzen muss, z. B.:

w = Führungsgröße ⇒ die Ausgangsspannung des Potenziometers (Sollwertvorgabe)
y = Stellgröße ⇒ die Motorspannung;
x = Steuergröße ⇒ die Drehzahl, bei der man einen gewünschten Sollwert und einen tatsächlichen Istwert zu unterscheiden hat;
z = Störgröße ⇒ die unterschiedlichen Lasten.

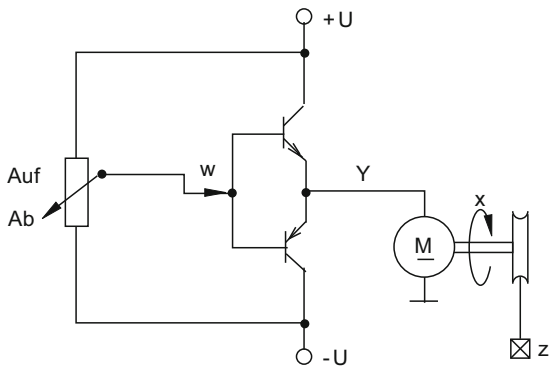


Abb. 67.1 Drehzahl-Steuerung

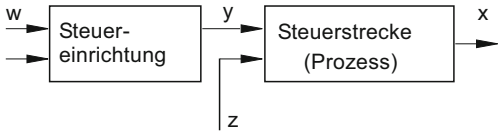


Abb. 67.2 Struktur einer Steuerung

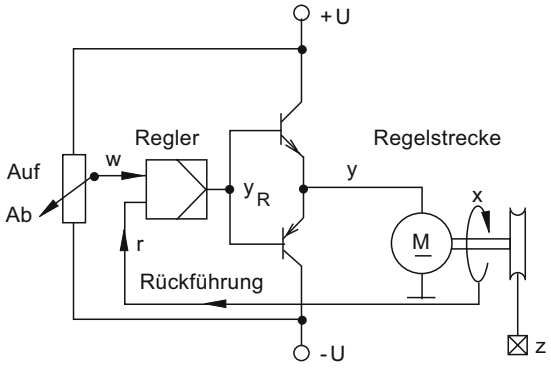


Abb. 67.3 Drehzahlregelung

Im Gegensatz zum Typ der Vorwärtssteuerung hat eine Regelung einen Informationsrückfluss, sodass ein geschlossener Wirkungsablauf entsteht, den man als Regelkreis bezeichnet. Beim Motor in Abb. 67.3 wird die tatsächliche Drehzahl zu einem Regler zurückgeführt. Hat sich die Drehzahl durch eine stärkere Belastung verringert, kann dies erkannt und nachgeregelt werden. Steuerungen mit Rückführung werden als *Regelung*en bezeichnet.

Regeln ist ein Vorgang, bei dem eine Größe, die man die *Regelgröße* nennt, fortlaufend erfasst und mit einer anderen Größe, die man als *Führungsgröße* bezeichnet, fortlaufend verglichen wird. Abhängig vom Ergebnis des Vergleichs muss das Regelungsprogramm eine Ausgangsgröße bilden, welche die Regelgröße im Sinne einer Angleichung an die Führungsgröße beeinflusst. Der sich dabei ergebende Wirkungsablauf findet in einem geschlossenen Kreis, dem *Regelkreis*, statt.

Das Erfassen der Regelgröße erfolgt durch Messen ihres aktuellen Wertes (Istwert). Die fortlaufende Erfassung bedeutet für eine SPS als Regler soviel wie eine hinreichend häufige Abtastung des Messwertes. Die Angleichung der Regelgröße an die Führungsgröße wird durch das ablaufende Regelungsprogramm erreicht.

Die SPS als Regler führt dabei einen Vergleich zwischen *Führungsgröße* w und *Regelgröße* x (genauer: *Rückführgröße* r) durch und ermittelt dabei die *Regeldifferenz* e. Aus der Regeldifferenz wird über die Reglerfunktion das *Reglerausgangssignal* y_R ermittelt. Dieses ist eine leistungsarme Signalgröße, die einem Stellgerät zugeführt wird, das den passenden Anschluss an die Regelstrecke herstellt. Das *Stellgerät* besteht oft aus einem *Stellantrieb* und einem *Stellglied*. Der Stellantrieb bildet aus dem Reglerausgangssignal y_R die *Stellgröße* y und betätigt das Stellglied, welches den Massen- oder Energiestrom für die geregelte Anlage dosiert. Die *Regelstrecke* ist derjenige Teil

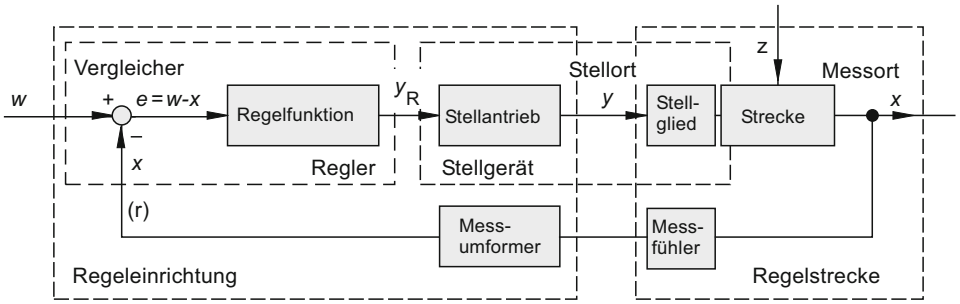


Abb. 67.4 Ausführlich dargestellte Struktur eines Regelkreises

des Regelkreises, in dem die Regelgröße geregelt wird. Sie beginnt am *Stellort*, also dort, wo die Stellgröße y in den Massen- oder Energiestrom eingreift und endet am *Messort*, wo sich der Messfühler zur Erfassung der Regelgröße x befindet. Der Stellantrieb wird zur Regeleinrichtung gezählt, weil er ein eigenes Zeitverhalten besitzt und damit das Gesamtzeitverhalten der Regeleinrichtung beeinflusst. Das Stellglied dagegen wird zur Regelstrecke gerechnet, weil es sich um deren Ventile, Schieber, Klappen oder elektronische Schalter (Transistoren, Thyristoren) handelt. Der Messfühler gehört ebenfalls zur Regelstrecke, denn er arbeitet nicht trägheitslos und beeinflusst damit das Zeitverhalten der Regelstrecke. Abb. 67.4 zeigt die Struktur einer Regelung in ausführlicher Darstellung. Gebräuchlich sind aber auch vereinfachte Darstellungen, wenn sie ihren Zweck erfüllen.

In den Lösungen von Automatisierungsaufgaben kommen Steuerungen und Regelungen zusammen vor. Geregelt werden muss immer dann, wenn Größen auf genaue Werte geführt oder gegen den Einfluss veränderlicher Störgrößen z konstant gehalten werden müssen. So ist z. B. eine Ablaufsteuerung vom Prinzip her eine Steuerung. Wenn aber während der Zeitdauer eines Schrittes z. B. die Temperatur eines Mischgutes bei unterschiedlichen Füllmengen im Behälter konstant gehalten werden muss, kann dies nur mit einer unterlagerten Regelung erreicht werden. Die SPS kann eine Vielzahl zeitgleich ablaufender Vorgänge steuern und regeln.

67.2 Regler-Technologien

In späteren Abschnitten werden Reglertypen nur noch im Sinne von ausführbaren Regelfunktionen, also recht abstrakt, beschrieben. Die technologische Ausführung von Reglern dagegen unterliegt der technischen Entwicklung.

Die ersten Regler waren schaltende Regler, die rein elektromechanische realisiert wurden und deren bekanntester Vertreter der noch heute verwendete *Bimetallkontaktregler* ist, wie er in einfachen Geräte, z. B. Toastern, vorkommt.

Es wurden dann Regler für anspruchsvollere Regelfunktionen benötigt, um sowohl analoge Eingangssignale verarbeiten als auch analoge Ausgangssignale ausgeben zu können. Analoge

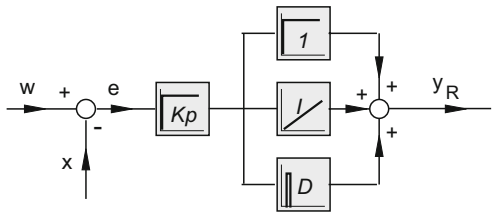


Abb. 67.5 PID-Regelfunktion

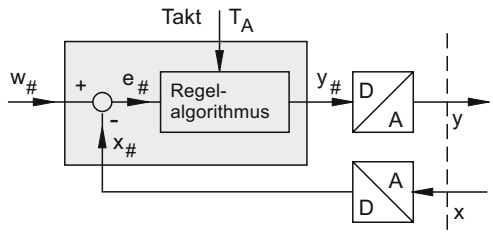


Abb. 67.6 Vereinfacht dargestellter Aufbau eines Digitalreglers. Mit # gekennzeichnete Größen sind die zu Regelungssignalen proportionalen Zahlenwerte

Signale sind solche, die innerhalb bestimmter Grenzen wie z. B. $\pm 10\,V$ jeden beliebigen Wert annehmen können. Diese sog. *Analogregler* bestehen technologisch betrachtet aus einem Verstärker und benötigen eine Energiequelle. Verstärken ist dabei ein Vorgang, bei dem mit möglichst kleinem Energieaufwand am Verstärkereingang eine möglichst kräftige Energieabgabe der Energiequelle an den Verstärkerausgang signalgesteuert erfolgt. Als Verstärker für analoge Regler kommen elektronische aber auch pneumatische und hydraulische Elemente in Frage. Bei den elektronischen Operationsverstärkern werden die typischen Reglerfunktionen, die man mit den Kennbuchstaben P für proportional, I für integral und D für differenzial bezeichnet, durch spezielle Beschaltungsmaßnahmen erreicht, die den Anwender jedoch nicht interessieren müssen. Den Schlusspunkt dieser Entwicklung bildet der Universalregler in der Ausführung als analoger PID-Regler, der alle Funktionen ausführen kann, wie in Abb. 67.5 schematisch dargestellt.

Mit der Verfügbarkeit von preiswerten Mikrocontrollern lassen sich heute sog. digitale Regler mit entsprechender Software für den Regelalgorithmus sowie AD- und DA-Umsetzern für die Signalumwandlungen realisieren. Die Signale außerhalb des di-

gitalen Reglers können nach wie vor analog sein, im Innern des Reglers werden nur digitale (zahlenmäßige) Signale verarbeitet. Abb. 67.6 zeigt den prinzipiellen Aufbau eines digitalen Reglers, der auch für die SPS zutrifft.

Der digitale Regler führt eine taktgesteuerte Abtastregelung aus, weil er neue Eingangswerte erst entgegennehmen kann, wenn er einen Berechnungszyklus abgeschlossen hat, für den er eine gewisse Rechenzeit benötigt. Das hat zur Folge, dass neue Eingangsgrößen nur zu diskreten Zeitpunkten eingelesen und die Stellgröße nur zur diskreten Zeitpunkten ausgegeben werden können. Im Regelungsprogramm, dass in gleichen Zeitabständen bearbeitet werden muss, wird aus dem Momentanwert der digitalisierten Regelgröße $x_\#$ und einer Führungsgröße $w_\#$ die Regeldifferenz $e_\#$ berechnet und daraus nach programmierten Funktionen die Stellgröße $y_\#$ gebildet. Die genau einzuhaltenden Zeitabstände werden als Abtastzeit T_A bezeichnet.

Da auch der Analog-Digital-Umsetzer zur Umwandlung eines Momentanwertes der Regelgröße eine gewisse Zeit benötigt, muss dieser Wert für den Wandlungszeitraum mit einem Halteglied konstant gehalten werden, um Umwandlungsfehler zu vermeiden. Bei der Digitalisierung der Regelgröße x im Analog-Digital-Umsetzer kann je nach seiner Stellenzahl nur eine bestimmte Unterscheidungsfähigkeit für Signalwerte erreicht werden. Bei einer Stellenzahl von 8 Bit stehen nur $2^8 = 256$ mögliche Zahlenwerte zur Verfügung, sodass sich bei einem Analogsignalbereich von 0 bis 10 V nur Spannungsstufen von ca. 40 mV unterscheiden lassen.

Auf Grund der wert- und zeitdiskreten Signalverarbeitung der digitalen Regler ergeben sich Grenzen hinsichtlich der Verarbeitung höherer Signalfrequenzen. Dafür können digitale Regler aber nicht nur die PID-Regelfunktion nachbilden, sondern ermöglichen auch die Realisierung neuartiger Reglertypen.

Regelstrecken

68

Günter Wellenreuther und Dieter Zastrow



68.1 Beispiele für Regelstrecken 826

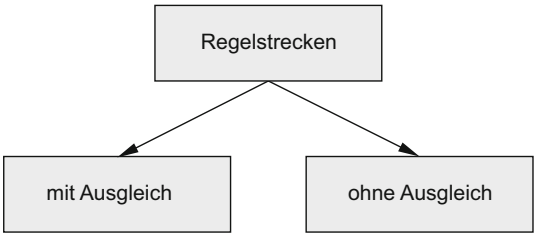
68.2 Beschreibungsmittel zur Darstellung von Regelstreckeneigenschaften 826

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_68

68.1 Beispiele für Regelstrecken

Das nachfolgend beschriebene Ordnungsschema soll helfen, um sich besser in der unübersehbaren Vielfalt möglicher Regelstrecken zurecht zu finden.

Erstes Ordnungskriterium Regelstrecken mit und ohne Ausgleich



Als Beispiel für eine *Regelstrecke mit Ausgleich* sei die Temperaturregelung einer Heizungsanlage genannt. Bei Veränderung der Mischventilstellung erreicht die Raumtemperatur auf verändertem Niveau wieder einen stabilen Wert.

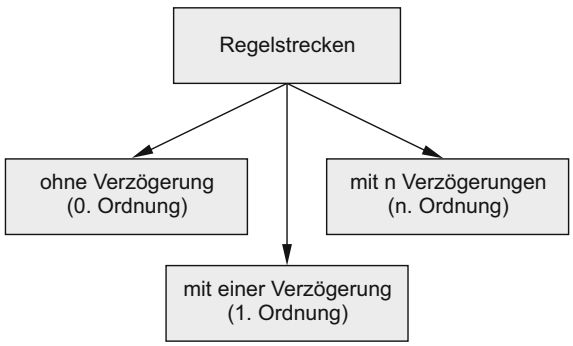
Ausgleich bedeutet, dass die Regelgröße x der Strecke nach einer sprungartigen Änderung der Stellgröße y innerhalb einer abzuwartenden Übergangszeit wieder einen stabilen Zustand, den sog. Beharrungszustand erreicht.

Regelstrecken mit Ausgleich und konstantem Übertragungsbeiwert K_{PS} haben einen proportionalen Charakter und werden deshalb auch *P-Strecken* genannt.

Bei einer *Regelstrecke ohne Ausgleich* würde die Regelgröße x nach einer sprungartigen Änderung der Stellgröße y keinen neuen Beharrungszustand finden. Dies ist z. B. der Fall bei einem Behälter mit dem Füllstand als Regelgröße x, wenn die Ablaufmenge in m^3/h durch eine Pumpe konstant gehalten wird. Jede Änderung der Zulaufmenge in m^3/h führt dann entweder zum Überlaufen oder Leerlaufen des Behälters.

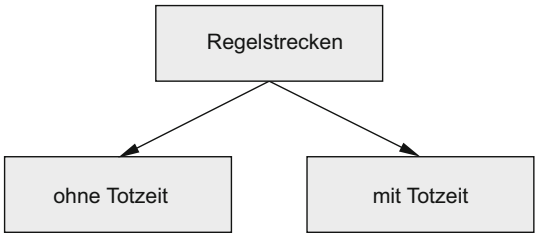
Regelstrecken ohne Ausgleich haben einen integralen Charakter und werden daher als *I-Strecken* bezeichnet.

Zweites Ordnungskriterium Regelstrecken mit und ohne Verzögerung



Verzögerung bedeutet, dass die Regelgröße x einer sprungartigen Änderung der Stellgröße y nicht sprunghaft folgen kann, sondern erst nach einer bestimmten Zeit einen neuen stabilen Wert erreicht. Verzögerungen treten bei technischen Prozessen immer auf, wenn Energie zu- oder abgeführt oder Massen beschleunigt oder abgebremst werden müssen.

Drittes Ordnungskriterium Regelstrecken mit und ohne Totzeit



Ein bekanntes Beispiel für eine Regelstrecke mit Totzeit ist das Förderband. Durch eine Schieberöffnung (Stellgrößenänderung Δy) gelang mehr Fördergut auf das Band. Die höhere Ausschüttmenge (Regelgrößenänderung Δx) wirkt sich am Ende des Förderbandes jedoch nicht sofort aus, sondern erst nach einer Totzeit, die von der Geschwindigkeit und der Länge des Bandes abhängt.

In der Praxis vorkommende Regelstrecken weisen zumeist Kombinationen von Eigenschaften auf. Die Kriterien Ausgleich, Verzögerung und Totzeit treten dann gemeinsam auf. Tab. 68.1 zeigt Beispiele für Regelstrecken mit den genannten Ordnungskriterien. Die definierten Streckenparameter und typischen Sprungantworten werden im Anschluss erläutert.

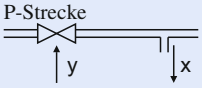
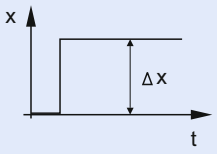
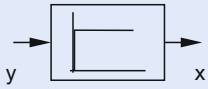
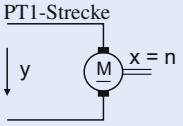
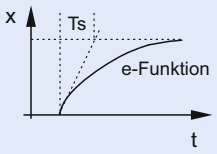
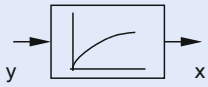
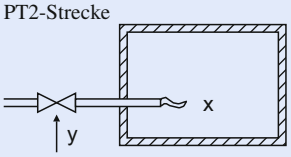
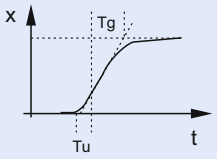
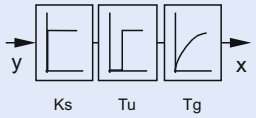
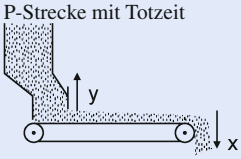
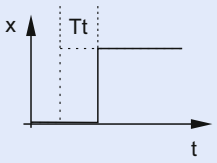
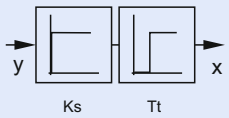
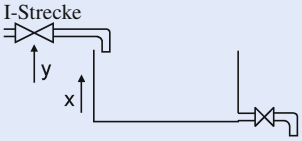
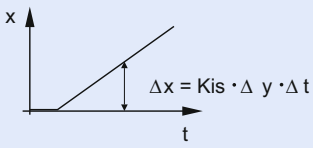
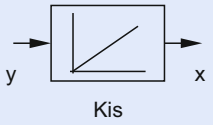
68.2 Beschreibungsmittel zur Darstellung von Regelstreckeneigenschaften

Für die Auswahl eines geeigneten Reglers muss das Verhalten der Regelstrecke bekannt sein. Man versucht das Verhalten der Regelstrecken durch die Angabe von Kennlinien und Zeitfunktionen zu beschreiben, um daraus die Regelstreckenparameter abzuleiten.

Kennlinien von Strecken mit Ausgleich, Übertragungsbeiwert K_{PS}

Die Kennlinie einer Regelstrecke ist die grafische Darstellung des Zusammenhangs zwischen Stellgröße y und Regelgröße x. Eine *Kennlinie* ist somit ein Diagramm, bei dem die Abhängigkeit einer physikalischen Größe von einer anderen physikalischen Größe (nicht der Zeit) dargestellt ist. Auf der senkrechten Achse können z. B. die Raumtemperatur und auf der waagrechten Achse die Stellung des Mischventils aufgetragen sein: $T = f(\alpha)$.

Tab. 68.1 Regelstrecken

Art der Strecke Beispiel	Sprungantwort	Streckenparameter	Signalflussplandarstellung
P-Strecke  Druck und Durchfluss in Flüssigkeits-rohrnetzen		K_{PS}	
PT1-Strecke  Drehzahl		K_{PS}, T_S	
PT2-Strecke  Ofentemperatur		K_{PS}, T_u, T_g	
P-Strecke mit Totzeit  Fördermenge		K_{PS}, T_t	
I-Strecke  Füllstand		K_{IS}	

Der *Übertragungsbeiwert* K_s der Regelstrecke ist durch den Quotienten der Änderung der Regelgröße x und der Änderung der Stellgröße y bestimmt. Im Signalflussplan entspricht dies dem Verhältnis der Ausgangsgröße bezogen auf die Eingangsgröße. Damit kann bei bekanntem Übertragungsbeiwert der Regelstrecke bei einer Änderung der Stellgröße Δy die zu erwartende Änderung der Regelgröße Δx bestimmt werden. Die Zeit für die Änderung spielt dabei insofern keine Rolle, weil die zu jeder Stellgrößenänderung sich ergebende Regelgrößenänderung abgewartet werden muss (Erreichen des Beharrungszustandes). Man spricht deshalb auch von der statischen Kennlinie der Regelstrecke.

Die Kennlinie zeigt den Zusammenhang zwischen Regelgröße x und Stellgröße y innerhalb des Laufbereichs beider Größen bei konstantem Störgrößeneinfluss. Zu jedem Stellgrößenwert y ist der zugehörige Regelgrößenwert x im Beharrungszustand aufgetragen, siehe Abb. 68.1.

Ist diese statische Kennlinie eine Gerade aus dem Achsenursprung, spricht man von einer P-Regelstrecke mit konstantem Übertragungsbeiwert K_{PS} (Index P für proportional, Index S für Strecke). Zur Berechnung von K_{PS} darf man deshalb auch den ganzen Regelbereich einsetzen:

$$K_{PS} = \frac{\Delta X_h}{\Delta Y_h}$$

mit X_h = Stellwirkung (Regelbereich der Strecke), Y_h = Stellbereich des Reglers.

Häufig haben Regelstrecken jedoch nichtlinearer Kennlinien, d. h. sie sind gekrümmt, sodass der Übertragungsbeiwert K_{PS} keine Konstante ist. Das möchte man vermeiden, weil solche Strecken viel schwieriger zu regeln sind. Abhilfe kann durch eine umgekehrt nichtlineare Kennlinie des Stellgliedes geschaffen werden. Dieser mögliche „Trick“ zeigt auch, warum das Stellglied zur Regelstrecke gehört.

Zeitverhalten von Regelstrecken mit Ausgleich

Ausgangspunkt der Betrachtung sei eine Regelstrecke mit Ausgleich und linearer Kennlinie. Damit ist die Regelstrecke noch nicht eindeutig beschrieben, denn es fehlt noch die Kenntnis über das Zeitverhalten der Strecke. Die zusätzliche Fragestellung lautet: Wie reagiert die Regelgröße x der Strecke auf eine plötzliche Änderung der Stellgröße um Δy dargestellt im zeitlichen Verlauf?

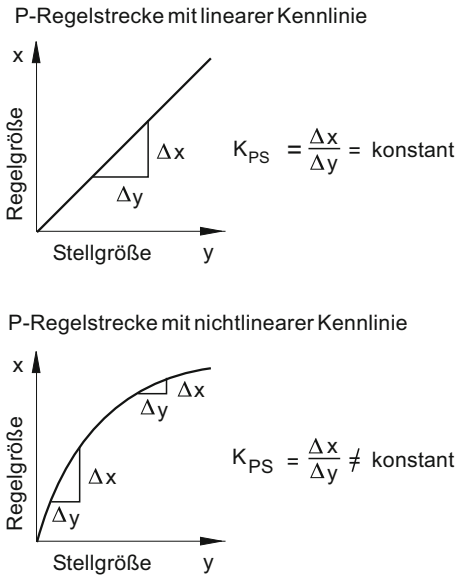


Abb. 68.1 Kennlinien von Regelstrecken mit Ausgleich

1. Regelstrecke mit Ausgleich 0. Ordnung: *P-Strecke*
   Kennzeichen einer P-Strecke ist die proportionale Reaktion der Regelgröße x auf die Änderung der Stellgröße y. Wird die Stellgröße sprungartig verändert, reagiert die Regelgröße ebenfalls sprungartig. In der Tab. 68.1 ist als Beispiel für eine P-Strecke ein Durchfluss in einer Rohrleitung angegeben. Wird durch stärkeres Öffnen eines Ventils am Anfang der Rohrleitung der Durchfluss erhöht, ist dies auch am Ende der Rohrleitung sofort in gleicher Weise der Fall, da die Flüssigkeit nicht komprimierbar ist.
   Die P-Strecke ist allein durch den Streckenparameter K_{PS} (Übertragungsbeiwert) gekennzeichnet.
2. Regelstrecke mit Ausgleich 1. Ordnung: *PT1-Strecke*
   Kennzeichen einer PT1-Strecke ist ein anfänglich schneller Anstieg der Regelgröße x, der sich dann aber immer mehr verzögert und erst nach längerer Zeit den neuen Beharrungszustand erreicht. In der Tab. 68.1 ist als Beispiel für eine PT1-Strecke die Drehzahländerung eines Motors angegeben. Der Motor hat auch ohne angeflanschte Arbeitsmaschine eine eigene Schwungmasse, die sich einer sofortigen Drehzahländerung widersetzt. Abstrakt ausgedrückt: Die PT1-Strecke ist eine Strecke mit Ausgleich und einen Energiespeicher, der die Verzögerung bewirkt.
   Die PT1-Strecke ist durch die Streckenparameter K_{PS} (Übertragungsbeiwert) und T_S (Verzögerungskonstante) gekennzeichnet. Die Verzögerungskonstante lässt sich durch Anlegen einer Tangente an die Sprungantwort ermitteln.
3. Regelstrecke mit Ausgleich 2. Ordnung: PT2-Strecke
   Kennzeichen einer PT2-Strecke ist ein anfänglich langsamer Anstieg der Regelgröße x, der sich dann aber beschleunigt und danach wieder verzögert bis der neue Beharrungszustand erreicht wird. Als Beispiel für eine PT2-Strecke wird in Tab. 68.1 eine Heizung angegeben. Es sind zwei Speicher vorhanden. Wird der Heißwasserdurchfluss sprungartig erhöht, muss sich erst der Heizkörper auf die neue Temperatur einstellen und dann erst wird mit dem Aufheizen der Raumluft begonnen.

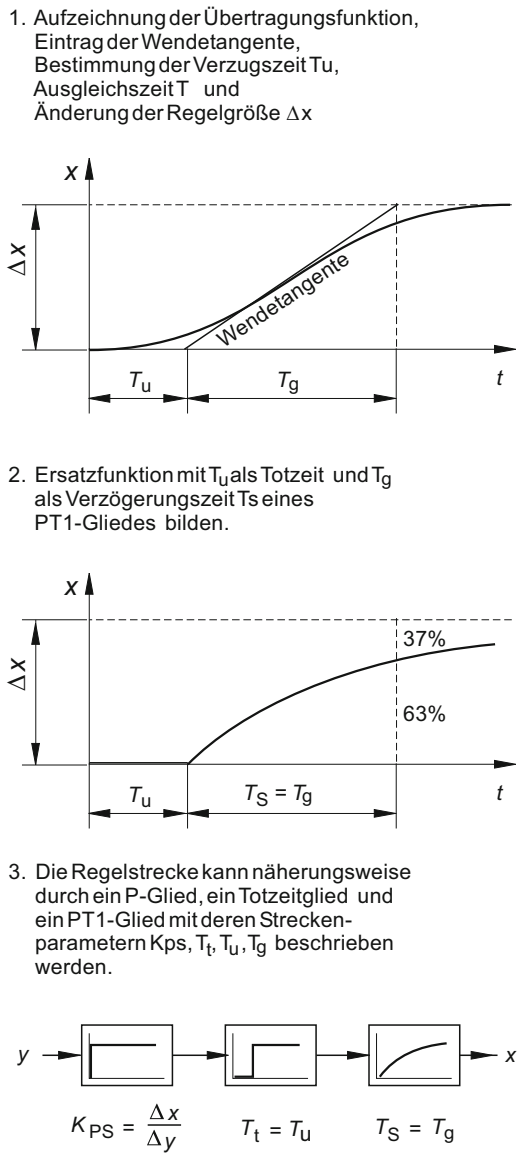


Abb. 68.2 Bestimmung der Regelparameter einer unbekannten Regelstrecke

Die Streckenparameter aus der Sprungantwort der PT2-Strecke sind K_{PS} (Übertragungsbeiwert) und die durch die eingetragene Wendetangente gebildeten Zeitkonstanten T_u (Verzugszeit) und T_g (Ausgleichszeit).
Das Ordnungszahlschema lässt sich entsprechend der Anzahl der in einer Regelstrecke vorkommenden Speicher noch weiter erhöhen.

4. Regelstrecke mit Ausgleich und Totzeit: PT_t
   Kennzeichen eines Totzeitverhaltens ist, dass sich die Regelgröße x zwar in gleicher Weise sprunghaft ändert wie die Stellgröße, jedoch erst um die Totzeit T_t zeitverzögert. In der Tab. 68.1 ist als Beispiel ein Förderband angegeben, bei dem eine gewisse Zeit vergeht, bis eine veränderte Schüttmenge am Ende des Förderbandes ankommt.
   Die Streckenparameter der PT_t-Strecke sind K_{PS} (Übertragungsbeiwert) und T_t (Totzeit).

5. Regelstrecke ohne Ausgleich 0. Ordnung: *I-Strecke*
   Bei den I-Strecken bildet sich die Regelgröße x durch Aufsummieren einer Stellgrößenänderung Δy über fortlaufende Zeitabschnitte Δt. Kennzeichen einer I-Strecke ist der zeitproportionale Anstieg oder Rückgang der Regelgröße, verbunden mit Überlauf oder Leerlauf, wenn nicht genau Abfluss gleich Zufluss ist. Als Beispiel für eine I-Strecke ist in der Tab. 68.1 ein offener Behälter mit freiem Zulauf angegeben.

Bestimmung der Regelstreckenparameter

Damit ein zur Regelstrecke passender Regler gefunden und günstig dimensioniert werden kann, müssen zuerst einmal die Streckenparameter der Regelstrecke bekannt sein. Dazu muss ein Signalflussplan mit dem zugehörigen Übertragungsbeiwert K_{PS} und den Zeitkonstanten ermittelt werden, der die zunächst noch unbekannte Regelstrecke möglichst gut beschreibt. Das kann auf experimentelle Art durch Aufnahme der Sprungantwort auf einen Stellgrößen- oder Störgrößensprung geschehen.

Der Ablauf des Experimentes gliedert sich in folgende Schritte:

1. Das Stellglied wird zum Zeitpunkt t_0 um den Wert Δy verstellt. Da es sich um eine Strecke mit Ausgleich handeln soll, wird die Regelgröße x einen um Δx verschobenen Beharrungszustand einnehmen. Die Antwort der Regelgröße x wird mit einem $y-t$-Schreiber aufgezeichnet und liegt dann als Liniendiagramm vor. An den Kurvenverlauf der Sprungantwort wird die Wendetangente eingezeichnet und aus den Achsenabschnitte die Verzugszeit T_u und die Ausgleichszeit T_g bestimmt.
2. Daraus kann eine Ersatzfunktion, bestehend aus einem Totzeitglied mit der Totzeit T_t und einem PT1-Glied mit der Verzögerungszeit T_S gebildet werden.
3. Die Übertragungsbeiwerte beider Glieder können als ein Übertragungsbeiwert K_{PS} in einem P-Glied zusammengefasst werden. Da es sich im einfachsten Fall um eine Regelstrecke mit Ausgleich und linearer Kennlinie handeln soll, lässt sich der Übertragungsbeiwert aus

$$K_{\mathrm{PS}} = \frac{\Delta x}{\Delta y}$$

   berechnen. Abb. 68.2 zeigt das Vorgehen anschaulich.

Erfahrungen haben ergeben, dass das Verhältnis von Verzugszeit T_u zu Ausgleichszeit T_g Auskunft über die Regelbarkeit der Strecke ergeben:

Gut regelbar	Noch regelbar	Schlecht regelbar
$\frac{T_u}{T_g} < \frac{1}{10}$	$\frac{T_u}{T_g} = \frac{1}{5}$	$\frac{T_u}{T_g} > \frac{1}{3}$

Regler

69

Günter Wellenreuther und Dieter Zastrow



69.1 P-Regler, P-Regelfunktion 832
69.2 I-Regler, I-Regelfunktion 834
69.3 PI-Regler, PI-Regelfunktion 834
69.4 PID-Regler, PID-Regelfunktion 834
69.5 Vergleich der verschiedenen Reglertypen 835
69.6 PID-Reglerbaustein für digitale Abtastregelung 835
69.7 SPS als kontinuierlicher PID-Abtastregler 836
69.8 Digitaler Schrittregler mit PI-Verhalten 837
69.9 Zweipunktregler, Zweipunkt-Regelfunktion 838
69.10 Regelgüte 838
Literatur 839

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_69

Aufgabe der Regler ist es, bei Abweichung der Regelgröße x von der Führungsgröße w die Reglerausgangsgröße y_R so zu verändern, dass die Regelgröße x in möglichst kurzer Zeit optimal an die Führungsgröße w angeglichen wird. Ursache für die Abweichung kann eine aufgetretene Störgröße z oder die veränderte Führungsgröße w sein.

In traditioneller Einteilung der Regler unterscheidet man

- stetige und unstetige Regler

oder

- schaltende Regler (Zweipunkt-, Dreipunktregler),
- analoge Regler und
- digitale Regler.

Aus Gründen der Umfangsbeschränkung soll hier auf die Darstellung der schaltungs- und gerätetechnischen Ausführungen von Reglern verzichtet werden. Es werden nur die typischen Eigenschaften bekannter Reglertypen anhand ihrer Regelfunktion und durch ihre sog. Sprungantworten beschrieben, siehe Übersicht in Abb. 69.2. Danach wird übergeleitet zum PID-Regelalgorithmus für den digitalen Abtastregler, der als Softwareregler in einer SPS realisiert werden kann.

69.1 P-Regler, P-Regelfunktion

Der P-Regler ist ein proportional wirkender Regler, der in seinem Arbeitsbereich, den man als Proportionalbereich bezeichnet, durch die Regelfunktion

$$y_R = K_{PR} \cdot e \quad \text{mit } K_{PR} = \text{Proportionalbeiwert}$$

beschrieben ist. Eine sprunghafte Änderung der Regeldifferenz e am Eingang führt ohne Zeitverzögerung zu einer proportionalen sprunghaften Änderung der Reglerausgangsgröße y_R am Ausgang, solange sich die Regeldifferenz e innerhalb des Proportionalbereichs bewegt.

Setzt man die Regeldifferenz $e = w - x$ in den Regelfunktion ein, so erhält man:

$$y_R = K_{PR} \cdot (w - x)$$

Diese Beziehung lässt erkennen, dass der P-Regler nicht das leisten kann, was von Reglern allgemein erwartet wird, nämlich die genaue Heranführung der Regelgröße x an die Führungsgröße w. Bei $x = w$ wird die Reglerausgangsgröße $y_R = 0$ und damit erhält die Regelstrecke kein Stellsignal y, um die Regelgröße x bilden zu können: Eine Heizung, der kein Wärmeträger zugeführt wird, kann keinen Temperaturwert bilden.

Andererseits kennzeichnet den P-Regler eine direkte Zuordnung von Regelgröße x zu Stellgröße y_R bei fest eingestelltem Proportionalbeiwert K_{PR} und vorgegebener Führungsgröße w, solange sich die Regelgröße x noch innerhalb des Proportionalbereich X_P befindet. Der Proportionalbereich X_P lässt sich aus dem Stellbereich Y_h und dem Proportionalbeiwert K_{PR} berechnen:

$$X_P = \frac{Y_h}{K_{PR}}$$

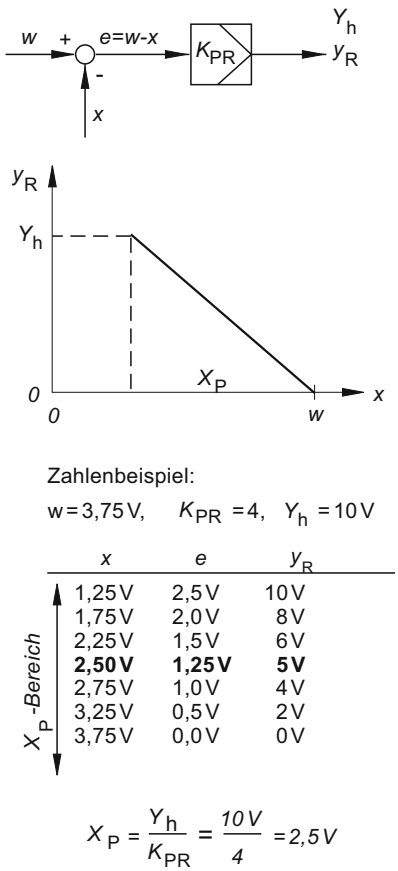


Abb. 69.1 Kennlinie des P-Reglers

Die feste Zuordnung von x und y_R kann anschaulich als statische Kennlinie des P-Reglers dargestellt werden, deren Achsenabschnitte sich aus

$$y_R = K_{PR} \cdot (w - x)$$

berechnen lassen:

$$\begin{aligned} &\text{Für } y_R = 0 && \Rightarrow \quad x = w \\ &\text{Für } y_R = Y_h = X_P \cdot K_{PR} && \Rightarrow \quad x = w - X_P \end{aligned}$$

Das Zahlenbeispiel zeigt, dass bei einem P-Regler die Eingangsgröße x und Ausgangsgröße y_R einander starr zugeordnet sind, solange sie sich im Proportionalbereich befinden. Wird in Folge einer Störeinwirkung z auf die Regelstrecke eine Änderung der Reglerausgangsgröße Δy_R erforderlich, dann ist dies nur bei einer gleichzeitigen Änderung der Regelgröße Δx möglich. Der P-Regler kann also eine Störung nicht völlig ausregeln. Der neue Beharrungszustand ist ein anderer Punkt auf der $y_R - x$-Kennlinie des P-Reglers, in deren fallenden Verlauf die Umkehrung des Wirkungssinns der Regelung zum Ausdruck kommt.

Die Änderung der Regelgröße Δx bei einer erforderlichen gleich großen Änderung der Reglerausgangsgröße Δy_R lässt sich ver-

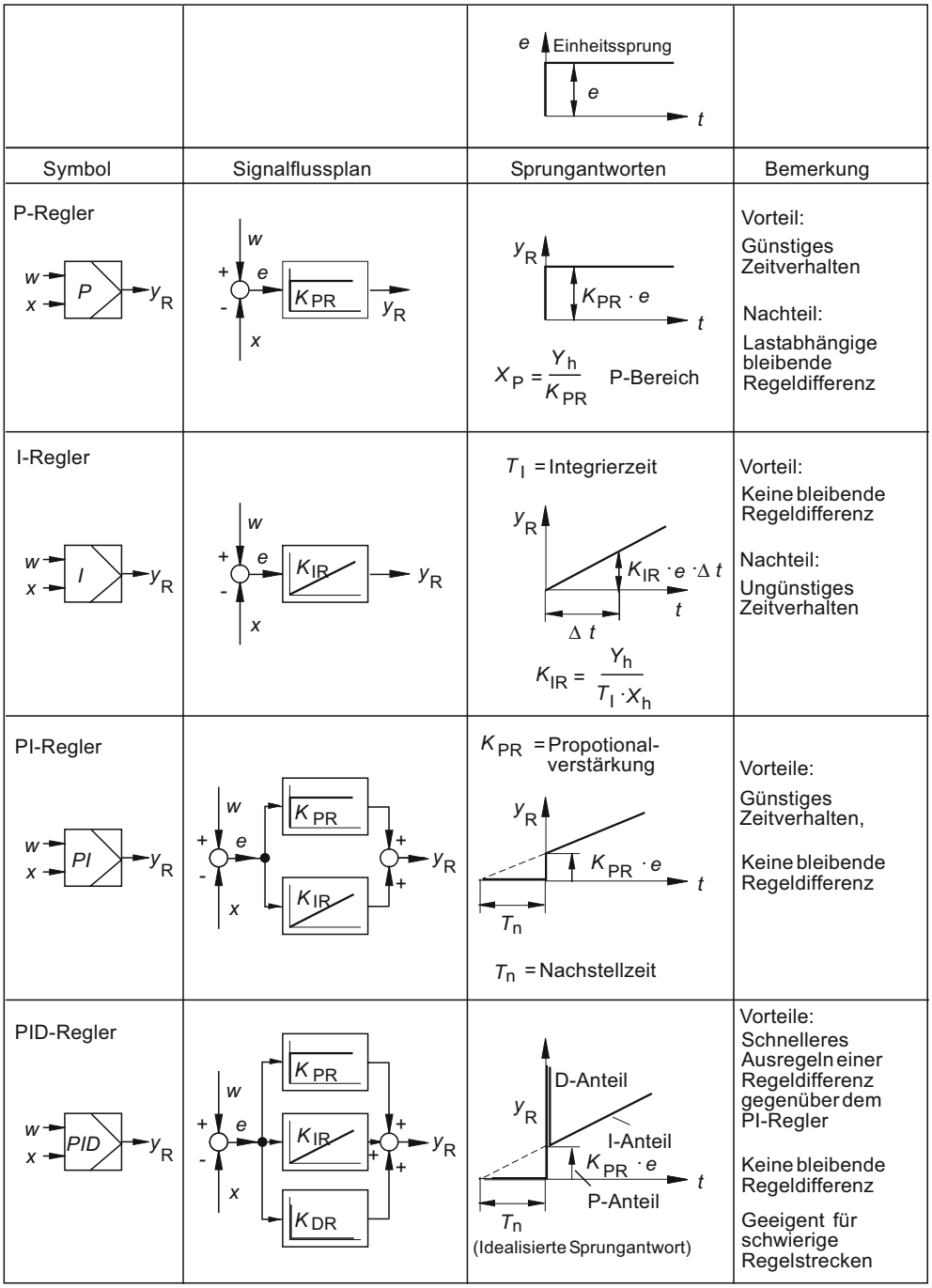


Abb. 69.2 Signalflusspläne und Sprungantworten der wichtigsten Reglertypen

ringern, wenn der Proportionalbereich X_P verkleinert wird, indem man K_P vergrößert. Dieser Verbesserung des P-Reglers sind aber Grenzen gesetzt, weil der Regelkreis instabil wird. Der P-Regler mit einem großen K_P-Wert reagiert überempfindlich auf kleinste Regelgrößenänderungen mit zu starken Ausschlägen bei der Stellgröße, die dann in extremer Weise ständig zwischen $y_R = 0$ und $y_R = Y_h$ umschaltet.

Der P-Regler als einfachster kontinuierlicher Regler mit sehr schnellem Reglereingriff kann überall dort gut verwendet werden, wo die Strecke einen kleinen Regler-Proportionalbereich erlaubt. Das ist der Fall, wenn entweder nur geringe Störgrößen z auftreten oder das Verhältnis der Streckenzeitkonstanten T_g/T_u groß ist und geringe bleibende Regeldifferenzen Δx hinnehmbar sind.

69.2 I-Regler, I-Regelfunktion

Der I-Regler ist ein integral wirkender Regler, der die Reglerausgangsgröße y_R in vielen kleinen Stellschritten solange ändert, bis eine aufgetretene Regeldifferenz $e = w - x$ vollständig beseitigt ist. Dabei ist die Änderungsgeschwindigkeit der Stellgröße direkt proportional zur noch vorhandenen Regeldifferenz e und einem Übertragungsbeiwert K_{IR}:

$$\frac{dy_R}{dt} = K_{\mathrm{IR}} \cdot e$$

In der Arbeitsweise gleicht der I-Regler einem Boten, dessen Laufgeschwindigkeit von der Dringlichkeit des Auftrags abhängig ist, wenn man die Regeldifferenz e als Dringlichkeitsmaß deutet. Solange noch eine Regeldifferenz besteht, ist der I-Regler damit beschäftigt, sie zu beseitigen.

Summiert man die unendlich vielen kleinen Stellschritte zur erzeugten Reglerausgangsgröße y_R, so erhält man die I-Regelfunktion:

$$y_R = K_{\mathrm{IR}} \cdot \int e \cdot dt$$

mit K_{IR} = Integrierbeiwert (Übertragungsbeiwert).

In der Regelfunktion ist K_{IR} als Integrierbeiwert nicht ganz leicht verständlich zu deuten. K_{IR} beinhaltet eine Zeit T_I, die man als Integrierzeit bezeichnet, und die angibt, wie lange es dauert, bis der I-Regler den vollen Stellbereich Y_h durchlaufen hat, wenn als Eingangsgröße die Regelgröße x um den vollen Bereich X_h geändert wird:

$$K_{\mathrm{IR}} = \frac{Y_h}{T_I \cdot X_h}$$

Die beiden Bereiche Y_h und X_h sind konstruktiv festliegende Werte, sodass als Einstellparameter des I-Reglers die Integrierzeit T_I in Frage kommt. Je kleiner die Integrierzeit T_I gewählt wird, umso schneller ändert sich die Reglerausgangsgröße y_R in Abhängigkeit von der Regeldifferenz e.

I-Regler arbeiten langsamer als P-Regler, lassen dafür aber keine bleibende Regeldifferenz bestehen. I-Regler sind ungeeignet für Regelstrecken mit großen Zeitkonstanten oder bei Strecken ohne Ausgleich. Für Regelstrecken, die nur durch eine Totzeit charakterisiert sind, ist der I-Regler von allen Reglertypen der am besten geeignete.

69.3 PI-Regler, PI-Regelfunktion

Die Eigenschaften von P-Reglern und I-Reglern ergänzen sich gut, weil P-Regler dynamisch besser sind als I-Regler und I-Regler statisch besser sind als P-Regler. Also ist es naheliegend, die beiden Regelverhalten zum PI-Regler zu kombinieren.

Die Regelfunktion des PI-Reglers ergibt sich aus der Addition der P-Regelfunktion und I-Regelfunktion:

$$y_R = K_{\mathrm{PR}} \cdot e + K_{\mathrm{IR}} \cdot \int e \cdot dt$$

Die Regelfunktion lässt erkennen, dass die Änderung der Reglerausgangsgröße y_R zum einen proportional zur Regeldifferenz e und zum anderen proportional zum Produkt aus Regeldifferenz und Zeit ist. Der erste Anteil sorgt für den schnellen und stabilen Reglereingriff und der zweite Anteil für die Beseitigung der bleibenden Regeldifferenz.

Eine bekanntere Darstellungsform der PI-Regelfunktion ersetzt den Integrierbeiwert K_{IR} durch einen Quotienten

$$K_{\mathrm{IR}} = \frac{K_{\mathrm{PR}}}{T_n}$$

mit dem Vorteil, den Proportionalbeiwert K_{PR} ausklammern zu können, sodass die PI-Regelfunktion nun in der Form

$$y_R = K_{\mathrm{PR}} \cdot \left(e + \frac{1}{T_n} \cdot \int e \cdot dt \right)$$

entsteht. Die darin neu erscheinende Größe T_n wird *Nachstellzeit* genannt. Ausdrücken möchte man, dass der PI-Regler im Prinzip ein I-Regler ist, dessen Wirkungsbeginn aber um die Nachstellzeit T_n vorverlegt erscheint, also schneller eingreift. Tatsächlich beruht der schnellere Eingriff aber auf dem P-Anteil.

K_{PR} und T_n sind die Regelparameter (Einstellgrößen) mit denen der PI-Regler an die Regelstrecke optimal angepasst werden kann. Beide Parameter können einzeln oder gleichzeitig verstellt werden:

Lässt man zunächst K_{PR} konstant und verändert nur T_n, bleibt der P-Anteil in der Reglerausgangsgröße gleich hoch, es lässt sich aber die Stellgeschwindigkeit beeinflussen. Vergrößert man T_n, verringert sich die Stellgeschwindigkeit bei der Reglerausgangsgröße y_R. Wird die Nachstellzeit T_n soweit vergrößert, dass sie gegen unendlich geht, erhält man das Verhalten eines P-Reglers. Umgekehrt erhöht sich die Stellgeschwindigkeit, wenn man die Nachstellzeit verringert. Dieses darf nicht zu weit getrieben werden, weil es die Schwingneigung und Instabilität des PI-Reglers erhöht.

Was geschieht, wenn die Nachstellzeit T_n fest eingestellt bleibt und die Proportionalverstärkung K_{PR} verändert wird? Durch die Änderung von K_{PR} ändern sich beide Anteile der Reglerausgangsgröße im gleichen Verhältnis. Bei einer Vergrößerung von K_{PR} erhöht sich der P-Anteil an der Gesamtverstellung und die Stellgeschwindigkeit.

Der PI-Regler ist für den Einsatz in allen Streckentypen gut geeignet.

69.4 PID-Regler, PID-Regelfunktion

Der PID-Regler ist ein PI-Regler mit einem zusätzlichen D-Anteil (D = Differenzial). Die Vorstellung von der Wirkung eines D-Anteils in einem Regler entspricht einem kurz geschlagenen Konterschlag gegen eine sehr starke Änderung der Regelgröße x, die sich infolge einer Störgröße z einstellen

würde. Der D-Anteil bewirkt eine sehr kräftige aber nur kurz wirkende Veränderung der Reglerausgangsgröße, die sich aus

$$\Delta y_R = K_{\mathrm{DR}} \cdot \frac{de}{dt} \quad \text{mit } K_{\mathrm{DR}} = \text{Differenzierbeiwert}$$

berechnet und dabei der zeitlichen Änderung der Regeldifferenz e proportional ist. In der Regelungstechnik wird die Eigenschaft eines Reglers, auf eine sich abzeichnende Entwicklung bereits im Voraus zu reagieren, als *Vorhalt* bezeichnet. Im Voraus bedeutet, der Regler beobachtet die Veränderung der Regeldifferenz bereits im aller ersten Ansatz bei ihrer Entstehung.

Die PID-Regelfunktion bildet sich aus der PI-Regelfunktion erweitert mit dem Vorhaltglied:

$$y_R = K_{\mathrm{PR}} \cdot \left(e + \frac{1}{T_n} \cdot \int e \cdot dt + T_v \cdot \frac{de}{dt} \right)$$

Die Vorhaltzeit T_v ist vergleichbar der Nachstellzeit T_n eine eingesparte Zeit. Damit ist der PID-Regler im Prinzip ein PI-Regler, dessen Wirkungsbeginn um die Vorhaltzeit T_v vorverlegt erscheint, allerdings kann dies nicht in der Sprungantwort dargestellt werden. Der Vorteil des PID-Reglers gegenüber dem PI-Regler ist die etwas schnellere Beseitigung einer aufgetretenen Regeldifferenz.

Weil der PID-Regler alle bisher besprochenen Regelparameter, nämlich die Proportionalverstärkung K_{PR} für den P-Anteil, die Integrierzeit T_I bzw. die Nachstellzeit T_n für den I-Anteil und die Vorhaltzeit T_v für den D-Anteil enthält, wird er auch als Universalregler bezeichnet. Der PID-Regler ist wegen der Vielzahl der möglichen Einstellkombinationen zur Anpassung an alle Regelstreckentypen geeignet, aber etwas schwierig in der Handhabung.

69.5 Vergleich der verschiedenen Reglertypen

Als Abschluss der Betrachtung verschiedener Reglertypen soll in Abb. 69.3 vergleichend gezeigt werden, wie sich die Regler bei optimaler Einstellung an einer Regelstrecke mit Ausgleich

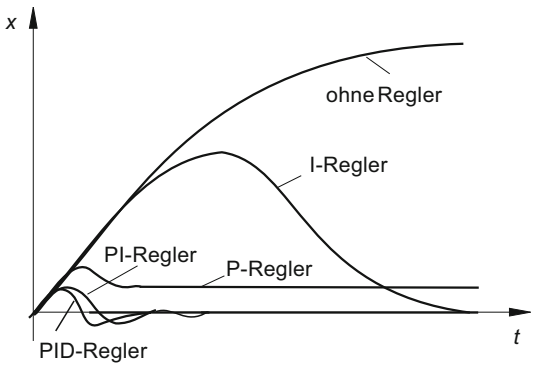


Abb. 69.3 Unterschiedliches Ausregeln einer Störung bei verschiedenen Reglertypen an einer Regelstrecke mit Ausgleich

und den Kenngrößen K_{PS}, T_u und T_g verhalten, wenn die Auswirkung einer aufgetretenen Störgröße $z$auf die Regelgröße x ausgeregelt werden soll.

69.6 PID-Reglerbaustein für digitale Abtastregelung

Eine Faustformel besagt, dass digitale Abtastregler gute Regelergebnisse liefern, wenn die größte Strecken-Zeitkonstante T_S etwa 10mal größer als die Abtastzeit T_A sein kann. Der Abtastzeitwert bestimmt sich aus der Programmbearbeitungszeit des Rechners einschließlich der Wandlungszeit des A/D-Umsetzers. Ein gut zu erreichender Abtastzeitwert bei einer SPS liegt bei 100 ms. Systembedingt kann sich bei Abtastreglern das Reglerausgangssignal y_R nur wertdiskret, also bestenfalls feinstufig, und zeitdiskret, nämlich im Takt der Abtastzeit T_A, ändern, wie Abb. 69.4 zeigt.

Zusätzlich muss die PID-Regelfunktion an die arithmetischen Möglichkeiten eines digitalen Rechners angepasst werden. Die originale PID-Regelfunktion muss ersetzt werden durch einen wirkungsähnlichen PID-Regelalgorithmus, bei dem das Integral des I-Anteils durch eine Summe und das Differenzial des D-Anteils durch einen Differenzenquotienten ersetzt wird. Die differenziell kleine Zeitspanne dt in der Regelfunktion wird durch die Abtastzeit T_A ersetzt. Eine SPS als Softwareregler verfügt dann mit ihren Additions-, Subtraktions-, Multiplikations- und Divisions-Befehlen über die notwendigen Rechenoperationen zur Durchführung des PID-Regelprogramms, der die Stellgröße y_R für jeden Abtastzeitpunkt k neu berechnet:

$$y_R(k) = K_{\mathrm{PR}} \cdot \left(e(k) + \frac{1}{T_n} \cdot \sum_{i=1}^{k} e(i) \cdot T_A + T_v \cdot \frac{e(k) - e(k-1)}{T_A} \right)$$

Die in der Formel noch vorhandene fortlaufende Summenbildung kann mit einem rekursiven Algorithmus ausgeführt

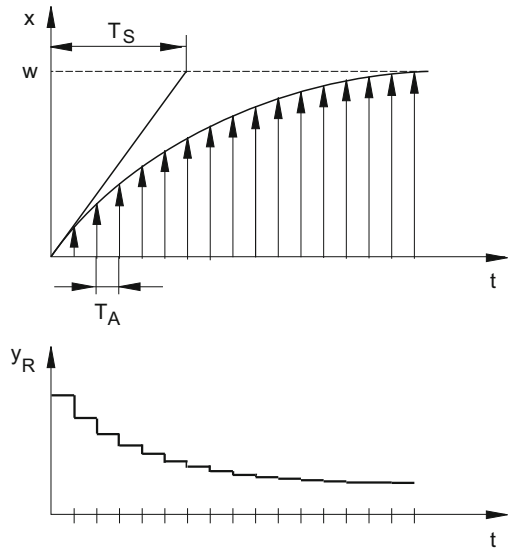


Abb. 69.4 Istwertabtastung und Regelspannungsausgabe

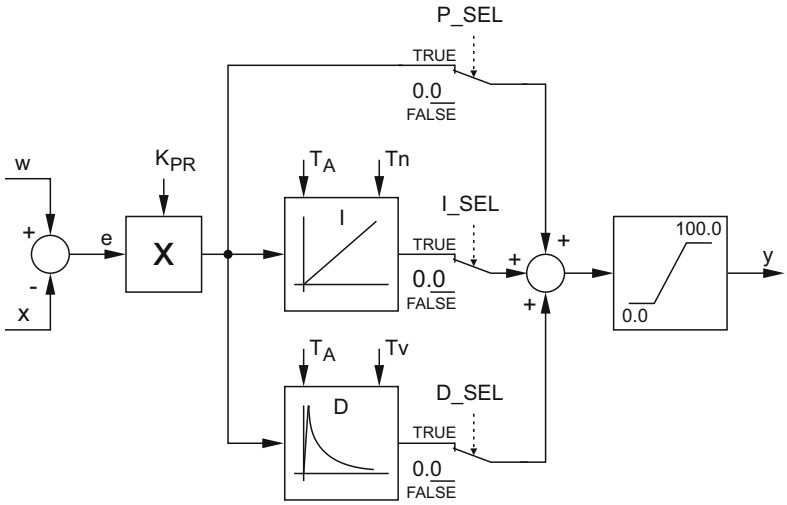


Abb. 69.5 Funktionsschema eines PID-Reglerbausteins für Stellungsalgorithmus (w = Sollwert, x = Istwert, e = Regeldifferenz, KPR = Proportionalverstärkung, T_A = Abtastzeit, Tn = Nachstellzeit, Tv = Vorhaltzeit, y = Stellgröße, P_SEL = P-Anteil, I_SEL = I-Anteil, D_SEL = D-Anteil)

werden, sodass eine Speicherung aller Regeldifferenzen e nicht erforderlich ist. Es ergibt sich eine Formel zur Berechnung der Reglerausgangsgröße $y_R(k)$ zu den bestimmten Zeitpunkten k und bezeichnet sie als *Stellungsalgorithmus* (Abb. 69.6). Die analoge Wertausgabe dieser Reglerausgangsgröße am Abtastreglerausgang der SPS, also nach der Digital-Analog-Umsetzung, muss auf ein P-Stellglied geführt werden, das die richtige Stellgröße y an die Regelstrecke abgibt:

$$y_R(k) = K_{\mathrm{PR}} \cdot \left(e(k) + \frac{T_A}{T_n} \cdot e_{\mathrm{SUM}(k)} + T_v \frac{e(k) - e(k-1)}{T_A} \right)$$

PID-Stellungsalgorithmus

Digitale Abtastregler steuern häufig auch sog. I-Stellglieder an. Das sind motorische Stellantriebe, die erst durch ihren Vorschub die Stellgröße y bilden. Da solchen I-Stellgliedern nicht der Stellgrößenwert, sondern nur die Stellengrößenänderung zu geführt werden darf, gibt es auch noch einen zweiten PID-Algorithmus, der nur die Änderung der Reglerausgangsgröße Δy_R pro Abtastzeit T_A berechnet und als *Geschwindigkeitsalgorithmus* bezeichnet wird. Dieses Signal kann zur Bildung von entsprechend langen oder kurzen Schritt-Stellsignalen für die sog. Schrittregler verwendet werden.

$$\Delta y = K_{\mathrm{PR}} \cdot \left(e(k) - e(k-1) + \frac{T_A}{T_n} \cdot e(k) + \frac{T_v}{T_A} (e(k) - 2e(k-1) + e(k-2)) \right)$$

PID-Geschwindigkeitsalgorithmus

Nachfolgend wird die Umsetzung des beschriebenen Stellungsalgorithmus in ein universelles PID-Reglerprogramm gezeigt, dessen P-Anteil, I-Anteil bzw. D-Anteil einzeln zu und abschaltbar sind, um quasikontinuierliche P-, PI- oder PID-Regelfunktionen ausführen zu lassen. Abb. 69.5 zeigt zunächst das Funktionsschema des zu realisierenden PID-Reglerbausteins.

Da die PID-Reglerfunktion über Speichereigenschaften verfügt, wird das Programm mit einem Funktionsbaustein realisiert. Über den Eingangsparameter EIN kann die Regelfunktion ein- oder ausgeschaltet werden. Im ausgeschalteten Zustand werden alle gespeicherten Werte zurückgesetzt. Weitere Eingangsparameter des Funktionsbausteins sind die Führungsgröße w (SW), die Regelgröße x (IW) sowie die Einstellwerte Proportionalbeiwert K_{PR}(KP), Nachstellzeit T_n (TN), Vorhaltzeit T_v (TV) und die Abtastzeit T_A (TA).

Es wird angenommen, dass die Führungsgröße w und die Regelgröße x als Gleitpunktzahlen im Bereich von 0,0 bis 100,0 vorliegen. Die Proportionalverstärkung KP und die Zeitwerte TN, TV und TA sind ebenfalls als Gleitpunktzahlen anzugeben. Bei den Zeitwerten TN, TV und TA entspricht dabei der Zahlenwert einem Zeitwert in Sekunden.

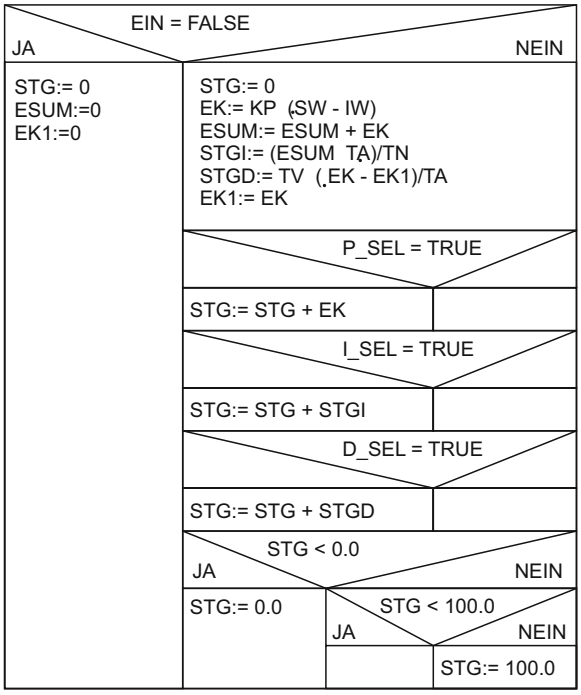
Mit den binären Eingangsparametern P_SEL, I_SEL und D_SEL können die einzelnen Anteile mit einem „1“-Signal zugeschaltet und mit einem „0“-Signal abgeschaltet werden.

Die Ausgangsgröße STG des PID-Reglerbausteins ist die kontinuierliche Stellgröße y_R in einem Gleitpunktzahlenbereich von 0,0 ... 100,0.

Der PID-Reglerbaustein muss entsprechend der gewählten Abtastzeit T_A in gleichen zeitlichen Abständen zur Bearbeitung aufgerufen werden.

69.7 SPS als kontinuierlicher PID-Abtastregler

Der in Abschn. 69.6 beschriebene PID-Reglerbaustein für den Stellungsalgorithmus ist noch kein gebrauchsfertiger digitaler SPS-Regler. Abb. 69.7 zeigt in vereinfachter Form, was noch



Lokale stationäre Variablen des Funktionsbausteins:

EK1 = Regeldifferenz des vorhergehenden Abtastzeitpunktes mit dem Proportionalbeiwert KP multipliziert.

ESUM = Summe aller Regeldifferenzen, die mit dem Proportionalbeiwert KP multipliziert sind.

Lokale temporäre Variablen des Funktionsbausteins:

EK = Regeldifferenz mit dem Proportionalbeiwert KP multipliziert. Der Wert entspricht dem P-Anteil.

STGI = Berechneter I-Anteil der Stellgröße STG.

STGD = Berechneter D-Anteil der Stellgröße STG.

Abb. 69.6 Struktogramm des PID-Reglerbausteins mit Stellungsalgorithmus

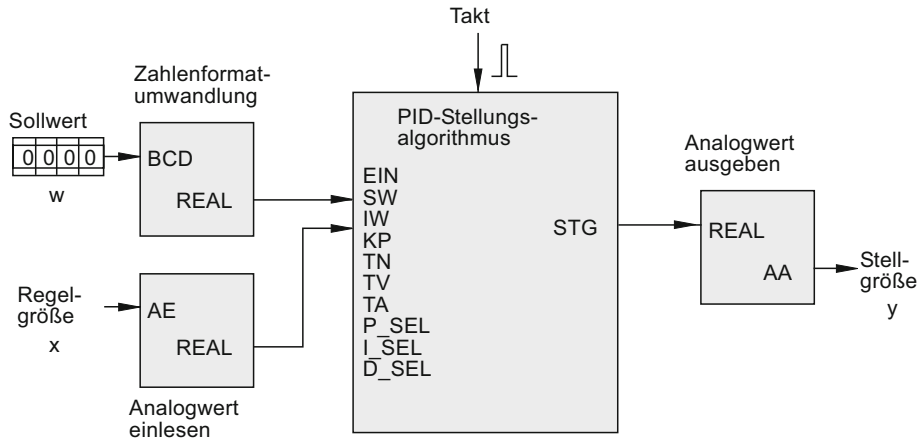


Abb. 69.7 SPS-Regelstruktur (vereinfacht)

alles benötigt wird. Die SPS muss neben der CPU-Baugruppe selbstverständlich auch über eine Analogeingabebaugruppe und eine Analogausgabebaugruppe verfügen. Programmmäßig ist nicht nur der Regelalgorithmus, sondern auch das Einlesen der analogen Regelgröße *x* und die Ausgabe der analogen Stellgröße *y* auszuführen, d. h. die Anbindung der SPS an den Messaufnehmer und das Stellglied der realen oder simulierten Regelstrecke zu realisieren. Praktisch umsetzen lässt sich eine solche SPS-Reglerlösung nur mit fertigen Programmierbausteinen aus der SPS-Bibliothek.

69.8 Digitaler Schrittregler mit PI-Verhalten

In der klassischen Analogregeltechnik wurde dieser Reglertyp mit einem Dreipunktregler und verzögerter Rückführung über dem Schaltverstärker realisiert, weshalb er dort auch als Dreipunkt-Schrittregler bezeichnet wurde. Die digitale Nachbildung dieses Reglertyps beruht nicht mehr auf dem Dreipunktregler, sondern auf dem PID-Geschwindigkeitsalgorithmus mit

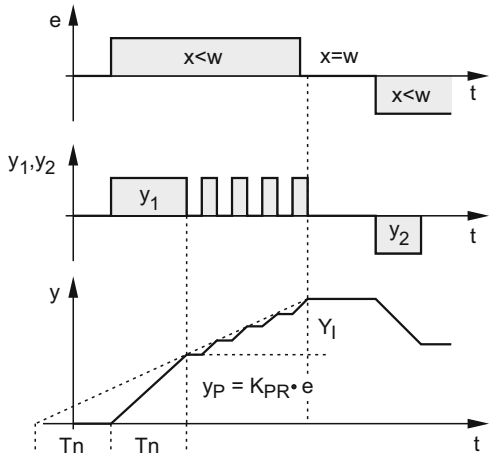


Abb. 69.8 Impulsverhalten eines PI-Schrittreglers

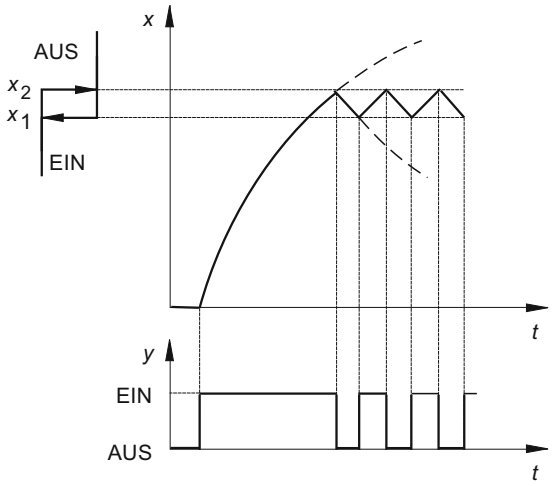


Abb. 69.9 Zweipunkt-Regelfunktion mit Verlauf der Regelgröße

einer nachgesetzten Impulsausgabe zur Ansteuerung eines motorischen Stellantriebs, der einen Schieber oder ein Ventil mit einer Folge von Stellschritten öffnet oder schließt. Die so gebildete Stellgröße ist quasi-kontinuierlich, da jede Stellgliedposition innerhalb des Stellbereichs erreichbar ist, wenn auch in kleinen Schritten. Werden keine Impulse ausgegeben, bleibt die Stellgröße unverändert. Mit den beiden binär wirkenden Impulsausgängen können die erforderlichen drei Schaltzustände „Rechtslauf“, „Linkslauf“ und „Aus“ für den Stellantrieb realisiert werden.

Auf eine sprungförmige Änderung der Regeldifferenz e reagiert der PI-Schrittregler sofort mit einem „langen“ Schritt. Das Stellsignal y_1 gibt dazu einen langen Impuls aus, der durch den P-Anteil des Regelalgorithmus verursacht wird. Die darauffolgenden kürzeren Impulse werden durch den I-Anteil des Regelalgorithmus gebildet, siehe Abb. 69.8. Der PI-Schrittregler wirkt solange, bis keine Regeldifferenz mehr besteht.

69.9 Zweipunktregler, Zweipunkt-Regelfunktion

Bei der Zweipunkt-Regelfunktion wird ein binäres Stellgrößensignal mit den Zuständen „EIN“ und „AUS“ erzeugt. Bei „EIN“ erfolgt 100 % Leistungszufuhr für die Regelstrecke und bei „AUS“ dagegen 0 %.

Die Kennlinie des Zweipunktreglers zeigt die Position der beiden Schaltpunkte auf der Regelgrößenachse x, deren Mittenlage durch die eingestellte Führungsgröße w festgelegt ist. Die Zuschaltung der Leistungszufuhr erfolgt, wenn die Regelgröße x den unteren Grenzwert unterschreitet. Die Leistungszufuhr wird abgeschaltet, sobald die Regelgröße x den oberen Grenzwert überschreitet.

Es ist charakteristisch für Zweipunktregler, dass es für die Regelgröße x keinen Beharrungszustand gibt, dem sie zustreben kann. Die Stellgröße schaltet ständig um und die Regelgröße verändert sich fortwährend im Rahmen einer noch zulässigen Schwankungsbreite, die durch die Differenz der beiden Schaltpunkte x_2 und x_1 festgelegt ist. Diese sog. *Schalthysterese* trägt zur Beruhigung des Regelvorgangs bei. Den Einfluss einer Störgröße z korrigiert der Zweipunktregler durch Änderung seines EIN-AUS-Schaltverhältnisses.

Grafisch lässt sich der Zweipunkt-Regelungsvorgang anschaulich darstellen, wie Abb. 69.9 zeigt.

69.10 Regelgüte

Ein Regler hat die Aufgabe, die Regelgröße x möglichst genau auf den durch die Führungsgröße w vorgegebenen Sollwert zu bringen und auch dort gegen den Einfluss einer Störgröße z zu halten. Außerdem wird gefordert, dass der Regelungsvorgang in möglichst kurzer Zeit und nur unter vorübergehenden noch zulässigem Überschwingen abgeschlossen wird. Diese Forderungen lassen sich nicht alle erfüllen, da einzelne Bedingungen sich auch widersprechen. Dämpft man beispielsweise die Überschwingweite, vergrößert sich die Ausregelzeit. Es ist daher nur möglich für eine vorliegende Regelungsaufgabe die günstigsten Reglereinstellungen im Sinne eines Kompromisses zu finden.

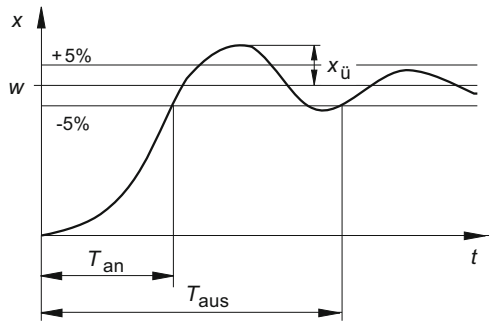


Abb. 69.10 Kriterien zur Regelgüte

Beurteilungsmaßstäbe für die Regelgüte sind die An- und Ausregelzeit sowie die Überschwingweite, wie in Abb. 69.10 dargestellt.

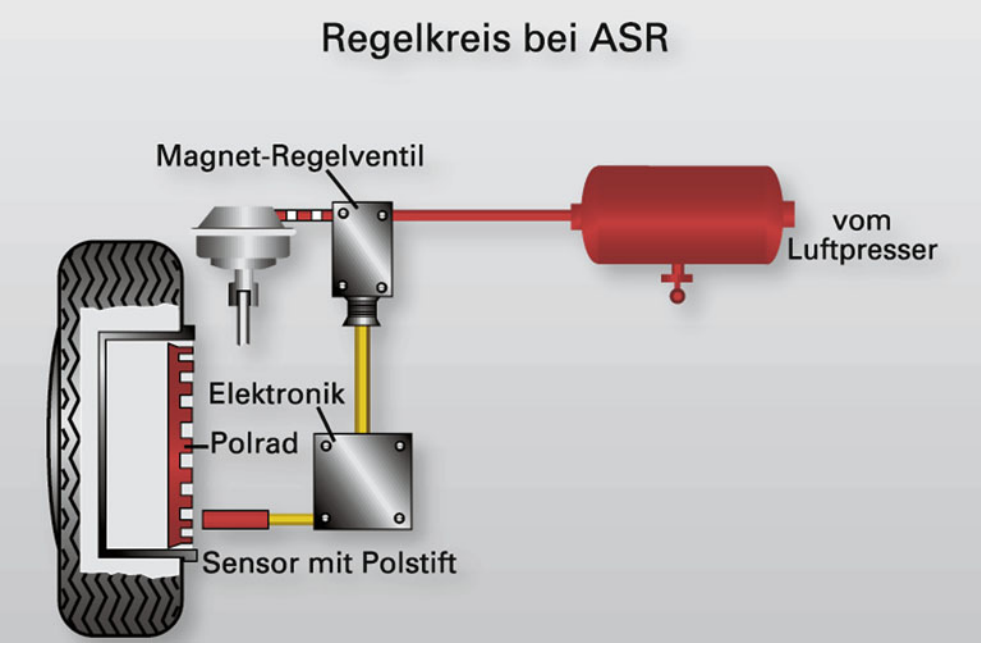
1. Die Anregelzeit T_{an} ist die Zeit, die bis zum erstmaligen Erreichen des Sollwertes der Regelgröße vergeht.
2. Die Ausregelzeit T_{aus} ist die Zeit, die vergeht bis die Regelgröße x endgültig in den Toleranzbereich einmündet und darin verbleibt.
3. Die Überschwingweite $x_{ü}$ wird in Prozent des Sollwertes der Regelgröße x festgelegt.

Literatur

1. DIN EN 61131-3, Speicherprogrammierbare Steuerungen, Teil 3: Programmiersprachen, Dezember 2003, Beuth Verlag Berlin
2. DIN 19226 Teil 4, Leittechnik; Regelungstechnik und Steuerungstechnik; Begriffe für Regelungs- und Steuerungssysteme, Februar 1994, Beuth Verlag Berlin
3. *Gehlen, P.:* Funktionale Sicherheit in Maschinen und Anlagen, Siemens AG (Hrsg.), Verlag: Publicis Corporate Publishing, Erlangen 2007
4. *Lepers, H.:* SPS-Programmierung nach IEC 61131-3, Franzis Verlag Poing, 2005
5. *Mann, H., Schiffelgen, H., Froriep, R.:* Einführung in die Regelungstechnik, Hanser Verlag München, 8. Auflage 1997
6. *Reuter, M., Zacher, S.:* Regelungstechnik für Ingenieure, Vieweg Verlag, 11. Auflage 2004
7. *Seitz, M.:* Speicherprogrammierbare Steuerungen, Fachbuchverlag Leipzig, 2003
8. *Wellenreuther, G., Zastrow, D.:* Automatisieren mit SPS, Theorie und Praxis, Verlag Vieweg Wiesbaden, 4. Auflage 2008
9. *Wellenreuther, G., Zastrow, D.:* Automatisieren mit SPS, Übersichten und Übungsaufgaben, Verlag Vieweg Wiesbaden, 4. Auflage 2009

Regelungstechnik

IX



70 Einführung 843

71 Strukturbild und Übertragungsglieder 847

72 Übertragungsverhalten linearer, zeitinvarianter Übertragungsglieder 853

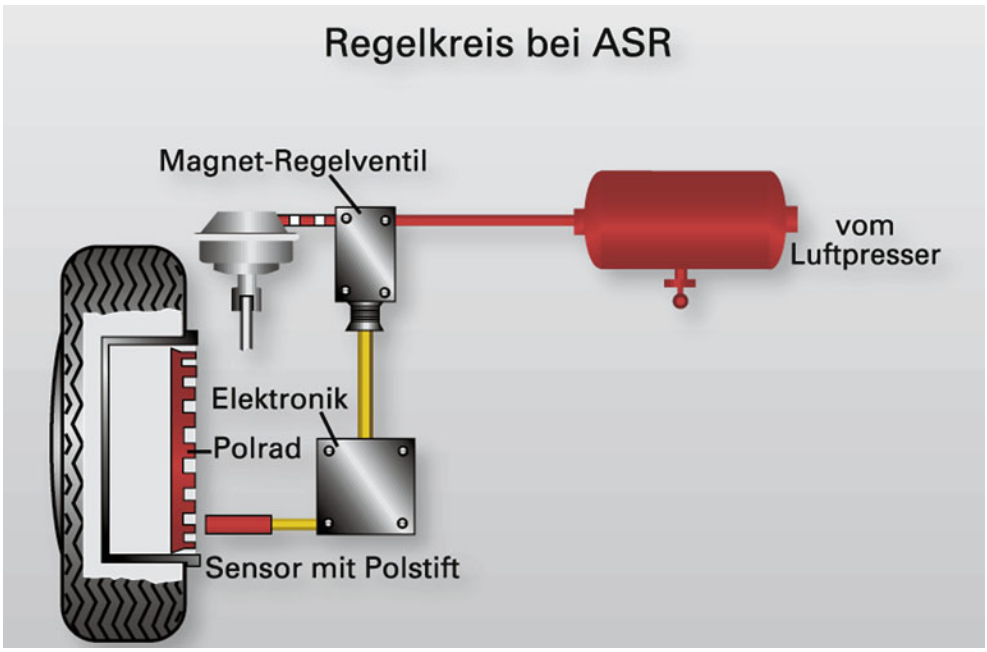
73 Stabilität 859

74 Entwurf von Regelkreisen 865

Einführung

70

Joachim Horn



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_70

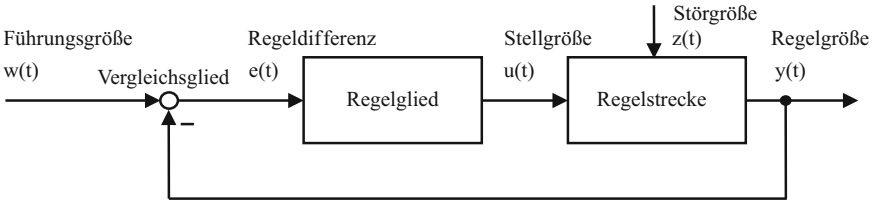


Abb. 70.1 Klassischer Regelkreis

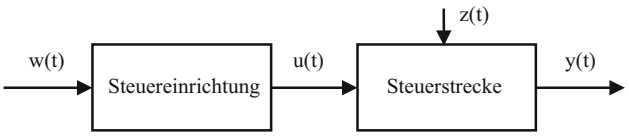


Abb. 70.2 Steuerkette

Regelungstechnik ist die Wissenschaft von der automatischen gezielten Beeinflussung dynamischer Prozesse [3]. *Dynamisch* bedeutet, dass die Prozessgrößen sich zeitlich ändern.

Prozesse laufen in *Systemen* ab, also Einheiten, die mit der Umwelt durch Ein- und Ausgangsgrößen verbunden sind. Der Zusammenhang von Ein- und Ausgangsgrößen wird durch mathematische Beziehungen, beispielsweise Differenzialgleichungen, beschrieben. Das zu beeinflussende System wird auch als *Regelstrecke* oder kurz *Strecke* bezeichnet, die Begriffe Prozess, System und Strecke werden im Folgenden synonym verwendet. Je nach Anwendungsfall kann es sich um ein technisches oder auch um ein nichttechnisches System, beispielsweise ein biologisches oder ökonomisches, handeln. Die im Folgenden behandelten Methoden der Regelungstechnik verwenden stets ein mathematisches Modell, also eine abstrahierte Beschreibung des Prozesses.

Abb. 70.1 zeigt die klassische Struktur zur gezielten Beeinflussung dynamischer Prozesse, den *Regelkreis*. Die zeitveränderliche Größe $y(t)$, die im Idealfall exakt einem vorgegebenen Verlauf folgen soll, wird gemessen und mit der Sollgröße $w(t)$ verglichen. $y(t)$ wird als *Regelgröße*, $w(t)$ als *Führungsgröße* bezeichnet. Die Differenz $e(t) = w(t) - y(t)$ heißt *Regeldifferenz*, *Regelabweichung* oder *Regelfehler*. Die Differenzbildung von Soll- und Istwert der Regelgröße wird als *Soll-Istwert-Vergleich* bezeichnet. Das *Regelglied* berechnet aus der Regeldifferenz $e(t)$ die *Stellgröße* $u(t)$, und zwar so, dass die Regelgröße der Führungsgröße möglichst gut folgt. Neben der von außen vorgebbaren Stellgröße $u(t)$ besitzt die Strecke noch eine weitere Eingangsgröße, die *Störgröße* $z(t)$. Im Gegensatz zur Führungsgröße kann die Störgröße nicht von außen vorgegeben und häufig auch nicht gemessen werden, ihr zeitlicher Verlauf ist also unbekannt.

Beim Regelungsentwurf, auch als Regelungssynthese bezeichnet, geht es nun darum, ein geeignetes Regelungsglied zu bestimmen, so dass die Regelgröße hinreichend genau der Führungsgröße folgt; die Wirkung der Störgröße muss hierzu hinreichend unterdrückt werden.

In der Norm DIN IEC 60050-351 [1] werden die Bestandteile des Regelkreises detaillierter gegliedert. Die Messung der Regelgröße erfolgt durch die *Messeinrichtung*, in der Literatur auch als *Messglied* bezeichnet. Das *Stellglied* als Teil der Strecke wandelt die Stellgröße vom niedrigen Leistungspegel des Regelglieds zu einem höheren Leistungspegel. Der *Regler* umfasst laut Norm das Vergleichsglied und das Regelglied; im üblichen Sprachgebrauch werden Regelglied und Regler jedoch synonym verwendet. Ist die Führungsgröße konstant, so spricht man von einer *Festwertregelung*, bei einer zeitabhängigen Führungsgröße von einer *Folgeregelung*.

Charakteristisch für eine Regelung ist die Rückführung oder *Rückkopplung* von Prozessgrößen, im klassischen Regelkreis gemäß Abb. 70.1 der Regelgröße $y(t)$. Wird durch die Störgröße eine Abweichung der Regelgröße von der Führungsgröße, also eine Regeldifferenz ungleich Null, verursacht, so kann diese, ein geeignetes Regelglied vorausgesetzt, im geschlossenen Kreis wieder ausgeregelt werden. Auch wenn die dynamischen Eigenschaften der Strecke nur näherungsweise bekannt sind, wird bei Verwendung eines geeigneten Regelglieds der Betrag der Regeldifferenz hinreichend klein sein. Diese Eigenschaft wird als *Robustheit* bezeichnet.

Im Gegensatz zu einem geschlossenen Regelkreis findet bei einer offenen Steuerkette oder kurz *Steuerung* gemäß Abb. 70.2 keine Rückführung von Prozessgrößen statt. Die Strecke wird

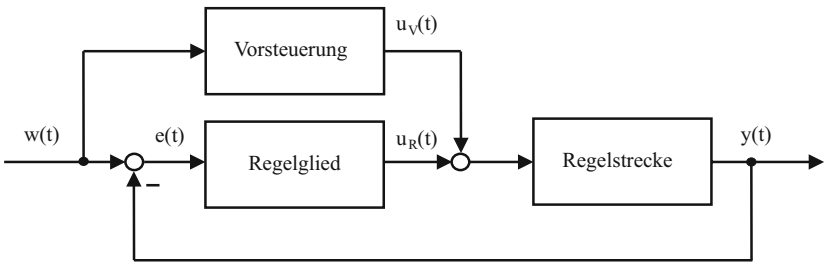


Abb. 70.3 Regelkreis mit Vorsteuerung

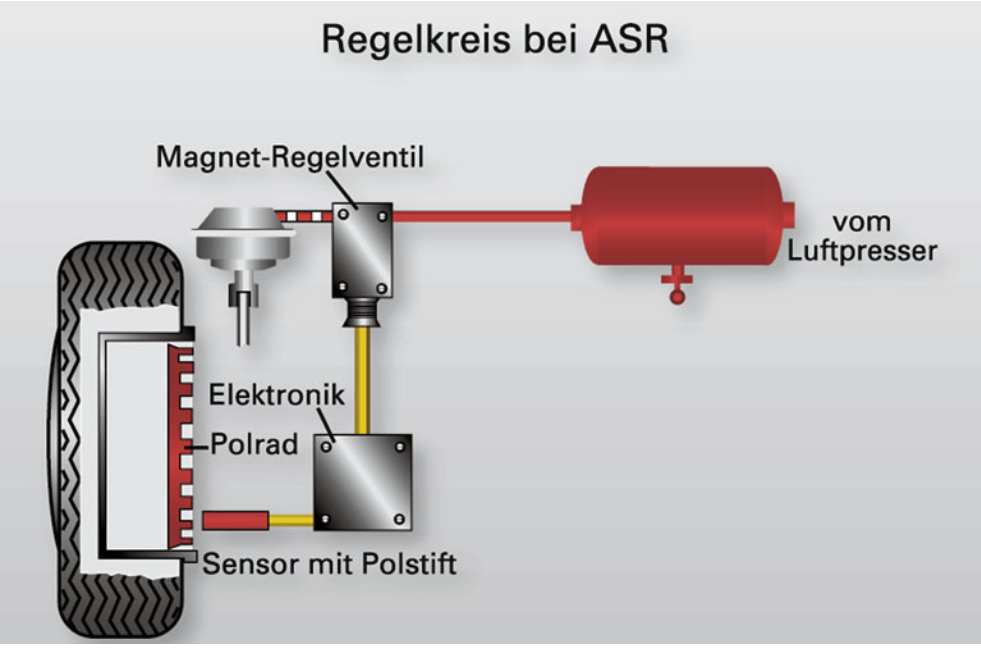
hier als Steuerstrecke bezeichnet, die steuernde Einheit als Steuereinrichtung oder kurz ebenfalls als Steuerung. (Hinweis: Der Begriff „Steuerung“ wird in der Automatisierungstechnik noch mit einer weiteren Bedeutung verwendet, nämlich als Kurzform von „ereignisdiskrete Steuerung“, die eine Rückführung wertediskreter Prozessgrößen (z. B. „Ventil geöffnet“ oder „Ventil geschlossen“) auf wertediskrete Stellgrößen bezeichnet.)

Eine offene Steuerkette kann dann angewendet werden, wenn das dynamische Verhalten und der Anfangszustand der Steuerstrecke genau bekannt sind sowie die Störgröße vernachlässigbar ist. Sind diese Voraussetzungen erfüllt, kann mit einer geeignet entworfenen Steuerung die Ausgangsgröße der Strecke schneller auf einen neuen Sollwert gebracht werden als mit einer Regelung. Für die Aufgabenstellung der Folgeregelung wird daher häufig die Regelung mit einer Steuerung, als *Vorsteuerung* bezeichnet, kombiniert. Wie in Abb. 70.3 dargestellt, setzt sich hierbei die Stellgröße $u(t)$ aus einem Vorsteuerungsanteil $u_V(t)$ und einem Regelungsanteil $u_R(t)$ zusammen. Der Regelungsanteil der Stellgröße kompensiert die Wirkung von Modellungenauigkeiten und Störungen, der Vorsteuerungsanteil gewährleistet, dass die Regelgröße schnell der Führungsgröße und damit der Solltrajektorie folgt.

Strukturbild und Übertragungsglieder

71

Joachim Horn

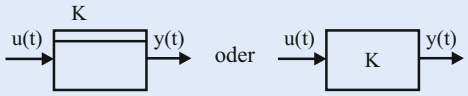
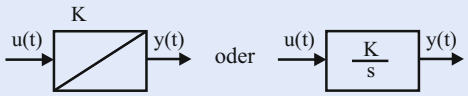
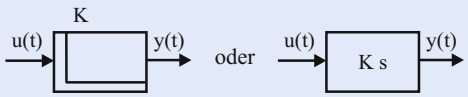
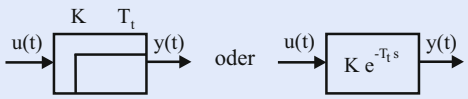
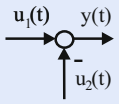
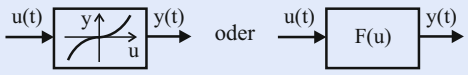
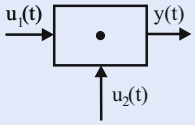
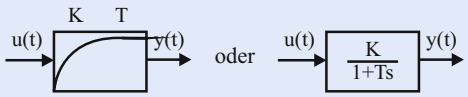
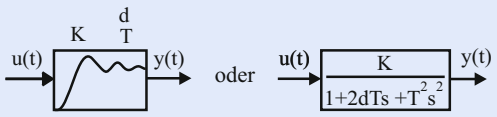


71.1 Strukturbild 848
71.2 Übertragungsglieder 848
71.3 Klassifikation von Übertragungsgliedern 849
71.4 Beschreibung von Übertragungsgliedern um einen Arbeitspunkt . 849
71.5 Impulsantwort und Sprungantwort einfacher Übertragungsglieder 850

Teil IX

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_71

Tab. 71.1 Funktionalbeziehung und Blocksymbol einfacher Übertragungsglieder

Übertragungsglied	Funktionalbeziehung	Blocksymbol
P-Glied (Proportionalglied)	$y(t) = K \cdot u(t)$	
I-Glied (Integrierglied)	$y(t) = K \int_0^t u(\tau) d\tau$	
D-Glied (Differenzierglied)	$y(t) = K \cdot \dot{u}(t)$	
TZ-Glied, T_t-Glied (Totzeitglied)	$y(t) = K \cdot u(t - T_t)$ mit $T_t > 0$	
S-Glied (Summierglied)	$y(t) = \pm u_1(t) \pm u_2(t)$	
KL-Glied (Kennlinienglied)	$y(t) = F(u(t))$	
M-Glied (Multiplizierglied)	$y(t) = K \cdot u_1(t) \cdot u_2(t)$	
PT_1-Glied, VZ_1-Glied (Verzögerungsglied erster Ordnung)	$T \cdot \dot{y}(t) + y(t) = K \cdot u(t)$	
PT_2-Glied, VZ_2-Glied (Verzögerungsglied zweiter Ordnung)	$T^2 \cdot \ddot{y}(t) + 2 \cdot d \cdot T \cdot \dot{y}(t) + y(t) = K \cdot u(t)$	

71.1 Strukturbild

Das *Strukturbild* eines dynamischen Systems, auch *Signalflussplan* genannt, ist eine grafische Darstellung der Funktionalbeziehungen der zeitveränderlichen Größen. Die zeitveränderlichen Größen werden dabei als Pfeile dargestellt. Blöcke charakterisieren die Funktionalbeziehung zwischen den Eingangsgrößen und den Ausgangsgrößen, beispielsweise durch eine Differenzialgleichung. Blöcke sind rückwirkungsfrei.

Das Strukturbild stellt eine anschauliche Beschreibung des dynamischen Systems dar. Rückkopplungen oder unabhängige Teilsysteme können in dieser grafischen Darstellung viel einfacher erkannt werden als direkt aus den Gleichungen der zeitveränderlichen Größen des Systems.

71.2 Übertragungsglieder

Ein Übertragungsglied wird durch die Funktionalbeziehung zwischen den Eingangsgrößen und den Ausgangsgrößen beschrieben. Übertragungsglieder werden im Strukturbild als Block dargestellt. In Tab. 71.1 sind einfache Übertragungsglieder mit den gebräuchlichen Blocksymbolen aufgelistet. Bei linearen, zeitinvarianten Übertragungsgliedern (siehe Abschn. 71.3.2 und 71.3.3) wird entweder die Sprungantwort $h(t)$, also die Ausgangsgröße $y(t)$ bei einem Einheitssprung $\sigma(t)$ als Eingangsgröße $u(t)$, im Blocksymbol dargestellt oder die in Abschn. 71.3.1 definierte Übertragungsfunktion $G(s)$ angegeben.

71.3 Klassifikation von Übertragungsgliedern

Die Funktionalbeziehung zwischen der Eingangsgröße $u(t)$ und der Ausgangsgröße $y(t)$ eines Übertragungsgliedes sei durch den Operator A gemäß

$$y(t) = A\{u(t)\}$$

beschrieben.

71.3.1 Statische und dynamische Übertragungsglieder

Bei einem *statischen* Übertragungsglied hängt die Ausgangsgröße zum Zeitpunkt t nur von der Eingangsgröße zum selben Zeitpunkt t ab. Das P-Glied mit der Funktionalbeziehung $y(t) = K \cdot u(t)$ und das KL-Glied mit der Funktionalbeziehung $y(t) = F(u(t))$ sind statische Übertragungsglieder.

Bei einem *dynamischen* Übertragungsglied hängt die Ausgangsgröße zum Zeitpunkt t von der Eingangsgröße zu einem anderen Zeitpunkt ab, typischerweise von der Eingangsgröße bis zum Zeitpunkt t. Das I-Glied mit der Funktionalbeziehung $y(t) = K \int_0^t u(\tau)d\tau$ ist ein dynamisches Übertragungsglied.

71.3.2 Lineare und nichtlineare Übertragungsglieder

Ein Übertragungsglied heißt *linear*, wenn es das Überlagerungsprinzip (Superpositionsprinzip),

$$A\{u_1(t) + u_2(t)\} = A\{u_1(t)\} + A\{u_2(t)\}$$

für beliebige $u_1(t)$ und $u_2(t)$, und das Verstärkungsprinzip,

$$A\{c \cdot u(t)\} = c \cdot A\{u(t)\}$$

für beliebige c und $u(t)$, erfüllt [3]. Andernfalls heißt das Übertragungsglied *nichtlinear*.

Mit

$$y_1(t) = A\{u_1(t)\}, \quad y_2(t) = A\{u_2(t)\}$$

und der Eingangsgröße

$$u(t) = c_1 \cdot u_1(t) + c_2 \cdot u_2(t)$$

lautet die Ausgangsgröße eines linearen Systems

$$y(t) = c_1 \cdot y_1(t) + c_2 \cdot y_2(t).$$

Das I-Glied ist ein lineares Übertragungsglied. Das KL-Glied – den Sonderfall des P-Gliedes ausgenommen – ist ein nichtlineares Übertragungsglied.

71.3.3 Zeitinvariante und zeitvariante Übertragungsglieder

Ein Übertragungsglied heißt *zeitinvariant*, wenn es das Verschiebungsprinzip,

$$y(t - t_0) = A\{u(t - t_0)\}$$

für jede Eingangsgröße $u(t)$ und jedes $t_0 > 0$, erfüllt [3]. Andernfalls heißt es *zeitvariant*.

Das Übertragungsverhalten eines zeitinvarianten Übertragungsgliedes ist also unabhängig von der Zeit, das Übertragungsverhalten eines zeitvarianten Übertragungsgliedes ist dagegen abhängig von der Zeit. Alle Übertragungsglieder in Tab. 71.1 sind zeitinvariant. Ein Übertragungsglied, dessen Übertragungsverhalten durch eine Differenzialgleichung mit zeitabhängigen Koeffizienten beschrieben wird, beispielsweise

$$a_0(t) \cdot y(t) + a_1(t) \cdot \dot{y}(t) = b_0(t) \cdot u(t) + b_1(t) \cdot \dot{u}(t),$$

ist dagegen zeitvariant.

Technische Systeme sind, verursacht unter anderem durch Alterung und Verschleiß, stets zeitvariant. Jedoch können sie, wenn man lediglich einen begrenzten Zeitraum betrachtet, in guter Näherung als zeitinvariant beschrieben werden.

71.3.4 Kausale und akausale Übertragungsglieder

Ein Übertragungsglied heißt *kausal*, wenn die Eingangsgröße zu einem beliebigen Zeitpunkt τ die Ausgangsgröße für $t < \tau$ nicht beeinflusst. Andernfalls heißt es *akausal*.

Alle Übertragungsglieder in Tab. 71.1 sind kausal. Ein TZ-Glied mit negativer Totzeit, $T_t < 0$, ist dagegen akausal.

Technische Systeme sind stets kausal. Damit ein Reglergesetz technisch realisiert werden kann, muss das Übertragungsverhalten kausal sein.

71.4 Beschreibung von Übertragungsgliedern um einen Arbeitspunkt

71.4.1 Arbeitspunkt eines Übertragungsgliedes

Im *Arbeitspunkt* eines Übertragungsgliedes, auch als Betriebspunkt, Ruhelage oder stationärer Zustand bezeichnet, sind alle zeitveränderlichen Größen konstant. Die zeitlichen Ableitungen

sind damit gleich Null. Wird das Übertragungsverhalten durch die Differenzialgleichung

$$f(u(t), \dot{u}(t), \ldots, \overset{(m)}{u}(t), y(t), \dot{y}(t), \ldots, \overset{(n)}{y}(t)) = 0 \qquad (71.1)$$

beschrieben, so gilt für die Ruhelage

$$f(u_R, 0, \ldots, 0, y_R, 0, \ldots, 0) = 0.$$

Zu einem vorgegebenen u_R können unendlich viele oder endlich viele Lösungen oder genau eine oder auch keine Lösung y_R existieren.

71.4.2 Linearisierung um den Arbeitspunkt

Betrachtet man das Übertragungsverhalten eines nichtlinearen Übertragungsgliedes in der Umgebung des Arbeitspunktes u_R, y_R, dann genügt es häufig, eine lineare Approximation zu verwenden. Man spricht dann von der *Linearisierung um den Arbeitspunkt*. Diese Vorgehensweise ist insbesondere dann angemessen, wenn mit einer Festwertregelung ein dynamisches System trotz Störungen im Arbeitspunkt gehalten werden soll.

Mit den Abweichungen

$$\Delta u(t) = u(t) - u_R,$$
$$\Delta y(t) = y(t) - y_R$$

erhält man durch Taylorreihenentwicklung und Vernachlässigung des Restgliedes aus (71.1) die lineare Näherung

$$a_0 \cdot \Delta y(t) + a_1 \cdot \Delta \dot{y}(t) + \ldots + a_n \cdot \Delta \overset{(n)}{y}(t)$$
$$= b_0 \cdot \Delta u(t) + b_1 \cdot \Delta \dot{u}(t) + \ldots + b_m \cdot \Delta \overset{(m)}{u}(t) \qquad (71.2)$$

mit

$$a_0 = \left.\frac{\partial f}{\partial y}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)},$$
$$a_1 = \left.\frac{\partial f}{\partial \dot{y}}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)},$$
$$\vdots$$
$$a_n = \left.\frac{\partial f}{\partial \overset{(n)}{y}}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)},$$
$$b_0 = -\left.\frac{\partial f}{\partial u}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)},$$
$$b_1 = -\left.\frac{\partial f}{\partial \dot{u}}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)},$$
$$\vdots$$
$$b_m = -\left.\frac{\partial f}{\partial \overset{(m)}{u}}\right|_{(u_R,0,\ldots,0,y_R,0,\ldots,0)}.$$

Als triviales Beispiel werde die nichtlineare Kennlinie $y = F(u)$ mit der Ruhelage $y_R = F(u_R)$ betrachtet. Mit

$$f(u, y) = y - F(u)$$

erhält man

$$a_0 = 1\big|_{(u_R,y_R)} = 1,$$
$$b_0 = \left.\frac{\partial F(u)}{\partial u}\right|_{(u_R,y_R)} = F'(u)\big|_{u_R}.$$

Aus (71.2) folgt somit die lineare Approximation

$$\Delta y(t) = F'(u)\big|_{u_R} \cdot \Delta u(t).$$

Das KL-Glied kann also in der Umgebung des Arbeitspunktes u_R, y_R durch ein P-Glied

$$\Delta y(t) = K \cdot \Delta u(t)$$

mit

$$K = F'(u)\big|_{u_R}$$

angenähert werden.

71.5 Impulsantwort und Sprungantwort einfacher Übertragungsglieder

Als *Impulsantwort* $g(t)$ eines Übertragungsgliedes bezeichnet man die Ausgangsgröße $y(t)$, die sich als Reaktion auf einen Dirac-Impuls $\delta(t)$ am Eingang $u(t)$ des Übertragungsgliedes ergibt. Als *Sprungantwort* $h(t)$ bezeichnet man die Ausgangsgröße $y(t)$, die sich als Reaktion auf einen Einheitssprung $\sigma(t)$ am Eingang $u(t)$ ergibt.

71.5.1 P-Glied

Für das *P*-Glied mit der Funktionalbeziehung

$$y(t) = K \cdot u(t)$$

lautet die Impulsantwort

$$g(t) = K \cdot \delta(t)$$

und die Sprungantwort (Abb. 71.1)

$$h(t) = K \cdot \sigma(t).$$

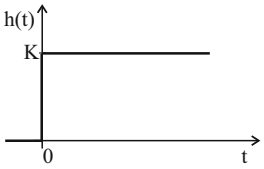


Abb. 71.1 Sprungantwort $h(t)$ des *P*-Gliedes

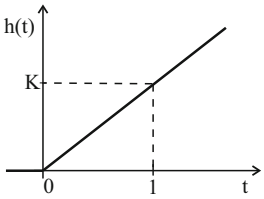


Abb. 71.2 Sprungantwort $h(t)$ des *I*-Gliedes

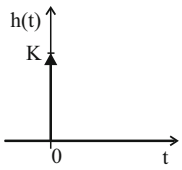


Abb. 71.3 Sprungantwort $h(t)$ des *D*-Gliedes

71.5.2 I-Glied

Für das *I*-Glied mit der Funktionalbeziehung

$$y(t) = K \int_0^t u(\tau) d\tau$$

lautet die Impulsantwort

$$g(t) = K \cdot \sigma(t)$$

und die Sprungantwort (Abb. 71.2)

$$h(t) = K \cdot t \cdot \sigma(t).$$

71.5.3 D-Glied

Für das *D*-Glied mit der Funktionalbeziehung

$$y(t) = K \cdot \dot{u}(t)$$

lautet die Impulsantwort

$$g(t) = K \cdot \dot{\delta}(t)$$

und die Sprungantwort (Abb. 71.3)

$$h(t) = K \cdot \delta(t).$$

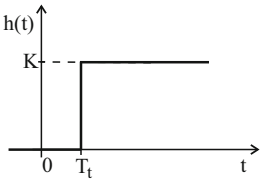


Abb. 71.4 Sprungantwort $h(t)$ des *TZ*-Gliedes

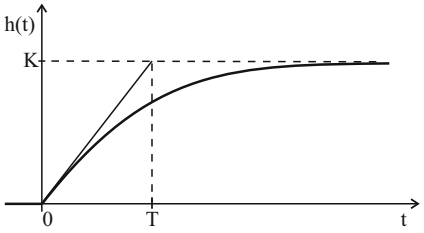


Abb. 71.5 Sprungantwort $h(t)$ des PT_1-Gliedes

71.5.4 TZ-Glied

Für das *TZ*-Glied mit der Funktionalbeziehung

$$y(t) = K \cdot u(t - T_t), \quad T_t > 0,$$

lautet die Impulsantwort

$$g(t) = K \cdot \delta(t - T_t)$$

und die Sprungantwort (Abb. 71.4)

$$h(t) = K \cdot \sigma(t - T_t).$$

71.5.5 PT_1-Glied

Für das PT_1-Glied mit der Funktionalbeziehung

$$T \cdot \dot{y}(t) + y(t) = K \cdot u(t), \quad T > 0$$

lautet die Impulsantwort

$$g(t) = \frac{K}{T} \cdot e^{-\frac{t}{T}} \cdot \sigma(t)$$

und die Sprungantwort (Abb. 71.5)

$$h(t) = K \cdot \left(1 - e^{-\frac{t}{T}}\right) \cdot \sigma(t)$$

71.5.6 PT_2-Glied

Für das PT_2-Glied mit der Funktionalbeziehung

$$T^2 \cdot \ddot{y}(t) + 2 \cdot d \cdot T \cdot \dot{y}(t) + y(t) = K \cdot u(t), \quad T, d > 0$$

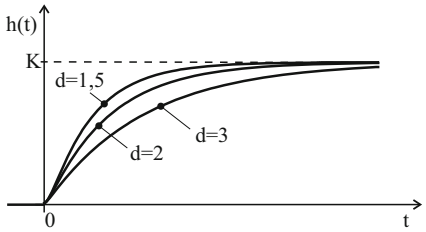


Abb. 71.6 Sprungantwort $h(t)$ des PT_2-Gliedes im aperiodischen Fall für verschiedene Dämpfungen d

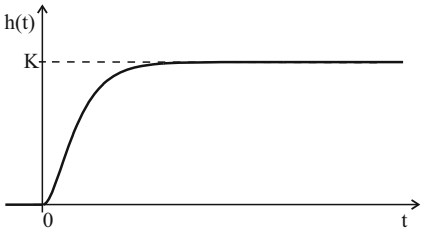


Abb. 71.7 Sprungantwort $h(t)$ des PT_2-Gliedes im aperiodischen Grenzfall

sind in Abhängigkeit von der Dämpfung d drei Fälle zu unterscheiden: der aperiodische Fall für $d > 1$, der aperiodische Grenzfall für $d = 1$ und der periodische Fall für $0 < d < 1$.

Im *aperiodischen Fall* für $d > 1$ lautet die Impulsantwort

$$g(t) = \frac{K}{T_1 - T_2} \cdot \left(e^{-\frac{t}{T_1}} - e^{-\frac{t}{T_2}} \right) \cdot \sigma(t)$$

mit

$$T_1 = T \cdot \left(d + \sqrt{d^2 - 1} \right),$$
$$T_2 = T \cdot \left(d - \sqrt{d^2 - 1} \right)$$

und die Sprungantwort (Abb. 71.6)

$$h(t) = \left(K - \frac{K}{T_1 - T_2} \cdot \left(T_1 \cdot e^{-\frac{t}{T_1}} - T_2 \cdot e^{-\frac{t}{T_2}} \right) \right) \cdot \sigma(t).$$

Im *aperiodischen Grenzfall* für $d = 1$ lautet die Impulsantwort

$$g(t) = \frac{K}{T^2} \cdot t \cdot e^{-\frac{t}{T}} \cdot \sigma(t)$$

und die Sprungantwort (Abb. 71.7)

$$h(t) = \left(K - K \cdot e^{-\frac{t}{T}} - \frac{K}{T} \cdot t \cdot e^{-\frac{t}{T}} \right) \cdot \sigma(t).$$

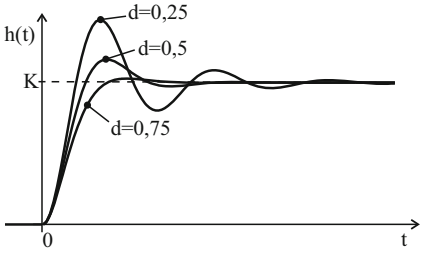


Abb. 71.8 Sprungantwort $h(t)$ des PT_2-Gliedes im periodischen Fall für verschiedene Dämpfungen d

Im *periodischen Fall* für $0 < d < 1$ lautet die Impulsantwort

$$g(t) = \frac{K}{T \cdot \sqrt{1 - d^2}} \cdot e^{-d \cdot \frac{t}{T}} \cdot \sin\left(\sqrt{1 - d^2} \cdot \frac{t}{T} \right) \cdot \sigma(t)$$

und die Sprungantwort (Abb. 71.8)

$$h(t) = \left(K - \frac{K}{\sqrt{1 - d^2}} \cdot e^{-d \cdot \frac{t}{T}} \cdot \sin\left(\sqrt{1 - d^2} \cdot \frac{t}{T} + \psi \right) \right) \cdot \sigma(t)$$

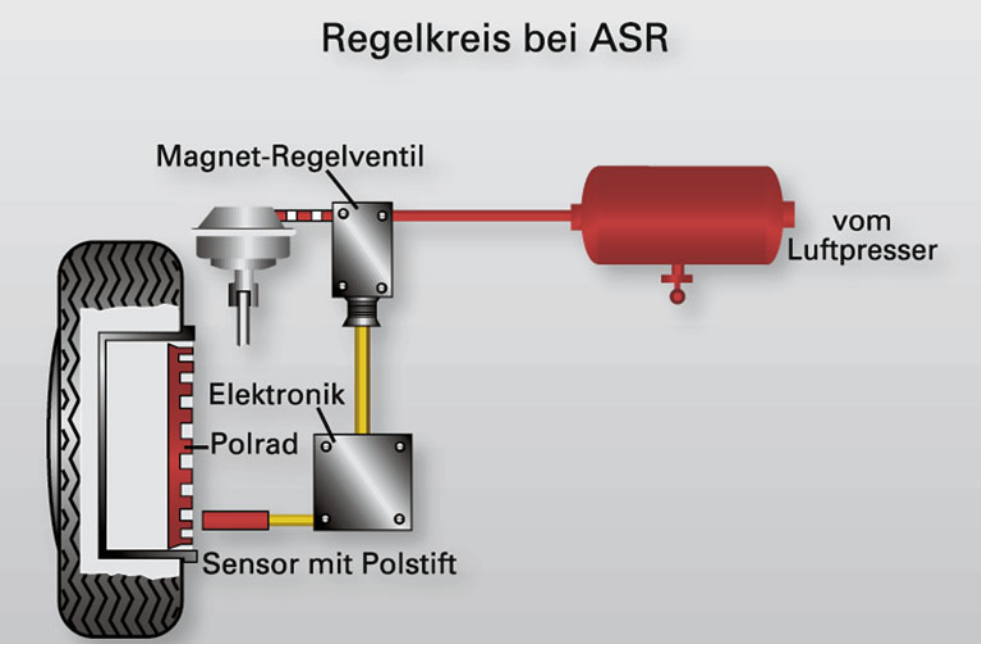
mit

$$\cos \psi = d, \quad 0 < \psi < \frac{\pi}{2}.$$

Übertragungsverhalten linearer, zeitinvarianter Übertragungsglieder

72

Joachim Horn



72.1 Übertragungsfunktion 854
72.2 Frequenzgangfunktion 854
72.3 Frequenzkennlinien elementarer Übertragungsfunktionen 855

Teil IX

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_72



Abb. 72.1 Lineares, zeitinvariantes Übertragungsglied

Betrachtet werde ein lineares, zeitinvariantes Übertragungsglied (*LZI*-Glied) mit der Eingangsgröße $u(t)$ und der Ausgangsgröße $y(t)$, siehe Abb. 72.1. Die Ausgangsgröße $y(t)$ berechnet sich bei verschwindenden Anfangswerten durch Faltung von Impulsantwort $g(t)$ und Eingangsgröße $u(t)$ gemäß

$$y(t) = \int_{\tau=0}^{t} g(t-\tau) \cdot u(\tau) d\tau. \tag{72.1}$$

72.1 Übertragungsfunktion

Durch Laplace-Transformation und Anwendung der Faltungsregel erhält man aus (72.1) im Bildbereich

$$Y(s) = G(s) \cdot U(s).$$

$G(s)$ wird als *Übertragungsfunktion* des linearen, zeitinvarianten Systems bezeichnet. Die Übertragungsfunktion $G(s)$ ist also die Laplace-Transformierte der Impulsantwort $g(t)$. Für eine beliebige Eingangsgröße $u(t)$ kann die Übertragungsfunktion gemäß

$$G(s) = \frac{Y(s)}{U(s)}$$

als Quotient der Laplace-Transformierten $Y(s)$ der Ausgangsgröße $y(t)$ und der Laplace-Transformierten $U(s)$ der Eingangsgröße $u(t)$ berechnet werden.

Wird das Übertragungsverhalten des linearen, zeitinvarianten Systems durch die Differenzialgleichung

$$\begin{aligned} &\overset{(n)}{y}(t) + a_{n-1} \cdot \overset{(n-1)}{y}(t) + \ldots + a_1 \cdot \dot{y}(t) + a_0 \cdot y(t) \\ &\quad = b_m \cdot \overset{(m)}{u}(t) + b_{m-1} \cdot \overset{(m-1)}{u}(t) + \ldots + b_1 \cdot \dot{u}(t) + b_0 \cdot u(t) \end{aligned}$$

beschrieben, so erhält man durch Laplace-Transformation und Anwendung der Differenziationsregel bei verschwindenden Anfangswerten im Bildbereich

$$\begin{aligned} &(s^n + a_{n-1} \cdot s^{n-1} + \ldots + a_1 \cdot s + a_0) \cdot Y(s) \\ &\quad = (b_m \cdot s^m + b_{m-1} \cdot s^{m-1} + \ldots + b_1 \cdot s + b_0) \cdot U(s). \end{aligned}$$

Die Übertragungsfunktion ist damit durch die gebrochen rationale Funktion

$$G(s) = \frac{b_m \cdot s^m + b_{m-1} \cdot s^{m-1} + \ldots + b_1 \cdot s + b_0}{s^n + a_{n-1} \cdot s^{n-1} + \ldots + a_1 \cdot s + a_0}$$

gegeben. Solche Übertragungsglieder werden als *R*-Glieder bezeichnet. Mit den *Polen* p_i und den *Nullstellen* n_j erhält man die Darstellung

$$G(s) = K \cdot \frac{\prod_{j=1}^{m} (s - n_j)}{\prod_{i=1}^{n} (s - p_i)}.$$

Als Pole der Übertragungsfunktion bezeichnet man die Nullstellen des Nennerpolynoms

$$N(s) = s^n + a_{n-1} \cdot s^{n-1} + \ldots + a_1 \cdot s + a_0,$$

als Nullstellen der Übertragungsfunktion die Nullstellen des Zählerpolynoms

$$Z(s) = b_m \cdot s^m + b_{m-1} \cdot s^{m-1} + \ldots + b_1 \cdot s + b_0,$$

wobei vorausgesetzt ist, dass Zähler- und Nennerpolynom keine gemeinsame Nullstelle besitzen. Die Anzahl m der Nullstellen ist kleiner oder gleich der Anzahl n der Pole. Die Polynomkoeffizienten $a_n, \ldots, a_0$ des Nennerpolynoms sind reell, daher treten ausschließlich reelle Pole oder konjugiert komplexe Polpaare auf. Entsprechendes gilt für die Polynomkoeffizienten $b_m, \ldots, b_0$ des Zählerpolynoms und die Nullstellen der Übertragungsfunktion.

72.2 Frequenzgangfunktion

Betrachtet man die Übertragungsfunktion $G(s)$ auf der imaginären Achse, d. h. für $s = j\omega$, so erhält man die *Frequenzgangfunktion*

$$F(j\omega) = G(s)\big|_{s=j\omega}.$$

Die Frequenzgangfunktion eines stabilen Übertragungsgliedes kann direkt gemessen werden: Schaltet man eine harmonische Schwingung mit der Frequenz ω und der Amplitude 1 auf den Eingang des linearen, zeitinvarianten Übertragungsgliedes,

$$u(t) = \sin(\omega \cdot t),$$

so erhält man am Ausgang nach dem Abklingen der Einschwingvorgänge ebenfalls eine harmonische Schwingung mit derselben Frequenz ω, einer Amplitude $A(\omega)$ und einer Phasenverschiebung $\varphi(\omega)$,

$$y(t) = A(\omega) \cdot \sin(\omega(t) + \varphi(\omega)).$$

Der Betrag der Frequenzgangfunktion für die Frequenz ω beträgt

$$|F(j\omega)| = A(\omega),$$

die Phase

$$|\underline{F}(j\omega) = \varphi(\omega),$$

und damit ergibt sich die Frequenzgangfunktion zu

$$F(j\omega) = A(\omega) \cdot e^{j\varphi(\omega)}$$

für die betrachtete Frequenz ω. Aufgrund der Linearität des Übertragungsgliedes kann die Messung mit einer beliebigen Eingangsamplitude $\hat{U}$ durchgeführt werden, die Ausgangsamplitude ergibt sich dann zu $\hat{Y} = A(\omega) \cdot \hat{U}$.

Wiederholt man das Experiment für verschiedene Frequenzwerte, erhält man die Frequenzgangfunktion $F(j\omega)$.

Von besonderer Bedeutung sind zwei grafische Darstellungen der Frequenzgangfunktion: Die *Ortskurve* zeigt $F(j\omega)$ für $\omega \geq 0$ in der komplexen Ebene. Im *Bode-Diagramm* werden Betrag und Phase von $F(j\omega)$ in zwei separaten Frequenzkennlinien dargestellt.

Stellt man die Frequenzgangfunktion $F(j\omega)$ als Produkt

$$F(j\omega) = F_1(j\omega) \cdot \ldots \cdot F_n(j\omega)$$

dar, so erhält man mit

$$\begin{aligned} |F(j\omega)| \cdot e^{j\underline{|F(j\omega)}} &= |F_1(j\omega)| \cdot e^{j\underline{|F_1(j\omega)}} \cdot \ldots \cdot |F_n(j\omega)| \cdot e^{j\underline{|F_n(j\omega)}} \\ &= |F_1(j\omega)| \cdot \ldots \cdot |F_n(j\omega)| \cdot e^{j(\underline{|F_1(j\omega)} + \ldots + \underline{|F_n(j\omega)})} \end{aligned}$$

für den Betrag

$$|F(j\omega)| = |F_1(j\omega)| \cdot \ldots \cdot |F_n(j\omega)|$$

und für die Phase

$$\underline{|F(j\omega)} = \underline{|F_1(j\omega)} + \ldots + \underline{|F_n(j\omega)}.$$

Geht man zum Betrag in Dezibel gemäß

$$|F(j\omega)|_{\text{dB}} = 20 \cdot \log |F(j\omega)|$$

über, so ergibt sich

$$|F(j\omega)|_{\text{dB}} = |F_1(j\omega)|_{\text{dB}} + \ldots + |F_n(j\omega)|_{\text{dB}}.$$

Betrags- bzw. Phasenkennlinie von $F(j\omega)$ erhält man also als Summe der Betrags- bzw. Phasenkennlinien von $F_1(j\omega)$ bis $F_n(j\omega)$. Die Frequenz wird im Bode-Diagramm logarithmisch aufgetragen.

72.3 Frequenzkennlinien elementarer Übertragungsfunktionen

72.3.1 Verstärkungsfaktor

Mit $F(j\omega) = K$, $K > 0$ ergibt sich der Betrag in Dezibel zu (Abb. 72.2)

$$|F(j\omega)|_{\text{dB}} = 20 \cdot \log K,$$

die Phase zu

$$\underline{|F(j\omega)} = 0°.$$

72.3.2 Integrierer

Mit $F(j\omega) = \frac{1}{j\omega}$ ergibt sich der Betrag in Dezibel zu (Abb. 72.3)

$$|F(j\omega)|_{\text{dB}} = -20 \cdot \log \omega,$$

die Phase zu

$$\underline{|F(j\omega)} = -90°.$$

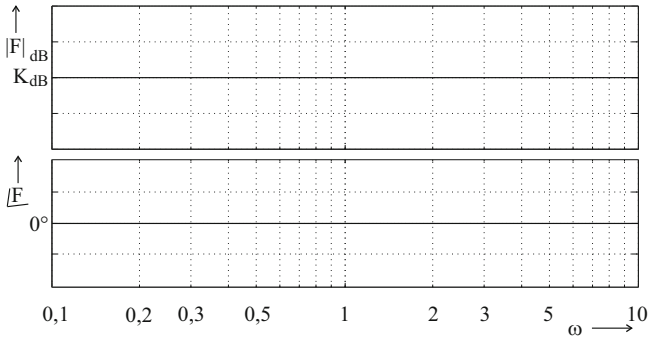


Abb. 72.2 Frequenzkennlinien von $F(j\omega) = K$

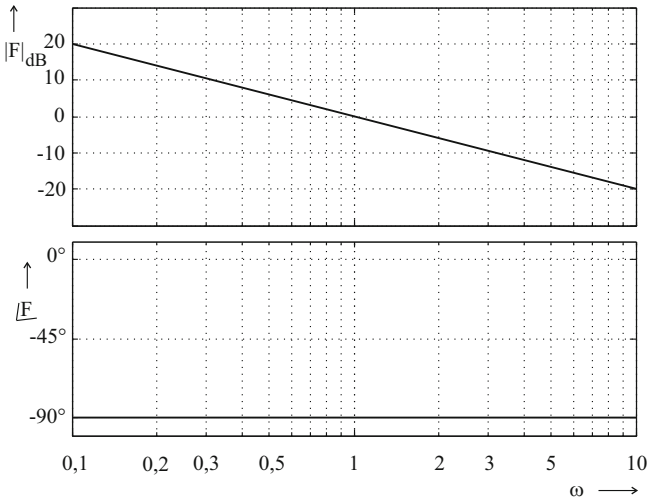


Abb. 72.3 Frequenzkennlinien von $F(j\omega) = \frac{1}{j\omega}$

72.3.3 Differenzierer

Mit $F(j\omega) = j\omega$ ergibt sich der Betrag in Dezibel zu (Abb. 72.4)

$$|F(j\omega)|_{\text{dB}} = 20 \cdot \log \omega,$$

die Phase zu

$$\underline{|F(j\omega)} = 90°.$$

72.3.4 Reeller Pol

Mit $F(j\omega) = \frac{1}{1+j\omega \cdot T}$, $T > 0$, ergibt sich der Betrag zu

$$|F(j\omega)| = \frac{1}{\sqrt{1 + \omega^2 \cdot T^2}}.$$

Für $\omega \ll \frac{1}{T}$ erhält man näherungsweise

$$|F(j\omega)| \approx 1$$

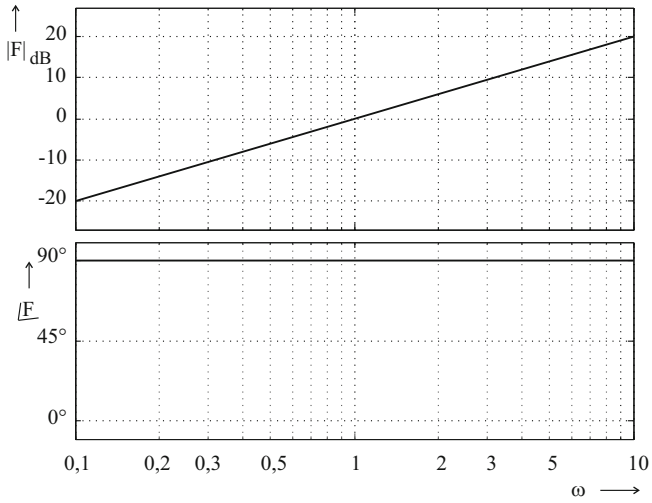


Abb. 72.4 Frequenzkennlinien von $F(j\omega) = j\omega$

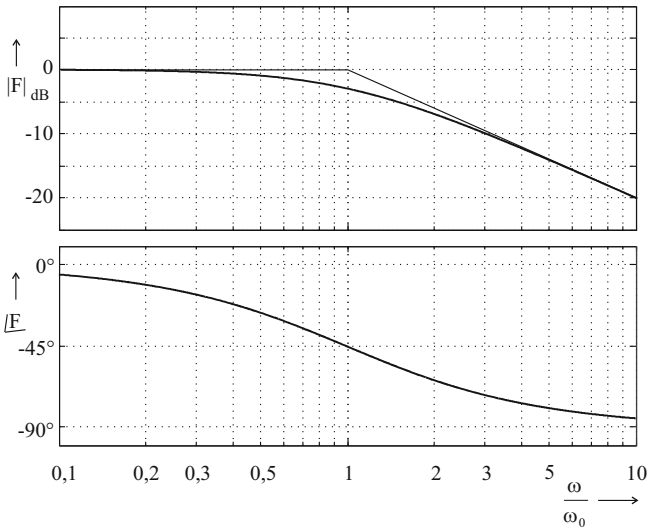


Abb. 72.5 Frequenzkennlinien von $F(j\omega) = \frac{1}{1+j\omega \cdot T}$

und damit

$$|F(j\omega)|_{\text{dB}} \approx 0.$$

Die 0-dB-Linie ist also die Asymptote für $\omega \ll \frac{1}{T}$.

Für $\omega \gg \frac{1}{T}$ erhält man näherungsweise

$$|F(j\omega)| \approx \frac{1}{\omega \cdot T}$$

und damit

$$|F(j\omega)|_{\text{dB}} = -20 \cdot \log \omega - 20 \cdot \log T.$$

Die Asymptote für $\omega \gg \frac{1}{T}$ ist also eine Gerade, die mit 20 dB pro Dekade abfällt. Die Betragskennlinie kann damit durch eine Knickgerade approximiert werden.

Abb. 72.5 zeigt den exakten Betrag und die Approximation als Knickgerade. Die Abweichung zwischen der exakten Betragskennlinie und der Approximation durch die Knickgerade ist bei der Knickfrequenz $\omega_0 = \frac{1}{T}$ maximal, sie beträgt dort 3 dB.

Die Phase ergibt sich zu

$$\angle F(j\omega) = -\arctan(\omega \cdot T).$$

72.3.5 Reelle Nullstelle

Die Frequenzkennlinie einer reellen Nullstelle mit

$$F(j\omega) = 1 + j\omega \cdot T, \quad T > 0,$$

erhält man aus den Frequenzkennlinien einer reellen Polstelle durch Vorzeichenumkehr, d. h. durch Spiegelung an der 0-dB- bzw. 0°-Linie (Abb. 72.6).

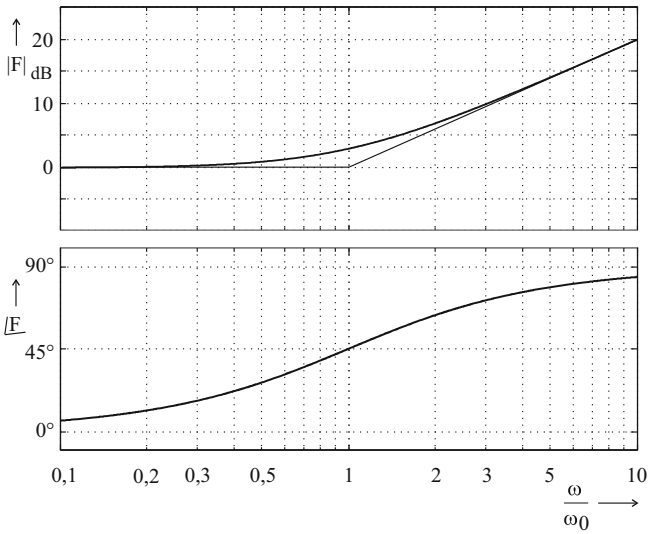


Abb. 72.6 Frequenzkennlinie von $F(j\omega) = 1 + j\omega \cdot T$

72.3.6 Konjugiert komplexes Polpaar

Mit $F(j\omega) = \frac{1}{1+2 \cdot d \cdot j\omega \cdot T + (j\omega \cdot T)^2}$, $0 < d < 1, T > 0$, ergibt sich der Betrag zu

$$|F(j\omega)| = \frac{1}{\sqrt{(1-(\omega \cdot T)^2)^2 + (2 \cdot d \cdot \omega \cdot T)^2}}.$$

Für $\omega \ll \frac{1}{T}$ erhält man näherungsweise

$$|F(j\omega)| \approx 1$$

und damit

$$|F(j\omega)|_{\text{dB}} \approx 0.$$

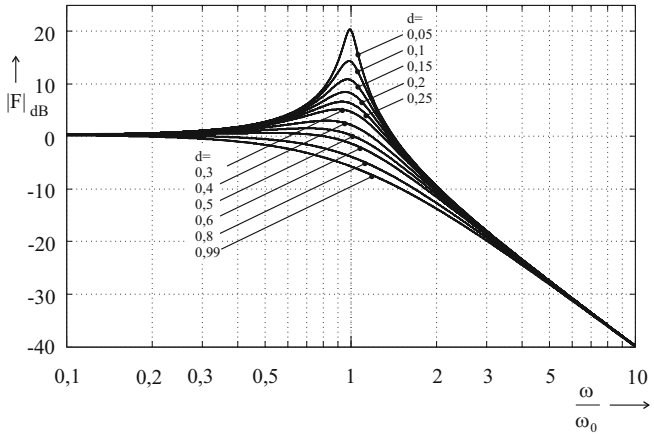


Abb. 72.7 Betragskennlinie von $F(j\omega) = \frac{1}{1+2\cdot d\cdot j\omega\cdot T+(j\omega\cdot T)^2}$ in Abhängigkeit von der Dämpfung d

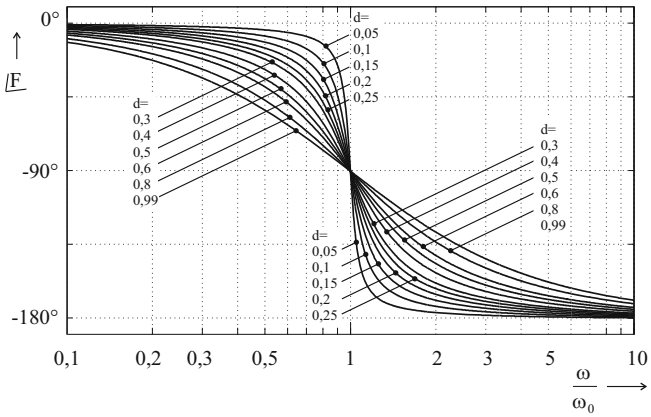


Abb. 72.8 Phasenkennlinie von $F(j\omega) = \frac{1}{1+2\cdot d\cdot j\omega\cdot T+(j\omega\cdot T)^2}$ in Abhängigkeit von der Dämpfung d

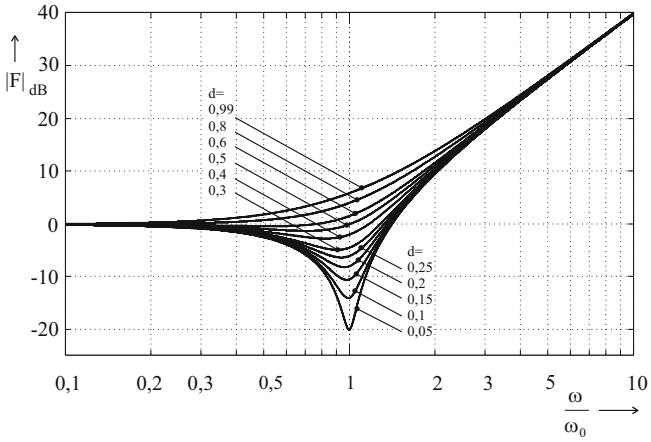


Abb. 72.9 Betragskennlinie von $F(j\omega) = 1 + 2\cdot d\cdot j\omega\cdot T + (j\omega\cdot T)^2$ in Abhängigkeit von der Dämpfung d

Die 0-dB-Linie ist also die Asymptote für $\omega \ll \frac{1}{T}$.

Für $\omega \gg \frac{1}{T}$ erhält man näherungsweise

$$|F(j\omega)| \approx \frac{1}{(\omega\cdot T)^2}$$

und damit

$$|F(j\omega)|_{\text{dB}} \approx -40\cdot\log\omega - 40\cdot\log T.$$

Die Asymptote für $\omega \gg \frac{1}{T}$ ist also eine Gerade, die mit 40 dB pro Dekade abfällt.

Abb. 72.7 zeigt die Betragskennlinie von

$$F(j\omega) = \frac{1}{1+2\cdot d\cdot j\omega\cdot T+(j\omega\cdot T)^2}$$

in Abhängigkeit von der Dämpfung d. Für $d < \frac{\sqrt{2}}{2}$

kommt es zur sogenannten *Resonanzüberhöhung*. Der Maximalwert der Betragskennlinie wird für die Resonanzfrequenz $\omega_R = \omega_0\cdot\sqrt{1-2\cdot d^2}$ mit $\omega_0 = \frac{1}{T}$ erreicht und beträgt

$$\max|F(j\omega)| = \frac{1}{2\cdot d\cdot\sqrt{1-d^2}}.$$

Diese Resonanzüberhöhung wird für $d \to 0$ unendlich groß.

Die Phase ergibt sich zu

$$|\underline{F}(j\omega) = -\arctan\frac{2\cdot d\cdot\omega\cdot T}{1-(\omega\cdot T)^2}.$$

Abb. 72.8 zeigt die Phasenkennlinie von $F(j\omega) = \frac{1}{1+2\cdot d\cdot j\omega\cdot T+(j\omega\cdot T)^2}$ in Abhängigkeit von der Dämpfung d.

72.3.7 Konjugiert komplexes Nullstellenpaar

Die Frequenzkennlinien eines konjugiert komplexen Nullstellenpaares mit $F(j\omega) = 1 + 2\cdot d\cdot j\omega\cdot T + (j\omega\cdot T)^2$, $0 < d < 1$, $T > 0$, erhält man aus den Frequenzkennlinien eines konjugiert komplexen Poolpaares durch Vorzeichenumkehr, d. h. durch Spiegelung an der 0-dB- bzw. 0°-Linie (Abb. 72.9 und 72.10).

72.3.8 Totzeit

Mit $F(j\omega) = e^{-j\omega\cdot T_t}$, $T_t > 0$, ergibt sich der Betrag zu (Abb. 72.11)

$$|F(j\omega)| = 1$$

und damit

$$|F(j\omega)|_{\text{dB}} = 0.$$

Teil IX

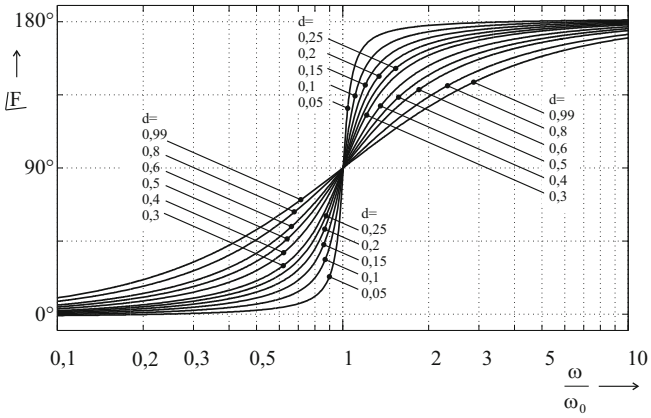


Abb. 72.10 Phasenkennlinie von $F(j\omega) = 1 + 2 \cdot d \cdot j\omega \cdot T + (j\omega \cdot T)^2$ in Abhängigkeit von der Dämpfung d

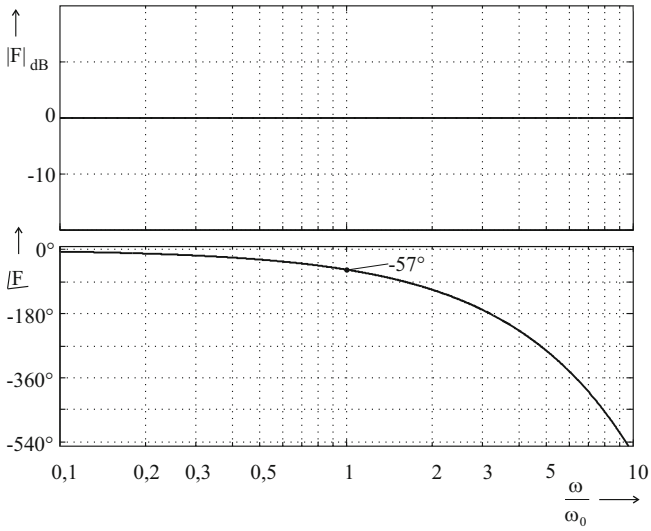


Abb. 72.11 Frequenzkennlinien von $F(j\omega) = e^{-j\omega \cdot T_t}$

Die Phase beträgt

$$|\underline{F}(j\omega) = -\omega \cdot T_t = -\frac{\omega}{\omega_0} \cdot \frac{180°}{\pi}$$

mit $\omega_0 = \frac{1}{T_t}$.

72.3.9 Allpass erster Ordnung

Mit $F(j\omega) = \frac{1-j\omega \cdot T}{1+j\omega \cdot T}$, $T > 0$, ergibt sich der Betrag zu (Abb. 72.12)

$$|F(j\omega)| = 1$$

und damit

$$|F(j\omega)|_{\text{dB}} = 0.$$

Die Phase beträgt

$$F(j\omega) = -2 \cdot \arctan(\omega \cdot T),$$

sie ist gleich der doppelten Phase eines reellen Pols mit der Zeitkonstanten T.

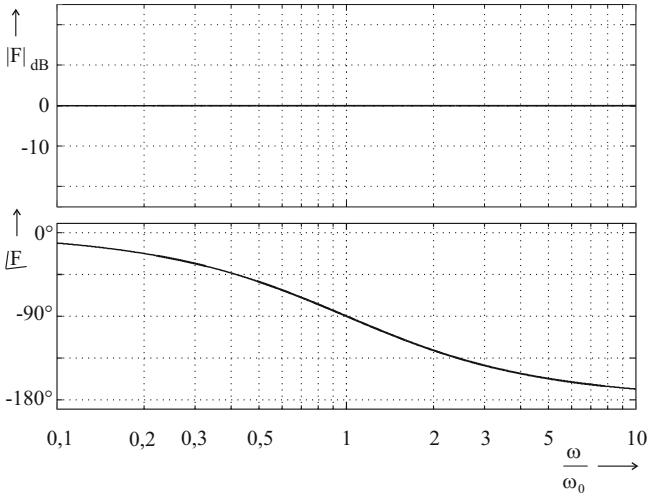


Abb. 72.12 Frequenzkennlinie von $F(j\omega) = \frac{1-j\omega \cdot T}{1+j\omega \cdot T}$

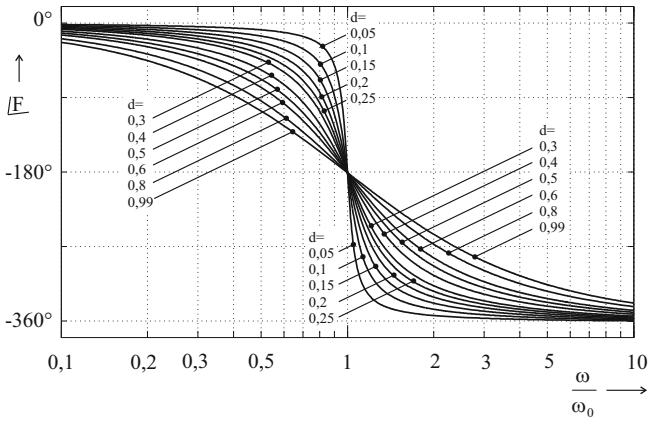


Abb. 72.13 Frequenzkennlinie von $F(j\omega) = \frac{1-2 \cdot d \cdot j\omega \cdot T+(j\omega \cdot T)^2}{1+2 \cdot d \cdot j\omega \cdot T+(j\omega \cdot T)^2}$

72.3.10 Allpass zweiter Ordnung

Mit $F(j\omega) = \frac{1-2 \cdot d \cdot j\omega \cdot T+(j\omega \cdot T)^2}{1+2 \cdot d \cdot j\omega \cdot T+(j\omega \cdot T)^2}$, $0 < d < 1, T > 0$, ergibt sich der Betrag zu (Abb. 72.13)

$$|F(j\omega)| = 1$$

und damit

$$|F(j\omega)|_{\text{dB}} = 0.$$

Die Phase beträgt

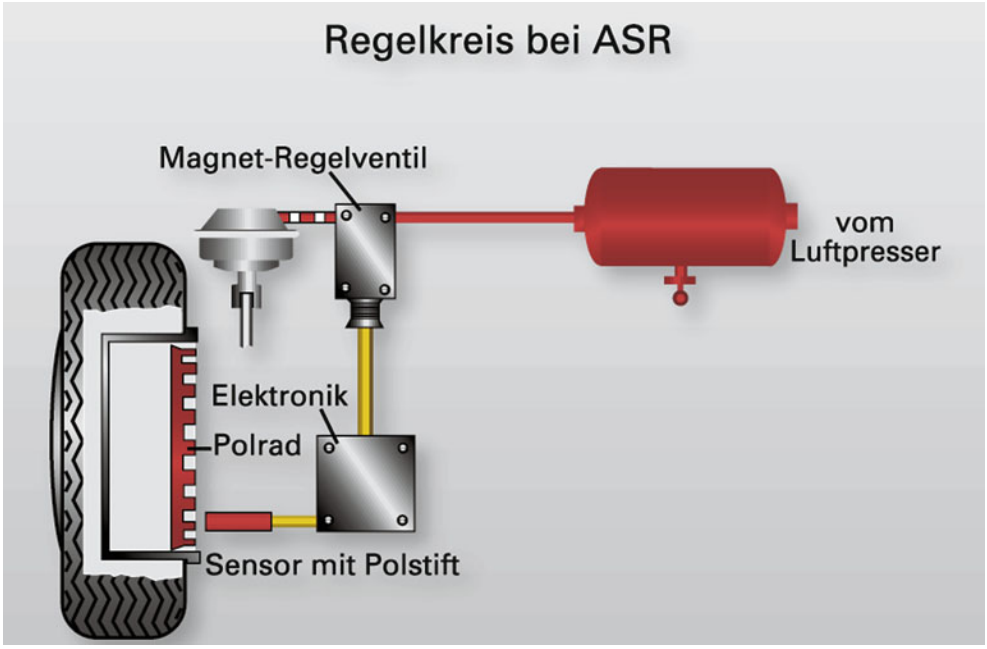
$$|\underline{F}(j\omega) = -2 \cdot \arctan \frac{2 \cdot d \cdot \omega \cdot T}{1 - (\omega \cdot T)^2},$$

sie ist gleich der doppelten Phase eines konjugiert komplexen Poolpaares mit der Dämpfung d und der Zeitkonstanten T.

Stabilität

73

Joachim Horn



73.1 Definition der Stabilität 860
73.2 Stabilitätskriterien 860

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_73

73.1 Definition der Stabilität

Für die Stabilität eines linearen, zeitinvarianten Übertragungsgliedes existieren verschiedene Definitionen. Bei der *asymptotischen Stabilität* wird gefordert, dass der Systemausgang nach einer kurzzeitigen Systemanregung mit wachsender Zeit wieder gegen Null geht. Bei der BIBO-Stabilität (Bounded Input-Bounded Output) wird gefordert, dass bei einem beschränkten Systemeingang der Systemausgang ebenfalls beschränkt ist.

73.1.1 Asymptotische Stabilität

Ein LZI-Glied heißt asymptotisch stabil, wenn seine Impulsantwort $g(t)$ für $t \to \infty$ gegen Null strebt. Andernfalls heißt es instabil.

Gleichwertig hierzu ist die Forderung, dass die Sprungantwort $h(t)$ für $t \to \infty$ gegen einen endlichen Wert strebt.

73.1.2 BIBO-Stabilität

Ein LZI-Glied heißt BIBO-stabil (Bounded Input-Bounded Output), wenn die Systemantwort auf jede beschränkte Eingangsgröße ebenfalls beschränkt ist.

Die BIBO-Stabilität wird auch als *Übertragungsstabilität* bezeichnet. Aus der BIBO-Stabilität folgt stets die asymptotische Stabilität, die BIBO-Stabilität stellt also eine strengere Forderung an das Übertragungsglied. Für den wichtigen Fall der *R*-Glieder stimmen jedoch asymptotische Stabilität und BIBO-Stabilität überein, so dass im Folgenden nur noch von Stabilität gesprochen wird.

73.2 Stabilitätskriterien

73.2.1 Grundlegendes Stabilitätskriterium

Ein R-Glied ist genau dann stabil, wenn alle Pole seiner Übertragungsfunktion links der imaginären Achse der komplexen Ebene liegen.

Betrachtet man ein *R*-Glied mit der Übertragungsfunktion

$$G(s) = \frac{Z(s)}{N(s)},$$

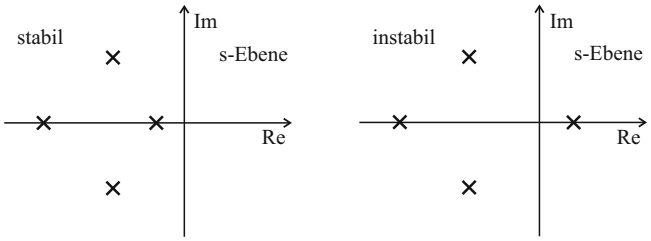


Abb. 73.1 Beispiel für die Polkonfiguration eines stabilen und eines instabilen *R*-Gliedes

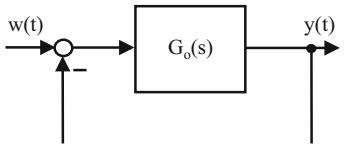


Abb. 73.2 Standardregelkreis

wobei Zählerpolynom $Z(s)$ und Nennerpolynom $N(s)$ *koprim* seien, d. h. keine gemeinsame Nullstelle besitzen, so sind die Pole der Übertragungsfunktion $G(s)$ gleich den Nullstellen des Nennerpolynoms $N(s)$. Die Pole eines stabilen *R*-Gliedes besitzen also alle einen negativen Realteil. Abb. 73.1 zeigt jeweils ein Beispiel für die Polkonfiguration eines stabilen und eines instabilen *R*-Gliedes.

Betrachtet man den Standardregelkreis gemäß Abb. 73.2, so ergibt sich die Führungsübertragungsfunktion $G_w(s) = \frac{Y(s)}{W(s)}$ zu

$$G_w(s) = \frac{G_o(s)}{1 + G_o(s)}.$$

Mit der Übertragungsfunktion $G_o(s)$ des offenen Kreises gemäß

$$G_o(s) = \frac{Z_o(s)}{N_o(s)}$$

erhält man

$$G_w(s) = \frac{Z_o(s)}{Z_o(s) + N_o(s)}.$$

Damit folgt das *grundlegende Stabilitätskriterium für Regelkreise*:

Ein Regelkreis ist genau dann stabil, wenn alle Nullstellen des Polynoms $Z_o(s) + N_o(s)$ links der imaginären Achse der komplexen Ebene liegen.

Die zur Bestimmung der Nullstellen zu lösende Gleichung

$$Z_o(s) + N_o(s) = 0$$

wird als *charakteristische Gleichung des Regelkreises* bezeichnet.

Für eine Stabilitätsaussage ist die Kenntnis der exakten Lage der Pole von $G(s)$ bzw. $G_w(s)$ nicht erforderlich. Es muss lediglich die Frage beantwortet werden, ob alle Pole links der imaginären Achse liegen. Hierfür wurden verschiedene Kriterien entwickelt. Im Folgenden werden als wichtige Vertreter das Hurwitz-Kriterium und das Nyquist-Kriterium vorgestellt.

73.2.2 Hurwitz-Kriterium

Betrachtet wird das Polynom

$$P(s) = a_n \cdot s^n + a_{n-1} \cdot s^{n-1} + \ldots + a_1 \cdot s + a_0$$

mit reellen Koeffizienten a_i, $i = 0, \ldots, n$, und

$$a_n > 0.$$

Zu untersuchen ist, ob alle Nullstellen des Polynoms $P(s)$ links der imaginären Achse liegen. Ein solches Polynom wird als *Hurwitz-Polynom* bezeichnet.

Eine *notwendige Bedingung* hierfür ist, dass alle Polynomkoeffizienten positiv sind, d. h.

$$a_i > 0 \quad \text{für } i = 0, \ldots, n.$$

Eine *notwendige und hinreichende Bedingung* erhält man, indem man die Hurwitz-Matrix

$$\underline{H} = \begin{bmatrix} a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & a_{n-9} & \cdots & 0 \\ a_n & a_{n-2} & a_{n-4} & a_{n-6} & a_{n-8} & \cdots & 0 \\ 0 & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots & \\ 0 & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots & \\ 0 & 0 & a_{n-1} & a_{n-3} & a_{n-5} & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ 0 & 0 & & & & \cdots & a_0 \end{bmatrix}$$

mit n Zeilen und n Spalten sowie $a_i = 0$ für $i < 0$ betrachtet. Der Index der Koeffizienten nimmt mit jeder Spalte um 2 ab und mit jeder Zeile um 1 zu, wobei $a_i = 0$ für $i > n$ gilt.

Aus der Hurwitz-Matrix $\underline{H}$ werden nun die Hurwitz-Determinanten $H_1, \ldots, H_n$ gebildet, die Hauptabschnittsdeterminanten oder „nordwestlichen“ Unterdeterminanten von $\underline{H}$ gemäß

$$\begin{aligned} H_1 &= a_{n-1}, \\ H_2 &= \begin{vmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{vmatrix}, \\ H_3 &= \begin{vmatrix} a_{n-1} & a_{n-3} & a_{n-5} \\ a_n & a_{n-2} & a_{n-4} \\ 0 & a_{n-1} & a_{n-3} \end{vmatrix}, \\ &\vdots \\ H_n &= \det \underline{H}. \end{aligned}$$

> Die Nullstellen des Polynoms $P(s)$ liegen genau dann alle links der imaginären Achse, wenn die Hurwitz-Determinanten $H_1, \ldots, H_n$ alle positiv sind.

Betrachtet man das R-Glied mit der Übertragungsfunktion $G(s) = \frac{Z(s)}{N(s)}$, so ist zur Stabilitätsuntersuchung gemäß dem grundlegenden Stabilitätskriterium zu prüfen, ob alle Pole, also alle Nullstellen des Nennerpolynoms $N(s)$, links der imaginären Achse liegen. Durch Anwendung des Hurwitz-Kriteriums erhält man folgenden Satz:

> Das R-Glied mit der Übertragungsfunktion $G(s) = \frac{Z(s)}{N(s)}$ ist genau dann stabil, wenn die Hurwitz-Determinanten $H_1, \ldots, H_n$ des Nennerpolynoms $N(s) = a_n \cdot s^n + \ldots + a_1 \cdot s + a_0$, $a_n > 0$, alle positiv sind.

Betrachtet man den Standardregelkreis gemäß Abb. 73.2 mit der Übertragungsfunktion $G_o(s) = \frac{Z_o(s)}{N_o(s)}$ des offenen Kreises, so lautet die Führungsübertragungsfunktion

$$G_w(s) = \frac{Z_o(s)}{Z_o(s) + N_o(s)}.$$

Damit ergibt sich das *Hurwitz-Kriterium für Regelkreise*:

> Ein Regelkreis mit der Übertragungsfunktion $G_o(s) = \frac{Z_o(s)}{N_o(s)}$ des offenen Kreises ist genau dann stabil, wenn die Hurwitz-Determinanten $H_1, \ldots, H_n$ des charakteristischen Polynoms $Z_o(s) + N_o(s)$ des Regelkreises alle positiv sind.

Mit

$$\begin{aligned} Z_o(s) &= b_n \cdot s^n + \ldots + b_1 \cdot s + b_0 \quad \text{und} \\ N_o(s) &= a_n \cdot s^n + \ldots + a_1 \cdot s + a_0 \end{aligned}$$

lautet das charakteristische Polynom des Regelkreises

$$\begin{aligned} Z_o(s) + N_o(s) = (a_n + b_n) \cdot s^n + \ldots \\ + (a_1 + b_1) \cdot s + (a_0 + b_0). \end{aligned}$$

Beim Hurwitz-Kriterium für Regelkreise gemäß obiger Formulierung ist vorausgesetzt, dass

$$a_n + b_n > 0$$

gilt.

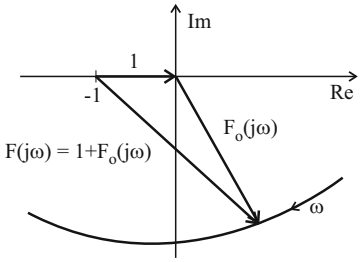


Abb. 73.3 Komplexer Zeiger vom kritischen Punkt −1 zur Ortskurve $F_o(j\omega)$ des offenen Kreises

73.2.3 Nyquist-Kriterium

Das Nyquist-Kriterium ermöglicht, aus dem Frequenzgang $F_o(j\omega)$ des offenen Kreises direkt auf die Stabilität des Standardregelkreises gemäß Abb. 73.2 zu schließen. Für eine einfache Formulierung des Kriteriums sei der Fall betrachtet, dass

$$G_o(s) = \frac{Z_o(s)}{N_o(s)}$$

mit

$$\lim_{s\to\infty} G_o(s) = 0$$

gilt. Der Grad des Zählerpolynoms $Z_o(s)$ ist also kleiner als der Grad n des Nennerpolynoms $N_o(s)$, der offene Kreis ist nicht sprungfähig. Nun betrachtet man den komplexen Zeiger vom *kritischen Punkt* −1 zur Ortskurve $F_o(j\omega)$ des offenen Kreises gemäß Abb. 73.3.

Mit

$$F(j\omega) = 1 + F_o(j\omega)$$

bezeichnet man die stetige Winkeländerung des komplexen Zeigers vom kritischen Punkt −1 zur Ortskurve $F_o(j\omega)$, wenn ω von 0 bis ∞ läuft, als $\Delta|\underline{F}(j\omega)$.

n_r sei die Anzahl der Pole des offenen Kreises rechts der imaginären Achse, n_a die Anzahl auf der imaginären Achse. Dann lautet das Nyquist-Kriterium:

Der geschlossene Kreis ist genau dann stabil, wenn

$$\Delta|\underline{F}(j\omega) = \left(n_r + \frac{n_a}{2}\right)\cdot\pi$$

gilt.

Ist der offene Kreis stabil, liegen also alle Pole des offenen Kreises links der imaginären Achse, so gilt $n_r = n_a = 0$. Der geschlossene Kreis ist also genau dann stabil, wenn $\Delta|\underline{F}(j\omega) = 0$ gilt. Für diesen Fall kann das Nyquist-Kriterium auch folgendermaßen formuliert werden [3]:

Ist der offene Kreis stabil, so ist der geschlossene Kreis genau dann stabil, wenn die Ortskurve des offenen Kreises den Punkt −1 weder umkreist noch durchdringt.

Bei einer einfachen Gestalt der Ortskurve kann folgende Formulierung angewandt werden:

Ist der offene Kreis stabil, so ist der geschlossene Kreis genau dann stabil, wenn der Punkt −1 links der Ortskurve des offenen Kreises liegt.

Zur Definition von „links“ ist die Ortskurve mit wachsender Frequenz ω zu durchlaufen.

Die zwei Formulierungen des Nyquist-Kriteriums für einen stabilen offenen Kreis können dahingehend erweitert werden, dass der offene Kreis neben den Polen links der imaginären Achse einen einfachen oder zweifachen Pol in $s = 0$ besitzen darf.

Alle genannten Formulierungen des Nyquist-Kriteriums gelten auch bei einer zusätzlichen Totzeit im offenen Kreis, also

$$G_o(s) = \frac{Z_o(s)}{N_o(s)} e^{-T_t\cdot s}.$$

Beispiele für die Anwendung des Nyquist-Kriteriums bei einem offenen Kreis mit Polen links der imaginären Achse sind in Abb. 73.4 dargestellt, bei einem offenen Kreis mit Polen links der imaginären Achse mit Ausnahme eines einfachen Poles in $s = 0$ in Abb. 73.5, sowie bei einem offenen Kreis mit Polen links der imaginären Achse mit Ausnahme eines zweifachen Poles in $s = 0$ in Abb. 73.6. Bei einer komplizierten Gestalt der Ortskurve ist es zweckmäßig, die erste Formulierung des Nyquist-Kriteriums, die auch als allgemeine Form bezeichnet wird, zu verwenden, und damit unklare Begriffe wie „umkreisen“ und „links“ zu vermeiden.

73.2.4 Phasenreserve und Amplitudenreserve

Interpretiert man das Nyquist-Kriterium unter Verwendung der Phasenreserve bzw. der Amplitudenreserve, welche aus dem Bode-Diagramm des offenen Kreises einfach abgelesen werden können, so erhält man zwei weitere Kriterien für die Stabilität des geschlossenen Kreises. Es sei der Fall

$$G_o(s) = K\cdot\frac{\prod_{j=1}^{m}(s-n_j)}{\prod_{i=1}^{n}(s-p_i)}$$

mit $K > 0$, $m < n$ und $\mathrm{Re}\{p_i\} < 0$, $i = 1,\ldots,n$, betrachtet. Weiterhin sei vorausgesetzt, dass die Ortskurve genau einen Schnittpunkt mit dem Einheitskreis um den Ursprung bzw. mit der negativen reellen Achse besitzt.

Abb. 73.7 zeigt zwei Ortskurven mit den zugehörigen Frequenzkennlinien. Im Fall a) liegt der kritische Punkt −1 links der

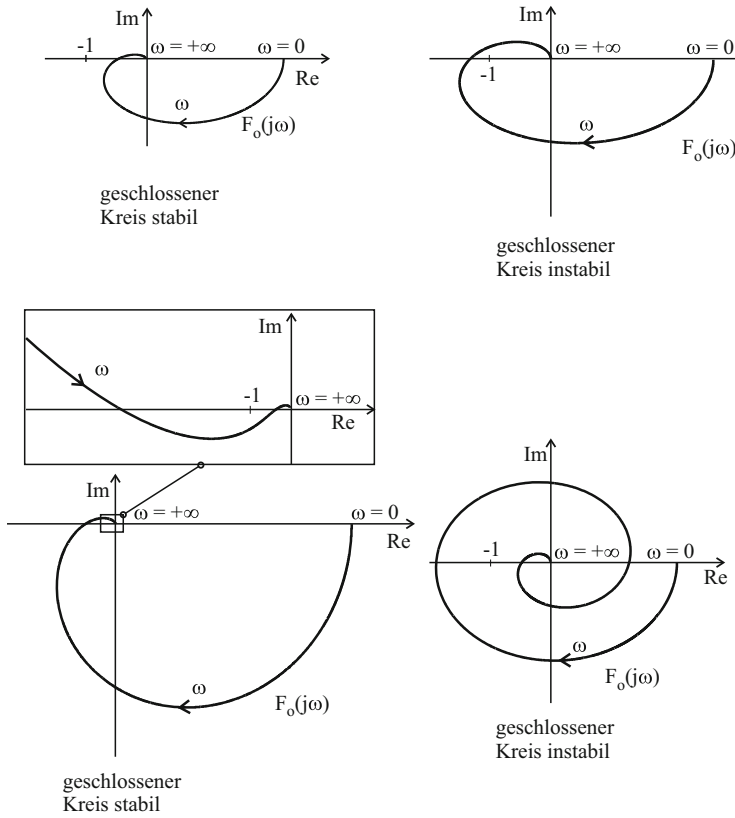


Abb. 73.4 Anwendung des Nyquist-Kriteriums bei einem offenen Kreis mit Polen links der imaginären Achse

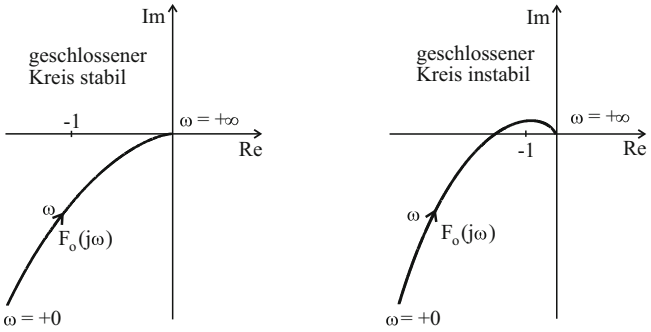


Abb. 73.5 Anwendung des Nyquist-Kriteriums bei einem offenen Kreis mit Polen links der imaginären Achse mit Ausnahme eines einfachen Poles in $s = 0$

Ortskurve des offenen Kreises, der geschlossene Kreis ist also stabil. Im Fall b) liegt der kritische Punkt -1 rechts der Ortskurve des offenen Kreises, der geschlossene Kreis ist also instabil. Betrachtet man den Schnittpunkt der Ortskurve mit dem Einheitskreis um den Ursprung, so ist die Phase im stabilen Fall größer als $-180°$, im instabilen Fall dagegen kleiner oder gleich $-180°$.

In der Frequenzkennliniendarstellung ist der Schnittpunkt der Ortskurve mit dem Einheitskreis durch die Bedingung

$$|F_o(j\omega_D)| = 1$$

bzw. in logarithmischer Darstellung durch

$$|F_o(j\omega_D)|_{\text{dB}} = 0$$

charakterisiert. Die zugehörige Frequenz ω_D wird als *Durchtrittsfrequenz* bezeichnet. Der geschlossene Kreis ist stabil für $|\underline{F_o}(j\omega_D) > -180°$ und instabil für $|\underline{F_o}(j\omega_D) \leq -180°$. Der Abstand von $|\underline{F_o}(j\omega_D)$ zu $-180°$ wird als *Phasenreserve* oder *Phasenrand*

$$\varphi_R = |\underline{F_o}(j\omega) + 180°$$

definiert. Anhand der Phasenreserve kann nun direkt das Stabilitätsverhalten des geschlossenen Kreises bestimmt werden:

> Der geschlossene Kreis ist genau dann stabil, wenn $\varphi_R > 0$ gilt.

Dabei sei zusätzlich vorausgesetzt, dass $-540° < |\underline{F_o}(j\omega) < 180°$ für $\omega < \omega_D$ erfüllt ist. Dieses Kriterium gilt auch dann, wenn $G_o(s)$ einen ein- oder zweifachen Pol in $s = 0$ besitzt und eine Totzeit enthält [3].

Betrachtet man den Schnittpunkt der Ortskurve mit der negativen reellen Achse, so ist der geschlossene Kreis stabil, wenn

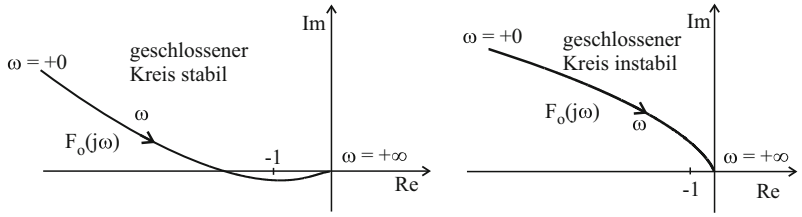


Abb. 73.6 Anwendung des Nyquist-Kriteriums bei einem offenen Kreis mit Polen links der imaginären Achse mit Ausnahme eines zweifachen Poles in $s = 0$

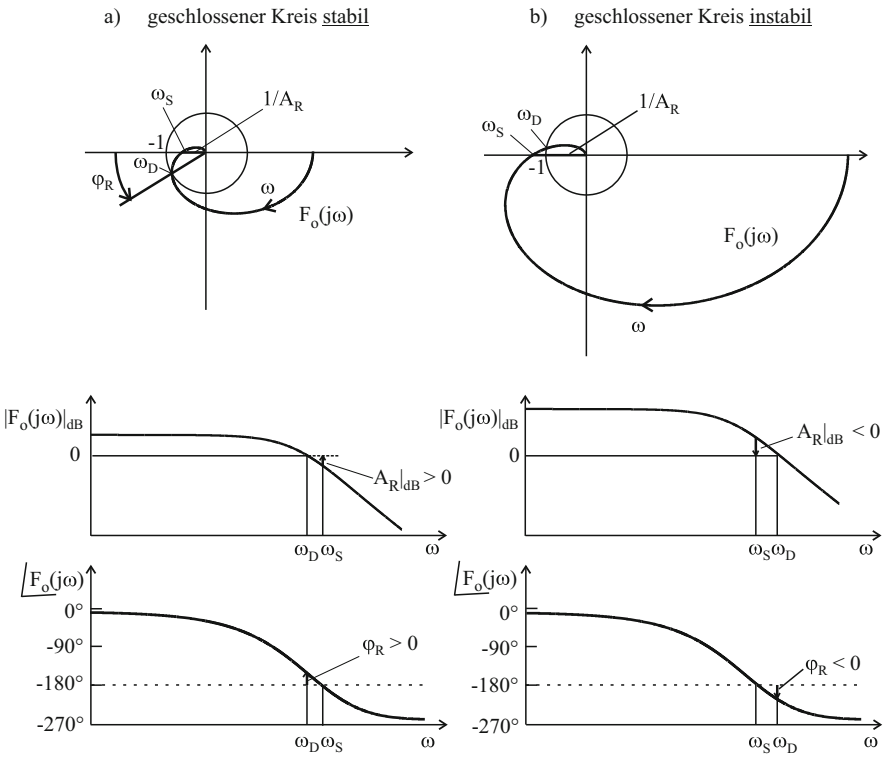


Abb. 73.7 Stabilitätsuntersuchung anhand der Frequenzkennlinien

der Betrag des Schnittpunkts kleiner als Eins ist, und instabil, wenn der Betrag größer oder gleich Eins ist. In der Frequenzkennliniendarstellung ist der Schnittpunkt der Ortskurve mit der negativen reellen Achse durch die Bedingung

$$\underline{|F_o(j\omega_S)} = -180^\circ$$

charakterisiert. Mit der zugehörigen Amplitude $|F_o(j\omega_s)|$ wird die *Amplitudenreserve* oder der *Amplitudenrand* als

$$A_R = \frac{1}{|F_o(j\omega_S)|}$$

definiert. Aus der Betragskennlinie kann die Amplitudenreserve in Dezibel gemäß

$$A_R\big|_{\mathrm{dB}} = -|F_o(j\omega_S)|_{\mathrm{dB}}$$

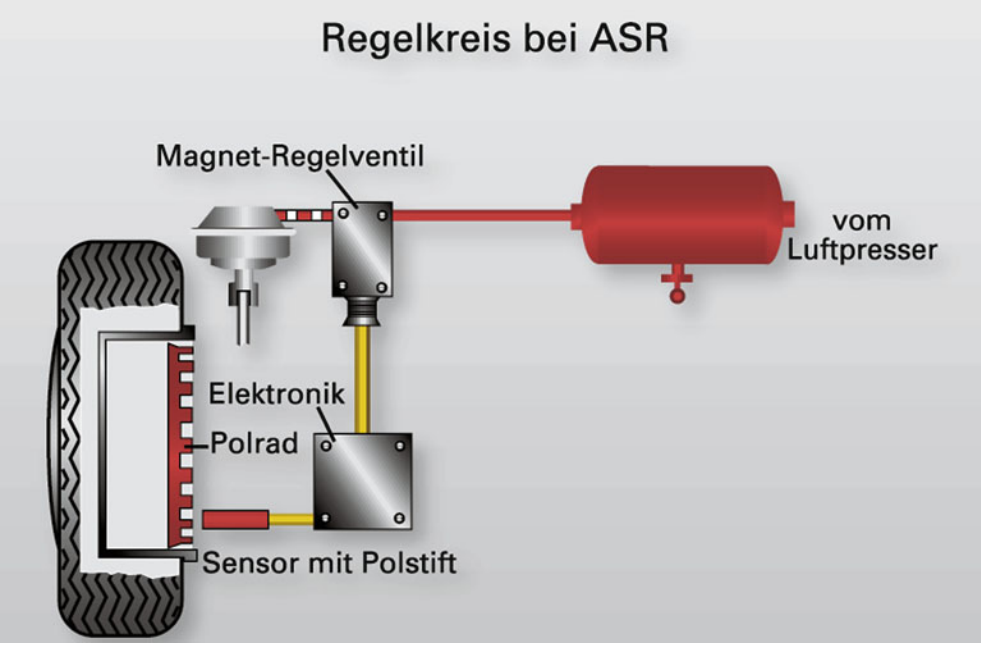
direkt abgelesen werden. Anhand der Betragsreserve in Dezibel kann nun direkt das Stabilitätsverhalten des geschlossenen Kreises bestimmt werden:

> Der geschlossene Kreis ist genau dann stabil, wenn $A_R\big|_{\mathrm{dB}} > 0$ gilt.

Entwurf von Regelkreisen

74

Joachim Horn



74.1 Forderungen an die Regelung 866
74.2 Reglertypen 867
74.3 Faustregeln für die Wahl der Reglerstruktur und der Reglerparameter 869
74.4 Reglerentwurf mit dem Frequenzkennlinienverfahren 870
74.5 Einstellregeln bei speziellen Regelstrecken 871
Literatur 872

Teil IX

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_74

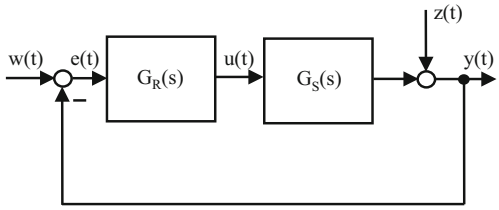


Abb. 74.1 Standardregelkreis

Betrachtet werde der Standardregelkreis gemäß Abb. 74.1. $G_S(s)$ bezeichnet die Übertragungsfunktion der Strecke, die bekannt sei. $G_R(s)$ bezeichnet die Übertragungsfunktion des Regelgliedes, dessen Struktur und Parameter so zu bestimmen sind, dass der geschlossene Regelkreis die gewünschten Eigenschaften besitzt. Zunächst sollen diese Forderungen an die Regelung dargestellt und die zu ihrer Erfüllung erforderlichen Bedingungen hergeleitet werden.

74.1 Forderungen an die Regelung

74.1.1 Stabilität

Die grundlegende Forderung nach Stabilität bedeutet, dass die Regelgröße $y(t)$ bei einer konstanten Führungsgröße $w(t) = w$ und einer konstanten Störgröße $z(t) = z$ asymptotisch gegen einen konstanten, endlichen Wert y_∞ strebt. Sie darf also insbesondere nicht gegen eine Dauerschwingung oder betragsmäßig ins Unendliche streben.

Die Laplace-Transformierte der Regelgröße $y(t)$ des Standardregelkreises ergibt sich zu

$$Y(s) = \frac{G_o(s)}{1 + G_o(s)} \cdot W(s) + \frac{1}{1 + G_o(s)} \cdot Z(s)$$

mit der Übertragungsfunktion des offenen Kreises

$$G_o(s) = G_R(s) \cdot G_S(s).$$

Die Führungsübertragungsfunktion

$$G_w(s) = \frac{G_R(s) \cdot G_S(s)}{1 + G_R(s) \cdot G_S(s)}$$

und der Störübertragungsfunktion

$$G_z(s) = \frac{1}{1 + G_R(s) \cdot G_S(s)}$$

besitzen dieselben Pole, nämlich die Lösungen der charakteristischen Gleichung

$$1 + G_R(s) \cdot G_S(s) = 0.$$

Gemäß dem grundlegenden Stabilitätskriterium muss daher die Übertragungsfunktion $G_R(s)$ des Regelgliedes so gewählt werden, dass diese Pole links der imaginären Achse liegen, damit der Regelkreis stabil ist.

74.1.2 Stationäre Genauigkeit

Die Regelgröße $y(t)$ eines stabilen Regelkreises strebt bei einer konstanten Führungsgröße $w(t) = w$ und einer konstanten Störgröße $z(t) = z$ gegen einen endlichen Wert y_∞. Der Regelkreis wird als *stationär genau* bezeichnet, wenn y_∞ gleich w ist, und zwar für beliebiges z. Ein stationär genauer Regelkreis ist somit durch

$$\lim_{t \to \infty} e(t) = 0$$

charakterisiert, die bleibende Regelabweichung ist also gleich Null.

Die Definition der stationären Genauigkeit kann auch auf nicht konstante Führungs- und Störgrößen, beispielsweise Rampen oder Sinusschwingungen, verallgemeinert werden.

Ein stabiler Regelkreis wird als *hinreichend stationär genau* bezeichnet, wenn

$$|\lim_{t \to \infty} e(t)| < \varepsilon$$

mit einem genügend kleinen $\varepsilon > 0$ für $w = 1$ erfüllt ist, wenn also der Betrag der bleibenden Regelabweichung hinreichend klein ist.

Betrachtet man das Führungsverhalten für $w(t) = w$, so erhält man mit

$$W(s) = \frac{w}{s}$$

und

$$Y(s) = \frac{G_o(s)}{1 + G_o(s)} \cdot W(s)$$

die Laplace-Transformierte der Regelabweichung $e(t)$ zu

$$E(s) = \frac{1}{1 + G_o(s)} \cdot \frac{w}{s}.$$

Mit dem Endwertsatz der Laplace-Transformation folgt daraus

$$\begin{aligned} \lim_{t \to \infty} e(t) &= \lim_{s \to 0} s \cdot E(s) \\ &= \lim_{s \to 0} \frac{1}{1 + G_o(s)} \cdot w. \end{aligned}$$

Besitzt die Übertragungsfunktion $G_o(s)$ des offenen Kreises einen Pol in $s = 0$, so folgt

$$\lim_{t \to \infty} e(t) = 0.$$

Ein stabiler Regelkreis mit einem I-Anteil in $G_o(s)$ ist also bezüglich des Führungsverhaltens stets stationär genau.

Besitzt der offene Kreis dagegen P-Verhalten mit $G_o(0) = K$, so folgt

$$\lim_{t \to \infty} e(t) = \frac{1}{K} \cdot w.$$

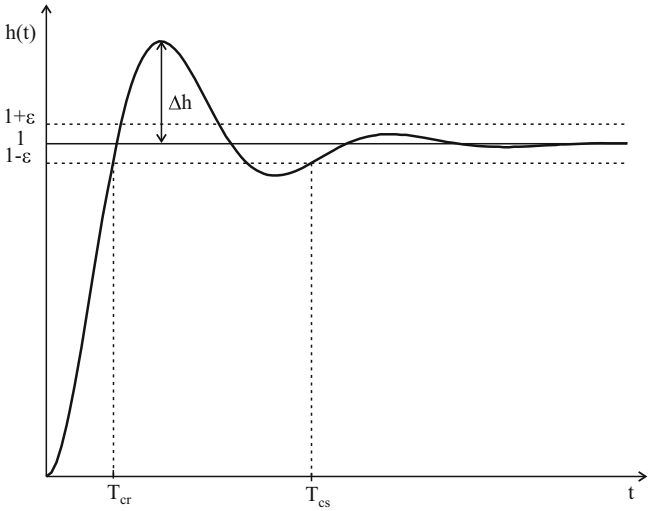


Abb. 74.2 Führungssprungantwort des Regelkreises und Kennwerte des dynamischen Verhaltens

Der Regelkreis ist somit nicht stationär genau. Um eine *hinreichende* stationäre Genauigkeit zu erzielen, muss die Verstärkung K genügend groß gewählt werden.

Betrachtet man das Störverhalten für $z(t) = z$, so erhält man dasselbe Ergebnis für die stationäre Genauigkeit. Betrachtet man einen Regelkreis, bei dem die Störung innerhalb der Strecke angreift, so bleibt das Ergebnis für die stationäre Genauigkeit unverändert, sofern die Störung *nach* dem I-Anteil eingreift.

Zusammengefasst ergibt sich:

> Ein stabiler Regelkreis ist genau dann stationär genau, wenn $G_o(s)$ einen I-Anteil enthält.

Falls die Strecke keinen I-Anteil besitzt, ist also ein Regler mit I-Anteil zu verwenden, um die stationäre Genauigkeit zu gewährleisten.

74.1.3 Hinreichende Dämpfung

Um ein zu starkes Über- oder Unterschwingen bei Führungs- oder Störgrößensprüngen zu vermeiden, muss der Regelkreis hinreichend gedämpft sein.

Anwendungsspezifisch kann eine zulässige Überschwingweite Δh, siehe Abb. 74.2, definiert werden. Für eine Vielzahl von Regelkreisen kann die Dämpfung anhand der in Abschn. 73.2.4 behandelten Phasenreserve bewertet werden. Eine hinreichende Dämpfung des Führungsverhaltens ist typischerweise für eine Phasenreserve von ca. 60° gewährleistet. Wird lediglich das Störverhalten betrachtet, so reicht meist eine Phasenreserve von 30° aus.

74.1.4 Hinreichende Schnelligkeit

Die Schnelligkeit des Regelkreises kann anhand verschiedener Kennwerte beurteilt werden. In Abb. 74.2 sind die Anregelzeit T_{cr} und die Ausregelzeit T_{cs} dargestellt:

Die Anregelzeit T_{cr} (control rise time) bezeichnet den Zeitpunkt, an dem die Sprungantwort $h(t)$ erstmals einen Toleranzstreifen der Breite 2ε um ihren Sollwert erreicht. Die Ausregelzeit T_{cs} (control settling time) bezeichnet den Zeitpunkt, ab dem die Sprungantwort vollständig in dem Toleranzstreifen der Breite 2ε verläuft, also

$$|h(t) - 1| \leq \varepsilon$$

für alle $t \geq T_{cs}$ gilt.

Die Schnelligkeit des Regelkreises kann anhand der in Abschn. 73.2.4 definierten Durchtrittsfrequenz ω_D bewertet werden, da der geschlossene Regelkreis näherungsweise durch ein Verzögerungsglied mit $T = \frac{1}{\omega_D}$ beschrieben werden kann [3]. Je größer die Durchtrittsfrequenz, umso schneller ist der Einschwingvorgang.

Als weitere Forderungen seien die Robustheit des Regelkreises gegenüber (kleinen) Variationen der Strecke gegenüber dem Nominalmodell sowie die hinreichende Unterdrückung des Messrauschens im Regelkreis genannt.

74.2 Reglertypen

Grundsätzlich kommt jede realisierbare Übertragungsfunktion $G(s)$ für den Regler in Frage. In der Praxis wird jedoch meist einer der in Abschn. 74.2.1 beschriebenen Standardregler eingesetzt. Die in Abschn. 74.2.2 dargestellten Korrekturglieder werden in der Regel für anspruchsvolle Regelungsaufgaben eingesetzt.

74.2.1 Standardregler

Einen PID-Regler erhält man gemäß Abb. 74.3 durch Parallelschaltung eines Proportional-, eines Integral- und eines

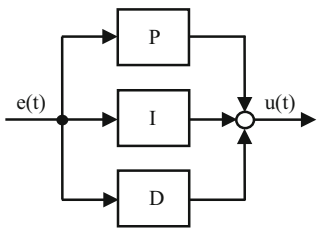


Abb. 74.3 Struktur eines PID-Reglers

Tab. 74.1 Auswahl von Standardreglern

Reglertyp	Übertragungsfunktion $G_R(s)$
P-Regler	K_R
PI-Regler	$K_R \cdot \frac{1+T_R \cdot s}{s}$
Realer PD-Regler	$K_R \cdot \frac{1+T_R \cdot s}{1+T_N \cdot s}$, $T_N \ll T_R$
Realer PID-Regler	$K_R \cdot \frac{(1+T_{R1} \cdot s) \cdot (1+T_{R2} \cdot s)}{s(1+T_N \cdot s)}$, $T_N \ll T_{R1}, T_{R2}$

Differenzialanteils. Der Zusammenhang zwischen der Regeldifferenz e und der Störgröße u wird im Zeitbereich durch

$$u(t) = K_P \cdot e(t) + K_I \cdot \int_0^t e(\tau) d\tau + K_D \cdot \frac{de(t)}{dt}$$

beschrieben. Aufgrund des P-Anteils führt eine große Regeldifferenz zu einer großen Stellgröße, aufgrund des I-Anteils führt eine anhaltende Regeldifferenz zu einem zusätzlichen Beitrag zur Stellgröße, und aufgrund des D-Anteils führt eine Änderung der Regeldifferenz zu einem zusätzlichen Beitrag zur Stellgröße. Die Übertragungsfunktion des PID-Reglers lautet

$$G_R(s) = K_p + \frac{K_I}{s} + K_D \cdot s.$$

Häufig wird diese Übertragungsfunktion in der Form

$$G_R(s) = K_P \cdot (1 + \frac{1}{T_I \cdot s} + T_D \cdot s)$$

dargestellt, wobei K_P als *Proportionalbeiwert* bezeichnet wird, $T_I = \frac{K_P}{K_I}$ als *Nachstellzeit* und $T_D = \frac{K_D}{K_P}$ als *Vorhaltezeit*.

Lässt man Anteile des PID-Reglers weg, so erhält man beispielsweise einen P-Regler, einen I-Regler, einen PI-Regler oder einen PD-Regler.

Für die Bestimmung geeigneter Reglerparameter ist die Darstellung der Übertragungsfunktion des PID-Reglers gemäß

$$G_R(s) = K_R \cdot \frac{(1 + T_{R1} \cdot s) \cdot (1 + T_{R2} \cdot s)}{s}$$

oder

$$G_R(s) = K_R \cdot \frac{1 + 2 \cdot d_R \cdot T_R \cdot s + T_R^2 \cdot s^2}{s}$$

mit $0 < d_R < 1$ besser geeignet, siehe Abschn. 74.3.

Eine Übertragungsfunktion mit einem Zählergrad größer als der Nennergrad ist weder realisierbar noch als Regler geeignet, da hochfrequentes Messrauschen die Stellgröße stark beeinflussen würde. Daher wird die „ideale“ Übertragungsfunktion eines Reglers mit D-Anteil im Nenner um den Faktor $(1 + T_N \cdot s)$ mit hinreichend kleinem T_N ergänzt. So erhält man beispielsweise aus dem „idealen“ PD-Regler den „realen“ PD-Regler (Tab. 74.1).

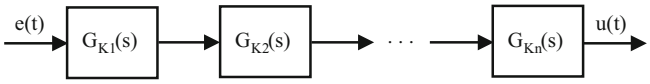


Abb. 74.4 Reglerstruktur als Reihenschaltung von Korrekturgliedern

74.2.2 Korrekturglieder

Einen Regler kann man gemäß Abb. 74.4 als Reihenschaltung von Korrekturgliedern erhalten. Die Auswahl geeigneter Korrekturglieder und die Bestimmung geeigneter Parameter erfolgt unter Verwendung des Bode-Diagramms der Strecke. Nachfolgend werden die wichtigsten Korrekturglieder vorgestellt.

74.2.2.1 Proportionalglied (P-Glied)

Das Einfügen eines Proportionalglieds mit der Übertragungsfunktion

$$G_K(s) = K, \quad K > 0,$$

und den Frequenzkennlinien gemäß Abb. 72.2 verschiebt die Betragskennlinie des offenen Kreises um den Wert

$$K\big|_{\text{dB}} = 20 \cdot \log K,$$

die Phasenkennlinie des offenen Kreises bleibt unverändert. Durch geeignete Wahl von K kann beispielsweise die gewünschte Phasenreserve eingestellt werden. Ist hierfür eine Verschiebung der Betragskennlinie um $\Delta A\big|_{\text{dB}}$ erforderlich, dann ist K gemäß

$$K = 10^{\frac{\Delta A|_{\text{dB}}}{20}}$$

zu wählen.

74.2.2.2 Integrierglied (I-Glied)

Das Einfügen eines Integrierglieds mit der Übertragungsfunktion

$$G_K(s) = \frac{1}{s}$$

und den Frequenzkennlinien gemäß Abb. 72.3 ist zur Gewährleistung der stationären Genauigkeit erforderlich, sofern die Strecke kein Integrierglied enthält, wie bereits in Abschn. 74.1.2 dargelegt.

74.2.2.3 Lead-Glied

Das Lead-Glied oder phasenanhebende Korrekturglied besitzt die Übertragungsfunktion

$$G_K(s) = \frac{1 + T_Z \cdot s}{1 + T_N \cdot s}, \quad T_Z > T_N.$$

Die Frequenzkennlinien sind in Abb. 74.5 für den Fall $T_Z = 10 \cdot T_N$ dargestellt.

Teil IX

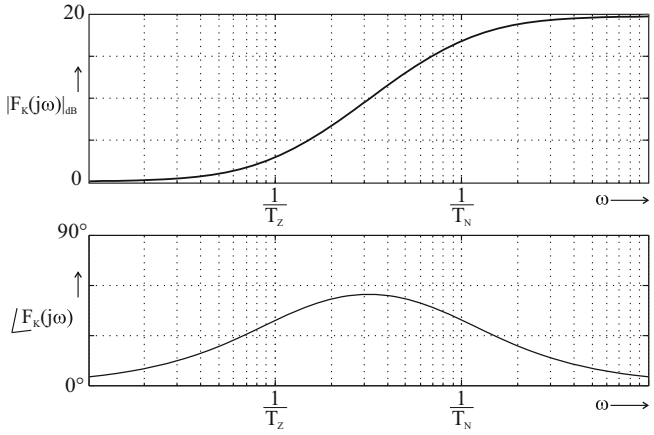


Abb. 74.5 Frequenzkennlinien von $F_K(j\omega) = \frac{1+j\omega \cdot T_Z}{1+j\omega \cdot T_N}$ für $T_Z = 10T_N$

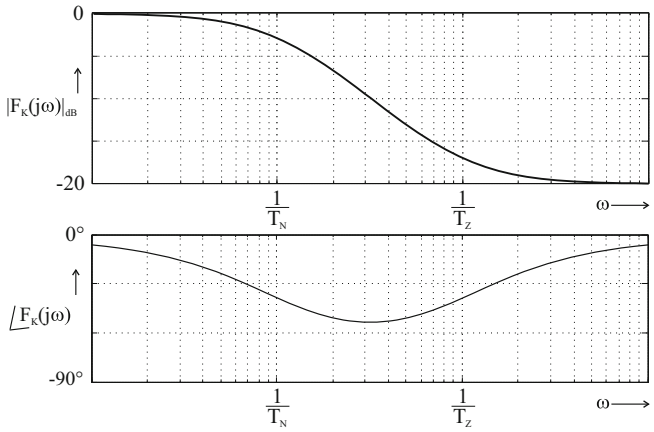


Abb. 74.6 Frequenzkennlinien von $F_K(j\omega) = \frac{1+j\omega \cdot T_Z}{1+j\omega \cdot T_N}$ für $T_Z = \frac{T_N}{10}$

Die maximale Phase tritt bei der Mittenfrequenz

$$\omega_M = \sqrt{\frac{1}{T_Z \cdot T_N}} \tag{74.1}$$

auf und beträgt

$$|\underline{F_K}(j\omega_M) = \arcsin \frac{\frac{T_Z}{T_N} - 1}{\frac{T_Z}{T_N} + 1}. \tag{74.2}$$

Mit dem Lead-Glied kann die Phasenkennlinie im gewünschten Frequenzbereich gezielt erhöht und damit die erforderliche Phasenreserve eingestellt werden.

74.2.2.4 Lag-Glied

Das Lag-Glied oder phasenabsenkende Korrekturglied besitzt die Übertragungsfunktion

$$G_K(s) = \frac{1 + T_Z \cdot s}{1 + T_N \cdot s}, \quad T_Z < T_N.$$

Die Frequenzkennlinien sind in Abb. 74.6 für den Fall $T_Z = \frac{T_N}{10}$ dargestellt.

Für $\omega \gg \frac{1}{T_Z}$ beträgt der Betrag in Dezibel näherungsweise

$$|F_K(j\omega)|_{\text{dB}} = 20 \cdot \log \cdot \frac{T_Z}{T_N}.$$

Mit dem Lag-Glied kann die Betragskennlinie für $\omega \gg \frac{1}{T_Z}$ gezielt abgesenkt und damit die gewünschte Betragsreserve eingestellt werden, ohne die Verstärkung des offenen Kreises zu reduzieren und damit die stationäre Regelabweichung bei einem offenen Kreis mit P-Verhalten zu erhöhen.

74.3 Faustregeln für die Wahl der Reglerstruktur und der Reglerparameter

Bei einer Strecke ohne *I*-Anteil empfiehlt sich die Verwendung eines Reglers mit *I*-Anteil, beispielsweise eines PI-Reglers, um stationäre Genauigkeit zu erzielen. Allerdings kann die Phasenabsenkung um 90° aufgrund des *I*-Anteils die Einstellung einer hinreichenden Phasenreserve erschweren. Daher sollte bei einer Strecke mit *I*-Anteil ein Regler ohne *I*-Anteil, beispielsweise ein realer PD-Regler, verwendet werden. Greift jedoch eine Störgröße *vor* dem *I*-Anteil der Strecke ein, so ist ein Regler mit *I*-Anteil für die stationäre Genauigkeit erforderlich. Besitzt die Strecke ein Doppel-I-Verhalten, also zwei *I*-Anteile, so wird meist eine Phasenanhebung, beispielsweise durch einen realen PD-Regler oder ein Lead-Glied, zur Stabilisierung des geschlossenen Kreises erforderlich sein.

Um eine hohe Dynamik des geschlossenen Kreises zu erzielen, ist es häufig sinnvoll, die langsamen Streckenanteile, d. h. die am weitesten rechts liegenden Streckenpole links der imaginären Achse, durch entsprechende Nullstellen des Reglers zu kompensieren. Hierzu wählt man beispielsweise bei einem PI-Regler T_R gleich der größten Nennerzeitkonstante der Strecke. Treten jedoch mehrere große Streckenzeitkonstanten auf, so empfiehlt es sich, T_R gleich der Summenzeitkonstante $T_\Sigma = \sum_{i=1}^{n} T_i$, d. h. der Summe der Zeitkonstanten im Nenner der Streckenübertragungsfunktion, zu wählen. Bei Verwendung eines realen PID-Reglers können zwei Pole der Strecke kompensiert werden, also entweder zwei reelle Pole oder ein konjugiert komplexes Polpaar. Besitzt die Strecke bereits Nullstellen in der Nähe von Polen, so tritt eine näherungsweise Pol-Nullstellen-Kompensation auf, die Pole sind also für das Übertragungsverhalten der Strecke von untergeordneter Bedeutung und müssen daher nicht durch den Regler kompensiert werden. Streckenpole auf oder rechts der imaginären Achse können in der Praxis nicht durch Nullstellen des Reglers kompensiert werden, da die Lage der Pole nicht exakt bekannt ist und auch durch Alterung oder Verschleiß der Strecke variiert. Durch Wahl der Reglerverstärkung K_R kann die Verstärkung des offenen Kreises so bestimmt werden, dass beispielsweise die geforderte Phasenreserve oder die hinreichende stationäre Genauigkeit bei einem offenen Kreis ohne *I*-Anteil gewährleistet wird.

74.4 Reglerentwurf mit dem Frequenzkennlinienverfahren

Der Reglerentwurf mit dem Frequenzkennlinienverfahren wird im Folgenden an zwei Beispielen vorgestellt. Das erste Beispiel behandelt den Entwurf eines PI-Reglers für ein Verzögerungssystem mit reellen Polen. Im zweiten Beispiel wird für eine doppelt integrierende Strecke ein Regler als Reihenschaltung geeigneter Korrekturglieder entworfen.

74.4.1 Entwurf eines PI-Reglers für ein Verzögerungssystem

Betrachtet werde ein Verzögerungssystem, also ein System, dessen Übertragungsfunktion keine Nullstellen besitzt. Die Polstellen seien alle reell, und die Übertragungsfunktion laute

$$G_S(s) = \frac{10}{(1+0{,}1\cdot s)\cdot(1+s)\cdot(1+10\cdot s)}.$$

Der „langsamste“ Dynamikanteil der Strecke wird durch die größte Streckenzeitkonstante $T_{S3} = 10$ beschrieben, also durch den Pol bei $s = -1/T_{S3} = -0{,}1$. Wählt man als Reglertyp einen PI-Regler mit der Übertragungsfunktion

$$G_R(s) = K_R \cdot \frac{1+T_R\cdot s}{s},$$

so kann dieser Streckenpol durch die Reglernullstelle kompensiert werden, indem $T_R = 10$ gewählt wird. Die Übertragungsfunktion des offenen Kreises lautet dann

$$G_o(s) = G_R(s)\cdot G_S(s) = \frac{10\cdot K_R}{s\cdot(1+0{,}1\cdot s)\cdot(1+s)}.$$

Aufgrund des I-Anteils des Reglers ist der geschlossene Kreis stationär genau. Der Parameter K_R wird nun so gewählt, dass die Phasenreserve 60° beträgt und somit eine hinreichende Dämpfung gewährleistet ist.

Abb. 74.7 zeigt die Frequenzkennlinien des offenen Kreises für $K_R = 1$. Zu einer Phase von $-120°$ liest man den Betrag von ca. 24,8 dB ab. Eine Phasenreserve von 60° erhält man also, wenn die Betragskennlinie um 24,8 dB nach unten verschoben wird. Hierzu muss K_R gemäß

$$K_R\big|_{\mathrm{dB}} = -24{,}8$$

gewählt werden, d. h.

$$K_R = 10^{-24{,}8/20} \approx 0{,}0575.$$

Abb. 74.8 zeigt die Sprungantwort des geschlossenen Kreises für die Reglerparameter $K_R = 0{,}0575$, $T_R = 10$.

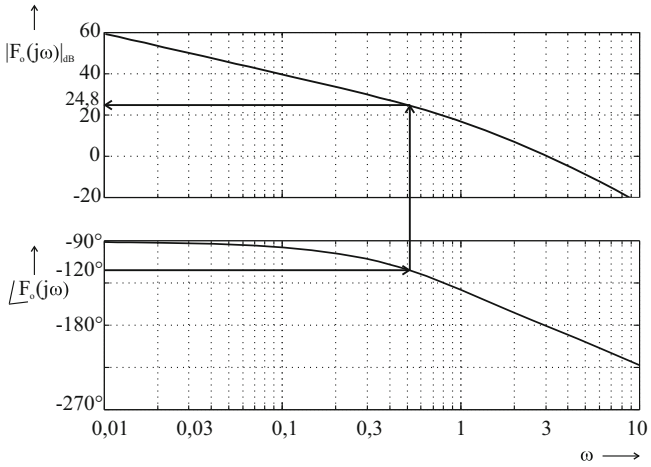


Abb. 74.7 Frequenzkennlinien von $F_0(j\omega) = \frac{10}{j\omega\cdot(1+j\frac{\omega}{10})\cdot(1+j\omega)}$

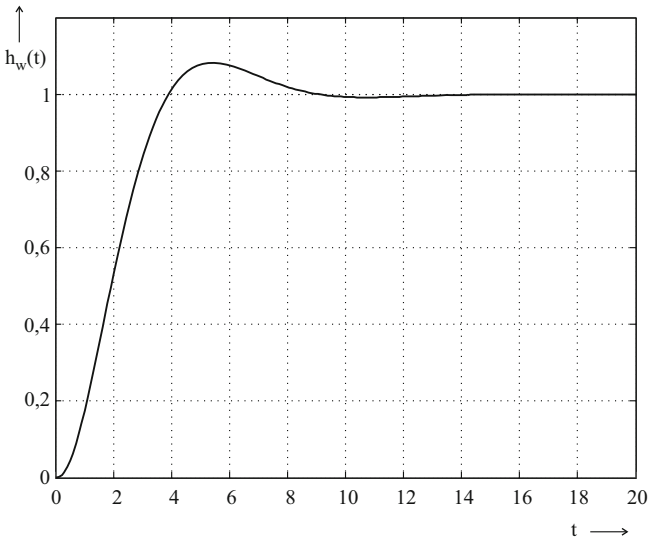


Abb. 74.8 Sprungantwort des geschlossenen Kreises für $G_0(s) = \frac{0{,}575}{s\cdot(1+0{,}1\cdot s)(1+s)}$

74.4.2 Reglerentwurf für eine doppelt integrierende Strecke

Betrachtet werde eine doppelt integrierende Strecke mit der Übertragungsfunktion

$$G_S(s) = \frac{1}{s^2},$$

für die ein Regler als Reihenschaltung geeigneter Korrekturglieder entworfen werden soll. Um eine hinreichende Dämpfung und eine hinreichende Schnelligkeit zu gewährleisten, sei eine Phasenreserve von $\varphi_R = 60°$ bei einer Durchtrittsfrequenz von $\omega_D = 10$ gefordert. Da die Phase der Strecke konstant $-180°$ ist, muss eine Phasenanhebung von 60° für $\omega = 10$ erfolgen. Hierfür wird ein Lead-Glied gemäß Abschn. 74.2.2.3 mit $\omega_M = 10$

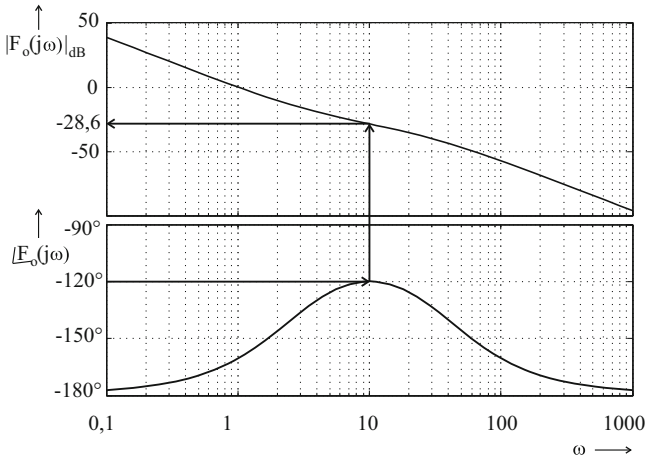


Abb. 74.9 Frequenzkennlinien von $F_0(j\omega) = \frac{1+j\omega\cdot 0{,}373}{(j\omega)^2\cdot(1+j\omega\cdot 0{,}0268)}$

und $|\underline{F_K}(j\omega_M) = 60°$ eingesetzt. Aus (74.1) folgt

$$T_Z \cdot T_N = \frac{1}{\omega_M^2}$$

und aus (74.2)

$$T_Z \cdot T_N = \frac{1+\sin|\underline{F_K}(j\omega_M)}{1-\sin|\underline{F_K}(j\omega_M)}.$$

Damit erhält man die Parameter des Lead-Glieds zu

$$T_Z = \frac{1}{\omega_M}\cdot\sqrt{\frac{1+\sin|\underline{F_K}(j\omega_M)}{1-\sin|\underline{F_K}(j\omega_M)}},$$

$$T_N = \frac{1}{\omega_M}\cdot\sqrt{\frac{1-\sin|\underline{F_K}(j\omega_M)}{1+\sin|\underline{F_K}(j\omega_M)}}.$$

Für das betrachtete Beispiel ergibt sich

$$T_Z = 0{,}373,$$
$$T_N = 0{,}0268\,.$$

Abb. 74.9 zeigt die Frequenzkennlinien des offenen Kreises bei Verwendung dieses Lead-Glieds, d. h. für die Übertragungsfunktion des offenen Kreises gemäß

$$G_o(s) = \frac{1+0{,}373\cdot s}{1+0{,}0268\cdot s}\cdot\frac{1}{s^2}.$$

Damit die Durchtrittsfrequenz wie gefordert $\omega_D = 10$ beträgt, muss die Betragskennlinie durch ein in Reihe geschaltetes Proportionalglied um ca. 28,6 dB angehoben werden. Aus

$$K\big|_{\text{dB}} = 28{,}6$$

folgt

$$K = 10^{28{,}6/20} \approx 26{,}9.$$

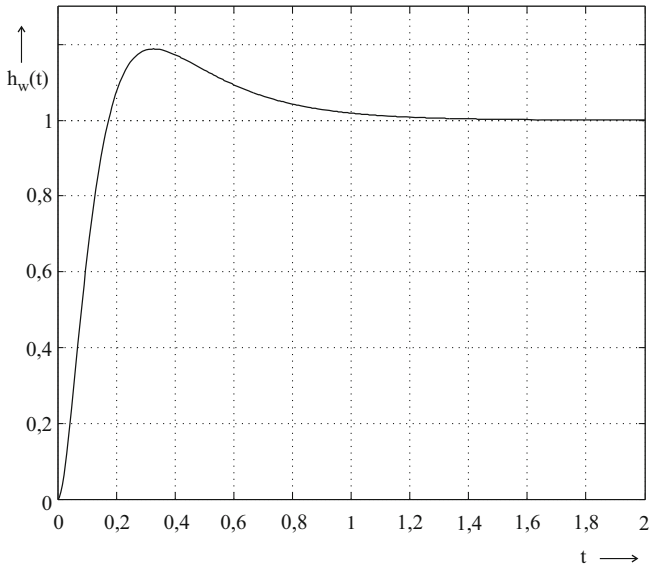


Abb. 74.10 Sprungantwort des geschlossenen Kreise für $G_0(s) = 26{,}9\cdot\frac{1+0{,}373\cdot s}{1+0{,}0268\cdot s}\cdot\frac{1}{s^2}$

Der Regler ergibt sich als Reihenschaltung eines Proportionalglieds und eines Lead-Glieds, die Übertragungsfunktion des Reglers lautet

$$G_R(s) = K\frac{1+T_Z\cdot s}{1+T_N\cdot s}$$

mit $T_N \ll T_Z$. Es handelt sich also um einen realen PD-Regler. Abb. 74.10 zeigt die Sprungantwort des geschlossenen Kreises.

74.5 Einstellregeln bei speziellen Regelstrecken

Für spezielle Regelstrecken wurden Einstellregeln entwickelt, die eine direkte Berechnung der Reglerparameter aus den Streckenparametern ermöglichen. Im Folgenden werden zwei Einstellregeln, das Betragsoptimum und das symmetrische Optimum, vorgestellt.

74.5.1 Betragsoptimum

Betrachtet wird ein Verzögerungssystem der Ordnung n mit ausschließlich reellen Polen, die Übertragungsfunktion der Strecke lautet also

$$G_S(s) = \frac{K_S}{\prod_{i=1}^{n}(1+T_i\cdot s)}.$$

In Tab. 74.2 sind für zwei Fälle die Einstellregeln des Betragsoptimums angegeben: im Falle einer großen Zeitkonstante die Einstellregeln für die Parameter eines PI-Reglers und im Falle zweier großen Zeitkonstanten die Einstellregeln für die

Tab. 74.2 Einstellregeln des Betragsoptimums

Strecke	Regler	Reglerparameter
$G_S(s) = \frac{K_S}{\prod_{i=1}^{n}(1+T_i)}$, $T_1 \gg \sum_{i=2}^{n} T_i$	$G_R(s) = K_R \cdot \frac{1+T_R \cdot s}{s}$, PI-Regler	$K_R = \frac{1}{2 \cdot K_S \cdot \sum_{i=2}^{n} T_i}$, $T_R = T_1$
$G_S(s) = \frac{K_S}{\prod_{i=1}^{n}(1+T_i \cdot s)}$, $T_1, T_2 \gg \sum_{i=3}^{n} T_i$	$G_R(s) = K_R \cdot \frac{(1+T_{R1} \cdot s) \cdot (1+T_{R2} \cdot s)}{s}$, idealer PID-Regler	$K_R = \frac{1}{2 \cdot K_S \cdot \sum_{i=3}^{n} T_i}$, $T_{R1} = T_1$, $T_{R2} = T_2$

Tab. 74.3 Einstellregeln des symmetrischen Optimums

Strecke	Regler	Reglerparameter
Allgemeiner Fall $G_S(s) = \frac{K_S}{\prod_{i=1}^{k}(1+T_i \cdot s) \cdot \prod_{j=1}^{l}(1+\tau_j \cdot s)}$ $T_1, \ldots, T_k \gg \tau_\Sigma = \sum_{j=1}^{l} \tau_j$	$G_R(s) = K_R \cdot \frac{(1+T_R \cdot s)^k}{s}$	$K_R = \frac{1}{2 \cdot K_S \cdot \tau_\Sigma} \cdot \frac{\prod_{i=1}^{k} T_i}{(4 \cdot k \cdot \tau_\Sigma)^k}$ $T_R = 4 \cdot k \cdot \tau_\Sigma$
Eine große Zeitkonstante $G_S(s) = \frac{K_S}{(1+T_1 \cdot s) \cdot \prod_{j=1}^{l}(1+\tau_j \cdot s)}$ $T_1 \gg \tau_\Sigma = \sum_{j=1}^{l} \tau_j$	$G_R(s) = K_R \cdot \frac{1+T_R \cdot s}{s}$ PI-Regler	$K_R = \frac{T_1}{8 \cdot K_S \cdot \tau_\Sigma^2}$ $T_R = 4 \cdot \tau_\Sigma$
Zwei große Zeitkonstanten $G_S(s) = \frac{K_S}{(1+T_1 \cdot s) \cdot (1+T_2 \cdot s) \cdot \prod_{j=1}^{l} 1+\tau_j \cdot s}$ $T_1, T_2 \gg \tau_\Sigma = \sum_{j=1}^{l} \tau_j$	$G_R(s) = K_R \cdot \frac{(1+T_R \cdot s)^2}{s}$ idealer PID-Regler	$K_R = \frac{T_1 \cdot T_2}{128 \cdot K_S \cdot \tau_\Sigma^3}$ $T_R = 8 \cdot \tau_\Sigma$

Teil IX

Parameter eines PID-Reglers. In [3] ist die Herleitung der Einstellregeln des Betragsoptimums angegeben, außerdem werden weitere Fälle betrachtet.

Im Falle einer großen Zeitkonstante T_1 wird der Streckenpol bei $-\frac{1}{T_1}$ durch die Reglernullstelle des PI-Reglers bei $-\frac{1}{T_R}$ kompensiert. Entsprechend werden im Falle zweier großen Zeitkonstanten T_1 und T_2 die Streckenpole bei $-\frac{1}{T_1}$ und $-\frac{1}{T_2}$ durch die Reglernullstellen des PID-Reglers bei $-\frac{1}{T_{R1}}$ und $-\frac{1}{T_{R2}}$ kompensiert.

74.5.2 Symmetrisches Optimum

Betrachtet wird wieder ein Verzögerungssystem mit ausschließlich reellen Polen, wobei die Zeitkonstanten in eine Gruppe großer Zeitkonstanten $T_1, \ldots, T_K$ und eine Gruppe kleiner Zeitkonstanten $\tau_1, \ldots, \tau_l$ mit

$$T_1, \ldots, T_K \gg \tau_\Sigma = \sum_{j=1}^{l} \tau_j$$

eingeteilt werden können. Die Übertragungsfunktion der Strecke lautet also

$$G_S(s) = \frac{K_S}{\prod_{i=1}^{k}(1 + T_i \cdot s) \cdot \prod_{j=1}^{l}(1 + \tau_j \cdot s)}.$$

Der Regler wird gemäß

$$G_R(s) = K_R \cdot \frac{(1 + T_R \cdot s)^k}{s}$$

angesetzt, der Zählergrad ist also gleich der Anzahl der großen Zeitkonstanten. In Tab. 74.3 sind die Einstellregeln des symmetrischen Optimums für ein allgemeines k, für den Fall einer großen Zeitkonstanten, d. h. $k = 1$, und den Fall zweier großer Zeitkonstanten, $k = 2$, angegeben. Für $k = 1$ erhält man einen PI-Regler, für $k = 2$ einen idealen PID-Regler. Die Herleitung dieser Einstellregeln des symmetrischen Optimums findet sich in [3].

Im Gegensatz zum Betragsoptimum findet beim symmetrischen Optimum keine Kompensation von Polstellen der Strecke durch Nullstellen des Reglers statt.

Literatur

1. DIN IEC 60050-351. Beuth. 2009.
2. *Dorf, R. C., Bishop, R. H.:* Moderne Regelungssysteme. Pearson Studium. 10., überarbeitete Auflage 2007.
3. *Föllinger, O.:* Regelungstechnik: Einführung in die Methoden und ihre Anwendung. Hüthig. 10., durchgesehene Auflage 2008.
4. *Horn, M., Dourdoumas, N.:* Regelungstechnik. Pearson Studium. 2004.
5. *Lunze, J.:* Regelungstechnik 1: Systemtheoretische Grundlagen, Analyse und Entwurf einschleifiger Regelungen. Springer. 8. Auflage 2010.
6. *Lutz, H., Wendt, W.:* Taschenbuch der Regelungstechnik: mit MATLAB und Simulink. Harry Deutsch. 8., ergänzte Auflage 2010.

Messtechnik

X



75 Grundlagen und Grundbegriffe der Messtechnik 875
76 Analog anzeigende Messgeräte 885
77 Oszilloskop 889
78 Schreibende Messgeräte 895
79 Digital anzeigende Messgeräte 897
80 Messverfahren zur Messung elektrischer Größen 901
81 Messverfahren zur Messung nichtelektrischer Größen 921
82 Messdatenaufbereitung 943
83 Bussysteme für die Messtechnik 947
84 Probleme bei der Digitalisierung analoger Messwerte 953
85 PC-gestützte Messverfahren und Messsignalanalyse 957

Grundlagen und Grundbegriffe der Messtechnik

75

Wilfried Plaßmann



75.1 Begriffe 876
75.2 Einheiten 876
75.3 Messabweichung, Messfehler 877
75.4 Abweichungsfortpflanzung, Fehlerfortpflanzung 879
75.5 Fehlerangaben von Messgeräten 881
75.6 Arithmetischer Mittelwert und Effektivwert von Wechselgrößen . . 882
75.7 Häufigkeitsverteilung, Vertrauensbereich 882
75.8 Bearbeitung und Auswertung von Messwerten 883

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_75

75.1 Begriffe

Es ist eine wesentliche Aufgabe der Messtechnik, technische Vorgänge quantitativ zu erfassen und anhand der gemessenen Größen Funktionsabläufe zu steuern. Als Beispiel sei ein Kraftwerk zur Energieerzeugung genannt, bei dem nur über die Messung von Temperaturen, Leistungen, Drücken und anderen Größen Aussagen über den momentanen Zustand möglich sind und bei Abweichungen vom Sollwert geeignete Eingriffe in das System erfolgen können. Damit eine eindeutige Kommunikation möglich wird, sind die in der Messtechnik verwendeten Begriffe, Messverfahren und Maßeinheiten in entsprechenden Normen oder Vorschriften festgelegt:

DIN 1301	Einheiten
DIN 1304	Allgemeine Formelzeichen
DIN 1313	Physikalische Größen und Gleichungen
DIN 1319	Grundbegriffe der Messtechnik
VDI/VDE 2600	Metrologie (Messtechnik)

Einige häufig verwendete Begriffe aus der Messtechnik sind im folgenden auszugsweise und zum Teil gekürzt nach DIN 1319 oder VDI/VDE 2600 wiedergegeben:

Messen ist der experimentelle Vorgang, durch den ein spezieller Wert einer physikalischen Größe als Vielfaches einer Einheit oder eines Bezugswerts ermittelt wird.

Messwert: Gemessener spezieller Wert einer Messgröße.

Messergebnis ist ein aus mehreren Messwerten einer physikalischen Größe oder aus Messwerten für verschiedene Größen nach einer festgelegten Beziehung ermittelter Wert oder Werteverlauf.

Messprinzip: Charakteristische physikalische Erscheinung, die bei der Messung benutzt wird.

Beispiele: Temperaturmessung; Änderung des elektrischen Widerstandes eines metallischen Leiters durch Temperaturänderung.

Messverfahren: Spezielle Art der Anwendung eines Messprinzips.

Beispiel: Abgleich einer Instrumentenanzeige auf Null. Wird z. B. bei Brückenschaltungen als *Nullabgleichverfahren* bezeichnet.

Messgerät: Liefert oder verkörpert Messwerte.

Messbereich: Bereich von Werten des Eingangssignals eines Messgerätes, der entsprechend der Kennlinie dieses Messgerätes eindeutig und innerhalb vorgegebener Fehlergrenzen durch Werte des Ausgangssignals abgebildet wird.

Empfindlichkeit: Verhältnis einer Änderung der Ausgangsgröße zu der sie verursachenden Eingangsgröße.

Kalibrieren: Ermitteln des gültigen Zusammenhanges zwischen dem Messwert oder dem Wert des Ausgangssignals und dem konventionell richtigen Wert der Messgröße.

Justieren: Einstellen oder Abgleichen eines Messgerätes mit dem Ziel, die Anzeige des Messgerätes möglichst nahe an den richtigen Wert der Messgröße anzugleichen.

Eichen: Von einer Eichbehörde nach den gesetzlichen Vorschriften und Anforderungen vorzunehmende Prüfung und Stempelung von Messgeräten.

Beispiele: Elektrizitätszähler, Waagen, Zapfsäulen.

Messumformer: Messgerät, das ein analoges Eingangssignal in ein eindeutig damit zusammenhängendes analoges Ausgangssignal umformt.

Beispiel: Temperaturmessung; eine Temperaturänderung wird in eine Widerstandsänderung umgeformt.

Messwandler: Am Ein- und Ausgang tritt die gleiche physikalische Größe auf; es wird keine Hilfsenergie benötigt.

Beispiele: Spannungswandler, Stromwandler.

Messumsetzer: Die Signalstruktur von Ein- und Ausgang ist entweder unterschiedlich (analog-digital oder digital-analog) oder nur digital.

Analoge Messverfahren: Der Messgröße wird ein Signal zugeordnet, das mindestens im Idealfall eine eindeutig umkehrbare Abbildung der Messgröße ist (häufig Skalenanzeige).

Digitale Messverfahren: Der Messgröße wird ein Signal zugeordnet, das eine mit fest gegebenen Schritten quantisierte Abbildung der Messgröße ist (häufig Ziffernanzeige).

75.2 Einheiten

Messen heißt „Vergleichen“. Die zu messende Größe wird damit als Vielfaches einer Einheit dargestellt. Hat eine Messung eine Spannung von 5 V ergeben (Größenwert), so heißt das, dass die gemessene Spannung das Fünffache der Einheit der Spannung 1 Volt beträgt.

$$\textit{Größenwert} = \text{Zahlenwert} \cdot \text{Einheit}$$

Wird in Gleichungen grundsätzlich das Produkt aus Zahlenwert und Einheit eingesetzt, spricht man von *Größengleichungen*. Sie gelten unabhängig von der Wahl der Einheiten und sind bevorzugt gegenüber *Zahlenwertgleichungen* anzuwenden, bei denen man die Einheiten vereinbart und nur noch Zahlenwerte in die Gleichungen einsetzt.

Für die in der Physik verwendeten Einheiten hat man die folgenden sieben **SI**-Basiseinheiten (**S**ystème **I**nternational d'Unités) nach Tab. 75.1 eingeführt:

Aus diesen Basiseinheiten werden weitere SI-Einheiten abgeleitet, von denen einige in Tab. 75.2 aufgeführt sind.

Die Darstellung, die Bewahrung und die Weitergabe der SI-Basiseinheiten und wichtiger abgeleiteter SI-Einheiten werden wegen des beträchtlichen apparativen Aufwandes national von Staatsinstituten wahrgenommen; in der Bundesrepublik von der Physikalisch-Technischen Bundesanstalt in Braunschweig.

Tab. 75.1 SI-Basiseinheiten

Basisgröße	Basiseinheit	
	Name	Einheitenzeichen
Länge	Meter	m
Masse	Kilogramm	kg
Zeit	Sekunde	s
Elektrische Stromstärke	Ampere	A
Thermodynamische Temperatur	Kelvin	K
Lichtstärke	Candela	cd
Stoffmenge	Mol	mol

Tab. 75.2 Abgeleitete SI-Einheiten (Auswahl)

Abgeleitete Größe	Name	Zeichen	Zusammenhang zu anderen SI-Einheiten
Kraft	Newton	N	$1\,\mathrm{N} = 1\,\mathrm{kg \cdot m/s^2}$
Energie	Joule	J	$1\,\mathrm{J} = 1\,\mathrm{N \cdot m}$
Leistung	Watt	W	$1\,\mathrm{W} = 1\,\mathrm{J/s}$
Elektrische Spannung	Volt	V	$1\,\mathrm{V} = 1\,\mathrm{W/A}$
Elektrischer Widerstand	Ohm	Ω	$1\,\Omega = 1\,\mathrm{V/A}$
Elektrische Kapazität	Farad	F	$1\,\mathrm{F} = 1\,\mathrm{A \cdot s/V}$
Induktivität	Henry	H	$1\,\mathrm{H} = 1\,\mathrm{V \cdot s/A}$
Magnetischer Fluss	Weber	Wb	$1\,\mathrm{Wb} = 1\,\mathrm{V \cdot s}$

75.3 Messabweichung, Messfehler

Eine *Messabweichung* ist der Unterschied zwischen dem *erhaltenen* und dem *wahren* Wert, oder, wenn der wahre Wert nicht bekannt ist, einem als richtig geltenden Wert, dem *konventionell richtigen Wert*. Der früher durchgehend gebrauchte Begriff *Fehler* sollte der Beschreibung von Messgeräten vorbehalten bleiben, die festgestellte systematische Abweichungen aufweisen (DIN 1319). In diesem Abschnitt wird der Übersichtlichkeit halber vorwiegend der Begriff „Abweichung“ verwendet.

Das Ziel eines Messvorganges ist es, den Wert der Messgröße so genau wie möglich zu bestimmen. Bei einem Messvorgang wird aber im allgemeinen nicht der gesuchte „wahre“ Wert gemessen, sondern ein Wert, der von diesem abweicht. Soll eine Aussage über diese Abweichung gemacht werden, muss der wahre Wert bekannt sein. Da dieser wahre Wert in der Regel aber nicht gefunden werden kann, wird ein als „richtig“ angenommener oder vereinbarter Wert für die Messung zugrunde gelegt, der *konventionell richtige Wert*. Sofern dies möglich ist, kann man ihn in der Praxis dadurch erhalten, dass man die zu messende Größe mit einem Messgerät mit sehr viel kleinerer Abweichung (1/10 oder kleiner, d. h. Präzisionsmessgerät) misst.

Die absolute Messabweichung F ist der Unterschied zwischen dem erhaltenen Messwert x_a und dem konventionell richtigen Wert x_r:

$$\textit{absolute Messabweichung} \quad F = x_a - x_r \tag{75.1}$$

Die relative Messabweichung f ergibt sich zu:

$$\textit{relative Messabweichung} \quad f = \frac{F}{x_r} = \frac{x_a - x_r}{x_r} \tag{75.2}$$

$$\textit{relative Messabweichung in \%} \quad f = \frac{F}{x_r} = \frac{x_a - x_r}{x_r} \cdot 100\,\% \tag{75.3}$$

Die Ursachen für die Messabweichungen lassen sich nach der Art ihrer Entstehung einteilen.

75.3.1 Systematische Abweichungen

Sie entstehen beispielsweise durch das Messinstrument selbst, das den Messkreis beeinflusst (Spannungsfall am Strommesser) und sind in der Regel nach Betrag und Vorzeichen bekannt und somit *korrigierbar*. Diese Abweichungen machen, wenn sie nicht korrigiert werden, ein Messergebnis *unrichtig*.

Bekannte systematische Abweichungen A_a werden durch die *Korrektion* K erfasst. Man erhält damit den *berichtigten Messwert* x_r:

$$\begin{aligned} &\textit{Korrektion} && K = -A \\ &\textit{berichtigter Messwert} && x_r = x_a + K \end{aligned} \tag{75.4}$$

75.3.2 Zufällige Abweichungen

Sie entstehen durch Einflüsse, die dem Zufall unterliegen, wie zum Beispiel Rauschen, Störungen, Abnutzung und Bauteiländerungen. Sie sind prinzipiell nicht bekannt und können deshalb auch nicht korrigiert werden. Diese Abweichungen machen ein Messergebnis *unsicher*.

Häufig kann man für die zufälligen Abweichungen die *Abweichungsgrenzen* G angeben, innerhalb derer der Messwert x „mit großer Wahrscheinlichkeit“ liegt:

$$\textit{Messwert mit Abweichungsgrenzen} \quad x_r \pm G \tag{75.5}$$

Das bedeutet, dass der Messwert x im Bereich $x_r - G \leq x \leq x_r + G$ liegt.

Beispiel 1 zeigt die unterschiedlichen Arten der Darstellung von Abweichungsgrenzen.

Beispiel 1

$5\,\mathrm{V} \pm 0{,}2\,\mathrm{V}$; $16{,}2\,(1 \pm 0{,}02)\,\mathrm{A}$; $(1 \pm 10\,\%)\,\mathrm{kW}$. Die Abweichungsgrenzen werden hier, den Erfahrungen aus der Praxis entsprechend, *symmetrisch* zum *konventionell richtigen Wert* angegeben. ◄

Mit der *maximal* auftretenden Abweichung lässt sich auch die *maximale relative* Messabweichung (75.2) berechnen (der Begriff „maximal“ entfällt meist):

(Maximale) relative Messabweichung

$$f = \frac{x_r \pm G - x_r}{x_r} = \frac{\pm G}{x_r} \qquad (75.6)$$

Die (maximale) relative Messabweichung wird in der Messtechnik bevorzugt angegeben.

Beispiel 2

Zwei unterschiedliche Spannungen werden gemessen: $U_1 = 8\,\text{V} \pm 0{,}1\,\text{V}$; $U_2 = 80\,\text{V} \pm 0{,}1\,\text{V}$. Die Abweichungsgrenzen sind für beide Messungen gleich. Erst die (maximale) relative Messabweichung zeigt, dass die Messung von U_2 genauer ist als die von U_1:

$$f_1 = \pm 0{,}1\,\text{V}/8\,\text{V} = \pm 0{,}0125 = \pm 1{,}25\,\%;$$
$$f_2 = \pm 0{,}1\,\text{V}/80\,\text{V} = \pm 0{,}00125 = \pm 0{,}125\,\%. \quad \blacktriangleleft$$

Zur Festlegung der Abweichungsgrenzen G werden praktische Erfahrungen, umfangreiche Messungen gleicher oder ähnlicher Art und die Statistik herangezogen. Überhaupt spielt die Statistik bei der Angabe von Abweichungen eine entscheidende Rolle. Da man den „wahren“ Wert niemals messen kann, lässt sich von einem Messwert nur angeben, mit welcher Wahrscheinlichkeit er innerhalb eines bestimmten Wertebereiches liegt. Dazu werden die zwei Begriffe *arithmetischer Mittelwert* und *(empirische) Standardabweichung* verwendet.

75.3.3 Arithmetischer Mittelwert, Erwartungswert

Der Einfluss von zufälligen Abweichungen auf das Messergebnis lässt sich in der Regel dadurch verringern, dass zunächst die Messung häufig wiederholt wird, eventuell sogar mit unterschiedlichen Messgeräten, und anschließend aus den n Einzelmessungen x_i der *arithmetische Mittelwert* gebildet wird:

$$\textit{arithmetischer Mittelwert} \quad \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad (75.7a)$$

n Anzahl der Messungen; x_i die einzelnen Messwerte.

Lässt man die Zahl der Messungen gegen unendlich gehen, geht der arithmetische Mittelwert $\bar{x}$ in den Erwartungswert μ über, von dem man annimmt, dass er den konventionell richtigen Wert mit großer Wahrscheinlichkeit repräsentiert.

$$\textit{Erwartungswert} \quad \mu = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{x_{in}}{n} \qquad (75.7b)$$

n Anzahl der Messungen; x_{in} die einzelnen Messwerte.

Der arithmetische Mittelwert kommt dem *konventionell richtigen Wert* meist schon sehr nahe, da in der Praxis die Abweichungen der Einzelmessungen sowohl positives als auch negatives Vorzeichen haben und sich damit teilweise aufheben. Trotzdem darf man nicht annehmen, dass der arithmetische Mittelwert gleich dem wahren Wert ist.

Beispiel 3

Eine wiederholt durchgeführte Spannungsmessung liefert die Einzelwerte: 5,0 V; 5,1 V; 4,9 V; 5,2 V; 5,0 V; 5,1 V. Der nach (75.7a) berechnete arithmetische Mittelwert ergibt 5,05 V. Eine zweite Messreihe ergibt die Einzelwerte: 5,0 V; 5,3 V; 4,8 V; 5,2 V; 5,2 V; 4,8 V. Der arithmetische Mittelwert ist wiederum 5,05 V. ◀

75.3.4 Standardabweichung

Die Erfahrung zeigt, dass trotz des zahlenmäßig gleichen arithmetischen Mittelwertes der Messreihen aus Beispiel 3 dem Mittelwert der ersten Messreihe eher der *konventionell richtige Wert* zugeordnet werden kann als dem zweiten, weil bei ihr die Einzelmessungen weniger voneinander abweichen. Deshalb wird die *(empirische) Standardabweichung* s (meist nur kurz als Standardabweichung bezeichnet) als Beurteilungskriterium für eine Messreihe gebildet, die die Differenzen zwischen den Einzelmessungen und dem arithmetischen Mittelwert erfasst und diese Differenzen vor der Summation quadriert, damit sie sich bei entgegengesetztem Vorzeichen nicht teilweise aufheben:

Standardabweichung

$$s = +\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (75.8a)$$

Der Ausdruck $(n-1)$ im Nenner anstatt n weist darauf hin, dass die Bildung der Standardabweichung erst für mehr als einen Messwert sinnvoll ist.

Das Quadrat der Standardabweichung wird mit Varianz bezeichnet:

$$\textit{Varianz} \quad s^2 \text{ oder auch } \sigma^2 \qquad (75.8b)$$

Dabei wird s^2 für eine endliche Anzahl von Messwerten verwendet, während σ^2 den Grenzwert für unendlich viele Messwerte bezeichnet, siehe (75.16), Gaußverteilung.

Beispiel 4

Die Standardabweichung s errechnet sich nach (75.8a) für die erste Messreihe von Beispiel 3 zu $s_1 = 0{,}105\,\text{V}$; für die zweite Messreihe ergibt sich $s_2 = 0{,}217\,\text{V}$. ◀

Teil X

Je größer die Standardabweichung s ist, desto mehr „streuen“ die Messwerte, und umso weniger wird man vom arithmetischen Mittelwert auf den konventionell richtigen Wert schließen können.

Die Behandlung von Messabweichungen oder Messfehlern geschieht nach Gesetzen der Wahrscheinlichkeitstheorie. Danach werden Messwerte als Werte aufgefasst, die mit einer bestimmten Wahrscheinlichkeit in einem gegebenen Wertebereich auftreten (Abschn. 75.7). Einzelheiten dazu sind z. B. in DIN 1319 aufgeführt.

75.4 Abweichungsfortpflanzung, Fehlerfortpflanzung

Werden mehrere unterschiedliche abweichungsbehaftete Größen gemessen (z. B. Spannung U und Strom I), um daraus eine weitere Größe rechnerisch zu bestimmen (z. B. den Widerstand R), gehen die Messabweichungen der Einzelgrößen in die rechnerisch bestimmte Größe in der nachfolgend dargestellten Weise ein. Gegeben ist $F = f(x, y, z, \dots)$. Darin steht F für den Gesamtfehler (z. B. von R), der durch die Einzelfehler der Größen $x, y, z, \dots$ (z. B. U, I) bestimmt wird. Die Einzelfehler erhält man durch partielles Differenzieren:

$$\frac{\partial F}{\partial x} = \frac{\partial f(x,y,z,\dots)}{\partial x}; \quad \frac{\partial F}{\partial y} = \frac{\partial f(x,y,z,\dots)}{\partial y}; \quad \frac{\partial F}{\partial z} = \frac{\partial f(x,y,z,\dots)}{\partial z}; \quad \dots \tag{75.9}$$

Bei den Abweichungen bzw. Fehlern handelt es sich um endliche Differenzen, so dass aus (75.9) folgt:

$$\Delta F_x = \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x;$$
$$\Delta F_y = \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y;$$
$$\Delta F_z = \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z; \quad \dots \tag{75.10}$$

Der Gesamtfehler ΔF berechnet sich aus der Summe der Einzelfehler:

$$\Delta F = \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x + \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y + \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z + \dots \tag{75.11a}$$

Dieser Ausdruck wird in der Mathematik als totales Differential bezeichnet.

Bei der Anwendung dieser Formel ist zu unterscheiden, ob es sich um systematische oder zufällige Abweichungen bzw. Fehler handelt.

Systematische Abweichungen bzw. *Fehler* sind nach Betrag und Vorzeichen bekannt und müssen deshalb *mit Vorzeichen* eingesetzt werden. Daraus kann sich ergeben, dass sich bei der Addition der Einzelterme eine Abweichungsverringerung oder sogar eine vollständige Abweichungsaufhebung ergibt, wie Beispiel 9 zeigt.

$$\Delta F = \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x + \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y + \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z + \dots \tag{75.11b}$$

absolute Abweichungsgrenze für systematische Abweichungen; $\Delta x, \Delta y, \Delta z, \dots$ sind vorzeichenrichtig einzusetzen.

$$\frac{\Delta F}{F} = \frac{1}{F} \cdot \left\{ \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x + \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y + \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z + \dots \right\} \tag{75.11c}$$

relative Abweichungsgrenze für systematische Abweichungen; $\Delta x, \Delta y, \Delta z, \dots$ sind vorzeichenrichtig einzusetzen.

Zufällige Abweichungen: Bei ihnen sind nur die Abweichungs*grenzen* bekannt.

Für $\Delta x, \Delta y, \Delta z, \dots$ werden die zugeordneten Abweichungsgrenzen $\pm G_x, \pm G_y, \pm G_z, \dots$ eingesetzt. Wichtig ist, dass die Vorzeichen der einzelnen Abweichungsgrenzen $(\Delta x, \Delta y, \Delta z, \dots)$ so gewählt werden, dass sich die größtmögliche Gesamtabweichungsgrenze ΔF ergibt. Die tatsächlich auftretende Gesamtabweichungsgrenze kann damit niemals größer werden als die berechnete. Formelmäßig lässt sich das dadurch erreichen, dass bei den Einzelabweichungen die Beträge gebildet und dann insgesamt addiert werden. Das „±“-Zeichen vor dem Ausdruck erfasst die größtmögliche Abweichungsgrenze sowohl in positiver als auch in negativer Richtung.

$$\Delta F = \pm \left\{ \left| \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x \right| + \left| \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y \right| + \left| \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z \right| + \dots \right\} \tag{75.11d}$$

absolute Abweichungsgrenze für zufällige Abweichungen.

$$\frac{\Delta F}{F} = \pm \frac{1}{F} \left\{ \left| \frac{\partial f(x,y,z,\dots)}{\partial x} \cdot \Delta x \right| + \left| \frac{\partial f(x,y,z,\dots)}{\partial y} \cdot \Delta y \right| + \left| \frac{\partial f(x,y,z,\dots)}{\partial z} \cdot \Delta z \right| + \dots \right\} \tag{75.11e}$$

relative Abweichungsgrenze für zufällige Abweichungen.

Beispiel 5

An einem Spannungsteiler werden die zwei Teilspannungen $U_1 = 12\,\text{V} \pm 0{,}2\,\text{V}$ und $U_2 = 16\,\text{V} \pm 0{,}1\,\text{V}$ gemessen. Daraus sollen die Gesamtspannung und der Gesamtfehler bestimmt werden. Bezogen auf (75.11a) gilt: $x \leftrightarrow U_1$; $y \leftrightarrow U_2 \Rightarrow F = f(U_1, U_2) = U_1 + U_2$; $\Delta x = \Delta U_1 = \pm 0{,}2\,\text{V}$; $\Delta y = \Delta U_2 = \pm 0{,}1\,\text{V}$. Mit (75.11a) folgt:

$$\begin{aligned}\Delta F &= \frac{\partial(U_1 + U_2)}{\partial U_1} \cdot \Delta U_1 + \frac{\partial(U_1 + U_2)}{\partial U_2} \cdot \Delta U_2 \\ &= 1 \cdot \Delta U_1 + 1 \cdot \Delta U_2\end{aligned}$$

Die Gesamtspannung aus diesen zwei Messwerten ergibt sich zu:

$$\begin{aligned}(12 + 16)\,\text{V} + \Delta F &= 28\,\text{V} \pm (0{,}2 + 0{,}1)\,\text{V} \\ &= 28\,\text{V} \pm 0{,}3\,\text{V}.\end{aligned}$$

Die relative Messabweichung beträgt:

$$\begin{aligned}[(28 \pm 0{,}3)\,\text{V}) - 28\,\text{V}]/28\,\text{V} &= \pm 0{,}011 \\ &= \pm 1{,}1\,\%.\end{aligned}$$

◀

Beispiel 6

Die Differenz zweier Ströme soll aus den zwei gemessenen Strömen $I_1 = 290\,\text{mA} \pm 5\,\text{mA}$ und $I_2 = 265\,\text{mA} \pm 5\,\text{mA}$ berechnet werden.

Mit den Zuordnungen entsprechend Beispiel 5 und (75.11a) erhält man: $x \leftrightarrow I_1$; $y \leftrightarrow I_2 \Rightarrow F = f(I_1, I_2) = I_1 - I_2$; $\Delta x = \Delta I_1 = \pm 5\,\text{mA}$; $\Delta y = \Delta I_2 = \pm 5\,\text{mA}$.

$$\begin{aligned}\Delta F &= \frac{\partial(I_1 - I_2)}{\partial I_1} \cdot \Delta I_1 + \frac{\partial(I_1 - I_2)}{\partial I_2} \cdot \Delta I_2 \\ &= 1 \cdot \Delta I_1 + 1 \cdot \Delta I_2\end{aligned}$$

Daraus folgt:

$$\begin{aligned}I_1 - I_2 &= (290 - 265)\,\text{mA} \pm (5 + 5)\,\text{mA} \\ &= 25\,\text{mA} \pm 10\,\text{mA}.\end{aligned}$$

Die relative Messabweichung beträgt

$$f = \pm 10\,\text{mA}/25\,\text{mA} = \pm 0{,}4 = \pm 40\,\%.$$

Das Vorzeichen der Messabweichungen wurde entsprechend der Anmerkung zu (75.11a) eingesetzt.

Die Messabweichung wird umso größer, je geringer die Differenz der Einzel-Messwerte ist.

Nach Beispiel 6 soll die rechnerische Differenzbildung vermieden werden, wenn sich zwei abweichungsbehaftete Messwerte nur wenig voneinander unterscheiden. ◀

Beispiel 7

Spannung und Strom an einem Widerstand werden gemessen: $U = 100\,\text{V} \pm 0{,}5\,\text{V}$; $I = 2{,}0\,\text{A} \pm 20\,\text{mA}$. Zu bestimmen sind die relativen Abweichungen für die Leistung an R und für den Widerstand R, wenn beide rechnerisch bestimmt werden. Die systematischen Fehler durch die Innenwiderstände der Messinstrumente werden vernachlässigt.

Wie in den Beispielen 5 und 6 ausführlich gezeigt, ergibt sich für P folgende Formel:

$$\Delta F = \frac{\partial(U \cdot I)}{\partial U} \cdot \Delta U + \frac{\partial(U \cdot I)}{\partial I} \cdot \Delta I = I \cdot \Delta U + U \cdot \Delta I.$$

Daraus folgt die Leistung:

$$\begin{aligned}P &= U \cdot I \pm (I \cdot \Delta U + U \cdot \Delta I) \\ &= 100\,\text{V} \cdot 2\,\text{A} \pm (2\,\text{A} \cdot 0{,}5\,\text{V} + 100\,\text{V} \cdot 20\,\text{mA}) \\ &= 200\,\text{W} \pm (1\,\text{W} + 2\,\text{W}) \\ &= 200\,\text{W} \pm 3\,\text{W}\end{aligned}$$

Die relative Messabweichung errechnet sich zu $\pm 3\,\text{W}/200\,\text{W} = \pm 1{,}5\,\%$. Für den Widerstand gilt:

$$\begin{aligned}\Delta F &= \frac{\partial(U/I)}{\partial U} \cdot \Delta U + \frac{\partial(U/I)}{\partial I} \cdot \Delta I \\ &= \Delta U/I - (U/I^2) \cdot \Delta I. \qquad (75.12) \\ R &= 100\,\text{V}/2\,\text{A} \pm ((0{,}5\,\text{V}/2\,\text{A}) + (100\,\text{V}/4\,\text{A}^2) \cdot 20\,\text{mA}) \\ &= 50\,\Omega \pm 0{,}75\,\Omega.\end{aligned}$$

Die relative Messabweichung beträgt $\pm 0{,}75\,\Omega/50\,\Omega = \pm 1{,}5\,\%$.

Das bei der mathematischen Berechnung in (75.12) auftretende Minuszeichen wird mit den Vorzeichen der Einzelmessabweichungen so verknüpft, dass sich wieder die größtmögliche Gesamtmessabweichung ergibt. ◀

Beispiel 8

Aus zwei parallel geschalteten Widerständen soll der Gesamtwiderstand R_p bestimmt werden.

Gegeben: $R_1 = 1000\,\Omega \pm 10\,\Omega$; $R_2 = 200\,\Omega \pm 2\,\Omega$.

$$\begin{aligned}\Delta F &= \frac{\partial\left(\frac{R_1 R_2}{R_1 + R_2}\right)}{\partial R_1} \cdot \Delta R_1 + \frac{\partial\left(\frac{R_1 R_2}{R_1 + R_2}\right)}{\partial R_2} \cdot \Delta R_2 \\ &= \frac{R_2^2}{(R_1 + R_2)^2} \cdot \Delta R_1 + \frac{R_1^2}{(R_1 + R_2)^2} \cdot \Delta R_2 \\ \Delta F &= \frac{4 \cdot 10^4 \cdot 10 + 10^6 \cdot 2}{(1{,}44 \cdot 10^6)}\,\Omega = 1{,}67\,\Omega\end{aligned}$$

Damit ergibt sich: $R_p = 167\,\Omega \pm 1{,}67\,\Omega$. Der relative Fehler beträgt $\pm 1{,}67\,\Omega/167\,\Omega = \pm 1\,\%$. ◀

Beispiel 9

Ein Widerstand R soll aus einer Spannungs- und einer Strommessung berechnet werden. Die beiden Messungen sind mit je einem *systematischen* Fehler behaftet:

$$U = 10\,\mathrm{V} + 10\,\mathrm{mV}; \quad I = 5\,\mathrm{mA} + 5\,\mu\mathrm{A}.$$

Für die absolute Fehlergrenze ergibt sich mit (75.11b):

$$\begin{aligned}\Delta R &= +\frac{\Delta U}{I} - \frac{U}{I^2}\cdot\Delta I \\ &= \frac{10\,\mathrm{mV}}{5\,\mathrm{mA}} - \frac{10\,\mathrm{V}}{(5\,\mathrm{mA})^2}\cdot 5\,\mu\mathrm{A} \\ &= 2\,\Omega - 2\,\Omega = 0.\end{aligned}$$

Man erhält rechnerisch den gleichen (korrekten) Wert aus $\frac{10\,\mathrm{V}}{5\,\mathrm{mA}}$ bzw. $\frac{10{,}01\,\mathrm{V}}{5{,}005\,\mathrm{mA}}$. ◀

75.5 Fehlerangaben von Messgeräten

Bei Messgeräten wird in der Regel die Fehlergrenze angegeben. Sie ist der maximal zugelassene prozentuale Fehler, der häufig auf den Messbereichsendwert bezogen wird (DIN 43780). Sie gilt außerdem nur bei Einhaltung gewisser Referenzbedingungen (z. B. Referenzfrequenz 45 bis 65 Hz ± 5 Hz).

75.5.1 Analog anzeigende Messgeräte

Bei analog (direkt wirkenden) anzeigenden Messgeräten sind Genauigkeitsklassen nach DIN 43780 festgelegt. Die dort angegebenen Klassenzeichen sind gleich dem Betrag der Fehlergrenzen in Prozent. Dabei ist der absolute Fehler auf einen Bezugswert bezogen, in der Regel auf den Messbereichsendwert. Wird dagegen ein Messwert abgelesen, so ist dieser Messwert der Bezugswert. Da der Bezugswert aber stets kleiner oder höchstens gleich dem Messbereichsendwert ist, wird der Fehler umso größer, je kleiner der gemessene Wert ist. Deshalb gilt:

Messwerte möglichst im oberen Drittel der Skala ablesen!

Folgende *Genauigkeitsklassen* sind nach DIN 43780 festgelegt: 0,05; 0,1; 0,2; 0,5; 1; 2,5; 5.

Beispiel 10

Ein Messgerät gehört der Klasse 1 an und hat den Messbereichsendwert 100 V. Dann ist die Fehlergrenze ±1 % vom Messbereichsendwert 100 V, also ±1 V. Diese Fehlergrenze gilt für jeden angezeigten Messwert. Beträgt der Messwert 20 V, so ist der relative Fehler $f = [(20 \pm 1)\,\mathrm{V} - 20\,\mathrm{V}]/20\,\mathrm{V} = \pm 5\,\%$. ◀

75.5.2 Digital anzeigende Messgeräte

Durch die Elektronik im Instrument setzt sich der Gesamtfehler aus mehreren Anteilen zusammen. In der Praxis findet man häufig die folgende Fehlerangabe für den maximal auftretenden (absoluten) Fehler $F_{\max}$, der hier an einem Beispiel gezeigt werden soll:

absoluter Fehler

$$F_{\max} = \pm(0{,}1\,\%\,\text{v.A.} + 0{,}1\,\%\,\text{v.E.} + 1\,\text{LSB}) \qquad (75.13)$$

v.A. von der Anzeige, also abgelesener Wert (of reading); v.E. vom Endwert, Anzeigebereichsendwert (of range); LSB Bit mit der geringsten Wertigkeit, in der Regel die Stelle ganz rechts, Quantisierungsfehler (least significant bit).

Beispiel 11

Angezeigter Messwert: $U = 7{,}50\,\mathrm{V}$; Anzeigebereichs-Endwert: $U_{\mathrm{sk}} = 9{,}99\,\mathrm{V}$; 3-stellige Anzeige. Absoluter Fehler $F_{\max}$ (mit den Angaben aus (75.13)):

$$\begin{aligned}F_{\max} &= \pm(0{,}001\cdot 7{,}50\,\mathrm{V} + 0{,}001\cdot 9{,}99\,\mathrm{V} + 0{,}01\,\mathrm{V}) \\ &= \pm 0{,}0275\,\mathrm{V}\end{aligned}$$

Relativer Anzeigefehler:

$$f_{\max} = \pm 0{,}0275\,\mathrm{V}/7{,}50\,\mathrm{V} = \pm 0{,}00367 = \pm 0{,}367\,\%$$

◀

Der Bezugswert für den relativen Anzeigefehler ist auch hier der angezeigte Messwert; der Fehler steigt mit kleiner werdendem Anzeigewert. Deshalb gilt auch hier:

Messwerte möglichst im oberen Drittel der Anzeige ablesen!

75.6 Arithmetischer Mittelwert und Effektivwert von Wechselgrößen

Der *Effektivwert* einer periodischen Wechselspannung (eines periodischen Wechselstromes) gibt die gleiche Leistung ab wie eine gleichgroße Gleichspannung (Gleichstrom). Deshalb soll auch in der Regel der Effektivwert gemessen werden. Er berechnet sich nach der Formel:

Effektivwert einer Wechselspannung $u(t)$:

$$U = \sqrt{\frac{1}{T}\int_0^T \{u(t)\}^2 dt} \tag{75.14a}$$

Effektivwert eines Wechselstromes $i(t)$:

$$I = \sqrt{\frac{1}{T}\int_0^T \{i(t)\}^2 dt} \tag{75.14b}$$

Der *arithmetische Mittelwert* oder *Gleichanteil* einer periodischen Wechselspannung oder eines periodischen Wechselstromes ergibt sich aus:

Arithmetischer Mittelwert einer Wechselspannung $u(t)$:

$$\bar{u} = \frac{1}{T}\int_0^T \{u(t)\} dt \tag{75.15a}$$

Arithmetischer Mittelwert eines Wechselstromes $i(t)$:

$$\bar{i} = \frac{1}{T}\int_0^T \{i(t)\} dt \tag{75.15b}$$

In der Messtechnik hat er eine gewisse Bedeutung, weil *Drehspulmessinstrumente* diesen Wert anzeigen und *Oszilloskope* in der Schalterstellung „AC“ für die Eingangsspannung diesen Signalanteil unterdrücken (Kap. 77 „Oszilloskop“).

75.7 Häufigkeitsverteilung, Vertrauensbereich

Trägt man die einzelnen Messwerte einer gemessenen Größe, der zufällige Fehler überlagert sind, in Form einer Häufigkeitsverteilung (Histogramm) auf, erhält man in der Regel einen charakteristischen Verlauf, der sich für eine gegen unendlich gehende Anzahl von Messwerten der Normal- oder Gauß-Verteilung („Gauß’sche Glockenkurve“) annähert. Sie ist wie folgt definiert:

Normalverteilung

$$p(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{75.16}$$

$\bar{x}$ siehe (75.7a), σ^2 siehe (75.8b).

Die Erfahrung hat gezeigt, dass der berechnete Mittelwert $\bar{x}$ einer Messreihe mit den Einzel-Messwerten x_i mit zunehmender Wahrscheinlichkeit dem konventionell richtigen Wert entspricht, je mehr sich die aus den Messwerten x_i gebildete Verteilung der Normalverteilung annähert und je größer die Anzahl der Messwerte ist. Obwohl der Mittelwert $\bar{x}$ nicht zwangsläufig gleich dem wahren Wert ist, kann man ein Intervall um $\bar{x}$ herum angeben, in dem der wahre Wert mit einer vorgegebenen Wahrscheinlichkeit $(1-\alpha)$ liegt. Die Grenzen dieses Intervalls heißen *Vertrauensgrenzen*. Das Intervall ist der *Vertrauensbereich*, der sich aus dem *Vertrauensniveau* $(1-\alpha)$ (Angabe in %, siehe Tab. 75.3) ergibt.

Obere Vertrauensgrenze

$$\bar{x} + \frac{t}{\sqrt{n}} \cdot s \tag{75.17a}$$

Untere Vertrauensgrenze

$$\bar{x} - \frac{t}{\sqrt{n}} \cdot s \tag{75.17b}$$

$\bar{x}$ siehe (75.7a), n Anzahl der Messungen, s Standardabweichung (75.8a), $t/\sqrt{n}$ siehe Tab. 75.3.

Tab. 75.3 zeigt Auszüge für t und $t/\sqrt{n}$ nach DIN 1319 Teil 3.

Tab. 75.3 Werte für t und $t/\sqrt{n}$ für verschiedene Werte des Vertrauensniveaus $(1-\alpha)$

Anzahl n der Einzelmessungen	$(1-\alpha) = 68{,}26\,\%$		$(1-\alpha) = 95\,\%$		$(1-\alpha) = 99{,}5\,\%$	
	t	$t/\sqrt{n}$	t	$t/\sqrt{n}$	t	$t/\sqrt{n}$
2	1,84	1,30	12,71	8,98	127,32	90,03
5	1,15	0,51	2,78	1,24	5,60	2,50
10	1,06	0,34	2,26	0,71	3,69	1,17
30	1,02	0,19	2,05	0,37	3,04	0,56
100	1,00	0,10	1,98	0,20	2,87	0,29

Tab. 75.4 Zu Beispiel 12

Messung Nr.	Spannung U_i in V	$(x_i - \bar{x})$ in V	$(x_i - \bar{x})^2$ in V	t für $(1-\alpha) = 95\,\%$	$t/\sqrt{n}$ für $(1-\alpha) = 95\,\%$
1	100,2	0,4	0,16	Aus Tab. 75.3	Aus Tab. 75.3
2	100,4	0,6	0,36		
3	99,3	−0,5	0,25		
4	99,0	−0,8	0,64		
5	100,1	0,3	0,09		
	$\bar{x} = 99{,}8\,\text{V}$	$\sum = 0\,\text{V}$	$\sum = 1{,}50\,\text{V}$ $s = 0{,}612\,\text{V}$	2,78	1,24

Beispiel 12

Eine Spannung wird fünfmal gemessen. Zu bestimmen sind die Vertrauensgrenzen für ein Vertrauensniveau von 95 %. Das Messergebnis ist anzugeben.

Daraus ergibt sich die obere Vertrauensgrenze zu: $99{,}8\,\text{V} + 1{,}24 \cdot 0{,}612\,\text{V} = 99{,}8\,\text{V} + 0{,}759\,\text{V}$; die untere zu $99{,}8\,\text{V} - 0{,}759\,\text{V}$. Das Messergebnis für die Spannung lautet: $99{,}8\,\text{V} \pm 0{,}759\,\text{V}$. ◀

75.8 Bearbeitung und Auswertung von Messwerten

Die Bearbeitung und Auswertung von Messwerten ist in der Regel rechenintensiv und wurde deshalb erst durch den Einsatz von Rechnern in größerem Umfang möglich. In diesem Zusammenhang haben Begriffe wie

- Chi-Quadrat-Test,
- Quantile, Median,
- Regression,
- Splines und
- Klassierung

an Bedeutung gewonnen. Sie werden im Kap. 85 „PC-gestützte Messverfahren und Messsignalanalyse“ dargestellt.

Analog anzeigende Messgeräte

76

Wilfried Plaßmann



76.1 Grundlagen 886
76.2 Drehspul-Messwerk 886
76.3 Dreheisen-Messwerk 887
76.4 Elektrodynamisches Messwerk 887
76.5 Symbole und Instrumentenbeschriftungen 888

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_76

76.1 Grundlagen

Die analog anzeigenden Messgeräte sind zum großen Teil von den digital anzeigenden abgelöst worden, weil letztere i. Allg. einen kleineren Messfehler und einen geringeren Eigenverbrauch besitzen und häufig Funktionen wie automatische Messbereichswahl, Speicherung von Messdaten und eine Schnittstelle zur Messdatenerfassung und -verarbeitung bieten. Die Analogmessgeräte sollen hier deshalb nur auszugsweise und kurz behandelt werden.

Die Wirkungsweise der im Folgenden vorgestellten Messgeräte beruht auf der Kraftwirkung auf einen stromdurchflossenen elektrischen Leiter im Magnetfeld:

$$\textit{Kraft} \quad \vec{F} = Q \cdot (\vec{v} \times \vec{B}) \quad \text{bzw.} \quad F = Q \cdot v \cdot B \qquad (76.1)$$

für $\vec{v} \perp \vec{B}$; $\vec{F} \perp \vec{B}$, $\vec{F} \perp \vec{v}$.

F Kraft in N, Q Ladung der den Leiter durchfließenden Elektronen in A s, v Geschwindigkeit der Elektronen in m/s, B magnetische Flussdichte in V s/m^2.

Es wird vorausgesetzt, dass v und B senkrecht zueinander auftreten. F steht senkrecht auf v und B, die Richtung ergibt sich aus der *„Rechten-Hand-Regel"*. Benutzt man die zwei Beziehungen $Q = I \cdot t$ und $v = l/t$ für den Stromfluss in einem elektrischen Leiter und setzt beides in (76.1) ein, folgt daraus eine Darstellung für die Kraft, die auch unter dem Namen *Lorentz-Kraft* bekannt ist:

$$\textit{Lorentz-Kraft} \quad F = B \cdot I \cdot l \qquad (76.2)$$

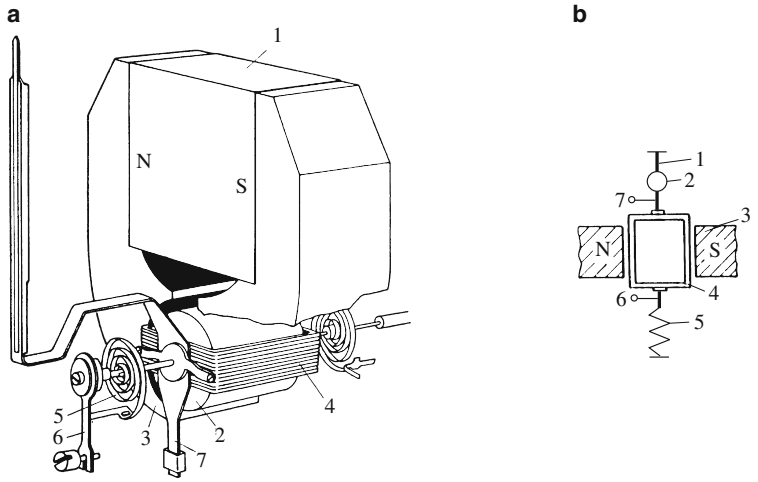
l Spulenlänge, I Leiter- bzw. Spulenstrom in A, B magnetische Flussdichte in V s/m^2.

Auch hier gilt für die Richtungsbeziehung der Größen B, I und F die *„Rechte-Hand-Regel"*. Die magnetische Flussdichte B kann durch einen Dauermagneten (Drehspulmesswerk), durch die stromdurchflossene Spule selbst (Dreheisenmesswerk) oder durch eine zweite Spule (Prinzip des Elektromagneten wie beim Elektrodynamischen Messwerk) zur Verfügung gestellt werden.

Im Folgenden werden nur die *Messwerke*, also die den Ausschlag direkt erzeugenden Elemente, betrachtet. Gehäuse, Anschlussklemmen, Bereichsumschalter, Vor- und Nebenwiderstände zur Messbereichserweiterung und andere Bauteile, die das Messwerk zum Messgerät ergänzen, bleiben unberücksichtigt.

76.2 Drehspul-Messwerk

Den prinzipiellen Aufbau zeigt Abb. 76.1a. Eine drehbar gelagerte Spule befindet sich im Feld eines Dauermagneten. Abhängig von der Richtung des durchfließenden Stromes dreht sie sich gemäß der *Lorentz-Kraft* nach (76.2) rechts oder links herum. Die Länge l ist die Spulenlänge „nach hinten". Durch eine Rückstellfeder (Spiralfeder), die gleichzeitig als Stromzuführung dient, wird eine Gegenkraft erzeugt. Der Zeiger kommt zum Stillstand, wenn das elektrische Moment und das mechanische Moment der Feder im Gleichgewicht sind. Die Skalenteilung ist linear, da das elektrische Moment linear vom Strom I und das mechanische Moment linear vom Drehwinkel a abhängt. Die Ausschlagrichtung hängt von der Stromrichtung ab (76.2). Die für einen Zeigerausschlag erforderliche Leistung (Eigenverbrauch) ist sehr gering. So benötigt das Messwerk des Messinstrumentes „UM" (AEG) für Vollausschlag einen Strom



1 Permanentmagnet
2 Weicheisenkern
3 Polschuhe
4 Drehspule
5 Spiralfeder und Stromzuführung
6 Nullpunkteinstellung
7 Äquilibrierarm

1 Spannband
2 Spiegel
3 Dauermagnet
4 Spule mit Rahmen
5 Spannfeder
6, 7 Stromzuführung

Abb. 76.1 **a** Drehspulmesswerk (H & B), **b** Spiegelgalvanometer

von 10 µA bei einer Spannung von 60 mV. Das entspricht einer Leistung von 0,6 µW.

Das Messwerk zeigt den arithmetischen Mittelwert (Gleichanteil) des fließenden Stromes an. Bei der Messung von Wechselströmen führt das bei den häufig vorkommenden sinusförmigen Stromverläufen zu der Anzeige Null (Voraussetzung: Frequenz $f \gg 1\,\text{Hz}$, d. h., das Instrument kann wegen seiner Trägheit dem Momentanwert des Stroms nicht mehr folgen). Damit das Messwerk *überhaupt* einen Ausschlag zeigt, wird der Strom durch das Instrument (nicht der zu messende Laststrom) durch einen Einweg- oder Zweiweggleicher so in der Form geändert, dass der arithmetische Mittelwert nicht mehr Null ist. Bei Wechselströmen soll in der Regel der Effektivwert des Stromes gemessen werden, der im Allgemeinen nicht mit dem arithmetischen Mittelwert übereinstimmt. Damit die Anzeige des Stromes wenigstens für *sinusförmigen* Stromverlauf richtig ist, wird die Skala des Instrumentes entsprechend korrigiert. Für sinusförmigen Stromverlauf ist der Effektivwert $I = \hat{i}/\sqrt{2}$, der arithmetische Mittelwert bei Einweggleichrichtung $\bar{i}_{\text{ew}} = \hat{i}/\pi$ bzw. bei Zweiweggleichrichtung $\bar{i}_{\text{zw}} = 2 \cdot \hat{i}/\pi$. Also wird die Skalenbeschriftung mit dem Wert $I/\bar{i}$ multipliziert. Dieser Wert heißt *Kurvenformfaktor* oder kurz *Formfaktor* und beträgt für Sinusform bei Einweggleichrichtung $\pi/1{,}41 = 2{,}22$ bzw. bei Zweiweggleichrichtung $\pi/(2 \cdot 1{,}41) = 1{,}11$. Bei nicht sinusförmigem Stromverlauf ist das Verhältnis von Effektivwert zu arithmetischem Mittelwert nicht mehr gleich dem in die Skala eingearbeiteten Wert und damit der vom Instrument angezeigte Wert *falsch!*

Die Messung von Gleich- und Wechselspannungen geschieht über das Ohmsche Gesetz. Eine angelegte Spannung verursacht bei bekanntem Messwerkwiderstand einen ihr proportionalen Strom.

Eine vor allem in der Präzisionsmesstechnik verwendete Sonderbauform ist das *Galvanometer*. Die Drehspule ist an einem dünnen Metallfaden mit rechteckigem Querschnitt befestigt, der einseitig fest eingespannt ist und am anderen Ende durch eine Feder gespannt wird, Abb. 76.1b. Damit tritt so gut wie keine Lagerreibung auf. Die Spule befindet sich auch hier im Feld eines Dauermagneten. Fließt ein Strom durch sie hindurch, dreht sie sich und tordiert (verdrillt) den Metallfaden. Dadurch entsteht die Rückstellkraft. An dem Metallfaden ist ein Spiegel angebracht (*Spiegelgalvanometer*), der sich mitdreht und dessen Drehwinkel proportional zum Spulenstrom ist. Über einen Lichtstrahl und Umlenkspiegel können so kleinste Ströme im nA-Bereich durch kleinste Winkelauslenkungen des Spiegels in großen Entfernungen (1 bis 10 m) als bequem ablesbare Lichtmarkenauslenkungen gemessen werden.

Daten des Spiegelgalvanometers BSGe (Ruhstrat):

Ein Strom von 40 nA ergibt auf einer Skala in 25 cm Entfernung vom Spiegel eine Lichtstrahlauslenkung von 1 mm.

76.3 Dreheisen-Messwerk

Beim *Dreheisen-Messwerk* befinden sich nach Abb. 76.2 ein festes und ein drehbar gelagertes Metallblech im Feld einer vom Messstrom durchflossenen Spule. Die beiden Bleche werden

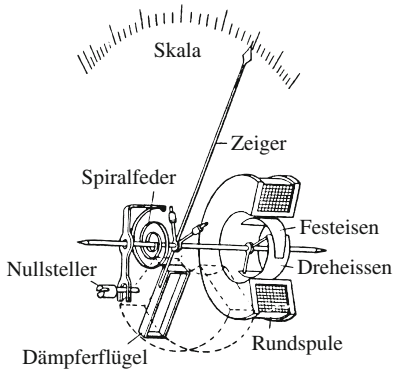


Abb. 76.2 Dreheisenmesswerk (H & B)

magnetisiert und stoßen sich umso mehr ab, je größer der Strom ist. Die Kraft F ist proportional zu I^2 (Prinzip „Elektromagnet“: $F \sim B^2 \sim I^2$). Damit ist die Ausschlagrichtung des Zeigers nicht mehr von der Stromrichtung abhängig. Bei der Messung von Wechselströmen bildet das Instrument bei Frequenzen des Stromes $\gg 1\,\text{Hz}$ wegen seiner Trägheit den Mittelwert aller quadrierten Momentanwerte, so dass der in (75.14b) unter der Wurzel stehende Ausdruck gebildet und damit das Quadrat des Effektivwertes angezeigt wird. Durch eine entsprechende Formgebung der Bleche kann man erreichen, dass sich eine näherungsweise lineare Skalenteilung für die Ablesung des Effektivwertes ergibt.

Theoretisch zeigt das Instrument den Effektivwert des Wechselstromes unabhängig von der Kurvenform an. In der Praxis dagegen wird der Messfehler bei der Messung von sinusförmigen Strömen mit Frequenzen oberhalb etwa (400 … 1000) Hz durch die Induktivität der Spule und das Verhalten des Eisens mit steigender Frequenz größer. Deshalb ist bereits das Messen *nichtsinusförmiger* Ströme mit Frequenzen weit unterhalb (400 … 1000) Hz mit erhöhten Messfehlern behaftet, weil die Oberschwingungsanteile oberhalb (400 … 1000) Hz zunehmend unberücksichtigt bleiben und damit der gemessene Stromverlauf vom tatsächlichen Stromverlauf abweicht (Fourierzerlegung des nichtsinusförmigen Stromverlaufes).

Da das magnetische Feld durch eine Spule und nicht durch einen Dauermagneten erzeugt wird und vorwiegend in Luft verläuft, ist die Flussdichte und damit die Kraft für die Zeigerauslenkung relativ klein. Daraus folgt, dass die für Vollausschlag erforderliche Leistung um den Faktor 10.000 bis 1.000.000 größer ist als bei Drehspulmesswerken.

Daten eines Dreheisen-Messwerkes (A-V-Multizet):

Strom für Vollausschlag 60 mA; Innenwiderstand 100 Ω. Daraus errechnet sich die Leistung für Vollausschlag zu $(60\,\text{mA})^2 \cdot (100\,\Omega) = 360\,\text{mW}$.

76.4 Elektrodynamisches Messwerk

Nach Abb. 76.3 besteht das *elektrodynamische Messwerk* aus einer feststehenden, vom Strom I_1 durchflossenen Spule S_1, die die Flussdichte B_1 erzeugt, und einer zweiten, drehbar gela-

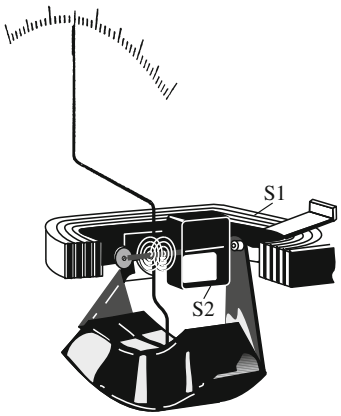


Abb. 76.3 Elektrodynamisches Messwerk (H & B)

gerten Spule S_2, die vom Strom I_2 durchflossen wird. Es gilt: $B_1 = k_1 \cdot I_1$; $F = k_2 \cdot B_1 \cdot I_2$ (mit (76.2)); $F = k_1 \cdot k_2 \cdot I_1 \cdot I_2$ bzw. $F = k_3 \cdot I_1 \cdot U_2$ mit $U_2 = k_4 \cdot I_2$ (k_i Konstante).

Die Kraft oder das Moment des Zeigers ist proportional dem *Produkt* zweier Ströme, oder, wenn man einen der beiden Ströme proportional zu einer Spannung macht, dem *Produkt* aus Spannung und Strom. Darin liegt die große Bedeutung dieses Messwerkes. Es wird hauptsächlich zur Leistungsmessung eingesetzt.

Die Wirkleistung P eines Verbrauchers berechnet sich bei sinusförmigen Spannungen und Strömen aus $P = U \cdot I \cdot \cos\varphi$. Schickt man den Strom des Verbrauchers durch die feststehende Spule (größerer Querschnitt, I_1 im Bereich 1 … 10 A) und macht den Strom I_2 proportional zur angelegten Spannung am Verbraucher, so zeigt das elektrodynamische Messwerk die Wirkleistung des Verbrauchers an. Das Messwerk muss aus den zugeführten Größen U und I das Moment für den Zeigerausschlag bilden. Damit wird der $\cos\varphi$ automatisch berücksichtigt.

Die feststehende Spule bildet den *Strompfad*, die bewegliche den *Spannungspfad*. Die Ausschlagrichtung des Zeigers hängt von dem Anschluss der beiden Pfade zueinander ab; beim Ausschlag des Zeigers in die falsche Richtung (an den Anschlag links) müssen die zwei Anschlüsse *eines* der beiden Pfade vertauscht werden. Auch aus diesem Grund kennzeichnen die Hersteller je einen Anschluss beider Pfade mit einem Stern oder einem Pfeil. Weitere Einzelheiten hierzu sind bei der Leistungsmessung aufgeführt.

Die Messung der Wirkleistung bei *nichtsinusförmigem* Spannungs- und Stromverlauf ist nur bedingt möglich. Hier gelten ähnliche Einschränkungen wie beim Dreheisenmesswerk.

Viele Hersteller haben für den Spannungspfad u. a. die Messbereiche 120, 240 und 480 V vorgesehen; für den Strompfad die Messbereiche 5, 1 A. Für eine Messbereichserweiterung mit Spannungs- und Stromwandler sind in der Regel die Messbereiche 100 V und 5 A erforderlich. Weitere Einzelheiten hierzu folgen im Abschn. 80.7.4 „Messbereichserweiterung bei der Leistungsmessung“.

76.5 Symbole und Instrumentenbeschriftungen

In Tab. 76.1 ist ein Auszug aus DIN 43780 der auf Instrumenten aufgedruckten Symbole und Beschriftungen enthalten. Besonders wird auf das Symbol „Dreieck mit innenliegendem Ausrufungszeichen“ hingewiesen. Es bedeutet, dass in der Betriebsanleitung wichtige Hinweise zum Betrieb des Messgeräts enthalten sind!

So kann ein Messgerät zwar einen Messbereich 5000 V haben, aber die Prüfspannung zwischen den Anschlussklemmen und dem Gehäuse ist geringer.

Tab. 76.1 Beschriftungen auf Messinstrumenten (Auszug aus DIN 43780)

Elektrodynamisches Meßwerk	
Eisengeschlossenes, elektrodynamisches Meßwerk	
Elektrodynamisches Quotientenmeßwerk	
Eisengeschlossenes, elektrodynamisches Quotientenmeßwerk	
Induktions-Meßwerk	
Bimetall-Meßwerk	
Elektrostatisches Meßwerk	
Vibrations-Meßwerk	
Thermoumformer allgemein	
Drehspul-Meßwerk mit Thermoumformer	
Isolierter Thermoumformer	
Gleichrichter	
Drehspul-Meßwerk mit Gleichrichter	

Wechselstrominstrument	
Gleich- und Wechselstrominstrument	
Drehstrominstrument mit einem Meßwerk	
Drehstrominstrument mit zwei Meßwerken	
Drehstrominstrument mit drei Meßwerken	
Senkrechte Gebrauchslage	
Waagrechte Gebrauchslage	
Schräge Gebrauchslage mit Angabe des Neigungswinkels	60°
Zeigernullstellvorrichtung	
Prüfspannungszeichen: die Ziffer im Stern bedeutet die Prüfspannung in kV (Stern ohne Ziffer 500 V Prüfspannung)	
Achtung (Gebrauchsanweisung beachten)	
Instrument entspricht bezüglich Prüfspannung nicht den Regeln	

Oszilloskop

77

Wilfried Plaßmann



77.1 Übersicht 890
77.2 Aufbau eines Oszilloskopes 890
77.3 Zusatzeinrichtungen 892

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_77

77.1 Übersicht

Das Oszilloskop wird häufig eingesetzt, um den zeitlichen Verlauf von Spannungen sichtbar zu machen. Aus dem Spannungsverlauf lassen sich weitere Größen bestimmen. Auch Größen, die keine Spannungen sind, lassen sich darstellen, wenn man sie in proportionale Spannungen umwandelt. Als Beispiel soll die Darstellung von Stromverläufen genannt werden, die über einen bekannten ohmschen Widerstand in Spannungsverläufe überführt werden können, ohne dass sich der zeitliche Verlauf ändert. Der Widerstandswert ist dann der Proportionalitätsfaktor. Bei entsprechendem technischem Aufwand sind auch mehr als ein Signal gleichzeitig auf dem Display darstellbar (2-Kanal-, 4-Kanal-Oszilloskope).

Bei den Oszilloskopen lässt sich der Leistungsumfang grob so klassifizieren:

- Erfassen periodischer Signale mit Frequenzen von wenigen Hz bis (20 … 50) MHz, ein bis maximal 2 Kanäle (Standardausführung)
- Erfassen periodischer Signale mit Frequenzen von Null bis (100 … 500) MHz und Erfassen einmaliger Signale, häufig 4 Kanäle, 2 davon schaltbar als Differenzeingang, farbiges Display, weitere nachrüstbare oder eingebaute Optionen, wie weiter unten dargestellt („Luxusausführung“)
- Erfassen periodischer Signale mit Frequenzen bis zu etlichen GHz, Sampling-Oszilloskope (Spezialausführung)

77.2 Aufbau eines Oszilloskopes

Das Prinzipschaltbild eines Zweikanal-Oszilloskopes zeigt Abb. 77.1. Wesentliche Baugruppen sind die folgenden Funktionseinheiten: Eingangsteiler-Verstärker, Triggerung, Analog-Digital-Umsetzer, Digitalspeicher (RAM), Multiplexer-Display-Ansteuerung, Farbdisplay. Eine Steuereinheit sorgt für den geordneten Ablauf der Vorgänge. Die Bezeichnungen sind häufig in englischer Sprache angegeben, manchmal kann man die Sprache der im Display eingeblendeten Bezeichnungen wählen und über die Bedienelemente auf dem Gehäuse eine Schablone in der unterstützten Sprache kleben.

- *Eingangskopplung*: Zunächst muss der Anwender einstellen, wie das darzustellende Signal in den Eingang des Oszilloskopes gelangen soll (Eingangskopplung):
  - DC (Gleichgröße, **D**irect **C**urrent): Das Eingangssignal wird original in das Gerät eingespeist.
  - AC (Wechselgröße; **A**lternating **C**urrent): Das Eingangssignal wird über einen in Reihe geschalteten Kondensator in das Gerät eingespeist. Damit wird der Gleichanteil (arithmetischer Mittelwert) des darzustellenden Signals zu Null erzwungen, übrig bleibt der Wechselanteil. Diese Betriebsart ist vorteilhaft, wenn z. B. einer Gleichspannung von +15 V eine Wechselspannung von einigen mV überlagert ist. Dann kann diese Wechselspannung z. B. mit einem Ablenkfaktor von 2 mV/Division „bildschirmfüllend“ dargestellt werden.
  - GND: In dieser Schalterstellung (GROUND oder kurz GND) wird der Verstärkereingang intern auf den Wert 0 V gelegt (Masse, Ground, Bezugspotential, Gehäuse). Hier kann man erkennen, wo sich die Nulllinie des Eingangssignals befindet, siehe Einstellmöglichkeit *Position* weiter unten. Bei vielen Oszilloskopen wird die Nulllinie des Signals intern ermittelt und auf dem Display eingeblendet.
- *Eingangsteiler*, *Eingangsverstärker*: Der Eingangsteiler-Verstärker passt den Spannungsbereich des Eingangssignals (z. B. −24 bis +45 mV oder −56 bis +12 V) an den Spannungsbereich des Analog-Digital-Umsetzers (z. B. −1 bis +1 V) an, indem das Eingangssignal entsprechend verstärkt oder abgeschwächt wird. Der Wert der Abschwächung bzw. Verstärkung wird vom Anwender über Links-Rechts-Impulsschalter eingestellt. Das Display blendet den zugehörigen *Y-Ablenkkoeffizienten* ein, der die Zuordnung zwischen der Amplitude der Eingangsspannung (in Volt) und der auf dem Schirm des Displays dargestellten Amplitude (in cm) festlegt. Er wird in V/cm oder in V/DIV (division, kurz DIV, Teilstrich) angegeben und lässt sich häufig in den Stufungen 1-2-5 von 1 mV/DIV bis 50 V/DIV fest einstellen. Damit ist eine direkte Zuordnung zwischen der Spannungsamplitude in V und der Darstellungslänge in cm möglich. Häufig ist auf dem Display ein Raster eingeblendet, oft im cm-Abstand. Um von der Rasterlänge unabhängig zu sein, wählt man die Angabe V/*Division*.
- *Triggerung*
  - Intern: Über einen Triggerbaustein kann der Anwender festlegen, bei welchem Spannungswert (z. B. +2,7 V, eingeblendet durch eine waagerechte farbige Linie) und welcher Steigung (z. B. positiv) die Signalaufzeichnung beginnen soll.
  - Pretrigger: Häufig bietet die Triggerschaltung die Möglichkeit der *Pretriggerung*: Das Oszilloskop beginnt mit der Signalaufzeichnung bereits *vor* dem eingestellten Wert (Beispiel oben: 2,7 V, positive Steigung), wobei die Vorlaufzeit einstellt werden kann. Das hat den Vorteil, dass z. B. bei Rechtecksignalen, bei denen man gezwungenermaßen auf die positive bzw. negative Flanke triggern muss, auch diese Flanke noch dargestellt werden kann. Andernfalls greifen Aufzeichnungsbeginn und Flanke ineinander, so dass der Beginn der Flanke nicht mehr ausreichend genau dargestellt werden kann.
  - LINE: In der Regel ist für die Triggerung noch die Betriebsart LINE einstellbar. Das Gerät triggert selbsttätig auf die Netzwechselspannung. Diese Triggerart ist nützlich für die Energietechnik und für Untersuchungen bei Einkopplung von 50-Hz-Netzwechselgrößen in Signalkreise.
  - AUTO: Das Oszilloskop legt die Triggerspannung selbsttätig fest, in der Regel bei 0 V.
  - EXTERN: Es gibt Signalverläufe, die sich innerhalb jeder Periode mehr oder weniger ausgeprägt ändern. Das kann z. B. eine modulierte Trägerschwingung sein. Dann ist es erforderlich, ein geeignetes Triggersignal extern zu erzeugen.
- *Position*: Über ein Potentiometer mit der Bezeichnung *Position* (*Y*-Position oder kurz *Y*-POS) lässt sich das Oszillogramm auf dem Display in vertikaler Richtung verschieben.

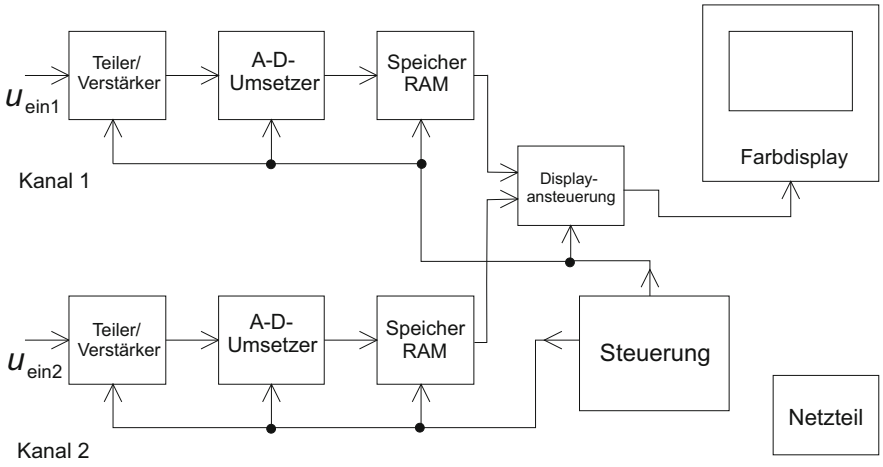


Abb. 77.1 Prinzipschaltbild eines Zweikanal-Oszilloskopes

Bei mehreren Signalen auf dem Display kann man den Ablenkfaktor und die Verschiebung so vorwählen, dass sich alle Signale untereinander in der gewünschten Reihenfolge und ohne Überlappung darstellen lassen. Allerdings geht dabei die Information darüber verloren, wo sich der Spannungswert 0 V bei den einzelnen Signalen befindet. Abhilfe schafft entweder die oben genannte Schalterstellung GND bei jedem Kanal oder es sind intern Schmitt-Trigger für jeden Kanal eingebaut, die den Wert 0 V erkennen und ihn entsprechend links als kleinen Pfeil einblenden.

- *Zeitachse (x-Achse)*
  Hier gibt es drei Betriebsarten.
  - (x-t)-Betrieb: Die Skalierung der x-Achse erfolgt linear in Zeitwerten. Angegeben wird ein Wert Zeit/Länge, wählbar von etwa 10 ns/DIV bis 10 s/DIV in den Abstufungen 1, 2, 5 innerhalb einer Dekade. Zu den Bezeichnungen gelten die gleichen Ausführungen wie unter dem Punkt „Eingangsteiler, Eingangsverstärker“ oben. Über einen Links-Rechts-Impulsschalter wird ein passender Wert eingestellt, der im Display eingeblendet wird. Intern bildet das Oszilloskop daraus die Zeitpunkte zur Bildung der Abtastwerte des A-D-Umsetzers für den Spannungswert der y-Achse. Damit gekoppelt ist die Zahl der auf dem Display dargestellten Perioden. Die Zeitpunkte für die Bildung der Abtastwerte liefert ein Frequenzgenerator, häufig in Form eines Zählers. Bei einem Ablenkfaktor von z. B. 1 ms/cm, einer Displaybreite von 10 cm und 1000 Abtastwerten pro waagerechtem (x-Achsen-)Durchlauf erhält der A-D-Umsetzer im Abstand von 10 µs einen Umsetzbefehl: $(1\,\text{ms})/(\text{cm}) \cdot (10\,\text{cm}) \cdot (1/1000) = 10\,\mu\text{s}$. Auf dem Display wird ein Zeitbereich von $(1\,\text{ms/cm}) \cdot (10\,\text{cm}) = 10\,\text{ms}$ abgebildet.
    Die Abtastung und Speicherung der y-Amplitudenwerte geschieht periodisch, und durch die Triggerung ergeben sich i. Allg. zu gleichen Zeitpunkten auf der x-Achse auch gleiche y-Achsen-Werte.
  - Single-Shot: Hiermit werden einmalige Vorgänge aufgezeichnet. Nach der Triggerung wird der Zeitbereich der x-Achse *einmal* durchlaufen und die zugehörigen Abtastwerte werden gespeichert. Ein erneutes Durchlaufen ist erst nach manuellem Rücksetzen und erneuter Triggerung möglich.
  - (x-y)-Betrieb: Das Oszilloskop besitzt einen zusätzlichen Eingang für eine Spannung mit der Bezeichnung x-Eingang mit nachfolgendem Eingangsteiler-Eingangsverstärker und nachgeschaltetem A-D-Umsetzer. Der Umsetzer erhält die gleichen Umsetzimpulse wie der A-D-Umsetzer für die y-Achse. Jedem Spannungswert auf der x-Achse ist damit ein Spannungswert auf der y-Achse eindeutig zugeordnet, beide werden gespeichert. Man erhält damit die funktionale Abhängigkeit $u_y = f(u_x)$. Die Spannung u_x kann eine periodische Spannung mit z. B. sinusförmigem oder sägezahnförmigem Verlauf sein, sie kann aber auch ein einmaliger Vorgang sein. dann muss die Frequenz der Umsetzimpulse lediglich der Frequenz des Signals u_x angepasst werden.
    Diese Betriebsart wird z. B. zur Darstellung einer Diodenkennlinie $i_D = f(u_D)$ eingesetzt. Diode und (niederohmiger) Widerstand liegen in Reihe und werden von einer periodischen Spannung u_G (häufig Sinus- oder Sägezahnverlauf) mit einer Frequenz im Bereich $(10 \ldots 1000)\,\text{Hz}$ gespeist. Die periodische Spannung gibt die Spannungswerte in x-Richtung vor, die Spannung am Widerstand, die dem Strom proportional ist, die zugeordneten Werte in y-Richtung.
- *Analog-Digital-Umsetzer*: Das in der Amplitude an den Analog-Digital-Umsetzer (A-D-Umsetzer) angepasste Signal wird digitalisiert, wobei die Abtastfrequenz vom Hersteller häufig so gewählt worden ist, dass sich für das dargestellte Signal 1000 Abtastwerte ergeben. Sie werden im Digitalspeicher (z. B. RAM) abgelegt, zusammen mit dem Zeitmaßstab und weiteren Kenngrößen. Bei Mehrkanaloszilloskopen erhält entweder jeder Kanal seinen eigenen A-D-Umsetzer mit zugehörigem Digitalspeicher, oder es stehen nur *ein* A-D-Umsetzer und *ein* Digitalspeicher zur Verfügung, die im Zeitmultiplexverfahren (siehe Kap. 100 „Mehrfachübertragung – Multiplexverfahren“) den einzelnen Kanälen nacheinander zugeordnet werden. Das beschriebene erste Verfahren ist kostenintensiver, ermöglicht aber mehr Umsetzungen pro Zeiteinheit als das zweite. Die

Grenzfrequenz des Oszilloskopes hängt entscheidend von der Zahl der Umsetzungen pro Sekunde des A-D-Umsetzers ab. Es besteht folgender prinzipieller Zusammenhang: Eine Grenzfrequenz des Oszilloskopes von 100 MHz erfordert eine Abtastfrequenz von etwas mehr als 200 MHz (Shannon, siehe Signal- und Systemtheorie, VIII Abtasttheorem), womit sich ein sinusförmiger Spannungsverlauf mit der Frequenz 100 MHz auf dem Display darstellen lässt. Soll dagegen ein unbekannter Signalverlauf mit der bekannten Frequenz 100 MHz einigermaßen originalgetreu dargestellt werden, sind erfahrungsgemäß alle Frequenzanteile bis zur 20. Harmonischen einschließlich zu erfassen, was eine Abtastfrequenz von etwas mehr als 4 GHz erfordert (2 GHz ist dann die erforderliche Grenzfrequenz des Oszilloskopes). Ein weiteres Kriterium für die Auswahl eines A-D-Umsetzers ist die bit-Zahl, mit der der digitalisierte Analogwert erfasst wird. Für die Darstellung auf dem Display (häufig 10 cm breit, 10 cm hoch) ist eine Umsetzung in einen mit 8 bit kodierten Wert durchaus ausreichend. Das ergibt einen Quantisierungsfehler von $\pm 0{,}125\,\%$. Sollen aber weitere Berechnungen (siehe „Mathematik-Option“ weiter unten) durchgeführt werden, kann ein A-D-Umsetzer mit einer höheren bit-Zahl erforderlich werden, was sich im Preis niederschlägt. Die A-D-Umsetzer sind demnach ein wesentlicher Kostenfaktor der Oszilloskope.

- *Weiterverarbeitung:* Die im Digitalspeicher oder in den Digitalspeichern abgelegten Werte können ausgelesen, in eine zur Anzeige geordnete Form gebracht, evtl. weiter aufbereitet, mit Zusatzinformationen versehen (Einblendung Cursor, Ablenkfaktoren, Triggerwerten, …) und dann dem Display zugeführt werden. Über Zusatzelektronik können die Digitalwerte auch z. B. über einen Netzwerkanschluss (LAN) in ein Netz eingespeist, abgerufen, weiterverarbeitet oder auch ausgedruckt werden. Hier bewährt sich die Maßnahme, dass zu jedem Oszillogramm auch die Kenndaten eingeblendet werden. Häufig können die Daten auch über einen USB-Anschluss auf einen Speicher-Stick übertragen werden.
- *Zusatzoptionen*: Die folgenden Optionen sind häufig insgesamt oder teilweise in die Oszilloskope der „Luxusausführung“ (s. o.) integriert:
  - Mathematik-Option: In das Gerät ist ein Rechner integriert. Damit lassen sich von dem aufgezeichneten Spannungsverlauf beispielsweise Kenngrößen wie Scheitelwert, Frequenz, arithmetischer Mittelwert und Effektivwert berechnen, statistische Kenngrößen und die Darstellung durch mathematische Funktion und ähnliches ermitteln und selbstdefinierte Berechnungen durchführen.
  - Farbige Displays: Es werden häufig LCD-Anzeigeeinheiten verwendet, wie sie auch bei Rechner-Bildschirmen verwendet werden. Bei den Mehrkanal-Oszilloskopen werden die Spannungsverläufe der einzelnen Kanäle farbig dargestellt, um die Übersichtlichkeit zu erhöhen. In das Schirmbild werden Amplituden, Ablenkfaktoren, Datum, Zeit, Triggerpegel, Nulllinien, Eingangskopplung usw. für die einzelnen Kanäle oder insgesamt ebenfalls in Farbe eingeblendet.
  - LAN-Anschlüsse: Zur Dokumentierung (Papierausdruck), Weiterverarbeitung, Speicherung und Versendung an andere Personen steht ein Netzwerkanschluss (LAN) zur Verfügung. Wegen seiner universellen und einfachen Verwendbarkeit hat er die Datenübertragung über spezielle Messdatenbussysteme teilweise abgelöst.
  - Cursoreinblendung: Mehrere Cursor für die Zeitachse (x-Achse) und die Spannungsamplitudenachse (y-Achse) ermöglichen das direkte Ablesen von Zeit- und Zeitdifferenz-Werten sowie Amplituden- und Amplitudendifferenz-Werten.
  - Tastkopf-Erkennung: Der Anschluss eines Tastkopfes bietet Vorteile (siehe weiter unten), hat aber den Nebeneffekt, dass die Spannung am Eingang des Oszilloskopes nur noch 1/10 oder 1/100 der tatsächlichen Spannung am Tastkopfeingang beträgt. Dieser verringerte Wert wird auch auf dem Schirm eingeblendet, was zu Fehlinterpretationen führen kann. Es gibt Tastköpfe mit einem Zusatzkontakt, der dem Oszilloskop (sofern dafür eingerichtet) den Teilungsfaktor mitteilt und das daraufhin die Werteeinblendung entsprechend korrigiert.
  - Differenzverstärker: Alle Signaleingänge eines Oszilloskopes haben einen gemeinsamen Bezugspunkt (Masse, Gehäuse, Abschirmung). Soll eine Spannung dargestellt werden, die keinen direkten Anschluss an diesem Bezugspunkt hat, wendet man Differenzverstärker an, siehe unten. Voraussetzung ist mindestens ein 2-Kanal-Oszilloskop, bei dem dann diese 2 Eingänge ihre Spannungswerte voneinander subtrahieren. Schaltungstechnisch erfordert diese Verstärker-Betriebsart einen erhöhten Aufwand.
  - Automatische Einstellung: Das Oszilloskop analysiert das anliegende Eingangssignal und stellt selbsttätig Amplituden- und Zeit-Ablenkfaktoren, Triggerschwelle und weitere Größen so ein, dass das Eingangssignal in der Regel nahezu optimal auf dem Schirm dargestellt wird. Diese Option ist hilfreich, wenn der Anwender Probleme mit der Einstellung des Gerätes hat. Der Leistungsumfang der Oszilloskope erfordert meist eine sehr intensive und damit lang andauernde Beschäftigung mit dem Gerät!
- *Netzteil*: Das Netzteil versorgt die Baugruppen mit Spannung und sorgt für die galvanische Trennung des Oszilloskopes vom 230-V-Energieversorgungsnetz. Es ist als Schaltnetzteil ausgelegt, das einen hohen Wirkungsgrad aufweist.

77.3 Zusatzeinrichtungen

- *Differenzverstärker*: Beim Einsatz von Mehrkanal-Oszilloskopen haben alle Eingangsspannungen das Gehäuse des Oszilloskopes als gemeinsames Bezugspotential. Ein Oszilloskop mit Differenzverstärker-Eingang (-Eingängen) ist prinzipiell ein Mehrkanal-Oszilloskop, bei dem jeweils 2 Kanäle in geeigneter Weise zusammengeschaltet werden. Die Kanäle CH 1 (Channel 1) und CH 2 (Channel 2) haben einen gemeinsamen Wahlschalter mit den Stellungen SINGLE/ADD. Einer oder beide Kanäle haben außerdem einen Wahlschalter mit den Stellungen NORM/INV (Normal/Invertiert). In der Stellung SINGLE werden beide Kanä-

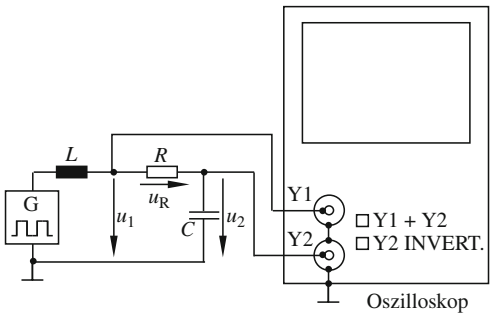


Abb. 77.2 Anwendung eines Differenzverstärkers

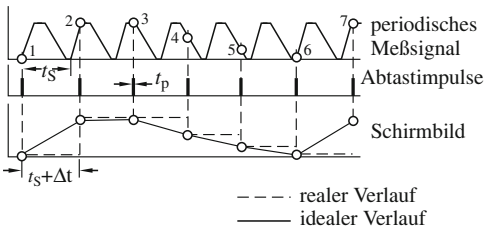


Abb. 77.3 Prinzip des Sampling-Oszilloskopes

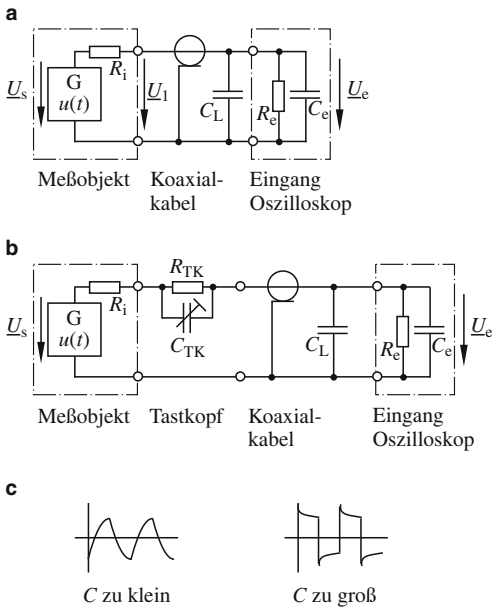


Abb. 77.4 Tastkopf. **a** Oszilloskop mit Signalquelle und Kabel **b** zusätzlich mit Tastkopf **c** dargestelltes Signal bei Tastkopfabgleich

le unabhängig voneinander verwendet, das Gerät arbeitet als Zweikanal-Oszilloskop. In der Schalterstellung INV wird das dargestellte Signal invertiert. In der Schalterstellung ADD werden CH 1 und CH 2 so verknüpft, dass die an den beiden Eingängen anliegenden Signale addiert werden. Stellt man den Schalter bei CH 2 auf INV, wird das Signal an CH 2 vom Signal an CH 1 subtrahiert. Damit lassen sich auch Signalverläufe darstellen, die keinen gemeinsamen Bezugspunkt, also eine gemeinsame Gleichtaktspannung, haben. Wichtig bei Messungen mit Differenzverstärker-Eingang ist die vom Hersteller angegebene maximal zulässige Gleichtaktspannung (COMMON-MODE-VOLTAGE).

In Abb. 77.2 soll die Spannung am Widerstand R dargestellt werden, wobei die Spannungsquelle und das Oszilloskop ein gemeinsames Bezugspotential (Masse, Gehäuse) haben, das nicht am Widerstand liegt. Die beiden Anschlüsse des Widerstandes werden mit den Eingängen CHANNEL 1 und CHANNEL 2 verbunden, die Betriebsweise wird auf ADD eingestellt und CHANNEL 2 invertiert. Dann bildet das Oszilloskop intern

$$u_1 + (-u_2) = u_R \tag{77.1}$$

und stellt damit die Spannung am Widerstand R dar.

- *Sampling-Oszilloskope*: Sampling-Oszilloskope werden dort eingesetzt, wo die Darstellung des Signals nicht mehr in Echtzeit erfolgen kann, also bei Signalfrequenzen oberhalb etwa 1000 MHz. Das Prinzip ist in Abb. 77.3 gezeigt. Das Signal wird abgetastet, und zwar mit Abtastimpulsen, deren Dauer t_p möglichst klein und deren zeitlicher Abstand t_a nur wenig größer (Δt) als die Periodendauer t_s des darzustellenden Signals ist ($t_a = t_s + \Delta t$). Die Momentanwerte der Spannungen gelangen auf ein *Abtast-Halte-Glied*, an dem ein formtreues Abbild des Originalsignals mit geringerer Frequenz steht und das dann auf dem Schirm angezeigt wird.

  In der Praxis kann der zeitliche Abstand zweier Abtastimpulse auch so gewählt werden, dass die Periodendauer t_s zwischen zwei Abtastzeitpunkten mehrmals vergangen ist: $t_a = n \cdot t_s + \Delta t$ mit $n = 1, 2, 3, \ldots$ Diese Betriebsart erfordert eine weniger kurze Abtastdauer t_p, dafür dauert ein Bildaufbau entsprechend länger.
- *Kalibrier-Spannungsquelle*: An einer zusätzlichen Anschlussbuchse mit der Bezeichnung CAL OUT kann in der Regel ein Rechtecksignal mit kalibrierter Amplitude, Tastgrad und Periodendauer entnommen werden. Häufig handelt es sich um ein symmetrisch zur Nulllinie gelegenes Rechtecksignal mit der Amplitude ± 1 V, einem Tastgrad 1 : 2 und der Frequenz 1 kHz. Damit kann, wenigstens teilweise, die Funktionstüchtigkeit des Y- und X-Verstärkers kontrolliert werden. Daneben dient dieses Signal auch zur Kalibrierung von *Tastköpfen*.
- *Tastköpfe*: Der Eingang eines Oszilloskopes stellt für die ihn speisende Quelle eine Last von etwa 1 MΩ (R_e) parallel zu etwa 50 pF (C_e) dar (Abb. 77.4a). Aus Gründen der Störunterdrückung werden die darzustellenden Signale dem Oszilloskopeingang über abgeschirmte Kabel, meist Koaxialkabel mit einem Wellenwiderstand von (50 … 75) Ω, zugeführt. Diese haben eine Kapazität von etwa (50 … 100) pF pro Meter (C_L/m). Selbst wenn man die Induktivität und die ohmschen Widerstände des Kabels vernachlässigt, bilden Leitungskapazität C_L und Eingangskapazität C_e mit dem Innenwiderstand R_i der Signalquelle einen Tiefpass.

Das Verhältnis der Spannungen $\underline{U}_e/\underline{U}_s$ berechnet sich für sinusförmige Spannungen mit $C = C_e + C_L$ zu:

$$\frac{\underline{U}_e}{\underline{U}_s} = \frac{\frac{\frac{1}{j\omega C}\cdot R_e}{\frac{1}{j\omega C}+R_e}}{\frac{\frac{1}{j\omega C}\cdot R_e}{\frac{1}{j\omega C}+R_e} + R_i} = \frac{1}{1 + R_i \cdot \frac{\frac{1}{j\omega C}+R_e}{\frac{R_e}{j\omega C}}}$$

$$= \frac{1}{1 + R_i \cdot \frac{1+R_e j\omega C}{R_e}} = \frac{1}{1 + \frac{R_i}{R_e} + R_i \cdot j\omega C} \qquad (77.2)$$

$$\frac{\underline{U}_e}{\underline{U}_s} \approx \frac{1}{1 + R_i \cdot j\omega C}$$

R_e wird mit 1 MΩ und R_i mit 50 Ω angenommen; damit ist der Summand R_i/R_e im Nenner vernachlässigbar. Die Grenzfrequenz f_g ergibt sich, wenn Imaginär- und Realteil im Nenner von (77.2) gleich sind: $1 = \omega_g \cdot C \cdot R_i$ und daraus

$$f_g = 1/(2 \cdot \pi \cdot C \cdot R_i) \quad \text{mit } C = C_L + C_e \qquad (77.3)$$

Beispiel 1

Ein Koaxialkabel von 2 m Länge und einer Kapazität von 50 pF/m ($C_L = 2\,\text{m} \cdot 50\,\text{pF/m} = 100\,\text{pF}$), eine Oszilloskop-Eingangskapazität von 50 pF und ein Signalquellen-Innenwiderstand von 50 Ω führen zu einer Grenzfrequenz nach (77.3) von etwa 21,2 MHz. Nimmt man an, dass bei nichtsinusförmigen Signalen mindestens noch die 10. Harmonische gemäß Fourierzerlegung erfasst werden muss, werden Signale mit Frequenzen oberhalb etwa 2,2 MHz bereits mehr oder weniger verfälscht dargestellt. ◀

Um diese Probleme zu umgehen, hat man die *Tastköpfe* entwickelt, deren interner Aufbau in Abb. 77.4b dargestellt ist. Der Kondensator C_{TK} ist von außen einstellbar und liegt parallel zu R_{TK}. R_{TK} mit C_{TK} und $C = C_L + C_e$ mit R_e bilden einen Spannungsteiler. Damit ist das Spannungsverhältnis:

$$\frac{\underline{U}_e}{\underline{U}_1} = \frac{\frac{\frac{1}{j\omega C}\cdot R_e}{\frac{1}{j\omega C}+R_e}}{\frac{\frac{1}{j\omega C}\cdot R_e}{\frac{1}{j\omega C}+R_e} + \frac{R_{TK}\cdot\frac{1}{j\omega C_{TK}}}{R_{TK}+\frac{1}{j\omega C_{TK}}}} = \frac{1}{1 + \frac{1+j\omega R_e C}{1+j\omega R_{TK} C_{TK}}} \qquad (77.4)$$

Dieses Verhältnis wird frequenz*un*abhängig, wenn gilt:

$$R_{TK} \cdot C_{TK} = R_e \cdot C \qquad (77.5)$$

Damit wird die Signalform eines *beliebigen* Eingangssignals nicht verändert, denn gemäß Fourierzerlegung dieses Signals werden *alle* vorkommenden Frequenzanteile um den gleichen Faktor herabgesetzt.

Da C in der Regel größer als 100 pF ist und C_{TK} als einstellbare Kapazität maximal etwa 30 pF gemacht werden kann, wird R_{TK} größer als R_e gewählt. Es werden Tastköpfe mit *Teilerverhältnissen* 10 : 1 und 100 : 1 angeboten.

Hinweise: Korrekterweise müsste für die *Teiler* der Wert 1 : 10 oder 1 : 100 angegeben werden, es wird aber der Kehrwert genommen.

Der am Oszilloskop eingestellte Y-Ablenkkoeffizient ist durch die Teilung mit dem Tastkopf nicht mehr gültig. Einige Oszilloskope erkennen den Anschluss eines Tastkopfes und geben eine entsprechende Mitteilung aus oder blenden den korrigierten Wert in die Anzeige oder den Schirm ein.

Beim Einsatz von Tastköpfen erhöht sich der Eingangswiderstand auf den Wert $R_e + R_{TK}$, während sich die Kapazität durch die Reihenschaltung von C_{TK} und $(C_L + C_e)$ verringert ($C_{TK} < C_L, C_e$).

Damit die Bedingung aus (77.5) für unterschiedliche Werte R_e und C_e von Oszilloskop-Eingängen mit dem jeweiligen Tastkopf erfüllt werden kann, schließt man an den Eingang des Tastkopfes die oben beschriebene Kalibrier-Spannungsquelle an und gleicht mit dem einstellbaren Kondensator das Bild auf dem Oszilloskopschirm auf ein Rechteck ab. Ist die Bedingung nach (77.5) *nicht* erfüllt, ergeben sich Signalverläufe, wie sie in Abb. 77.4c dargestellt sind.

Schreibende Messgeräte

78

Wilfried Plaßmann



78.1 Betriebsarten 896
78.2 Schreibstiftauslenkung 896

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_78

Die schreibenden Messgeräte werden hauptsächlich eingesetzt, um Vorgänge aufzuzeichnen und zu dokumentieren, die sich im Minuten- und Stunden-Bereich abspielen. Sie sind weitgehend durch Oszilloskope oder elektronische Geräte ähnlich einem Oszilloskop ersetzt worden, die entsprechend für Langzeitaufnahmen ausgelegt sind. Der Vorteil dieser Geräte gegenüber den schreibenden Messgeräten ist der, dass die Daten dauerhaft gespeichert sind, einschließlich der notwendigen Kenngrößen wie Zeitmaßstab, Amplitude, Datum, Aufzeichnungsgerät usw., und sich in dieser Form beliebig oft ausdrucken und vor allem direkt weiterverarbeiten lassen. Weiterhin entfallen die Probleme mit der Mechanik, dem Papier und den Schreibstiften. Deshalb folgt hier eine nur kurze Darstellung der Geräte.

78.1 Betriebsarten

Schreibende Messgeräte haben, wie ein Oszilloskop, die (y-t)- und die (x-y)-Betriebsart.

(y-t)-Betriebsart: Ein Schreibstift wird in y-Richtung proportional zur darzustellenden Spannung u von der Ruhelage in der Mitte (0 V) nach hinten ($u > 0$) oder nach vorn ($u < 0$) ausgelenkt. Unter ihm wird ein Papierstreifen von in der Regel 20 oder 30 cm Höhe (für die y-Auslenkung) und einer Länge ab 10 m bis 100 m mit konstantem einstellbarem Vorschub von einer Rolle links ab- und auf eine zweite Rolle rechts wieder aufgewickelt (x-Richtung, Zeitmaßstab bzw. Geschwindigkeit, z. B. 1 cm/min). Dazu werden meist Schrittmotoren eingesetzt. Die Rollen haben einen Abstand von ebenfalls 20 oder 30 cm, was für diese Betriebsart aber ohne Bedeutung ist. Der Schreibstift zeichnet den Spannungsverlauf auf dem vorbeiziehenden Papier und damit als Funktion der Zeit auf. Es kann sich um periodische oder einmalige Signale handeln. Die Aufzeichnungsdauer wird begrenzt durch die Vorschubgeschwindigkeit (den Zeitmaßstab) und die Länge der Papierrolle. Die Frequenz des Signals sollte unter etwa 5 Hz liegen, weil der Schreibstift als bewegtes mechanisches Gebilde wegen seiner Massenträgheit diesen Wert vorgibt. Der Einsatz mehrerer farbiger Stifte bei Mehrkanalschreibern ist möglich, aber aufwändig.

(x-y)-Betriebsart: Hierbei wird das Papier nicht bewegt, es kann auch ein Papierblatt von 20 cm mal 20 cm oder 30 cm mal 30 cm auf den Schreiber fest aufgelegt werden. Es wird ein zusätzlicher Kanal benötigt, der für die Auslenkung des Schreibstiftes in x-Richtung sorgt. Der andere, bereits vorhandene Kanal sorgt, wie bei der (y-t)-Betriebsart, wiederum für die Auslenkung in y-Richtung. Damit erhält man die Darstellung $u_y = f(u_x)$, wie beim Oszilloskop.

Die zwei Kanäle sind in ihren Eigenschaften und Einstellmöglichkeiten denen beim Oszilloskop sehr ähnlich (-AC, -DC, -GND, Eingangsteiler, Eingangsverstärker, Position, Vorschubgeschwindigkeit beim Schreiber (cm/min) entspricht dem Kehrwert der x-Achsen-Skalierung beim Oszilloskop (ms/cm)).

78.2 Schreibstiftauslenkung

Die Umsetzung der Ausgangsspannung des Y-Verstärkers allgemein und des X-Verstärkers in der x-y-Betriebsart in eine Schreibstiftauslenkung erfolgt im Wesentlichen auf zwei Arten:

- *Direkte Auslenkung*: Das Prinzip lässt sich an einem Drehspulmesswerk zeigen (Abb. 76.1, Abschn. 76.2). Die Verstärker-Ausgangsspannung wird an die Drehspule angeschlossen, verursacht einen proportionalen Strom und damit eine proportionale Zeigerauslenkung. An der Spitze des Zeigers ist der Schreibstift befestigt, der auf dem darunter vorbeilaufenden Papier den Momentanwert der Spannung markiert.
- *Kompensationsverfahren*: Der Schreibstift kann durch eine Schlittenführung über einen Elektromotor und einen Seilzug in Y-Richtung über das Papier geführt werden. An der Schlittenführung ist in voller Papierbreite eine Potentiometerbahn befestigt. Am Schreibstift sitzt der Schleifer dieses Potentiometers, so dass der abgegriffene Spannungswert der Stellung des Schreibstiftes proportional ist. Das Ausgangssignal des Y-Verstärkers und der am Potentiometer abgegriffene Spannungswert werden einem Differenzverstärker zugeführt. Ist die Differenz nicht Null, wird der Schreibstift über den Elektromotor entsprechend verstellt, bis die Differenz zu Null geworden ist. Dann hat der Schreibstift seine Sollstellung erreicht.

Der Vorteil dieses Verfahrens gegenüber dem der direkten Auslenkung liegt zum einen in einer höheren Grenzfrequenz, da das Erreichen der Sollstellung des Schreibstiftes mit relativ großer Leistung über den Elektromotor erfolgen kann. Zum anderen ist es mit einem kleineren Fehler behaftet, weil kleine Abweichungen von der Sollstellung, z. B. durch Reibung des Stiftes auf dem Papier, vom Kompensationsverfahren erkannt und korrigiert werden; bei der direkten Auslenkung fehlt diese Rückmeldung.

Für die Auslenkung des Schreibstiftes zusätzlich in X-Richtung wird ebenfalls bevorzugt das Kompensationsverfahren eingesetzt.

Digital anzeigende Messgeräte

79

Wilfried Plaßmann



79.1 Digitalvoltmeter 898
79.2 Digitalmultimeter 898
79.3 Messung von Kapazitäten, Frequenzen und Stromverstärkungen . 899
79.4 Messung von Temperaturen 899

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_79

Tab. 79.1 Vorteile von analog und digital anzeigenden Messgeräten

Analog anzeigende Messgeräte	Digital anzeigende Messgeräte
Meist keine Hilfsenergie (Batterie) erforderlich	Geringer Messfehler, vor allem bei Gleichspannungsmessung ($\|f\| < 10^{-3}$)
Tendenzen/Änderungen leicht erkennbar (einfacher Abgleich auf Extremwert)	Hoher Instrumenten-Innenwiderstand bei Spannungsmessung (10 MΩ bis 100 GΩ)
Erfassen mehrerer Messwerte gleichzeitig (Flugzeugkanzel, Leitzentrale in Kernkraftwerken)	Einfache Messwert-Speicherung und -Weiterverarbeitung
	Mehrere Messungen (10 bis 1000) pro Sekunde möglich
	Erweiterung zu Multimetern (Messung von Widerständen, Frequenzen, Kapazitäten, Temperaturen, ...)

Die im Kap. 76 behandelten analog anzeigenden Messgeräte werden zunehmend durch digital anzeigende Geräte ergänzt oder ersetzt. In der Tab. 79.1 sind einige Vorteile beider Verfahren einander gegenübergestellt.

Um die Vorteile beider Verfahren zu kombinieren, werden digital anzeigende Messgeräte mit einer zusätzlichen Balkenanzeige mit bis zu 100 Einzelelementen gefertigt. Auf diese Weise wird eine quasi-analoge Anzeige nachgebildet.

79.1 Digitalvoltmeter

Digital anzeigende Messgeräte haben in der Regel einen Analog-Digital-Umsetzer für Gleichspannungen eingebaut. Für diese Umsetzer werden, je nach messtechnischer Anforderung, unterschiedliche Verfahren verwendet. So sind diejenigen nach dem Prinzip des *Mehrrampenverfahrens* besonders für die Präzisionsmesstechnik geeignet, da sie mit maximalen relativen Messfehlern im Bereich $\pm 10^{-3}$ bis $\pm 10^{-6}$ behaftet sind. Allerdings benötigen sie für einen Messzyklus Zeiten zwischen 0,1 s (Fehler ca. $\pm 10^{-3}$) und 10 s (Fehler ca. $\pm 10^{-6}$). Für Fehler $\geq |\pm 10^{-5}|$ werden diese Umsetzer als Integrierte Bausteine gefertigt und haben häufig einen Eingangs-Spannungsbereich von 0 bis ±200 mV oder 0 bis ±2 V. Maximale Fehler und Stellenzahl der Anzeige werden in der Regel einander angepasst, so dass die Anzeige dann den Wertebereich von 0 bis ±1999 (Fehler ca. $\pm 10^{-3}$), 0 bis ±19.999 (Fehler ca. $\pm 10^{-4}$) oder 0 bis ±199.999 (Fehler ca. $\pm 10^{-5}$) erfasst. Im ersten Fall spricht man von einer $3\frac{1}{2}$-stelligen Anzeige, und entsprechend für die anderen Fälle von $4\frac{1}{2}$- oder $5\frac{1}{2}$-stelliger Anzeige. In die Anzeige muss ein Komma eingefügt werden, wenn der angezeigte Zahlenwert der Spannung in Volt entsprechen soll. Daneben gibt es auch Digitalvoltmeterbausteine mit $3\frac{3}{4}$-, $4\frac{3}{4}$- oder $5\frac{3}{4}$-stelliger Anzeige mit einem Wertebereich von 0 bis ±3999, 0 bis ±39.999 oder 0 bis ±399.999.

Zur Messung von Wechselspannungen werden zwei Verfahren eingesetzt:

a) Die zu messende Spannung wird gleichgerichtet und dann deren Scheitelwert gemessen. Anschließend wird dieser Wert durch 1,41 geteilt und angezeigt. Da aber nur für sinusförmigen Spannungsverlauf der Faktor 1,41 zwischen Effektivwert und Scheitelwert besteht, ist die Messung von Wechselspannungen auf diesen Signalverlauf beschränkt. Man findet dieses Verfahren häufig bei den Digitalvoltmetern mit $3\frac{1}{2}$- oder $3\frac{3}{4}$-stelliger Anzeige, seltener bei denen mit $4\frac{1}{2}$- oder $4\frac{3}{4}$-stelliger Anzeige.

b) Der Effektivwert wird tatsächlich gemessen. Das geschieht wiederum nach zwei Verfahren:

b1) Der *analoge* Spannungsverlauf wird quadriert, dann über eine Periode integriert, auf diese Periode bezogen, aus dem so erhaltenen Wert die Wurzel gezogen und dann angezeigt. Damit wird die mathematische Darstellung des Effektivwertes schaltungstechnisch realisiert:

$$\textit{Effektivwert} \quad U = \sqrt{\frac{1}{T}\int_0^T \{u(t)\}^2 dt} \qquad (79.1)$$

b2) Der analoge Spannungsverlauf wird pro Periode 100mal oder mehr abgetastet und über einen Analog-Digital-Umsetzer in Digitalwerte umgesetzt. Diese digital kodierten Spannungswerte werden dann, um (79.1) zu erfüllen, quadriert, die Werte einer Periode summiert, durch die Anzahl der Werte geteilt, aus dem so erhaltenen Wert die Wurzel gezogen und angezeigt. Da hier Umsetzer mit einer Umsetzzeit verwendet werden müssen, die sich aus der Periodendauer der Wechselspannung, geteilt durch die Zahl der Abtastwerte pro Periode, ergibt, kommen nur Umsetzer nach dem *Flash-Verfahren* in Frage. Diese haben Umsetzzeiten im Bereich von einigen Mikrosekunden oder weniger, die erreichten relativen Fehler liegen häufig bei $> |\pm 10^{-3}|$. Bedingt durch die zusätzlichen mathematischen Operationen nach (79.1) haben Digitalvoltmeter in den Gleichspannungsmessbereichen kleinere Fehler als in den Wechselspannungsmessbereichen.

Die Messbereiche integrierter Digitalvoltmeterbausteine können zur Messung größerer Spannungen durch ohmsche Widerstandsteiler (Abb. 80.1) erweitert werden. Der Eingangswiderstand dieser Digitalvoltmeterbausteine liegt in der Größenordnung 100 MΩ bis 1 GΩ; beim Einsatz von Widerstandsteilern zur Messbereichserweiterung wird der resultierende Eingangswiderstand des Digitalvoltmeters durch den resultierenden Widerstand dieses Teilers bestimmt. Er liegt für Gleich- und Wechselspannungen häufig bei 10 MΩ.

79.2 Digitalmultimeter

In Anlehnung an die analogen Multimeter wird auch bei Digitalvoltmetern durch Zusatzeinrichtungen die Messung von Strömen und Widerständen möglich und damit das Instrument zum Digitalmultimeter erweitert.

- *Messung von Widerständen*: In das Instrument wird eine Konstantstromquelle eingebaut, deren Strom den unbekannten Widerstand durchfließt. Die an dem Widerstand entstehende Spannung ist dem Strom proportional und wird gemessen. Der Strom der Stromquelle und der Messbereichsendwert für den jeweiligen Widerstandsmessbereich werden so gewählt, dass sich beispielsweise ein Spannungswert von 200 mV ergibt: Im 200-Ω-Messbereich wird ein Strom für die Stromquelle von 1 mA eingestellt; im Messbereich 2 kΩ entsprechend 0,1 mA. Bei einigen Messgeräten kann der Strom um den Faktor 10 höher gewählt werden (Bezeichnung „HIGH“ im Gegensatz zu „LOW“), so dass sich eine maximale Spannung von 2 V ergibt. Das ist vorteilhaft, wenn der Widerstand von Dioden oder Transistoren in Durchlassrichtung gemessen werden soll, da dann auch die Schwellspannung von Siliziumhalbleitern überschritten wird.
  Die Messbereiche sind meist dekadisch gestuft und reichen von 200 Ω bis 20 MΩ. Der Fehler des Messgerätes bei Widerstandsmessungen liegt deutlich höher als bei Gleichspannungsmessungen, da die Stromquelle mit einem entsprechend höheren Fehler behaftet ist.
  Bei der Messung von Widerständen $< 10\,\Omega$ gehen sowohl die Zuleitungswiderstände zum Messinstrument als auch die Art des Anschlusses am zu messenden Widerstand in das Messergebnis ein. Um dieses Problem zu umgehen, bieten Digitalmultimeter für Widerstands-Präzisionsmessungen die *Vierpolmessung* an. Dazu muss der Widerstand Potentialanschlüsse haben, an denen die Spannung über dem Widerstand unabhängig von der Art des Anschlusses und der Länge der Messleitungen gemessen wird. Der Strom wird in die Stromanschlüsse eingespeist. Damit ist die Länge der Zuleitungen ohne Bedeutung. Besitzt der Widerstand keine Potentialanschlüsse, kann man versuchen, sie durch Anklemmen oder Anlöten zweier zusätzlicher Anschlüsse nachzubilden.
- *Messung von Strömen*: Sie wird auf eine Spannungsmessung zurückgeführt, indem der zu messende Strom I einen in das Instrument eingebauten bekannten Widerstand R_0 durchfließt. Damit ist die gemessene Spannung U dem Strom I proportional ($U = I \cdot R_0$). Die Messbereiche sind in der Regel mit dem Faktor 100 gestuft und ergeben sich damit zu 2 mA, 200 mA und 20 A. Die Widerstände haben die Werte 100, 1 und 0,01 Ω, damit sich beim Messbereichsendwert eine Spannung von 200 mV einstellt. Der 20-A-Messbereich ist in der Regel über getrennte Buchsen erreichbar und über eine Feinsicherung abgesichert. Da Sicherungen in den anderen niedrigeren Strommessbereichen nicht vorgesehen sind, besteht die Gefahr einer Überlastung der Widerstände durch zu hohe Ströme, was entweder zu ihrer Zerstörung oder aber zur Änderung ihres Wertes und damit zur Erhöhung des Messfehlers führen kann.
  Es können sowohl Gleich- als auch Wechselströme gemessen werden. Da die Strommessung auf eine Spannungsmessung zurückgeführt wird, gelten bei der Messung von Wechselströmen die im Abschn. 80.4 „Messung von Wechselspannungen“ aufgeführten Verfahren und Probleme hier gleichermaßen. Der Messfehler ist bei Gleichstrommessungen um ein bis zwei Größenordnungen größer als bei Gleichspannungsmessungen, weil die Widerstände einen entsprechend höheren Fehler aufweisen, der sich durch Überlastung noch vergrößern kann. Bei der Wechselstrommessung kommen noch zusätzlich die bereits bei der Messung von Wechselspannungen gegenüber Gleichspannungen vorhandenen höheren Messfehler hinzu.

79.3 Messung von Kapazitäten, Frequenzen und Stromverstärkungen

Viele Digitalmultimeter sind noch zusätzlich mit einer oder mehreren der folgenden Messmöglichkeiten ausgestattet.

- *Frequenzmessung*: Für die Frequenzmessung wird der für die Analog-Digital-Umsetzung erforderliche interne Taktgeber, meist ein Quarzoszillator, zur Bereitstellung der Torzeit des Zählers herangezogen. Der Zähler selbst ist wegen des Mehrrampenverfahrens bei der Analog-Digital-Umsetzung ebenfalls schon im Baustein vorhanden. Als Messbereiche stehen häufig 20 und 200 kHz zur Verfügung, manchmal auch zusätzlich 2 MHz; der relative Fehler liegt bei ca. ±1 %.
- *Kapazitätsmessung*: Zur Messung der Kapazität eines Kondensators wird die für die Widerstandsmessung erforderliche Stromquelle benutzt. Bei Konstantstromladung eines Kondensators C mit dem Strom I_0 über eine festgelegte Zeit Δt ergibt sich eine Spannungszunahme ΔU am Kondensator. Den Zusammenhang zwischen diesen Größen erhält man über die dem Kondensator zugeführte Ladung Q ($Q = C \cdot \Delta U = I_0 \cdot \Delta t$). Daraus folgt für C: $C = I_0 \cdot \Delta t / \Delta U$. Die Messbereiche sind häufig zu 2 nF, 200 nF und 2 µF gewählt. Der Fehler liegt in der Größenordnung ($\pm 1 \ldots \pm 5$) %.
- *Stromverstärkung*: Für die Messung der *Stromverstärkung* h_{21} von Transistoren wird in die Basis des Transistors ein Strom aus der für Widerstandsmessungen erforderlichen Stromquelle eingespeist und der Kollektorstrom gemessen. Wählt man den Strom der Stromquelle zu 0,01 mA und wird der Kollektorstrom zu 1,53 mA gemessen, muss nur das Komma in der Anzeige unterdrückt werden, damit die Stromverstärkung 153 angezeigt wird.

Für die hier angesprochenen Messungen von Kapazitäten, Frequenzen und Stromverstärkungen gilt, dass der Fehler meist im Bereich von ±1 % bis ±5 % liegt, und zwar relativ unabhängig von der Anzahl der Stellen des Digitalmultimeters.

79.4 Messung von Temperaturen

Manchmal ist einem Digitalmultimeter eine Temperaturmesseinrichtung in Form eines Widerstandsthermometers mit Platin-Widerstandselement (Nennwert meist 1000 oder 10.000 Ω) beigefügt. Zur Messung der Temperatur wird entweder die bereits mehrfach erwähnte Stromquelle oder eine Brückenschaltung eingesetzt. Daneben werden auch Thermoelement-Temperaturaufnehmer eingesetzt. Die Vergleichsstelle (Ver-

gleichstemperatur) wird durch die Umgebungstemperatur gebildet, so dass die Messgeräte nur in einem eng begrenzten Umgebungstemperaturbereich betrieben werden sollen (Beispiel: (+10 … +30) °C; siehe dazu Betriebsanleitung des Messgerätes). Innerhalb dieses Temperaturbereiches findet häufig eine Korrektur über die Temperaturabhängigkeit der Schwellspannung einer Diode statt. Die maximal messbare Temperatur liegt hier höher (ca. (400 … 600) °C) als bei den Platin-Widerstandsthermometern (ca. 300 °C).

Einzelheiten dazu im Abschn. 81.1.6 „Thermische Aufnehmer“.

Messverfahren zur Messung elektrischer Größen

80

Wilfried Plaßmann



80.1 Messung von Gleichspannungen 902
80.2 Messung von Gleichströmen 903
80.3 Messbereichserweiterung 904
80.4 Messung von Wechselspannungen 904
80.5 Messung von Wechselströmen 906
80.6 Widerstands- und Impedanzmessung 906
80.7 Leistungsmessung 911
80.8 Messung der Arbeit 916
80.9 Messung von L, C, Gütefaktor und Verlustfaktor 917
80.10 Messung magnetischer Größen 918

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_80

80.1 Messung von Gleichspannungen

Die Messung von Gleichspannungen kann mit analog oder digital anzeigenden Messgeräten erfolgen.

80.1.1 Analog anzeigende Spannungsmessgeräte

Je nach maximal zulässigem Fehler werden entweder Dreheisen-Messinstrumente der Klassen 1, 2,5 und 5 oder Drehspul-Messinstrumente der Klassen 0,2, 0,5 und 1, für Präzisionsmessungen der Klassen 0,05 und 0,1 eingesetzt (Definition des Begriffs „Klasse“ in Abschn. 75.5). Die Messinstrumente sind entweder für nur einen Messbereich ausgelegt oder verfügen durch den Einbau einer Messbereichserweiterung über mehrere Messbereiche. Die am häufigsten anzutreffenden Stufungen bei mehreren Messbereichen sind 1-3-10-30 ... und 2-5-20-50 ...

Bei der Spannungsmessung wird das Messinstrument von einem der Spannung proportionalen Strom durchflossen, der wiederum den proportionalen Zeigerausschlag bewirkt. Dieser Strom wird dem Messkreis entnommen und verändert den Wert der zu messenden Spannung. Um die Belastung durch das Messinstrument erfassen zu können, wird häufig dessen *Kennwiderstand* r_K in Ω/V angegeben. Er muss mit dem Messbereichsendwert multipliziert werden, um den Widerstand des Instrumentes (Innenwiderstand) zu erhalten. Bei Drehspulmessinstrumenten liegt der Kennwiderstand im Bereich von $20\,\mathrm{k\Omega/V}$ bis $200\,\mathrm{k\Omega/V}$, bei Dreheisenmessinstrumenten dagegen nur bei etwa $200\,\Omega/\mathrm{V}$ (siehe auch Beispiel 1).

80.1.2 Digital anzeigende Spannungsmessgeräte

Digital anzeigende Spannungsmesser (Digitalvoltmeter) besitzen in der Regel mehrere Messbereiche, die dekadisch gestuft sind: 0,2, 2, 20, 200 und 1000 V oder 0,4, 4, 40, 400 und 1000 V. Als höchster Messbereich wird 1000 V gewählt, weil es sich bei Spannungen größer 1000 V um Hochspannung handelt und damit besondere Isolationsmaßnahmen erforderlich werden.

Der Innenwiderstand ist unabhängig vom gewählten Messbereich und wird vom Eingangsspannungsteiler zur Messbereichserweiterung gebildet (Abb. 80.1). Er liegt häufig bei $10\,\mathrm{M\Omega}$, bei Präzisionsmessgeräten ist er ausnahmsweise im kleinsten Messbereich durch Abtrennen des Eingangsspannungsteilers erheblich größer, z. B. $10\,\mathrm{G\Omega}$. Der Eingangswiderstand des Analog-Digital-Umsetzers ist $\geq 10\,\mathrm{G\Omega}$ und beeinflusst damit nicht das Teilerverhältnis des Widerstandsteilers beim Umschalten des Messbereiches. Der Messfehler liegt erfahrungsgemäß bei $\leq |\pm 5 \cdot 10^{-3}|$, wenn Geräte der unteren Preisklasse unberücksichtigt bleiben.

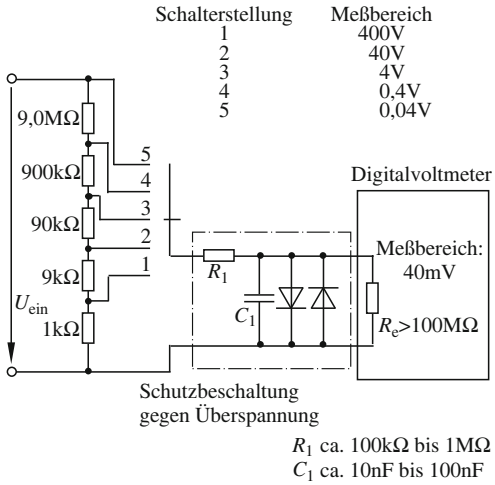


Abb. 80.1 Typischer Eingangsspannungsteiler eines Digitalvoltmeters

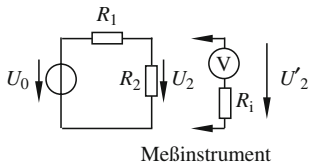


Abb. 80.2 Messabweichung bei der Spannungsmessung

80.1.3 Messabweichung durch den Innenwiderstand des Spannungsmessers

Ist der Innenwiderstand des Spannungsmessers (analog oder digital anzeigend) nicht unendlich groß, wird die zu messende Spannung durch das Messinstrument verändert. In der Schaltung Abb. 80.2 soll die Spannung über dem Widerstand R_2 gemessen werden. Das Instrument hat den Innenwiderstand R_i. Ohne angeschlossenes Messinstrument beträgt die Spannung U_2 an R_2:

$$U_2 = U_0 \cdot R_2/(R_1 + R_2)$$

mit angeschlossenem Messinstrument:

$$U'_2 = U_0 \cdot (R_2 \parallel R_i)/(R_1 + R_2 \parallel R_i)$$

Die gemessene Spannung ist damit kleiner als der tatsächliche Wert ohne Instrument. Die daraus resultierende systematische Messabweichung lässt sich berechnen, indem U_2 und U'_2 in die Formel für den relativen Fehler eingesetzt werden:

$$f = (x_a - x_r)/x_r = (U'_2 - U_2)/U_2 = \frac{U'_2}{U_2} - 1$$

$$= \frac{U_0 \cdot \frac{R_i \cdot R_2}{R_i + R_2}}{\frac{R_i \cdot R_2}{R_i + R_2} + R_1} \cdot \frac{R_1 + R_2}{U_0 \cdot R_2} - 1$$

Teil X

Nach Gleichungsumformung folgt daraus:

Systematische Messabweichung

$$f = -\frac{1}{1 + \frac{R_i}{R_i \| R_2}} \tag{80.1}$$

Die durch das Messinstrument verursachte Messabweichung wird umso kleiner, je größer dessen Innenwiderstand R_i im Verhältnis zur Parallelschaltung der Widerstände R_i und R_2 ist. Deshalb gilt allgemein:

Der Innenwiderstand von Spannungsmessern sollte so groß wie möglich sein!

Beispiel 1

Ein analog anzeigendes Messinstrument mit dem Kennwiderstand $r_K = 100\,\mathrm{k\Omega/V}$ hat im Messbereich 0,3 V einen Innenwiderstand von $100\,\mathrm{k\Omega/V} \cdot 0{,}3\,\mathrm{V} = 30\,\mathrm{k\Omega}$; im Messbereich 300 V einen Innenwiderstand von $100\,\mathrm{k\Omega/V} \cdot 300\,\mathrm{V} = 30\,\mathrm{M\Omega}$. Ein Digitalvoltmeter hat den messbereichsunabhängigen Innenwiderstand 10 MΩ. Dann ist der Innenwiderstand des analog anzeigenden Messgerätes in den Messbereichen > 100 V größer als der des Digitalvoltmeters. ◄

80.1.4 Spannungsmessung mit dem Kompensator

Die Messabweichung durch den endlichen Innenwiderstand des Spannungsmessers lässt sich vermeiden, wenn der für die Messung notwendige Strom von einer Hilfsenergiequelle geliefert wird. In Abb. 80.3 ist das Prinzip gezeigt. Die Spannung U_N ist bekannt, sie wird z. B. von einem Normalelement geliefert. Ihr Innenwiderstand R_{iN} ist ebenfalls bekannt. Der Widerstandsteiler ist ein Präzisionsteiler mit dem Gesamtwiderstand R_N und dem einstellbaren Teilwiderstand R_K. Er wird häufig als einstellbarer Stufenwiderstand mit mehreren Dekaden ausgeführt. Die zu messende Spannung ist U_x, der Innenwiderstand ist R_i. Zur Messung wird der Abgriff am Widerstand R_N solange verändert, bis der Nullindikator den Strom Null anzeigt. Dann gilt:

$$U_x = U_a = U_K = U_N \cdot R_K/(R_{iN} + R_N) \tag{80.2}$$

Mit diesem Verfahren ist es möglich, auch die Leerlaufspannung von Spannungsquellen mit sehr großem Innenwiderstand zu messen. Für die zu messende Spannung gilt: $U_x < U_N$.

Der Fehler des Verfahrens hängt davon ab, wie genau die Spannung U_N, deren Innenwiderstand R_{iN} und die Widerstände R_K und R_N bekannt sind und wie empfindlich der Nullindikator ist.

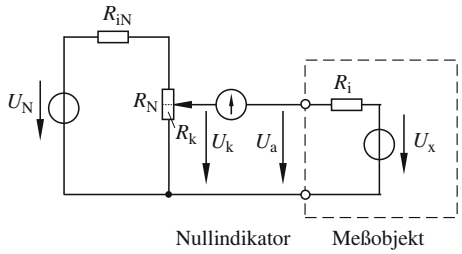


Abb. 80.3 Prinzip des Kompensators

Mit Präzisionskompensatoren lassen sich Spannungen mit einem Fehler $f < |\pm 10^{-4}|$ messen.

Allerdings verliert dieses Messverfahren an Bedeutung, weil Präzisions-Digitalvoltmeter im 0,2-V- und 2-V-Messbereich Fehler in der gleichen Größenordnung aufweisen und häufig Innenwiderstände im Bereich 1 GΩ bis 10 GΩ besitzen. Damit kann die Messabweichung nach (80.1) meist vernachlässigt werden.

80.2 Messung von Gleichströmen

80.2.1 Analog anzeigende Strommessgeräte

Drehspulmesswerke benötigen einen Strom von etwa 10 µA bis 100 µA für Vollausschlag. Der Innenwiderstand liegt bei 100 Ω bis 10 kΩ, so dass sich im kleinsten Strommessbereich ein Spannungsfall über dem Messwerk von 10 bis 100 mV ergibt. Für größere Ströme wird dem Messwerk ein entsprechend kleinerer Widerstand parallelgeschaltet (Prinzip Stromteiler).

Der Fehler der Strommessung ist durch die Klasse des Instrumentes gegeben und entspricht dem Fehler bei der Spannungsmessung.

80.2.2 Digital anzeigende Strommessgeräte

Der zu messende Strom durchfließt einen bekannten Widerstand und verursacht an ihm einen Spannungsfall. Besitzt der Widerstand dekadische Werte (1, 10, 100 Ω, ...), ist der Zahlenwert der gemessenen Spannung durch entsprechendes Setzen eines Kommas in der Anzeige als Stromwert in µA, mA oder A ablesbar. Meist wird eine Stufung in Anlehnung an die Spannungsmessbereiche, aber mit dem Multiplikator 100, verwendet (200 µA, 20 mA, 2 A). Der Fehler der Strommessung ist mit $f \geq |\pm 10^{-2}|$ deutlich höher als bei der Spannungsmessung.

Für ein $4\frac{1}{2}$-stelliges Digitalmultimeter beträgt der Spannungsfall beim Maximalstrom unabhängig vom Messbereich typisch 200 mV, der Fehler liegt bei ±(0,4 % of reading + 3 LSB), siehe dazu auch Kap. 79.

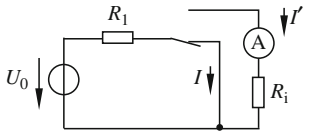


Abb. 80.4 Messabweichung bei der Strommessung

80.2.3 Messabweichung durch den Innenwiderstand des Strommessers

In Abb. 80.4 soll der Strom durch den Widerstand R_1 gemessen werden.

Ohne Strommesser ergibt sich der Strom durch den Widerstand R_1 zu: $I = U_0/R_1$, mit Strommesser dagegen zu $I' = U_0/(R_1 + R_i)$. Eingesetzt in die Fehlerformel folgt daraus:

$$f = (x_a - x_r)/x_r = (I' - I)/I$$
$$= \frac{\frac{U_0}{R_i+R_1} - \frac{U_0}{R_1}}{\frac{U_0}{R_1}} = -\frac{1}{1 + \frac{R_1}{R_i}} \qquad (80.3)$$

Die Messabweichung durch den endlichen Innenwiderstand des Strommessers ist umso kleiner, je kleiner er im Verhältnis zum Widerstand R_1 ist.

Der Innenwiderstand von Strommessern sollte möglichst klein sein!

80.3 Messbereichserweiterung

80.3.1 Spannungsmessung

Soll mit einem analog oder digital anzeigenden Spannungsmesser, dessen Innenwiderstand in einem Messbereich (Wahl des Messbereichs siehe unten) mit R_{iU} gegeben ist, eine höhere Spannung U_x gemessen werden, schaltet man einen Widerstand R_v in Reihe zu dem Spannungsmesser (Abb. 80.5a). Dieser Spannungsteiler liefert die Beziehung:

$$U_x/U_U = (R_{iU} + R_v)/R_{iU}$$

Darin ist U_U die am Instrument abgelesene Spannung. Da man U_x im Allgemeinen nicht kennt, sind R_v und der Messbereich des Spannungsmessers so einzustellen, dass sich ein Ausschlag oder eine Anzeige möglichst im oberen Drittel des Messbereichs ergibt, da dann der Fehler klein ist. In der Praxis wird man für R_v einen Stufenwiderstand mit mehreren Dekaden wählen, den höchsten Wert einstellen und dann den Widerstand stufenweise verringern, bis der genannte Ausschlag erreicht ist. Achtung: Für den Dekadenwiderstand muss eine Ausführung mit unterbrechendem Schalter verwendet werden, da der Spannungsmesser sonst im Umschaltmoment durch den kurzgeschlossenen Reihenwiderstand überlastet wird!

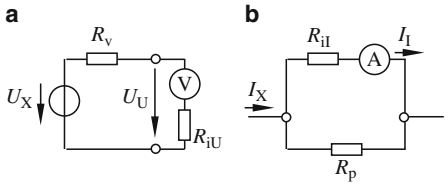


Abb. 80.5 Messbereichserweiterung. **a** Spannungsmessung, **b** Strommessung

80.3.2 Strommessung

Hier wird zum Instrument ein Widerstand R_p parallel geschaltet (Abb. 80.5b), sofern ein größerer Strom als dem Messbereich entsprechend gemessen werden soll. R_{iI} ist der Innenwiderstand des Strommessers in einem Messbereich (Wahl siehe unten), der zu messende Strom ist I_x. Ist I_I der am Strommesser abgelesene Strom, folgt mit dem Stromteiler:

$$\frac{I_x}{I_I} = \frac{R_{iI}}{R_{iI} \parallel R_p}$$

In der Praxis wird man für R_p einen Stufenwiderstand mit mehreren Dekaden wählen, den kleinsten Wert einstellen und diesen solange erhöhen, bis sich ein Ausschlag oder eine Anzeige im oberen Drittel des Messbereichs ergibt. Achtung: Für den Dekadenwiderstand muss eine Ausführung mit nichtunterbrechendem Schalter verwendet werden, da der Strommesser sonst im Umschaltmoment durch den fehlenden Parallelwiderstand überlastet wird!

80.4 Messung von Wechselspannungen

80.4.1 Analog anzeigende Wechselspannungsmessgeräte

Zur Messung von Wechselspannungen werden Messgeräte mit Dreheisenmesswerk, Drehspulmesswerk und Thermoumformermesswerk eingesetzt.

80.4.1.1 Spannungsmesser mit Dreheisenmesswerk

Die den Zeigerausschlag verursachende Kraft lässt sich durch die Lorentzkraft darstellen ($F = B \cdot I \cdot I$). Die Flussdichte B wird ebenfalls durch den Strom I erzeugt, so dass $F = k_1 \cdot I^2 = k_2 \cdot U^2$ gesetzt werden kann. Durch die mechanische Trägheit des Zeigers wird daraus der Mittelwert der Spannung bei Frequenzen $\gg 1$ Hz gebildet. Berücksichtigt man die Wurzelbeziehung beim Aufbringen der Skalenteilung, wird theoretisch der Effektivwert U angezeigt:

$$\textit{Effektivwert} \quad U = \sqrt{\frac{1}{T}\int_0^T \{u(t)\}^2 dt} \qquad (80.4)$$

In der Praxis gilt das aber nur für sinusförmige Spannungen mit Frequenzen bis etwa (400 ... 1000) Hz. Oberhalb dieser Frequenzen nimmt der Zeigerausschlag ab. Ein Grund dafür ist die mit steigender Frequenz zunehmende Impedanz ($j\omega L$) der Spule. Bei nichtsinusförmigen Spannungen wird der Signalverlauf nach Fourier zerlegt. Dabei muss die höchste noch zu berücksichtigende Signalfrequenz, das ist in der Regel die 10. bis 20. Harmonische, an der oberen Grenze des genannten Frequenzbereiches liegen. Auf diese Weise können bereits nichtsinusförmig verlaufende Spannungen mit einer Frequenz von 50 Hz mit einem erhöhten Messfehler durch Verfälschung der Kurvenform behaftet sein.

80.4.1.2 Spannungsmesser mit Drehspulmesswerk

Das Drehspulmesswerk zeigt den arithmetischen Mittelwert der Wechselspannung an. In der Praxis interessiert aber der Effektivwert, weil er für die Leistung maßgebend ist. Da aber arithmetischer Mittelwert und Effektivwert zahlenmäßig nicht übereinstimmen und der Umrechnungsfaktor zwischen beiden von der Kurvenform abhängt, ist dieses Instrument nicht zur Messung von Wechselspannungen geeignet. Es besitzt jedoch besondere Vorteile, nämlich geringen Eigenverbrauch und lineare Skalenteilung. Deshalb hat man, allerdings nur für sinusförmige Spannungen, die Skala so beschriftet, dass der korrekte Wert abgelesen wird. Da der arithmetische Mittelwert eines sinusförmigen Signals Null ist, wird der Spannungsverlauf mit einem Einweg- oder einem Brückengleichrichter so verändert, dass der arithmetische Mittelwert ungleich Null ist und damit ein Ausschlag erfolgt. Der *Kurvenformfaktor* oder kurz *Formfaktor* ist das Verhältnis von Effektivwert zu arithmetischem Mittelwert bei Einweg- bzw. Brückengleichrichtung. Er beträgt bei sinusförmigem Verlauf und Einweggleichrichtung 2,22 und bei Brückengleichrichtung 1,11. Mit diesen Faktoren werden die Werte auf der Skala multipliziert. Es sind Spannungen mit Frequenzen bis zu einigen 10 kHz messbar.

> Messinstrumente mit Drehspulmesswerk und Gleichrichter zeigen nur bei sinusförmigen Wechselspannungen den Effektivwert richtig an!

80.4.1.3 Spannungsmesser mit Thermoumformermesswerk

Bei diesem Messwerk durchfließt ein der Spannung proportionaler Strom einen sehr dünnen (einige µm) Widerstandsdraht. Der Draht erwärmt sich; seine Temperaturdifferenz zur Umgebungstemperatur ist dem Effektivwert des Stromes proportional. Die Temperatur wird mit einem Thermoelement gemessen (Abschn. 81.1.6).

Der Vorteil dieses Verfahrens liegt darin, dass der Frequenzbereich für sinusförmige Spannungen von etwa 0,2 Hz bis 65 MHz reicht, bei Sonderausführungen bis 300 MHz bzw. 10 GHz. Die untere Frequenzgrenze ist dadurch gegeben, dass die Temperatur des Widerstandsdrahtes unterhalb 0,2 Hz dem Momentanwert des Stromes folgt. Gleichspannungen sind als dem Strom proportionale Größe messbar. Ein Nachteil ist, dass der Strom durch den Widerstandsdraht einige mA betragen muss, um Drahttemperaturen oberhalb 100 °C und damit ausreichend große Spannungen am Thermoelement zu bewirken. Damit der Messkreis nicht unzulässig hoch belastet wird, hat man elektronische Lösungen gefunden, die den Strom durch den Widerstandsdraht einer Hilfsspannungsquelle entnehmen, die durch die zu messende Spannung gesteuert wird.

80.4.2 Digital anzeigende Wechselspannungsmessgeräte

Zur Messung des Effektivwertes einer Spannung werden unterschiedliche Messverfahren eingesetzt.

1. Verfahren: Vor allem bei Digitalmultimetern der unteren Preisklasse wird bei der Spannungsmessung der Scheitelwert gemessen, durch 1,41 geteilt und angezeigt. Diese Umrechnung zwischen Scheitel- und Effektivwert gilt aber nicht allgemein für beliebige Spannungsverläufe, so dass der Effektivwert nur bei sinusförmigem Spannungsverlauf richtig angezeigt wird. Die Scheitelwertmessung ist wegen des hohen Eingangswiderstandes des Analog-Digital-Umsetzers einfach möglich. Verglichen mit der Gleichspannungsmessung ist der Eingangswiderstand gleich, der Fehler ist um den Faktor 5 bis 10 größer.

2. Verfahren: Der Effektivwert wird tatsächlich gemessen; deshalb spricht man auch von einer „echten" Effektivwertmessung. Verwendet werden elektronische Schaltkreise, die nach (80.4) den analogen Spannungswert am Eingang zunächst quadrieren, diesen dann integrieren, auf die Periodendauer beziehen und anschließend die Wurzel ziehen. Der Fehler dieses Verfahrens ist $> | \pm 10^{-3}|$, weil die erforderlichen mathematischen Operationen u. a. durch Logarithmieren und Potenzieren realisiert werden.

Vor allem bei Geräten der Präzisionsmesstechnik wird die zu messende Spannung pro Periode 100mal oder mehr abgetastet und in Digitalwerte umgesetzt. Anschließend werden diese digital kodierten Werte den erforderlichen Rechenoperationen nach (80.4) zur Bildung des Effektivwertes unterzogen. Der so bestimmte Effektivwert ist mit einem kleineren Fehler behaftet als beim rein analogen Verfahren, erfordert aber einigen Aufwand. Es ist ein Analog-Digital-Umsetzer mit 8 bit oder mehr und einer Umsetzzeit erforderlich, die nur ein Bruchteil der Periodendauer der zu messenden Spannung ist (Umsetzzeit = Periodendauer/Abtastwerte pro Periode).

Bei Messgeräten mit „Echt-Effektivwertmessung" gibt es Ausführungen, die zwar den Effektivwert tatsächlich messen, an das zu messende Signal aber die Bedingung knüpfen, dass der arithmetische Mittelwert Null ist. Das liegt daran, dass im Eingangskreis des Messgerätes ein Kondensator in Reihe zu den Messanschlüssen eingefügt ist. Dadurch vereinfacht sich zwar der elektronische Aufwand innerhalb des Messgerätes, schränkt aber seinen Einsatz wesentlich ein. Nicht immer weist die Betriebsanleitung eindeutig auf diese Tatsache hin.

Eine wichtige Kenngröße bei der Messung von Wechselspannungen und -strömen ist der *Crestfaktor* als Verhältnis von Scheitelwert zu Effektivwert. Er darf einen bestimmten Wert nicht überschreiten und liegt in der Größenordnung 4 bis 10. Bei einem Crestfaktor von 4 kann demnach eine rechteckförmige Spannung mit einem Tastgrad 5 nur noch mit erhöhtem Fehler gemessen werden.

80.5 Messung von Wechselströmen

80.5.1 Analog anzeigende Wechselstrommessgeräte

Hier werden die gleichen Messwerke verwendet wie bei der Wechselspannungsmessung. Da bei ihnen der Ausschlag stromproportional ist, können die Aussagen über die Wechselspannungsmessung entsprechend übernommen werden.

Strommesser mit Dreheisenmesswerk: Theoretisch wird der Effektivwert unabhängig von der Kurvenform angezeigt, in der Praxis dagegen wird der Zeigerausschlag bei Frequenzen oberhalb 400 bis 1000 Hz geringer. Die Innenwiderstände sind gleich denen bei der Gleichstrommessung.

Strommesser mit Drehspulmesswerk: Die Anzeige ist ebenfalls nur für sinusförmige Ströme richtig, die Frequenzgrenze für sinusförmige Ströme liegt auch bei einigen 10 kHz. Die Innenwiderstände entsprechen denen bei der Gleichstrommessung, sofern die Instrumente einen elektronischen Präzisionsgleichrichter eingebaut haben.

Strommesser mit Thermoumformermesswerk: Die Innenwiderstände entsprechen denen bei der Gleichstrommessung, der Frequenzbereich reicht von 0,2 Hz bis ca. 65 MHz oder 300 MHz bzw. 10 GHz.

80.5.2 Digital anzeigende Wechselstrommessgeräte

Die Wechselstrommessung wird, wie bei der Gleichstrommessung, auf eine Spannungsmessung zurückgeführt. Deshalb sind die Innenwiderstände in beiden Fällen gleich. Entsprechend gilt analog zur Wechselspannungsmessung:

Der Fehler ist größer als bei der Wechselspannungsmessung, weil der Fehler des Strommesswiderstandes in das Messergebnis eingeht.

80.6 Widerstands- und Impedanzmessung

Für die Widerstands- und Impedanzmessung werden hauptsächlich die Brückenschaltung, der Vergleich mit einem bekannten Widerstand und die kombinierte Messung aus Strom und Spannung eingesetzt.

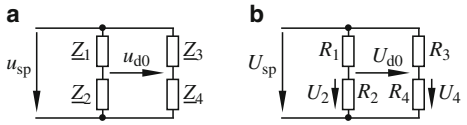


Abb. 80.6 Messbrücke. **a** allgemein, **b** Wheatstone-Messbrücke für ohmsche Widerstände

80.6.1 Gleichstrom-Messbrücken zur Widerstandsmessung

Eine Messbrücke besteht nach Abb. 80.6a aus vier Impedanzen oder Widerständen Z_1 oderR_1 bis Z_4 oder R_4 und wird von einer Spannungsquelle mit der Spannung u_{sp} gespeist. Zur Messung von Widerständen kann u_{sp} sowohl eine Gleich- als auch eine Wechselspannung sein, zur Impedanzmessung wird in der Regel eine sinusförmige Spannung verwendet. Man spricht dann von Gleich- oder Wechselstrom-Messbrücken.

80.6.1.1 Grundlagen

Mit der Gleichstrom-Messbrücke lassen sich nur ohmsche Widerstände messen. Eventuell vorhandene induktive Anteile (in Reihe liegend durch die wendelförmige Widerstandsbahn-Anordnung) ergeben den Wert Null, kapazitive Anteile (2 voneinander isolierte Folien) den Wert unendlich.

Die Spannung U_{d0} in der Brückendiagonalen (Abb. 80.6b) ergibt sich zu:

$$U_{d0} = U_2 - U_4 \tag{80.5}$$

R_1 und R_2 sowie R_3 und R_4 bilden einen Spannungsteiler. Der Strom in der Brückendiagonalen ist Null (Index „0“ für U_{d0}), und es werden folgende Teilspannungen abgegriffen:

$$U_2 = U_{sp} \cdot \frac{R_2}{R_1 + R_2}; \quad U_4 = U_{sp} \cdot \frac{R_4}{R_3 + R_4} \tag{80.6}$$

Für U_{d0} folgt daraus:

$$\begin{aligned} U_{d0} &= U_{sp} \cdot \left(\frac{R_2}{R_1 + R_2} - \frac{R_4}{R_3 + R_4} \right) \\ &= U_{sp} \cdot \frac{R_2R_3 + R_2R_4 - R_1R_4 - R_2R_4}{(R_1 + R_2)(R_3 + R_4)} \\ &= U_{sp} \cdot \frac{R_2R_3 - R_1R_4}{(R_1 + R_2)(R_3 + R_4)} \end{aligned} \tag{80.7}$$

Für die Anwendung der Brückenschaltung zur Widerstandsmessung gibt es zwei Verfahren: 1. *Abgleichverfahren*: Die Spannung U_{d0} wird durch Abgleich zu Null gemacht. 2. *Ausschlagverfahren*: Die Spannung U_{d0} ist ein Maß für die Abweichung eines zu messenden Widerstandes von einem Sollwert.

80.6.1.2 Wheatstone-Messbrücke im Abgleichverfahren

Der unbekannte Widerstand ist R_1. Mindestens einer der Widerstände R_2 bis R_4 wird solange verändert, bis U_{d0} Null ist. Dann

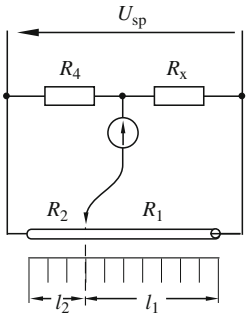


Abb. 80.7 Schleifdraht-Messbrücke

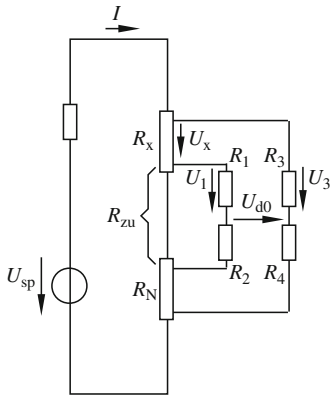


Abb. 80.8 Thomson-Messbrücke

gilt nach (80.7):

$$R_1 = R_2 \frac{R_3}{R_4} \tag{80.8}$$

Die Brücke ist abgeglichen. In der Praxis ist R_2 stetig veränderbar, während das Verhältnis R_3/R_4 in dekadischen Stufen eingestellt werden kann und damit der Messbereich für den zu messenden Widerstand gewählt wird. Die Messunsicherheit dieser Messbrücken ist $< |\pm 10^{-3}|$. Sie lässt sich auf $< |\pm 10^{-4}|$ verringern, wenn auch für den Widerstand R_2 Stufenwiderstände eingesetzt werden, die mehrere Dekaden überstreichen und deren kleinste Stufung bei 1 oder 0,1 Ω liegt. Es können Widerstände im Bereich von etwa 10 Ω bis 10 MΩ gemessen werden.

Bei der *Schleifdraht-Messbrücke* nach Abb. 80.7 werden die Widerstände R_1 und R_2 durch einen homogenen Widerstandsdraht mit konstantem Querschnitt gebildet. Damit sind die Widerstände den jeweiligen Drahtlängen l proportional.

$$R_1 = \frac{\rho \cdot l_1}{A}; \quad R_2 = \frac{\rho \cdot l_2}{A} \tag{80.9}$$

ρ spezifischer elektrischer Widerstand in Ω · m, l Drahtlänge in m, A Drahtquerschnittsfläche in m^2.

Daraus folgt die Abgleichbedingung:

$$R_x = R_4 \frac{R_1}{R_2} = R_4 \cdot \frac{l_1}{l_2} \tag{80.10}$$

Der Schleifdraht wird häufig entweder als ausgestreckter Draht von 1 m Länge oder als Schleifdrahtwendel auf einem walzenförmigen Körper ausgeführt.

80.6.1.3 Thomson-Messbrücke im Abgleichverfahren

Die Messung von Widerständen kleiner 10 Ω erfordert die Anwendung von *Potentialanschlüssen*, um den Einfluss von Übergangswiderständen an den Anschlussklemmen zu verringern, da diese in der Größenordnung der zu messenden Widerstände liegen. Die dann verwendete *Thomson*-Messbrücke ist nach Abb. 80.8 aufgebaut. R_x ist der zu messende Widerstand, R_N ein bekannter Widerstand. R_1 bis R_4 sind im Vergleich zu R_x hochohmig. Für die Parallelschaltung aus den Zuleitungswiderständen R_{zu} und $(R_1 + R_2)$ wird R_p eingeführt:

$$R_p = \frac{(R_1 + R_2) \cdot R_{zu}}{R_1 + R_2 + R_{zu}} \tag{80.11}$$

Damit erhält man für die Brückendiagonalspannung U_{d0}:

$$\begin{aligned} U_{d0} &= U_3 - U_x - U_1 \\ &= I(R_x + R_p + R_N) \cdot \frac{R_3}{R_3 + R_4} - I \cdot R_x - I \cdot R_p \cdot \frac{R_1}{R_1 + R_2} \\ U_{d0} &= I\left(\frac{R_x \cdot R_3}{R_3 + R_4} + \underbrace{\frac{R_p \cdot R_3}{R_3 + R_4}}_{A} + \frac{R_N \cdot R_3}{R_3 + R_4} - R_x - \underbrace{\frac{R_p \cdot R_1}{R_1 + R_2}}_{B} \right) \end{aligned} \tag{80.12}$$

Die *Thomson*-Brücke ist für $U_{d0} = 0$ abgeglichen. Diese Bedingung wird unabhängig von R_p und damit von R_{zu}, wenn sich die beiden oben gekennzeichneten Summanden A und B aufheben:

$$\frac{R_3}{R_3 + R_4} = \frac{R_1}{R_1 + R_2} \tag{80.13}$$

Das lässt sich z. B. durch eine entsprechende mechanische Kopplung der Widerstände R_1 und R_3 sowie R_2 und R_4 erreichen.

Setzt man die Abgleichbedingung aus (80.13) in (80.12) ein, so folgt durch Auflösen des Ausdruckes nach dem unbekannten Widerstand R_x:

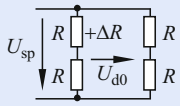
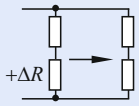
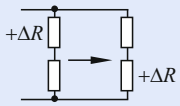
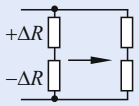
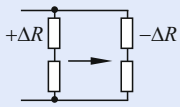
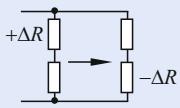
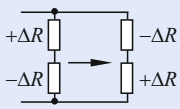
$$R_x = R_N \frac{R_3}{R_4} \tag{80.14}$$

Es können Widerstände bis herab zu etwa 10^{-5} Ω gemessen werden.

80.6.1.4 Wheatstone-Messbrücke im Ausschlagverfahren

Bei der Messbrücke im Ausschlagverfahren wird die Spannung U_{d0} gemessen. Dann lässt sich der Widerstand R_1 nach

Tab. 80.1 Viertel-, Halb- und Vollbrücken mit $+\Delta R$ oder $\pm\Delta R$

Brückenanordnung	Exakt	Näherung
	$U_{d0} = -U_{sp}\frac{\Delta R}{4R+2\Delta R}$	$\approx -U_{sp}\frac{\Delta R}{4R}$
	$U_{d0} = U_{sp}\frac{\Delta R}{4R+2\Delta R}$	$\approx U_{sp}\frac{\Delta R}{4R}$
	$U_{d0} = -U_{sp}\frac{\Delta R}{2R+\Delta R}$	$\approx -U_{sp}\frac{\Delta R}{2R}$
	$U_{d0} = -U_{sp}\frac{\Delta R}{2R}$	
	$U_{d0} = -U_{sp}\frac{2R\cdot\Delta R}{4R^2-(\Delta R)^2}$	$\approx -U_{sp}\frac{\Delta R}{2R}$
	$U_{d0} = -U_{sp}\frac{\Delta R(2R-\Delta R)}{4R^2-(\Delta R)^2}$	$\approx -U_{sp}\frac{\Delta R}{2R}$
	$U_{d0} = -U_{sp}\frac{\Delta R}{R}$	

(80.7) bestimmen, wenn R_2 bis R_4 bekannt sind. Im Gegensatz zum Abgleichverfahren ist jetzt aber der Widerstand R_1 veränderlich. Als Beispiel dient ein temperaturabhängiger Widerstand, bei dem die Temperatur über den Widerstandswert erfasst wird. In der Praxis ergibt sich häufig die Forderung, dass eine Widerstandsänderung gegenüber einem Arbeitspunkt (Sollwert) gemessen werden soll. Für diesen Fall wird die Brücke so ausgelegt, dass die vier Widerstände im Arbeitspunkt gleich sind, die Brücke also abgeglichen ist (temperaturabhängiger Widerstand bei 0 °C). Die Abweichungen des Widerstandes vom Arbeitspunkt (Sollwert) werden dann zweckmäßigerweise mit ΔR bezeichnet.

Setzt man in (80.7) $R_2 = R_3 = R_4 = R$ und $R_1 = R + \Delta R$ ein, so folgt für U_{d0}:

$$\begin{aligned} U_{d0} &= U_{sp}\cdot\left(\frac{R^2-(R+\Delta R)\cdot R}{(R+\Delta R+R)\cdot 2R}\right) \\ &= U_{sp}\cdot\frac{-\Delta R\cdot R+R^2-R^2}{(2R+\Delta R)\cdot 2\cdot R} \\ U_{d0} &= U_{sp}\cdot\frac{-\Delta R}{(4R+2\cdot\Delta R)} \end{aligned} \qquad (80.15)$$

ΔR ist *mit Vorzeichen* einzusetzen! Das bedeutet, dass bei einer Widerstands*abnahme* von R_1, d. h. $R_1 - \Delta R$, die Spannung U_{d0} positiv ist (Abb. 80.6b). Die Dimensionierung $R_1 = R_2 =$

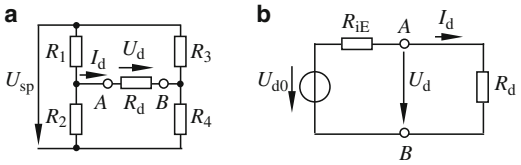


Abb. 80.9 Brückenersatzschaltung mit Widerstand in der Brückendiagonalen. **a** Schaltbild, **b** Ersatzschaltbild

$R_3 = R_4 = R$ für $\Delta R = 0$ ist sinnvoll, weil hierfür gilt: Die Spannung in der Brückendiagonalen ist für $\Delta R \neq 0$ maximal. Zwischen der Spannung U_{d0} und der Widerstandsänderung ΔR besteht kein linearer Zusammenhang. Je nach Anwendungsfall und zulässiger Messabweichung kann unter Umständen mit einem linearen Zusammenhang gerechnet werden:

$$U_{d0} \approx U_{sp}\frac{-\Delta R}{4R} \quad \text{für } |\Delta R| \ll R \qquad (80.16)$$

Die beschriebene Brücke wird auch als *Viertelbrücke* bezeichnet, weil nur *ein* Widerstand veränderlich ist. In der Praxis werden Brückenschaltungen häufig so eingesetzt, dass zwei oder vier Widerstände veränderlich sind und Widerstandsänderungen mit positivem und negativem Vorzeichen gleichzeitig auftreten. Für diese *Halb-* oder *Vollbrücken* sind die Änderungen in positiver und negativer Richtung in der Regel gleich groß ($|-\Delta R| = |+\Delta R|$). In Tab. 80.1 sind die verschiedenen Brückenanordnungen zusammengestellt. Ist der Zusammenhang zwischen U_{d0} und ΔR nichtlinear, kann er häufig durch eine lineare Abhängigkeit angenähert werden.

80.6.1.5 Wheatstone-Messbrücke im Ausschlagverfahren mit Widerstand in der Brückendiagonalen

Bei der Berechnung der Spannung U_{d0} wurde der Strom in der Brückendiagonalen bisher zu Null angenommen. Ist dagegen ein Widerstand R_d vorhanden (Abb. 80.9a), empfiehlt sich zur Berechnung der Spannung in der Brückendiagonalen das Verfahren mit Ersatzspannungsquelle und Ersatzinnenwiderstand, bezogen auf die Anschlusspunkte A und B (Abb. 80.9b). Die Ersatzspannung U_E ist die Leerlaufspannung und damit die Spannung U_{d0}, die für einen unendlich großen Widerstand R_d berechnet worden ist (80.15). Der Ersatzinnenwiderstand R_{iE} ergibt sich, wenn die Spannungsquelle „kurzgeschlossen" und der Gesamtwiderstand zwischen A und B bestimmt wird. Er lässt sich direkt ablesen.

$$R_{iE} = (R_1 \parallel R_2) + (R_3 \parallel R_4) \qquad (80.17)$$

Für den Fall, dass alle 4 Widerstände gleich groß sind (R) und dass eventuelle Widerstandsänderungen $\pm\Delta R$ unberücksichtigt bleiben, was wegen der kleinen Werte zulässig ist, folgt für R_{iE}:

$$R_{iE} = (R \parallel R) + (R \parallel R) = R \qquad (80.18)$$

Wird jetzt ein Widerstand R_d in die Brückendiagonale zwischen die Anschlüsse A und B eingefügt, berechnen sich der Strom I_d

in R_d und die Spannung U_d zu:

$$I_d = \frac{U_{d0}}{R_{iE} + R_d}; \quad U_d = I_d \cdot R_d \tag{80.19}$$

Diese Art der Berechnung ist dann von Vorteil, wenn bei der Messung im Ausschlagverfahren in der Brückendiagonalen empfindliche Strommesser, z. B. Galvanometer oder Digitalvoltmeter im kleinsten Spannungsmessbereich eingesetzt werden.

80.6.2 Wechselstrom-Messbrücken zur Widerstands- und Impedanzmessung

Wechselstrommessbrücken dienen dazu, entweder ohmsche Widerstände oder Impedanzen zu messen. Die die Brücke speisende Spannung hat zweckmäßigerweise sinusförmigen Verlauf, weil nach Fourier keine Oberschwingungen auftreten und die Werte der Impedanzen eindeutig angegeben werden können. Außerdem ermöglichen Wechselspannungsverstärker für *eine* Frequenz eine erhöhte Störspannungsunterdrückung (siehe auch folgender Abschnitt).

80.6.2.1 Messung von ohmschen Widerständen

Die Brückenschaltung kann sowohl im Abgleich- als auch im Ausschlagverfahren eingesetzt werden. Es wird vorausgesetzt, dass die Widerstände keinen Blindanteil enthalten. Dann können (80.5) bis (80.8) und (80.14) mit dem Zusatz übernommen werden, dass alle Spannungen sinusförmigen Verlauf haben, also auch $\underline{U}_{d0}$. Besonders bei der Verwendung der Brückenschaltung im Ausschlagverfahren liegen die auszuwertenden Spannungen in der Brückendiagonalen häufig im Bereich µV … mV und sind aufgrund von Störspannungen, die in der gesamten Messeinrichtung auftreten, schwierig zu messen. Ist dagegen u_{d0} sinusförmig mit der Frequenz f_0, wird eine Messeinrichtung verwendet, die nur Signale der Frequenz f_0 auswertet und damit die Wirksamkeit von Störspannungen und Rauschen auf einen sehr begrenzten Frequenzbereich um f_0 herum einschränkt. Driftvorgänge beispielsweise durch Temperaturänderungen oder Alterung besitzen Frequenzen im Bereich unterhalb 1 Hz und treten damit nicht mehr störend in Erscheinung. Die Frequenz f_0 wird in der Regel zu 420 Hz, 5 kHz oder 10 kHz gewählt. Die Frequenz von 420 Hz wird eingesetzt, damit man etwa in der Mitte zwischen der 7. und 8. Oberschwingung der Netzfrequenz von 50 Hz liegt (Fourierzerlegung z. B. von nichtsinusförmigen Verbraucherströmen am 50-Hz-Netz).

80.6.2.2 Messung von Impedanzen

Mit der Wechselstrommessbrücke lassen sich Impedanzen messen, die aus einer Induktivität L oder einer Kapazität C *mit* oder *ohne* ohmschen Anteil R bestehen. Dabei kann R parallel oder in Reihe zu L oder C angeordnet sein. Die Spannung $\underline{U}_{d0}$ in der Brückendiagonalen ist nur dann Null, wenn sie nach Betrag und Phase Null ist. Damit muss sowohl ein Betrags- als auch ein Phasenabgleich durchgeführt werden. Messbrücken dieser Art besitzen also immer zwei Abgleichelemente, die wechselseitig solange einzustellen sind, bis die Spannung $\underline{U}_{d0}$ Null ist. Da zwei Abgleichbedingungen vorhanden sind, können auch zwei unbekannte Größen für die zu messende Impedanz bestimmt werden.

Angewendet wird dieses Verfahren zur Bestimmung der Elemente von realen Induktivitäten und Kapazitäten. Das Ersatzschaltbild einer realen Induktivität besteht aus der Reihenschaltung der Induktivität mit einem ohmschen Widerstand, der im Wesentlichen durch den Widerstand der Wicklung verursacht wird. Ein realer Kondensator wird durch den Kondensator mit einem parallelgeschalteten Widerstand dargestellt, wobei der Widerstand den Isolationswiderstand erfasst. Neben diesen anschaulichen Ersatzschaltbildern werden noch zwei weitere, weniger anschauliche, verwendet: In Reihe zu C wird ein Widerstand R bzw. parallel zu L ein Widerstand R_p geschaltet. Beide Darstellungsarten lassen sich ineinander umrechnen.

Für den Abgleichfall gilt prinzipiell die gleiche Ableitung wie bei der Gleichstrom-Messbrücke mit dem Unterschied, dass an die Stelle der Widerstände teilweise oder insgesamt Impedanzen treten. Zur Berechnung setzt man für die Impedanzen die Summe aus einem Wirk- und einem Blindanteil ein:

$$\begin{aligned} \underline{Z}_1 &= R_1 + jX_1; \quad \underline{Z}_2 = R_2 + jX_2; \\ \underline{Z}_3 &= R_3 + jX_3; \quad \underline{Z}_4 = R_4 + jX_4 \end{aligned} \tag{80.20}$$

Die Abgleichbedingung lautet:

$$\underline{Z}_1 \cdot \underline{Z}_4 = \underline{Z}_2 \cdot \underline{Z}_3$$

$$\Rightarrow \quad (R_1 + jX_1) \cdot (R_4 + jX_4) = (R_2 + jX_2) \cdot (R_3 + jX_3) \tag{80.21}$$

$$\begin{aligned} &R_1 \cdot R_4 - X_1 \cdot X_4 + j(R_1 \cdot X_4 + R_4 \cdot X_1) \\ &\quad = R_2 \cdot R_3 - X_2 \cdot X_3 + j(R_3 \cdot X_2 + R_2 \cdot X_3) \end{aligned} \tag{80.22}$$

Real- und Imaginärteil müssen jeweils Null sein.

$$\begin{aligned} \text{Realteil} = 0: &\quad R_1 \cdot R_4 - X_1 \cdot X_4 = R_2 \cdot R_3 - X_2 \cdot X_3 \\ \text{Imaginärteil} = 0: &\quad R_1 \cdot X_4 + R_4 \cdot X_1 = R_3 \cdot X_2 + R_2 \cdot X_3 \end{aligned} \tag{80.23}$$

Damit können beispielsweise Realteil R_1 und Imaginärteil X_1 von $\underline{Z}_1$ aus den bekannten Größen R_2 bis R_4 und X_2 bis X_4 bestimmt werden.

Ist der reale Kondensator als Parallelschaltung aus R_p und $1/j\omega C_p$ gegeben, lässt sie sich in eine äquivalente Reihenschaltung aus R_r (Realteil) und C_r (Imaginärteil) umrechnen und in (80.20) bis (80.23) einsetzen. Dazu wird der Ausdruck der Parallelschaltung konjugiert komplex erweitert und in die Summe

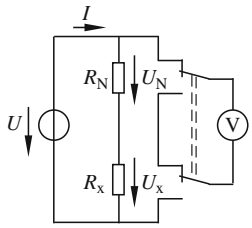


Abb. 80.10 Widerstandsmessung durch Spannungsvergleich

aus Real- und Imaginärteil zerlegt:

$$\underline{Z} = \frac{R_p \cdot \frac{1}{j\omega C_p}}{R_p + \frac{1}{j\omega C_p}} = \frac{R_p}{R_p j\omega C_p + 1} = \frac{R_p\left(1 - j\omega R_p C_p\right)}{R_p^2\omega^2 C_p^2 + 1}$$

$$\underline{Z} = \frac{R_p}{R_p^2\omega^2 C_p^2 + 1} - j\frac{R_p^2\omega C_p}{R_p^2\omega^2 C_p^2 + 1}$$

$$= \frac{R_p}{R_p^2\omega^2 C_p^2 + 1} + \frac{1}{j\omega} \cdot \underbrace{\left[\frac{R_p^2\omega^2 C_p}{R_p^2\omega^2 C_p^2 + 1}\right]}_{\frac{1}{C_r}}$$

$$\Rightarrow \quad R_r = \frac{R_p}{R_p^2\omega^2 C_p^2 + 1}; \quad C_r = \frac{R_p^2\omega^2 C_p^2 + 1}{R_p^2\omega^2 C_p} \qquad (80.24)$$

Die Abgleichbedingungen können auch in der Exponentialform dargestellt werden.

$$\underline{Z} = R + jX = |\underline{Z}| \cdot e^{j\varphi} = Z \cdot e^{j\varphi} \qquad (80.25)$$

$$\Rightarrow \quad Z_1 \cdot Z_4 = Z_2 \cdot Z_3 \quad \text{und} \quad \varphi_1 + \varphi_4 = \varphi_2 + \varphi_3 \qquad (80.26)$$

Die Anwendungen von Messbrücken zur Widerstands- und Impedanzmessung werden in den Abschn. 81.1 und 81.2 gezeigt.

80.6.3 Vergleich mit bekanntem Widerstand – Spannungsvergleich

Das vor allem in der Präzisionsmesstechnik angewendete Verfahren ist in Abb. 80.10 dargestellt. Ein bekannter Widerstand R_N und der unbekannte Widerstand R_x sind in Reihe geschaltet und werden vom gleichen Strom durchflossen. Dann ergeben sich die folgenden Beziehungen:

$$U_N = I \cdot R_N; \quad U_x = I \cdot R_x$$

ineinander eingesetzt folgt:

$$R_x = R_N \frac{U_x}{U_N} \qquad (80.27)$$

Voraussetzung bei dieser Ableitung ist, dass der Innenwiderstand des Spannungsmessers groß gegenüber den Widerständen R_N und R_x ist. Kann er nicht vernachlässigt werden, ist die hierdurch verursachte Messabweichung umso kleiner, je mehr die zwei Widerstände in ihrem Wert übereinstimmen.

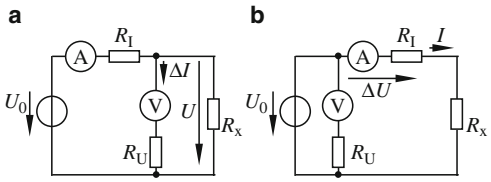


Abb. 80.11 Widerstandsbestimmung durch Strom- und Spannungsmessung. **a** spannungsrichtige Schaltung **b** stromrichtige Schaltung

80.6.4 Messung von Strom und Spannung

Werden an einem unbekannten Widerstand Strom und Spannung gleichzeitig gemessen, kann der Widerstandswert mit dem Ohm'schen Gesetz berechnet werden. Abb. 80.11 zeigt das Verfahren. Zur Erfassung von U und I sind zwei Messschaltungen möglich.

In der *spannungsrichtigen Schaltung* (Abb. 80.11a) wird die am Widerstand anliegende Spannung U richtig gemessen, der Strommesser erfasst zusätzlich den Strom ΔI durch den Spannungsmesser mit dem Innenwiderstand R_U. Die hierdurch verursachte systematische relative Messabweichung berechnet sich wie folgt:

Angezeigter (gemessener) Wert $x_a = R_x \parallel R_U$; richtiger Wert $x_r = R_x$; eingesetzt folgt:

$$f = \frac{R_x \parallel R_U}{R_x} - 1 = \frac{R_x \cdot R_U}{R_x + R_U} \cdot \frac{1}{R_x} - 1 = \frac{R_U - R_x - R_U}{R_x + R_U}$$

systematische relative Messabweichung

$$f = \frac{1}{1 + \frac{R_U}{R_x}} \qquad (80.28)$$

Die Messabweichung ist als systematische Messabweichung korrigierbar und umso kleiner, je größer der Innenwiderstand R_U des Messgerätes im Verhältnis zu dem zu messenden Widerstand R_x ist. Damit eignet sich dieses Verfahren besonders zum Messen kleiner Widerstände.

In der *stromrichtigen Schaltung* (Abb. 80.11b) wird der Strom richtig gemessen, der Spannungsmesser misst zusätzlich den Spannungsfall ΔUam Innenwiderstand R_i des Strommessers. Die relative systematische Messabweichung berechnet sich zu:

Angezeigter (gemessener) Wert $x_a = R_i + R_x$; richtiger Wert $x_r = R_x$.

Setzt man diese Ausdrücke in die Formel für die relative Messabweichung ein, folgt:

Relative Messabweichung

$$f = \frac{R_i + R_x}{R_x} - 1 = \frac{R_i}{R_x} \qquad (80.29)$$

Die Messabweichung ist als systematische Messabweichung korrigierbar und umso kleiner, je größer der unbekannte Widerstand R_x im Verhältnis zum Innenwiderstand R_i des Strommessers ist. Deshalb ist dieses Verfahren bevorzugt zur Messung großer Widerstände einsetzbar.

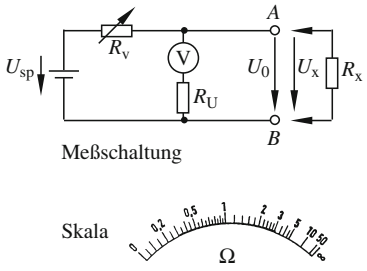


Abb. 80.12 Widerstandsmessung nach dem Prinzip des Spannungsvergleiches

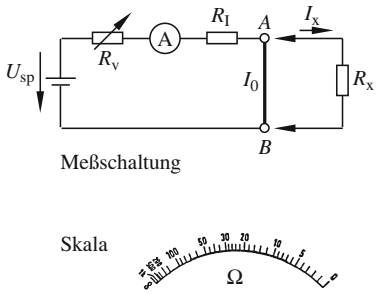


Abb. 80.13 Widerstandsmessung nach dem Prinzip des Stromvergleiches

80.6.5 Widerstandsmessung mit analogen Multimetern

Analoge Multimeter bieten die Möglichkeit, den Wert von Widerständen wenigstens überschlägig zu messen. Dazu werden zwei Verfahren benutzt.

Spannungsvergleich oder *Parallelverfahren*: Zunächst wird, siehe Abb. 80.12, bei Leerlauf an den Messklemmen A und B ($R_x \to \infty$) mit dem Potentiometer R_v Vollausschlag am Messinstrument eingestellt. R_U ist der Innenwiderstand des Messinstrumentes, U_{sp} wird von einer eingebauten Batterie geliefert. Dann gilt für U_0:

$$U_0 = U_{sp} \cdot \frac{R_U}{R_U + R_v} \tag{80.30}$$

Wird der unbekannte Widerstand R_x angeschlossen, ergibt sich für die Spannung U_x:

$$U_x = U_{sp} \cdot \frac{R_U \parallel R_x}{R_U \parallel R_x + R_v} \tag{80.31}$$

Löst man (80.31) nach R_x auf und setzt für U_{sp} den entsprechend umgeformten Ausdruck aus (80.30) ein, so folgt

$$R_x = \frac{R_U \cdot R_v}{R_U + R_v} \cdot \frac{\frac{U_x}{U_0}}{1 - \frac{U_x}{U_0}} \tag{80.32}$$

Mit den Abkürzungen $\beta = U_x/U_0$ und $R_U \parallel R_v = R_{iU}$ ergibt sich (80.33)

$$R_x = R_{iU} \cdot \frac{\beta}{1 - \beta} \tag{80.33}$$

In Abb. 80.12 ist auch der daraus folgende Skalenverlauf dargestellt. Für $U_x = U_0$ (Vollausschlag) geht $R_x \to \infty$; für $U_x = 0$ ist $R_x = 0$. Eine überschlägige Messung von R_x ist im Bereich $0{,}1 \cdot R_{iU} < R_x < 10 \cdot R_{iU}$ möglich.

Durch entsprechende Wahl von R_v und R_{iU} (z. B. Vorwiderstände) lassen sich mehrere „Messbereiche“ einstellen. Nach dem Abtrennen des zu messenden Widerstandes ist das Instrument auszuschalten, weil sonst ein ständiger Strom durch das Instrument fließt und die Lebensdauer der Batterie herabsetzt.

Stromvergleich oder *Reihenverfahren*: Bei diesem Verfahren (Abb. 80.13) wird der Strommesser bei Messbeginn und bei Kurzschluss zwischen den Messklemmen A und B durch Verändern des Widerstandes R_v auf Vollausschlag eingestellt. Dann ergibt sich für I_0:

$$I_0 = \frac{U_{sp}}{R_v + R_I} \tag{80.34}$$

R_I ist der Innenwiderstand des Strommessers und U_{sp} die Spannung einer eingebauten Batterie. Mit dem unbekannten Widerstand R_x zwischen den Anschlüssen A und B fließt dann der Strom

$$I_x = \frac{U_{sp}}{R_v + R_I + R_x} \tag{80.35}$$

Gleichung (80.35) wird nach R_x aufgelöst, und mit (80.34) folgt:

$$R_x = (R_v + R_I)\frac{1 - \frac{I_x}{I_0}}{\frac{I_x}{I_0}} \tag{80.36}$$

Mit den Abkürzungen $I_x/I_0 = \beta$ und $(R_v + R_I) = R_{iI}$ ergibt sich:

$$R_x = R_{iI}\frac{1 - \beta}{\beta} \tag{80.37}$$

Der sich daraus ergebende Skalenverlauf ist ebenfalls in Abb. 80.13 dargestellt.

Der unbekannte Widerstand R_x lässt sich überschlägig im Bereich $0{,}1 \cdot R_{iI} < R_x < 10 \cdot R_{iI}$ messen. Nach Abtrennen des zu messenden Widerstandes ist der Strom durch das Instrument Null, so dass die Lebensdauer der Batterie nicht dadurch verkürzt wird, dass das Instrument nicht ausgeschaltet wurde.

80.7 Leistungsmessung

Neben der Wirkleistung P sind vor allem in der Energietechnik noch die Blindleistung Q und die Scheinleistung S von Bedeutung. Zwischen ihnen besteht die Beziehung

$$S^2 = P^2 + Q^2 \tag{80.38}$$

Hier wurde vorausgesetzt, dass die Spannungs- und Stromverläufe sinusförmig sind.

80.7.1 Wirkleistungsmessung

Die Wirkleistung P ist für beliebige Spannungs- und Stromverläufe definiert über

$$\textit{Wirkleistung} \quad P = \frac{1}{T}\int_0^T u(t)\cdot i(t)\,dt \tag{80.39}$$

Für den Fall, dass Spannung und Strom sinusförmigen Verlauf haben, ergibt sich

$$P = U\cdot I\cdot\cos\varphi \tag{80.40}$$

U, I Effektivwerte von Spannung und Strom in V bzw. A, $\cos\varphi$ Phasenwinkel zwischen U und I.

Ein positiver Winkel bedeutet, dass die Spannung dem Strom voreilt. Damit besitzen ideale Induktivitäten einen Phasenwinkel $\varphi = 90°$, ideale Kapazitäten $\varphi = -90°$.

Die Leistungsmessung erfolgt entweder analog mit einem elektrodynamischen Messwerk oder durch Auswertung der Beziehung nach (80.39) durch elektronische Baugruppen. Die Klemmenbezeichnungen sind DIN 43807 entnommen und in Abb. 80.14 eingetragen. Die vom Strom durchflossene Spule wird im Schaltzeichen als waagerechter, durchgehender, dicker Leiter eingezeichnet und als *Strompfad* bezeichnet, der senkrechte dünnere Strich entspricht der Spannungsspule und heißt *Spannungspfad*.

80.7.1.1 Wirkleistungsmessung bei Wechselstrom

Strom und Spannung werden nach Abb. 80.14 gleichzeitig gemessen, so dass sich, wie bei der Widerstandsmessung im Abschn. 80.6.4, Abb. 80.11, zwei mögliche Schaltungen ergeben: *Stromrichtige Schaltung* (Abb. 80.14a) und *spannungsrichtige Schaltung* (Abb. 80.14b). Die durch die Messschaltung entstehenden systematischen Messabweichungen nach (80.28) und (80.29) sind auch hier gültig. Die Lastimpedanzen $\underline{Z}_L$ sind in der Regel vergleichsweise klein, so dass man nach Abschn. 80.6.4 die spannungsrichtige Schaltung verwenden sollte. In der Praxis dagegen wird vorwiegend die stromrichtige Schaltung eingesetzt, weil je ein Spannungs- und ein Strompfadanschluss am Leistungsmesser gekennzeichnet sind (Stern, Pfeil, ...) und die Vereinbarung gilt, dass die gekennzeichneten Anschlüsse zu verbinden und an einen der Außenleiter L1, L2 oder L3 anzuschließen sind. Diese Regelung ist erforderlich, um bei der Aron-Schaltung eine Aussage über das Vorzeichen der gemessenen Leistung zu bekommen (Abschn. 80.7.1.4).

Beispiel 2

Ein Leistungsmesser mit den Messbereichen 100 V und 5 A wird im Spannungspfad mit 50 V und im Strompfad mit 10 A betrieben. Damit wird der Strompfad überlastet, obwohl der Leistungsmesser gerade Vollausschlag zeigt. ◄

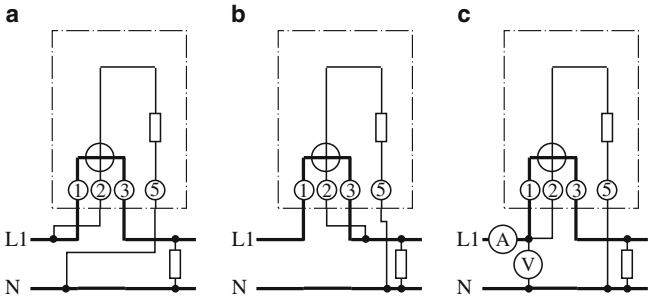


Abb. 80.14 Leistungsmesserschaltungen. **a** Stromrichtig, **b** spannungsrichtig, **c** vollständige Messschaltung

Um die Überlastung eines Leistungsmessers ohne eingebauten Überlastschutz erkennen zu können, wird die *vollständige Messschaltung* nach Abb. 80.14c eingesetzt. Dazu schaltet man einen Strommesser in Reihe zum Strompfad und einen Spannungsmesser parallel zum Spannungspfad. Diese *Kontrollinstrumente* können die Klasse 2,5 oder 5 besitzen, da sie nur zum Erkennen der ungefähren Größe von Spannung und Strom dienen und gegebenenfalls eine Überlastung anzeigen. Da die Netzspannung relativ konstant ist, wird oft auf den zusätzlichen Spannungsmesser verzichtet. Leistungsmesser mit elektrodynamischem Messwerk besitzen in der Regel keinen Überlastschutz, so dass hier häufig die vollständige Messschaltung zum Einsatz kommt. Wird die Leistung auf elektronischem Wege gemessen, erübrigt sich der Einsatz der Kontrollinstrumente dann, wenn diese Geräte eine eingebaute Überlasterkennung besitzen.

Hinweis: Die Begriffe „spannungsrichtig“ und „stromrichtig“ beziehen sich durchgehend auf die Verbraucherseite.

80.7.1.2 Wirkleistungsmessung in Drehstromsystemen

Drehstromsystem ist die übliche Bezeichnung für ein dreiphasiges Wechselstromsystem. Die drei Spannungen haben untereinander einen Phasenwinkel von 120° bzw. 240°. Hier ist zu unterscheiden, ob das Drehstromsystem symmetrisch belastet ist (gleiche Impedanz der drei im Stern und/oder im Dreieck geschalteten Verbraucher), oder aber unsymmetrisch belastet ist (die drei an die Außenleiter angeschlossenen Lastimpedanzen sind nicht gleich). Bei symmetrischer Last kann der *Null-* oder *Neutralleiter* entfallen!

80.7.1.3 Symmetrisch belastetes Drehstromsystem

Ist die Last im Stern geschaltet und der Sternpunkt zugänglich, wird der nicht gekennzeichnete Anschluss des Spannungspfades an den Sternpunkt angeschlossen. Andernfalls wird eine Messschaltung nach Abb. 80.15 eingesetzt. Die zwei externen Widerstände R_U sind gleich dem Widerstand des Spannungspfades und bilden einen künstlichen Sternpunkt (Abb. 80.15). Der Null- oder Neutralleiter kann entfallen.

Auch hier wird in kritischen Fällen die *vollständige Messschaltung* eingesetzt, wobei die Kontrollinstrumente grundsätzlich so

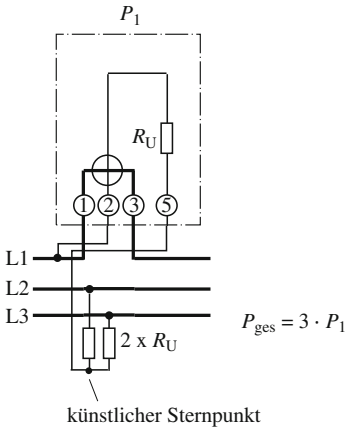


Abb. 80.15 Leistungsmesserschaltung bei symmetrischer Last

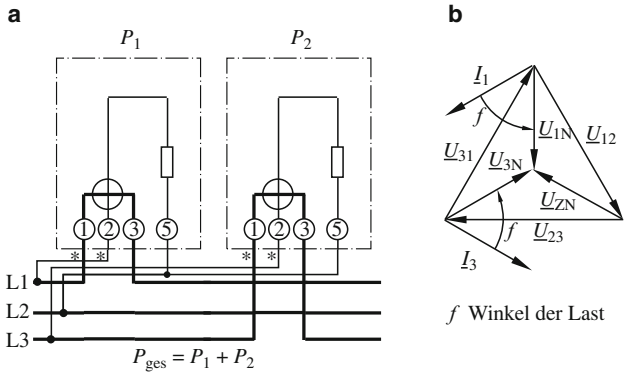


Abb. 80.16 Zwei-Leistungsmesser-Verfahren nach Aron. **a** Anordnung der Leistungsmesser, **b** Zeigerdiagramm für den Sonderfall der symmetrischen Last in Sternschaltung

eingeschaltet werden, dass ein Strommesser in Reihe zu jedem Strompfad eines Leistungsmessers und ein Spannungsmesser parallel zu jedem Spannungspfad in die Schaltung eingefügt wird.

Da in der Regel die gesamte Wirkleistung P_{ges} gemessen werden soll, muss der vom Leistungsmesser gemessene Wert P_1 mit 3 multipliziert werden. Es gibt Leistungsmesser, die diese Multiplikation intern vornehmen und damit die Gesamtwirkleistung anzeigen.

80.7.1.4 Beliebig belastetes Dreileiter-Drehstromsystem

Hier wird vorwiegend das *Zwei-Leistungsmesser-Verfahren* nach *Aron* eingesetzt. Die in Abb. 80.16 dargestellte Anordnung der zwei Leistungsmesser in der Messschaltung kann man anhand der Momentanwerte von Strom und Spannung ableiten. Die Last ist im Stern geschaltet, der Index „$*$“ kennzeichnet den Sternpunkt.

$$\begin{aligned} p &= u_{1*} \cdot i_1 + u_{2*} \cdot i_2 + u_{3*} \cdot i_3 \\ i_1 + i_2 + i_3 &= 0 \quad \Rightarrow \quad i_2 = -i_1 - i_3 \\ p &= i_1 \cdot (u_{1*} - u_{2*}) + i_3 \cdot (u_{3*} - u_{2*}); \\ u_{1*} - u_{2*} &= u_{12}; \quad u_{3*} - u_{2*} = u_{32} \\ p &= i_1 \cdot u_{12} + i_3 \cdot u_{32} \end{aligned} \qquad (80.41)$$

Aus (80.41) folgt, dass kein Null- oder Neutralleiter angeschlossen sein darf, weil sonst die verwendete Beziehung $i_1 + i_2 + i_3 = 0$ nicht mehr gilt. Der Übergang von den Momentanwerten zu den Effektivwerten wird z. B. beim elektrodynamischen Messwerk durch die Trägheit des mechanischen Systems, bestehend aus Drehspule mit Zeiger, bewirkt oder aber in elektronischen Messgeräten durch Mittelwertbildung.

Ausgedrückt durch U, I und $\cos\varphi$ ergibt sich die gesamte umgesetzte Wirkleistung P_{ges} für den allgemeinen Lastfall zu:

$$P_{ges} = U_{12} \cdot I_1 \cdot \cos\varphi_1 + U_{32} \cdot I_3 \cdot \cos\varphi_3 \qquad (80.42a)$$

Formelgrößen siehe Text

Darin ist φ_1 der Winkel zwischen $\underline{U}_{12}$ und $\underline{I}_1$ und entsprechend φ_3 der zwischen $\underline{U}_{32}$ und $\underline{I}_3$. Die Ausschläge der zwei Leistungsmesser sind *vorzeichenrichtig* zu addieren, da auch negative Ausschläge möglich sind (siehe unten). In der Praxis ist der Fall der unsymmetrischen Last in Sternschaltung kritisch, da die Höhe der Spannungen an den Lastimpedanzen vom Wert der einzelnen Lastimpedanzen abhängt.

In Abb. 80.16 ist das Zeigerdiagramm für symmetrische Last in Sternschaltung dargestellt. Die Phasenwinkel der drei Lastimpedanzen sind gleich. Bezeichnet man sie mit φ_L und zerlegt die zwei Winkel aus (80.42a) in die Summe oder Differenz aus dem Phasenwinkel φ_L der Last und dem Winkel von 30° (zwischen Leiterstrom und Mittelpunktspannung), so erhält man

$$P_{ges} = U_{12} \cdot I_1 \cdot \cos(\varphi_L + 30°) + U_{32} \cdot I_3 \cos(\varphi_L - 30°); \qquad \text{symmetrische Last} \qquad (80.42b)$$

φ_L Winkel der Last.

Hinweis: $\underline{U}_{32}$ hat die entgegengesetzte Zeigerrichtung von $\underline{U}_{23}$.

Anhand des Lastfalles „Symmetrische Last“ wird auf einige Besonderheiten bei der Leistungsmessung nach *Aron* hingewiesen:

1. Nach (80.42b) hat *ein* Leistungsmesser einen negativen Ausschlag bei $\varphi_L > 60°$ oder $\varphi_L < -60°$. Bei Zeigerinstrumenten muss dann der Spannungspfad umgepolt werden. Eine Umpolung des Strompfades ist möglich, aber ungünstig, da Ströme im Ampere-Bereich geschaltet werden müssen und die Unterbrechung von induktiven Verbrauchern wegen möglicher Spannungsüberhöhungen problematisch sein kann. Viele Leistungsmesser besitzen deshalb einen entsprechenden Umschalter im Spannungspfad. Die so abgelesene Leistung ist bei der Berechnung der Gesamtwirkleistung negativ einzusetzen.
2. Auch der Ausschlag Null ist bei *einem* Leistungsmesser möglich, wenn gilt: $\varphi_L = 60°$ oder $\varphi_L = -60°$.
3. Beim Anschluss von idealen Blindwiderständen sind die Anzeigen der zwei Leistungsmesser *nicht Null*, da sie nicht den Winkel der Last mit $+90°$ oder $-90°$ erhalten, sondern einen um $+30°$ oder $-30°$ veränderten Winkel. Allerdings sind

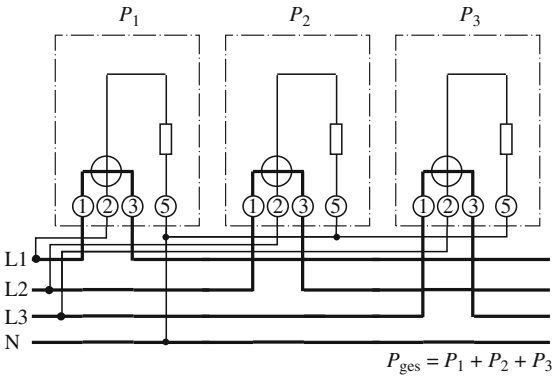


Abb. 80.17 Leistungsmesserschaltung im beliebig belasteten Vierleitersystem

die Ausschläge der zwei Leistungsmesser entgegengesetzt gleich groß, so dass sich als Gesamtwirkleistung nach vorzeichenrichtiger Addition der korrekte Wert Null ergibt.

80.7.1.5 Beliebig belastetes Vierleiter-Drehstromsystem

Die Leistungsmessung geschieht durch drei Leistungsmesser, die nach Abb. 80.17 angeschlossen werden. Dabei kann die Last aus einer beliebigen Kombination von Verbrauchern bestehen (Verbraucher im Dreieck und/oder im Stern geschaltet und/oder Wechselstromverbraucher). Die gesamte umgesetzte Wirkleistung P_{ges} ergibt sich aus der Summe der drei von den Leistungsmessern angezeigten Einzelwirkleistungen: $P_{ges} = P_1 + P_2 + P_3$.

80.7.2 Blindleistungsmessung

Für sinusförmigen Spannungs- und Stromverlauf berechnet sich die Blindleistung Q zu

$$\textit{Blindleistung} \quad Q = U \cdot I \cdot \sin\varphi_L \qquad (80.43)$$

φ_L Winkel der Last; U, I Effektivwerte von Spannung und Strom in V bzw. A.

In größeren Firmen muss die meist durch Motoren verursachte induktive Blindleistung kompensiert werden, weil die zugehörigen Ströme das Netz zusätzlich belasten.

80.7.2.1 Blindleistungsmessung bei Wechselstrom

Sie wird in der Praxis selten angewendet. Bei den analog anzeigenden Blindleistungsmessern wird ein Phasenschieber eingebaut, der im Spannungspfad eine Phasenverschiebung von 90° erzeugt ($\sin\varphi = \cos(90° - \varphi)$). Bei den digital anzeigenden Leistungsmessern wird der Phasenwinkel elektronisch gemessen und die Blindleistung nach (80.43) rechnerisch ermittelt.

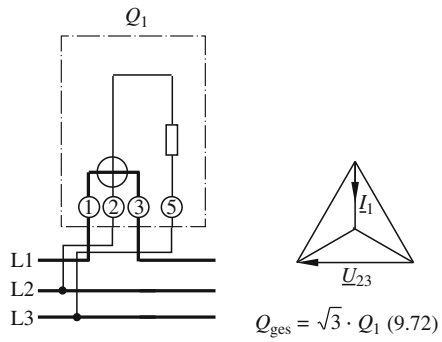


Abb. 80.18 Blindleistungsmessung im symmetrisch belasteten Dreileiter-Drehstromsystem; Zeigerbild für ohmsche Last

80.7.2.2 Blindleistungsmessung in symmetrisch belasteten Dreileiter-Drehstromsystemen

Im Zeigerbild sind Spannung und Strom für einen ohmschen Verbraucher R_L eingetragen (Abb. 80.18). Damit ist der Phasenwinkel der Last Null. Der Leistungsmesser erhält aber die Spannung $\underline{U}_{23}$ und den Strom $\underline{I}_1$, die den Winkel 90° zueinander haben. Er zeigt damit die Blindleistung für diesen Fall korrekt zu Null an. Andere Phasenwinkel der Last werden für die Blindleistungsmessung ebenfalls richtig erfasst. Wegen der symmetrischen Last kann der Nullleiter entfallen.

Die gesamte Blindleistung erhält man mit folgender Beziehung:

$$Q_{ges} = \sqrt{3} \cdot Q_1; \quad \text{symmetrische Last} \qquad (80.44)$$

Q_1 vom Leistungsmesser angezeigte Blindleistung in Var; Q_{ges} gesamte Blindleistung in Var.

Da die Spannung am Spannungspfad um den Faktor $U_{23}/U_{1N} = \sqrt{3}$ zu groß ist, muss zur Berechnung der gesamten Blindleistung Q_{ges} zunächst durch $\sqrt{3}$ geteilt und anschließend die gemessene Blindleistung Q_1 mit 3 multipliziert werden. Das ergibt den Faktor $3/\sqrt{3} = \sqrt{3}$ in (80.44).

80.7.2.3 Blindleistungsmessung in beliebig belasteten Vierleiter-Drehstromsystemen

Die Messschaltung nach Abb. 80.19 entspricht der *Aron*-Schaltung bei der Wirkleistungsmessung. Die *Gesamt*-Blindleistung wird unabhängig von der Art der angeschlossenen Last richtig gemessen, wenn folgende Grundsätze beachtet werden:

1. Die angezeigten Blindleistungen Q_1 und Q_2 der zwei Leistungsmesser sind vorzeichenrichtig zu addieren und anschließend mit dem Faktor $\sqrt{3}$ zu multiplizieren.
2. Beim Anschluss der zwei Leistungsmesser ist auf richtige Polung zu achten (Abb. 80.19).

Der Faktor $\sqrt{3}$ entsteht auf die gleiche Weise wie im symmetrisch belasteten Drehstromsystem, jetzt aber über den zu groß gemessenen Strom.

$$Q_{ges} = \sqrt{3} \cdot (Q_1 + Q_2), \quad \text{beliebige Last} \qquad (80.45)$$

Q_1, Q_2 von den Leistungsmessern angezeigte Blindleistung in Var; Q_{ges} gesamte Blindleistung in Var.

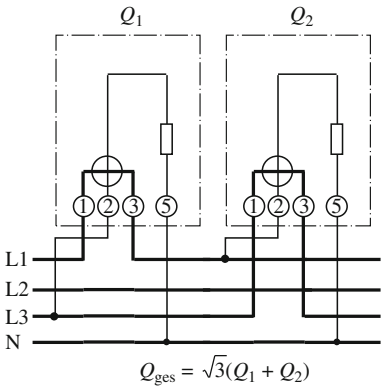


Abb. 80.19 Blindleistungsmessung im beliebig belasteten Drehstromsystem

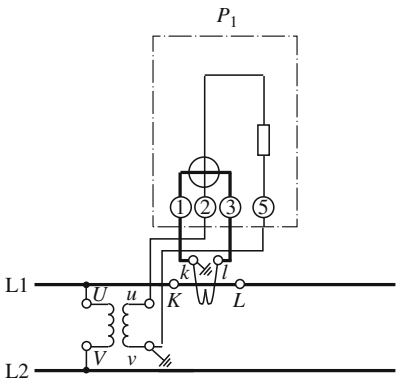


Abb. 80.20 Leistungsmesserschaltung mit Spannungs- und Stromwandler

80.7.3 Scheinleistungsmessung

Die Scheinleistung S ist das Produkt aus Spannung und Strom für $\varphi_L = 0°$:

$$\textit{Scheinleistung} \quad S = U \cdot I \tag{80.46}$$

Sie wird gemessen, indem entweder U und I getrennt gemessen und dann multipliziert werden, oder aber Spannungs- und Strompfad erhalten je einen Gleichrichter mit anschließender Glättung im Spannungspfad. Dann werden beide Größen multipliziert, entweder analog im elektrodynamischen Messwerk oder analog oder digital auf elektronischem Wege im digital anzeigenden Leistungsmesser.

80.7.4 Messbereichserweiterung bei der Leistungsmessung

Für die Messbereichserweiterung von Spannungs- und Strompfad sind grundsätzlich die im Abschn. 80.3 aufgeführten Maßnahmen mit Vor- und Parallelwiderstand möglich. Allerdings wird der Einsatz von Spannungs- und Stromwandlern bevorzugt, weil diese neben einer günstigeren Leistungsbilanz (keine Verluste an Vor- und Nebenwiderständen) eine galvanische Trennung zwischen Leistungskreis und Messkreis bewirken. Abb. 80.20 zeigt eine Leistungsmesserschaltung mit Spannungs- und Stromwandler.

Eine Kenngröße für den Spannungswandler ist die Nennübersetzung k_{NU}.

Nennübersetzung des Spannungswandlers

$$k_{NU} = \frac{U_{Np}}{U_{Ns}} \tag{80.47}$$

U_{Np} primärseitige Nennspannung in V; U_{Ns} sekundärseitige Nennspannung in V.

Die sekundärseitige Nennspannung ist in der Regel zu 100 V festgelegt, wenn die primärseitige Nennspannung anliegt.

Nennübersetzung des Stromwandlers

$$k_{NI} = \frac{I_{Np}}{I_{Ns}} \tag{80.48}$$

I_{Np} primärseitiger Nennstrom in A; I_{Ns} sekundärseitiger Nennstrom in A.

Der sekundärseitige Nennstrom beträgt in der Regel 5 A bei primärseitigem Nennstrom.

Bei der Leistungsmessung mit Spannungs- und Stromwandlern ergibt sich die Beziehung zwischen der mit dem Leistungsmesser gemessenen Leistung P_1 und der tatsächlich in der Lastimpedanz $\underline{Z}$ umgesetzten Wirkleistung P_Z aus

$$P_Z = k_{NU} \cdot k_{NI} \cdot P_1 \tag{80.49}$$

P_Z Wirkleistung in der Impedanz $\underline{Z}$ in W; k_{NU}, k_{NI} siehe (80.47), (80.48); P_1 vom Leistungsmesser angezeigte Wirkleistung in W.

Wichtig beim Einsatz von Stromwandlern ist, dass sie niemals sekundärseitig im Leerlauf betrieben werden dürfen, weil sonst sehr hohe lebensgefährliche Spannungen anliegen. Pro Spannungswandler können sekundärseitig mindestens drei Instrumente (Spannungspfade, Spannungsmesser) parallelgeschaltet werden; pro Stromwandler ebenfalls mindestens drei (Strompfade, Strommesser) in Reihe. Beide Wandler sind sekundärseitig zu erden, damit bei Fehlern im Wandler (sich verringernde Übergangswiderstände zwischen Primär- und Sekundärwicklung) sekundärseitig keine unzulässig hohen Spannungen auftreten.

80.7.5 Leistungsfaktormessung

Der Leistungsfaktor $\cos\varphi$ ist bei sinusförmigen Spannungen und Strömen gleich dem Kosinus des Winkels φ zwischen $\underline{U}$ und $\underline{I}$. Er wird entweder rechnerisch über zwei der drei gemessenen Größen P, Q und S über die Beziehungen $\varphi = \arctan(Q/P)$ oder $\varphi = \arccos(P/S)$ bestimmt, oder er wird mit einem entsprechend konstruierten $\cos\varphi$-Messgerät mit Kreuzspulmesswerk gemessen. Digital anzeigende Leistungsmesser setzen häufig das rechnerische Verfahren ein.

80.8 Messung der Arbeit

Die elektrische Arbeit W berechnet sich bei konstanter Leistung P zu:

$$\textit{Arbeit} \quad W = P \cdot t \quad \text{für } P = \text{const} \tag{80.50a}$$

W elektrische Arbeit in W s bzw. kW h, P Leistung in W bzw. kW, t Zeit in s bzw. h.

Ändert sich die Leistung mit der Zeit, müssen die Momentanwerte addiert werden. Mathematisch ausgedrückt heißt das, dass über die Momentanwerte $p(t)$ der Leistung integriert werden muss:

$$\textit{Arbeit} \quad W = \int_0^T p(t)dt \tag{80.50b}$$

W elektrische Arbeit in Ws bzw. kWh, $p(t)$ Momentanwert der Leistung in W bzw. kW (zeitabhängig), T Dauer der Erfassung in s bzw. h.

Die Messung der elektrischen Arbeit geschieht in jedem Haushalt zum Zweck der Gebührenerfassung mit einem *Elektrizitätszähler*, der nach dem Prinzip des *Induktionszählers* nach Abb. 80.21 aufgebaut ist. Der von der Stromspule SP1 hervorgerufene sinusförmige magnetische Fluss Φ_1 und der von der Spannungsspule SP2 verursachte magnetische Fluss Φ_2 bewirken ein resultierendes Drehmoment für die Aluminiumscheibe, das dem Momentanwert der Wirkleistung proportional ist. Der Bremsmagnet BM erzeugt in der Scheibe Wirbelströme, die ein Gegendrehmoment proportional zur Drehzahl bewirken. Sind beide Drehmomente im Gleichgewichtszustand, ist der Momentanwert der Drehzahl der Scheibe dem Momentanwert der Wirkleistung proportional. Über ein Zählwerk werden die Momentanwerte addiert und damit die vom Elektrizitäts-Versorgungsunternehmen (EVU) gelieferte Arbeit gemessen. Sie wird aus praktischen Erwägungen nicht in Wattsekunden, sondern in Kilowattstunden angegeben.

Der dargestellte Induktionszähler ist für die Messung in Wechselstromkreisen ausgelegt, kann aber auf die Erfassung der Arbeit in Drehstromsystemen dadurch erweitert werden, dass die Spulenanordnung dreimal (für jeden Außenleiter getrennt) angebracht wird. Damit sind dann drei Spannungs- und drei Stromspulen vorhanden. Der Anschluss geschieht, wie bei der Leistungsmessung in Wechselstromkreisen oder Drehstromsystemen, nach Abb. 80.14 oder Abb. 80.17. Intern ist der Elektrizitätszähler für die stromrichtige Messschaltung verdrahtet. Die Proportionalitätskonstante zwischen der Anzahl der Umdrehungen der Aluminiumscheibe und der Arbeit heißt *Zählerkonstante* und wird auf dem Typenschild des Zählers in Umdrehungen/kWh angegeben.

Für die Elektrizitätszähler werden auch elektronische Geräte eingesetzt, die (80.50b) mathematisch auswerten. $u(t)$ und $i(t)$ werden als Momentanwerte in allen 3 Leitern erfasst, multipliziert und integriert. Die Entwicklung dieser elektronischen Elektrizitätszähler war insofern schwierig, weil sie die geringe Ausfallrate, den geringen Messfehler auch bei nichtsinusförmigen Spannungen und vor allem Strömen und den geringen Preis der herkömmlichen Induktionszähler erreichen sollten. Ihr Vorteil liegt darin, dass der Momentanwert der Leistung unmittelbar zur Verfügung steht, dass Leistung und Arbeit jederzeit abrufbar und speicherbar sind und dass eine aktuelle Gebührenerfassung möglich ist. Im Bedarfsfall können die gemessenen Größen dazu dienen, Schaltvorgänge auszulösen (Großverbraucher in kritischen Versorgungsfällen zu- oder abschalten). Auch Blind- und Scheinleistung lassen sich erfassen, speichern und auswerten.

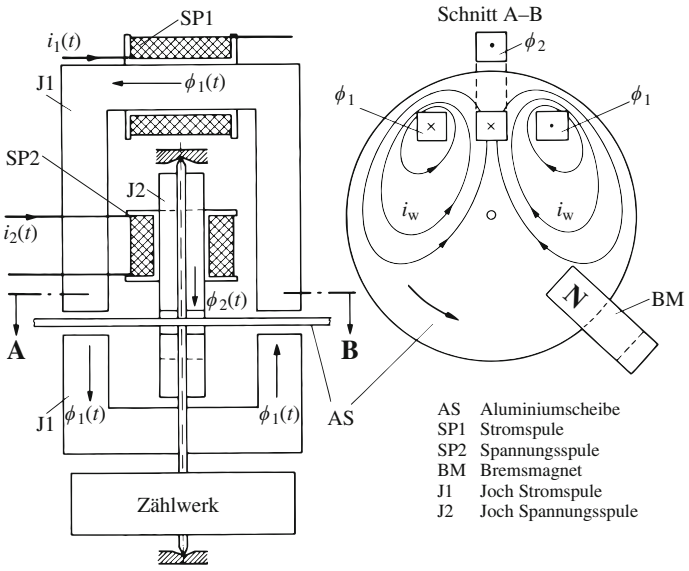


Abb. 80.21 Prinzip eines Induktionszählers für Wechselstrom

Teil X

80.9 Messung von L, C, Gütefaktor und Verlustfaktor

In der Praxis hergestellte Induktivitäten und Kapazitäten sind keine idealen Bauelemente in dem Sinne, dass sie einen Phasenwinkel von genau $+90°$ oder $-90°$ zwischen $\underline{U}$ und $\underline{I}$ besitzen. Vielmehr treten Verluste auf, die in einem einfachen Ersatzschaltbild durch ohmsche Widerstände dargestellt werden (Abb. 80.22). In Reihe zu der als ideal angenommenen Induktivität L liegt der Wirkwiderstand R_L, der im Wesentlichen den Wicklungswiderstand der Spule erfasst. Es wird vorausgesetzt, dass die Spule keinen Eisenkern besitzt, weil dann durch die nichtlineare Beziehung zwischen der magnetischen Feldstärke H des Eisens und der Flussdichte B die Induktivität L vom Strom abhängt. Parallel zum Kondensator C liegt der Widerstand R_C, der den Isolationswiderstand darstellt. Damit ergeben sich die Zeigerdiagramme in Abb. 80.22.

Eine reale Spule kommt der idealen Spule umso näher, je kleiner R_L im Verhältnis zur Impedanz $X_L = \omega \cdot L$ ist. Für dieses Verhältnis gilt $(\omega \cdot L)/R_L = \tan\varphi$. Damit könnte man durch die Angabe von φ oder $\tan\varphi$ die Eigenschaft einer Spule charakterisieren mit der Tendenz, dass eine reale Spule der idealen umso näher kommt, je mehr sich φ dem Wert 90° oder $\tan\varphi$ dem Wert ∞ nähert. Aus praktischen Erwägungen heraus hat man nicht den Winkel φ, sondern den Ergänzungswinkel zu 90°, δ, gewählt. $\tan\delta$ wird als *Verlustfaktor* bezeichnet, der Kehrwert $1/\tan\delta$ als *Gütefaktor* Q. Für reales L und C ergibt sich

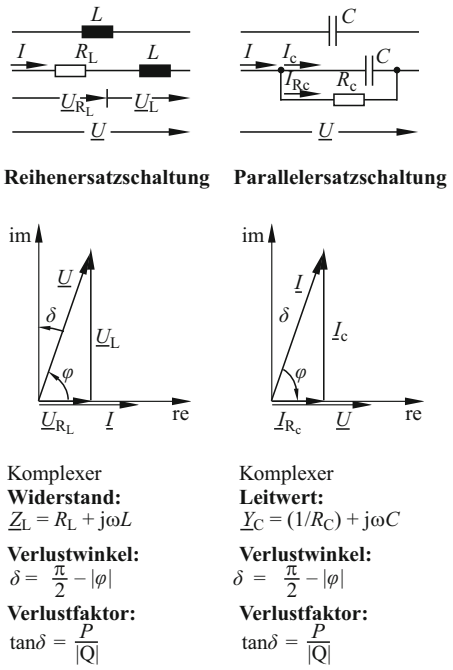


Abb. 80.22 Ersatzschaltbild und Zeigerdiagramm realer Spulen und Kondensatoren

Gütefaktor einer Spule

$$Q = \frac{1}{\tan\delta} = \omega \cdot L/R_L$$

Gütefaktor eines Kondensators

$$Q = \frac{1}{\tan\delta} = \omega \cdot C \cdot R_C \tag{80.51}$$

Zur Messung der Induktivität oder der Kapazität gibt es zwei prinzipiell unterschiedliche Messverfahren.

80.9.1 Messung von $|\underline{Z}_L|$ oder $|\underline{Z}_C|$

Dieses Messverfahren wird hauptsächlich in Vielfach-Messinstrumenten oder L- und C-Messgeräten der unteren Preisklasse eingesetzt. In den zu messenden Kondensator oder in die Spule wird ein sinusförmiger Strom bekannter Frequenz (z. B. 1 kHz) eingespeist. Die sich einstellende Spannung wird gemessen und ist dem Betrag der Impedanz proportional.

$$U_L = I \cdot \sqrt{R_L^2 + (\omega L)^2} = I \cdot Z_L$$
$$U_C = I \cdot \frac{1}{\sqrt{(\frac{1}{R_C})^2 + (\omega C)^2}} = I \cdot Z_C \tag{80.52}$$

Die gemessenen Impedanzen stimmen mit denen von Spule bzw. Kondensator umso mehr überein, je kleiner R_L und je größer R_C ist. Da bei Kondensatoren häufig $R_C \gg 1/(\omega \cdot C)$ ist, ergeben sich dann relativ genaue Werte für C. Bei Induktivitäten sind in vielen Fällen Überschlagsmessungen für L möglich.

80.9.2 Messung von Z_L, Z_C, Gütefaktor und Verlustfaktor

Die verwendeten Messverfahren messen in der Regel L und R_L bzw. C und R_C. Daraus wird Q oder $\tan\delta$ entweder manuell errechnet oder durch ein in das Messgerät eingebautes Rechenprogramm bestimmt. Grundlage für die Messverfahren ist die Brückenschaltung, die nach Abschn. 80.6.2.2 und (80.23) die Möglichkeit bietet, zwei Größen einer Impedanz zu messen:

$$\begin{aligned} &\text{Realteil} && = 0: R_1 \cdot R_4 - X_1 \cdot X_4 && = R_2 \cdot R_3 - X_2 \cdot X_3 \\ &\text{Imaginärteil} && = 0: R_1 \cdot X_4 + R_4 \cdot X_1 && = R_3 \cdot X_2 + R_2 \cdot X_3 \end{aligned} \tag{80.53}$$

In Abb. 80.23 sind drei Brückenanordnungen dargestellt. Allgemein gilt, dass die zu messende Größe mit dem Index „x“ versehen ist und an der Stelle von $\underline{Z}_1$ angeordnet wurde. Die jeweils zwei Abgleichelemente sind wechselseitig solange zu verstellen, bis die Spannung in der Brückendiagonalen Null ist. Dann lassen sich mit (80.53) die zwei unbekannten Elemente bestimmen.

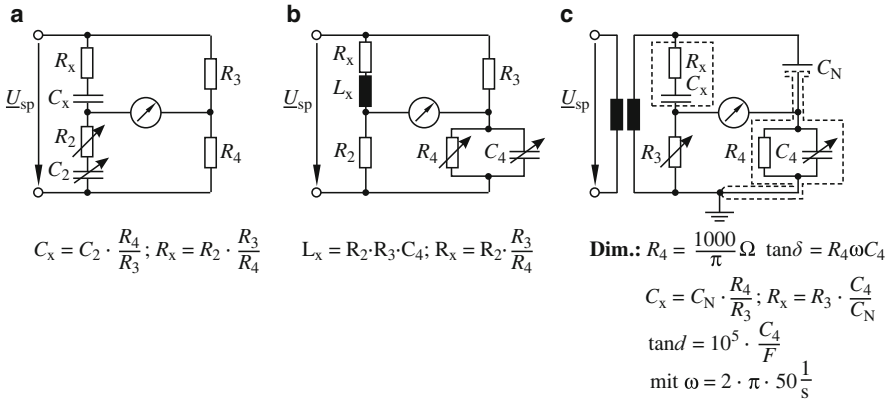


Abb. 80.23 Beispiele für *L*- und *C*-Messbrücken. **a** Kapazitätsmessbrücke, **b** Induktivitätsmessbrücke nach Maxwell-Wien, **c** Schering-Messbrücke

Bei der Kapazitätsmessbrücke ist das Reihenersatzschaltbild für den Kondensator angegeben, das sich aber in das anschaulichere Parallelersatzschaltbild umrechnen lässt. Dieses Ersatzschaltbild hat den Vorteil, dass die folgende einfache Zuordnung der Elemente von (80.53) nach den Elementen der Abb. 80.23a gilt:

$$R_1 \to R_x; \quad X_1 \to -1/(\omega \cdot C_x);$$
$$R_2 \to R_2; \quad X_2 \to -1/(\omega \cdot C_2);$$
$$R_3 \to R_3; \quad R_4 \to R_4; \quad X_3 = X_4 = 0$$

Beispiel 3

Umrechnung eines Reihen-Ersatzschaltbildes in ein Parallel-Ersatzschaltbild.

$$\underline{Z} = R_r - j\frac{1}{\omega C_r} = \frac{R_r \cdot \omega \cdot C_r - j}{\omega C_r}$$
$$\frac{1}{Z} = \frac{\omega C_r}{R_r \cdot \omega \cdot C_r - j} = \frac{\omega C_r (R_r \cdot \omega \cdot C_r + j)}{(R_r \cdot \omega \cdot C_r)^2 + 1}$$
$$= \frac{R_r \cdot \omega^2 \cdot C_r^2}{(R_r \cdot \omega \cdot C_r)^2 + 1} + j\frac{\omega C_r}{(R_r \cdot \omega \cdot C_r)^2 + 1}$$
$$= \frac{1}{R_p} + j\omega C_p$$
$$\Rightarrow \quad R_p = \frac{(R_r \cdot \omega \cdot C_r)^2 + 1}{R_r \cdot \omega^2 \cdot C_r^2}; \quad C_p = \frac{C_r}{(R_r \cdot \omega \cdot C_r)^2 + 1}$$

◀

Die *Schering*-Messbrücke dient zum Messen von Hochspannungskondensatoren für die Energietechnik. Damit ist die Frequenz zu 50 Hz festgelegt, die Brückenspeisespannung kann mehrere kV betragen. Durch die angegebene Dimensionierung von R_4 lässt sich der Verlustfaktor direkt messen. Dieser darf bei Hochspannungskondensatoren bestimmte Werte nicht überschreiten, denn die Erfahrung hat gezeigt, dass der Verlustfaktor eine Aussage liefert über die im Kabel enthaltenen Fehlstellen. Diese Fehlstellen können dazu führen, dass sich das Kabel z. B. stark erwärmt und sich damit selbst zerstört.

Moderne Messgeräte verwenden intern häufig die Brückenschaltungen zur Bestimmung von Real- und Imaginärteil des zu messenden Bauelementes. Der Nullabgleich geschieht vollautomatisch. Güte- und Verlustfaktor werden berechnet, das Ersatzschaltbild des Kondensators oder der Spule kann sowohl in der Paralleldarstellung (C_p und R_{Cp} oder L_p und R_{Lp} parallel) als auch in der Reihendarstellung (C_r und R_{Cr} oder L_r und R_{Lr} in Reihe) auf einem LCD-Bildschirm dargestellt werden. Die einzelnen Größen einschließlich Messfrequenz und Messstrom werden ebenfalls zahlenmäßig eingeblendet. Die Angabe des Messstromes ist wichtig für Spulen mit Eisenkern: Durch die Hysteresekennlinie hängt die Induktivität vom Messstrom ab; streng genommen ist die Induktivität eine Funktion des Momentanwertes des Stromes. Angegeben werden kann nur ein wie auch immer gebildeter „Mittelwert" über eine Stromperiode.

Ein anderes Verfahren speist einen konstanten sinusförmigen Wechselstrom vorgegebener Frequenz in das zu messende Bauelement ein. Zwei Spannungsmesser mit phasenempfindlichem Gleichrichter messen die Spannung über dem Bauelement und filtern jeweils die Inphase-Komponente der Spannung mit dem Strom (entspricht dem Wirkanteil) und die ±90°-Komponente (entspricht dem induktiven oder kapazitiven Blindanteil) heraus. Daraus lassen sich dann die zu messenden Größen berechnen und anzeigen.

80.10 Messung magnetischer Größen

Die am häufigsten gemessenen magnetischen Größen sind der magnetische Fluss Φ in Weber ($1\,\mathrm{Wb} = 1\,\mathrm{V\,s}$), die magnetische Flussdichte B in Tesla ($1\,\mathrm{T} = 1\,\mathrm{V\,s/m^2} = 1\,\mathrm{Wb/m^2}$), die magnetische Feldstärke H in A/m und die Permeabilität ($\mu = \mu_0 \cdot \mu_r = B/H$).

80.10.1 Magnetischer Fluss

Zeitlich veränderlicher Fluss: Handelt es sich um einen sinusförmig veränderlichen Fluss, der in seiner Gesamtheit durch

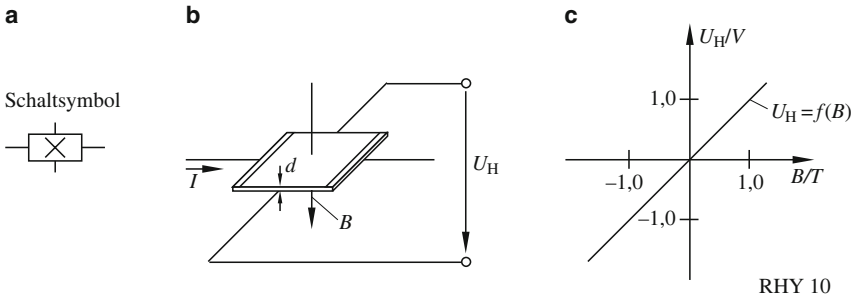


Abb. 80.24 Hallsonde. **a** Schaltsymbol, **b** Aufbau, **c** Kennlinie RHY 10 (Siemens)

eine bestimmte Fläche hindurchtritt (kein Streufluss), legt man um diese Fläche eine Spule mit N Windungen. In der Spule wird dann eine Spannung $u(t)$ induziert:

$$u(t) = -N \cdot \frac{d\Phi(t)}{dt} \tag{80.54}$$

Mit $\Phi(t) = \Phi_m \cdot \sin \omega t$ und $d(\Phi(t))/dt = \omega \cdot \Phi_m \cdot \cos \omega t$ folgt für die Beziehung der Scheitelwerte u_m und Φ_m

$$u_m = -N \cdot \omega \cdot \Phi_m \tag{80.55}$$

Dieses Verfahren hat den Vorteil, dass an Form und Größe der Fläche, durch die der Gesamtfluss hindurchtritt, keine Bedingungen gestellt werden, da die Spule nur diese Fläche voll umfassen muss.

Zeitlich konstanter Fluss: Ist der zu messende magnetische Fluss zeitlich konstant, kann man den Strom, der den Fluss verursacht, ein- oder ausschalten. Durch diese Flussänderung wird eine Spannung in der benutzten Spule induziert. Der Momentanwert der induzierten Spannung hängt natürlich von dem zeitlichen Verlauf des Schaltvorgangs und damit vom zeitlichen Verlauf des Flusses $\Phi(t)$ ab. Integriert man aber die Momentanwerte der Spannung mit einem elektronischen Integrierer, so ist der erhaltene Spannungswert u_{ges} dem Gesamtfluss Φ_{ges} proportional.

Eine weitere Möglichkeit zur Flussmessung unabhängig vom zeitlichen Verlauf ist mit der bei der Messung der magnetischen Flussdichte B beschriebenen *Hall-Sonde* möglich. Diese misst zwar die magnetische Flussdichte B, wenn aber B auf der Fläche A konstant ist, erhält man Φ durch Multiplikation mit der Fläche A ($\Phi = B \cdot A$). Ist B nicht konstant auf A, teilt man die Fläche in kleine Teilflächen auf, misst die jeweilige Flussdichte B und summiert die erhaltenen Werte zum Gesamtfluss auf.

80.10.2 Magnetische Flussdichte

Die magnetische Flussdichte B wird zunehmend mit der *Hallsonde* gemessen (Abb. 80.24). Der Strom I durchfließt ein Plättchen der Dicke d, wobei d klein ist gegenüber der Länge und der Breite. Das Plättchen wird senkrecht in der eingezeichneten Weise von einem Magnetfeld mit der Flussdichte B durchsetzt. Dann kann nach Abb. 80.24 eine Spannung, die sogenannte *Hallspannung*, abgegriffen werden. Sie entsteht durch die *Lorentzkraft*, die auf die in einem Magnetfeld bewegten Elektronen eine Kraft ausübt und eine Ladungstrennung bewirkt.

$$\textit{Hallspannung} \quad U_H = (k_H \cdot I \cdot B)/d \tag{80.56}$$

U_H Hallspannung in V; k_H Hallkonstante in $m^3/A\,s$; I Strom in A; B magnetische Flussdichte in $V\,s/m^2$; d Plattendicke in m.

Die *Hallkonstante* k_H ist besonders groß bei den technologisch gut beherrschten Materialien Indiumantimonid und Galliumarsenid. Sie liegt in der Größenordnung $(10^2 \ldots 10^3)\,cm^3/A\,s$.

Die Hallsonde vom Typ RHY 10 (Siemens) hat einen Gehäusedurchmesser von 2,5 mm. Der Nennstrom ist 100 mA. Bei einer Flussdichte von 1 T ergibt sich eine Leerlauf-Hallspannung von etwa 100 mV. Der Innenwiderstand der Hallspannungsquelle und der Stromeinspeisungsseite betragen jeweils etwa 2 Ω. Häufig wird die Leerlauf-Empfindlichkeit angegeben: Ausgangsspannungsänderung zu Stromänderung × Flussdichteänderung in der Einheit V/A T. Damit können unterschiedliche Hallsonden miteinander verglichen werden. Für die Sonde RHY 10 ergibt sich ein Wert von $0{,}1\,V/(0{,}1\,A \cdot 1\,T) = 1\,V/A\,T$. Das Vorzeichen der Hallspannung hängt von der Richtung des Magnetfeldes ab.

Weiterhin lassen sich zur Messung *magnetisch steuerbare Widerstände* (Feldplatten) verwenden. Werden sie von einem Magnetfeld durchsetzt, nimmt ihr Widerstand R mit zunehmender Flussdichte B zu. Die Abhängigkeit ist nichtlinear.

Die Feldplatte FP 30 N 60 E (Siemens) hat die äußeren Maße 3,2 mm × 1,2 mm. Der Widerstand beträgt bei $B = 0$ etwa 50 Ω; bei $B = 0{,}3$ T ist er auf den Wert $1{,}4 \cdot 50\,\Omega$ angestiegen, bei $B = 1$ T auf $5 \cdot 50\,\Omega$. Durch eine Wechsel-Vormagnetisierung lassen sich Flussdichten im µT-Bereich messen. Der Widerstand ist ein reiner Wirkwiderstand, der *nicht* von der *Richtung* des Magnetfeldes abhängt.

80.10.3 Magnetische Feldstärke

In Materialien, in denen die Permeabilitätszahl μ_r konstant ist ($\mu_r \neq f(i)$), wird die magnetische Feldstärke dadurch bestimmt,

dass man die magnetische Flussdichte B mit einem der oben genannten Verfahren misst und über die Beziehung $B = \mu_0 \cdot \mu_r \cdot H$ die Feldstärke errechnet.

In ferromagnetischen Materialien ändert sich die Permeabilitätszahl μ_r mit der Flussdichte B. Zur Messung nutzt man aus, dass in parallelen magnetischen Kreisen die magnetische Spannung gleich groß ist. Der eine Kreis ist der Kreis mit dem ferromagnetischen Material, der andere parallele Kreis ein Luftkreis. Wird in beiden Kreisen ein homogenes Feld erzeugt, bestimmt man die Feldstärke im Luftkreis und hat damit gleichzeitig den Wert für den anderen Kreis. Die Feldstärke im Luftkreis wird über die induzierte Spannung gemessen, indem von (80.55) und $\Phi = B \cdot A = \mu \cdot H \cdot A$ Gebrauch gemacht wird.

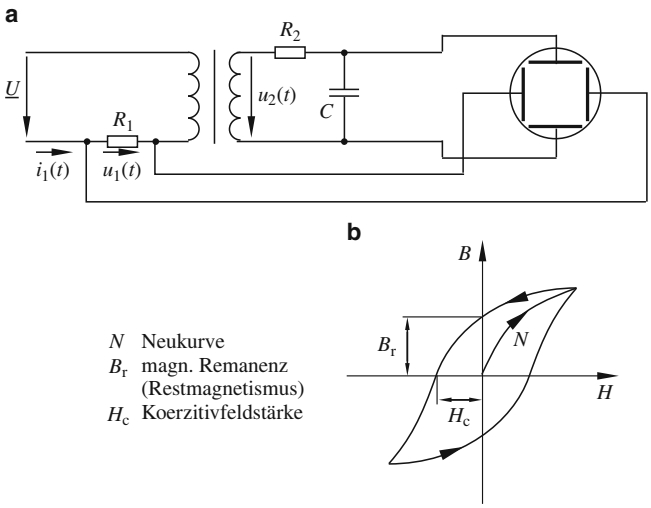


Abb. 80.25 Messung der Hystereseschleife. **a** Messschaltung, **b** Hystereseschleife

80.10.4 Permeabilität

Die magnetischen Eigenschaften ferromagnetischer Werkstoffe werden häufig durch die Magnetisierungskurven $B = f(H)$ beschrieben. Ändert man die Feldstärke periodisch zwischen einem positiven und einem gleich großen negativen Wert, bekommt man die *Hystereseschleife* nach Abb. 80.25.

Im Teil a) ist die Messschaltung zu deren Aufnahme dargestellt. Die magnetische Feldstärke $H(t)$ ist dem Strom $i_1(t) = I_m \cdot \sin \omega t$ in der Primärspule proportional. Dieser Strom wird als proportionaler Spannungsfall $u_1(t) = i_1(t) \cdot R_1$ dem X-Verstärker eines Oszilloskops zugeführt. Damit ist $u_1(t) = k \cdot H(t)$. Die Flussdichte $B(t)$ wird über $\Phi(t)$ ermittelt. Für die in der Sekundärspule induzierte Spannung $u_2(t)$ gilt nach (80.54): $u_2(t) = -N \cdot d\Phi(t)/dt$. Für $\Phi(t) = \Phi_m \cdot \sin \omega t$ ergibt sich $u_2(t) = -N \cdot \omega \cdot \Phi_m \cdot \cos \omega t = -N \cdot \omega \cdot B_m \cdot A \cdot \cos \omega t$.

Zwischen $B(t)$ und $H(t)$ besteht aber noch eine Phasenverschiebung von 90°. Deshalb wird die Spannung $u_2(t)$ zunächst über einen Tiefpass aus R_2 und C geschickt und anschließend dem Y-Verstärker des Oszilloskops zugeführt. Dieser Tiefpass wirkt dann als Integrierer mit einer Phasenverschiebung von nahezu −90°, wenn seine Grenzfrequenz wesentlich geringer ist als die Frequenz der den Transformator speisenden Spannung. Mit $R_2 = 1\,\text{M}\Omega$ und $C = 1\,\mu\text{F}$ ergibt sich eine Grenzfrequenz von etwa 1 Hz, und damit ist diese Bedingung bei einer Transformatorspeisung mit 50 Hz erfüllt.

Über die Messung der Scheitelwerte von B und H nach einem der oben beschriebenen Verfahren kann auf der X- und Y-Achse eine Skalierung angebracht werden. Der dargestellten Hystereseschleife lässt sich die flussdichteabhängige Permeabilität gemäß $B = \mu \cdot H$ entnehmen. Die von der Hystereseschleife eingeschlossene Fläche ist der für die Wirbelströme aufzubringenden Arbeit proportional. Weiterhin lassen sich die *Koerzitivfeldstärke* H_c und die *Remanenzflussdichte* B_r ablesen.

Messverfahren zur Messung nichtelektrischer Größen

81

Wilfried Plaßmann



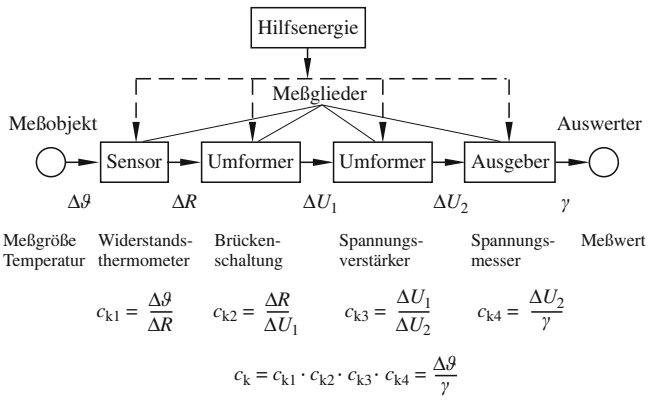
81.1 Messaufnehmer 923
81.2 Messverfahren 930

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_81

Tab. 81.1 Vor- und Nachteile des elektrischen Messens nichtelektrischer Größen

Vorteile	Nachteile
Hohe Messgenauigkeit	Meist Hilfsenergie erforderlich
Große Empfindlichkeit	Manchmal höherer Preis
Meist geringe Rückwirkung auf den Messkreis (sowohl elektrisch als auch mechanisch)	
Gute Anpassungsmöglichkeit an die Messgröße	
Einfache Messwertübertragung über 2-Draht-Leitung oder drahtlos	
Einfache Speicherung, Registrierung und Weiterverarbeitung	
Grenzfrequenz der Messeinrichtung wesentlich höher als die der zu messenden Größe	
Große Ausgangsleistung durch elektrische Verstärker möglich	
Geringere mechanische Abmessungen möglich	



$$c_{k1} = \frac{\Delta\vartheta}{\Delta R} \quad c_{k2} = \frac{\Delta R}{\Delta U_1} \quad c_{k3} = \frac{\Delta U_1}{\Delta U_2} \quad c_{k4} = \frac{\Delta U_2}{\gamma}$$

$$c_k = c_{k1} \cdot c_{k2} \cdot c_{k3} \cdot c_{k4} = \frac{\Delta\vartheta}{\gamma}$$

Abb. 81.1 Prinzip einer Messkette

Die elektrische Messung nichtelektrischer Größen gewinnt zunehmend an Bedeutung, da einerseits sehr leistungsfähige Messverfahren entwickelt worden sind und andererseits die Messdatenverarbeitung auf elektrischem Wege in Rechnern geschieht und damit die Umsetzung der nichtelektrischen Größen in elektrische Größen zwingend notwendig ist. In Tab. 81.1 sind einige Vor- und Nachteile der elektrischen Messung nichtelektrischer Größen gegenüber der nichtelektrischen Messung dargestellt.

In Abb. 81.1 ist eine Messeinrichtung zur Messung nichtelektrischer Größen mit den verwendeten speziellen Bezeichnungen dargestellt. Ein Teil der Begriffe wurde schon im Abschn. 75.1 erläutert.

Messkette: Die gesamte Messeinrichtung wird als *Messkette* bezeichnet. Sie beginnt bei dem *Messobjekt*, dessen nichtelektrische Größe, hier am Beispiel der Temperaturmessung dargestellt, gemessen werden soll. Die *Messgröße* Temperatur wird von einem *Sensor* oder *Messaufnehmer* erfasst, der das *erste* von mehreren *Messgliedern* ist. Der Sensor kann entweder *aktiv* oder *passiv* sein. Ein aktiver Sensor setzt die thermische Energie in elektrische Energie um. So steht beim *Thermoelement* eine reale Spannungsquelle mit *temperaturabhängiger* Leerlaufspannung und Innenwiderstand zur Verfügung, der ein Strom entnommen werden kann. Beim passiven Sensor wird eine elektrische Größe durch die nichtelektrische Größe *beeinflusst* (in ihrem Wert geändert), wie das bei den temperaturabhängigen Widerständen, beispielsweise Widerstandsthermometern, NTC- und PTC-Widerständen, der Fall ist. Der sich ändernde Widerstandswert kann nicht direkt ausgewertet werden, sondern muss erst über einen Spannungsteiler oder einen eingeprägten konstanten Strom in eine auswertbare Spannungsänderung umgesetzt werden. Das geschieht in einem weiteren Messglied.

Manche Messeinrichtungen erfordern drei oder mehr Messglieder, wie z. B. die Kraftmessung mit *Dehnungsmessstreifen* (Abschn. 81.2.1). Die zu messende Kraft wird zunächst in eine Längenänderung, dann in eine Widerstandsänderung, diese in eine Spannungsänderung und schließlich in einen Zeigerausschlag umgesetzt. Ziel der Messtechnik ist es, bei aktiven Sensoren die nichtelektrische Energie direkt, also ohne Zwischenschalten weiterer Energieformen, in elektrische Energie umzusetzen. Bei passiven Sensoren versucht man, die Zahl der Messglieder so klein wie möglich zu halten. Am Ende der Messkette steht der Ausgeber, der das Signal zur Anzeige und/oder zur Auswertung oder Weiterverarbeitung (Rechner) bereitstellt.

Messglied, *Messkettenkoeffizient*: Die Eigenschaften der einzelnen Messglieder werden durch den *Messgliedkoeffizienten* c_{ki} erfasst, der definiert ist als das Verhältnis von Eingangsgröße zu Ausgangsgröße. Das Verhalten der gesamten Messeinrichtung ergibt sich dann durch Multiplikation der einzelnen Messgliedkoeffizienten zum *Messkettenkoeffizienten* c_k (siehe Beispiel 1).

Beispiel 1

Temperaturmessung mit Widerstandsthermometer; Abb. 81.1 und Abschn. 81.1.1. $R(\vartheta) = R_0 \cdot (1 + \alpha \cdot \Delta\vartheta) = R_0 + R_0 \cdot \alpha \cdot \Delta\vartheta$; $R(\vartheta) - R_0 = \Delta R = R_0 \cdot \alpha \cdot \Delta\vartheta$. Daraus ergibt sich der Messgliedkoeffizient c_{k1} mit $\alpha = 4 \cdot 10^{-3}/°\text{C}$ und $R_0 = 100\,\Omega$ zu: $c_{k1} = \Delta\vartheta/\Delta R = 1/(R_0 \cdot \alpha) = 2{,}5\,°\text{C}/\Omega$.

Wird ΔR in einer Viertelbrücke gemessen, so gilt für c_{k2} mit $\Delta U_1 \approx U_{sp} \cdot \Delta R/(4 \cdot R)$ (Näherungslösung, siehe Tab. 80.1; im Folgenden als exakte Beziehung verwendet): $c_{k2} = \Delta R/\Delta U_1 = 4 \cdot R/U_{sp} = 40\,\Omega/\text{V} = 40\,\text{m}\Omega/\text{mV}$ (mit $U_{sp} = 10\,\text{V}$).

Ein eventuell erforderlicher Spannungsverstärker mit der Ausgangsspannung ΔU_2 ist gekennzeichnet durch $c_{k3} = \Delta U_1/\Delta U_2 = 1\,\text{mV}/1000\,\text{mV}$ („Verstärkung“ $1/c_{k3} = 1000$).

Abweichend von Abb. 81.1 wird der Zusammenhang zwischen ΔU_2 des Verstärkers und dem Strom ΔI_i eines Zeigerinstrumentes über c_{k4} zugeordnet: $c_{k4} = \Delta U_2/\Delta I_i = 1\,\text{mV}/10\,\text{mA}$ („Innenwiderstand“ $c_{k4} = 100\,\Omega$). Schließlich werden Strom und Ausschlag $\Delta\gamma$ über c_{k5} erfasst: $c_{k5} = 10\,\text{mA}/1\,\text{Skt}$.

Den Zusammenhang zwischen der Temperatur und den Skalenteilen erhält man durch Multiplikation der 5 Messgliedkoeffizienten zum Messkettenkoeffizienten c_k:

$$c_k = c_{k1} \cdot c_{k2} \cdot c_{k3} \cdot c_{k4} \cdot c_{k5}$$
$$c_k = (2{,}5\,°\mathrm{C}/\Omega) \cdot (40\,\mathrm{m}\Omega/\mathrm{mV}) \cdot (1\,\mathrm{mV}/1000\,\mathrm{mV}) \cdot (1\,\mathrm{mV}/10\,\mathrm{mA}) \cdot (10\,\mathrm{mA}/1\,\mathrm{Skt})$$
$$c_k = 10^{-4}\,°\mathrm{C}/\mathrm{Skt}$$

In Abb. 81.1 wird ΔU_2 auf einen Spannungsmesser gegeben. Dann gilt $c_{k4} = \Delta U_2/\gamma$, $c_k = \Delta\vartheta/\gamma$.

Hinweise:

1. Die Messgliedkoeffizienten werden in der Regel auch dann mit Einheiten angegeben, wenn sie dimensionslos sind, siehe c_{k3}.
2. Teile oder Vielfache der Einheiten sind beliebig wählbar, sollten sich aber an der Praxis orientieren: $c_{k4} = 1\,\mathrm{mV}/10\,\mu\mathrm{A} = 1\,\mathrm{V}/10\,\mathrm{mA} = 100\,\mathrm{V}/1\,\mathrm{A}$. Hier entspricht die Angabe 1 mV/10 mA am ehesten den in der Schaltung auftretenden Größen.
3. Die Bedeutung (Definition) eines Messglied- oder Messkettenkoeffizienten lässt sich in der Regel an seinen Einheiten ablesen.
4. Auch wenn „Skt." (Skalenteile) keine Einheit ist, wurde sie hier zum besseren Verständnis verwendet. ◄

81.1 Messaufnehmer

Zunächst werden die physikalischen Prinzipien einiger wichtiger Aufnehmer aus messtechnischer Sicht dargestellt.

81.1.1 Ohm'sche Aufnehmer

Es werden Widerstände betrachtet, die im hier vorkommenden Frequenzbereich weder eine kapazitive noch eine induktive Komponente besitzen. Der elektrische Widerstand R berechnet sich zu

$$R = \frac{l}{\gamma \cdot A} = \frac{\rho \cdot l}{A} \qquad (81.1)$$

l Länge des Leiters in m, A Querschnittsfläche des Leiters in m^2, γ elektrische Leitfähigkeit in $1/(\Omega \cdot \mathrm{m})$, ρ spezifischer elektrischer Widerstand in $\Omega \cdot \mathrm{m}$.

Um den Wert des Widerstandes zu ändern, gibt es mehrere Möglichkeiten:

1. *Ändern der Länge l*: Über einen Schleifkontakt wird nur ein Teil des Gesamtwiderstandes R erfasst. Das Prinzip ist vom Potentiometer her bekannt, wobei das Verändern der Schleiferstellung entweder durch eine Drehbewegung (Drehpotentiometer mit Teilwiderstand R_α, Abb. 81.2a) oder eine Linearbewegung (Linearpotentiometer, „Flachbahnpotentiometer" mit Teilwiderstand R_x, Abb. 81.2a, b) erfolgen kann. Im einfachen Fall besteht der Widerstand aus einer Kohleschicht oder aus leitfähigem Kunststoff. Über einen Schleifkontakt aus Metall wird der Teilwiderstand abgegriffen.

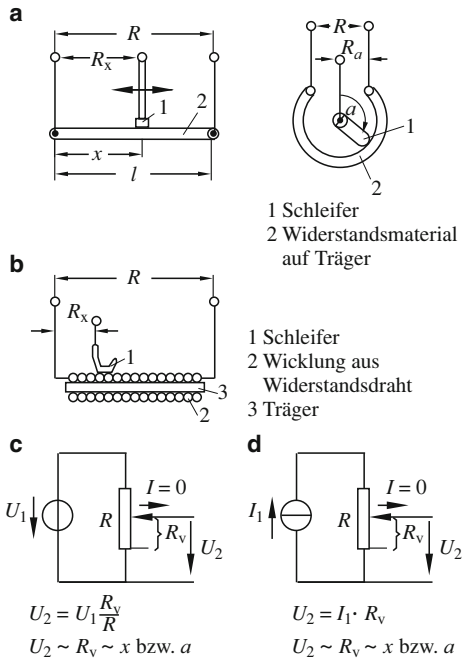


Abb. 81.2 Potentiometer. **a** Schichtpotentiometer, **b** Drahtpotentiometer, **c** Auswerteschaltung mit Spannungsquelle, **d** Auswerteschaltung mit Stromquelle

Besonders zuverlässig sind Drahtpotentiometer nach Abb. 81.2b. Allerdings können die über den Schleifer abgegriffenen Widerstandswerte nur stufig eingestellt werden. Die messtechnische Erfassung des Widerstandswertes geschieht entweder über einen Spannungsteiler (Abb. 81.2c), eine Stromquelle (Abb. 81.2d) oder in einer Brückenschaltung (Abschn. 80.6.1, Tab. 80.1 und (80.16)). Hat das Spannungsmessgerät (Abb. 81.2c, d) einen entsprechend hohen Eingangswiderstand ($I \approx 0$), besteht eine lineare Beziehung zwischen Widerstandswert und zugeordneter Spannung. Damit kann man mit diesen Potentiometern Längen und Drehwinkel erfassen.

Beispiel 2

Ein Draht-Potentiometer mit $R = 1000\,\Omega$ besteht aus $N = 1000$ Windungen. Dann kann der über den Schleifer abgegriffene Widerstand in Stufen von $1000\,\Omega/N = 1\,\Omega$ eingestellt werden. Dabei wird angenommen, dass der

Tab. 81.2 Messung elektrischer und nichtelektrischer Größen durch Widerstandsänderungen

Einflussgröße	Widerstandsänderung
Temperatur T	$R(T)$ Widerstand bei der Temperatur T R_0 Widerstand bei der Temperatur T_0, häufig 0 °C Abhängigkeit des Widerstandes von der Temperatur: 1. $R(T) = R_0(1 + A(T - T_0) + B(T - T_0)^2)$, gilt allgemein für metallische Leiter, z. B. Cu, Pt, Ni. Näherung: $R(T) \approx R_0(1 + A(T - T_0))$, für A wird hier auch α verwendet. A (bzw. α), B Materialkonstanten; A im Bereich $+(3 \ldots 6) \cdot 10^{-3}\,\mathrm{K}^{-1}$, B im Bereich $(-1 \ldots +10) \cdot 10^{-6}\,\mathrm{K}^{-2}$; Silizium-Temperatursensoren $A \approx 7{,}7 \cdot 10^{-3}\,\mathrm{K}^{-1}$, $B \approx 19 \cdot 10^{-6}\,\mathrm{K}^{-2}$; R_0 im Bereich $(1 \ldots 10)\,\mathrm{k\Omega}$. 2. $R(T) = R_0 \exp(B(1/T - 1/T_0))$, gilt für spezielle Materialien, hier NTC-Widerstand, B Materialkonstante in der Größenordnung 3000 bis 4000 K. 3. $R(T) = R_N \exp(\alpha(T - T_N))$, gilt für spezielle Materialien, hier PTC-Widerstand, näherungsweise in einem begrenzten, technisch ausgenutzten Temperaturbereich $T_{\max} \geq T \geq T_N$ gültig; Typenspektrum für T_N im Bereich von ca. −20 bis 350 °C und für R_N im Bereich von ca. 1 Ω bis 100 kΩ; α Materialkonstante im Bereich $(0{,}1 \ldots 0{,}7)\,\mathrm{K}^{-1}$ oder meist angegeben in der Form $(10 \ldots 70)\,\%/\mathrm{K}$
Licht, hier Beleuchtungsstärke E	$R(E) = R_0(E/E_0)^{-\gamma}$; $R(E)$ Widerstand bei der Beleuchtungsstärke E; R_0 Widerstand bei der Beleuchtungsstärke E_0; γ Materialkonstante im Bereich 0,5 … 1,2
Kraft, hier mechanische relative Dehnung $\Delta l/l$ eines elektrischen Leiters	$\Delta R/R = K \cdot \Delta l/l$; K Materialkonstante, mit „K-Faktor“ bezeichnet; bei Dehnungsmessstreifen (DMS) gilt: $K \approx 2\Delta R/R$ relative Widerstandsänderung des DMS; $\Delta l/l$ relative Längenänderung von DMS und Werkstück, auf den der DMS aufgeklebt ist
Magnetfeld, hier magnetische Flussdichte B	$R_B = R_0(1 + k \cdot B^2)$; R_B Widerstand bei der Flussdichte B; R_0 Widerstand bei der Flussdichte 0; k Materialkonstante, angegeben wird meist R_B/R_0, liegt bei $B = 1\,\mathrm{T}$ im Bereich 5 bis 15; R_0 im Bereich ca. 10 Ω bis 1 kΩ

Schleifer ideal von Windung zu Windung übergreift. Man beachte, dass der relative Fehler durch diese Stufigkeit dann besonders groß ist, wenn kleine Widerstandswerte abgegriffen werden und ein Widerstandswert genau in der Mitte zwischen zwei abgreifbaren Werten eingestellt werden soll. Die maximal mögliche Abweichung zwischen Ist- und Sollwert beträgt $\pm 0{,}5\ R/N$, hier also $\pm 0{,}5\,\Omega$, unabhängig vom abgegriffenen Widerstandswert. Der maximal mögliche relative Fehler nimmt also mit abnehmendem abgegriffenem Widerstandswert zu. So ergibt sich bei einem einzustellenden Widerstandswert von 20,5 Ω ein relativer Fehler von $\pm 0{,}5\,\Omega/20{,}5\,\Omega \approx \pm 2{,}4\,\%$. Deshalb werden manchmal Potentiometer mit einem Drehwinkel von 270° eingesetzt, wobei für die eigentliche Messung nur der Bereich von 180° bis 270° ausgenutzt wird. Dadurch bleibt der Fehler durch die Stufigkeit in vorgegebenen Grenzen. ◄

2. *Ändern von γ oder ρ*: Die elektrische Leitfähigkeit bzw. der spezifische Widerstand kann durch Temperatureinwirkung (Widerstandsthermometer, NTC-, PTC-Widerstände), Lichteinwirkung (Fotowiderstände), mechanische Dehnung oder Stauchung (Dehnungsmessstreifen) und Magnetfeldeinwirkung (magnetfeldabhängige Widerstände) geändert werden (Tab. 81.2).
3. *Ändern der Fläche A*: Hier ist im Wesentlichen der bereits im vorigen Abschnitt genannte und im Abschn. 81.2.1 näher beschriebene *Dehnungsmessstreifen* zu nennen. Die relative Dehnung $\Delta l/l$ eines elektrischen Leiters hat auch eine relative Flächenänderung zur Folge, die über die Querkontraktionszahl im „*K-Faktor*“ erfasst ist.

81.1.2 Kapazitive Aufnehmer

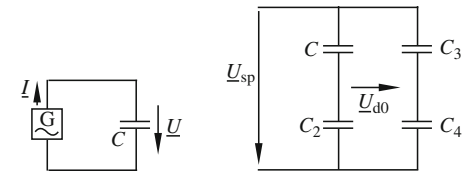
Die Kapazität eines Plattenkondensators berechnet sich unter Vernachlässigung der Randeffekte nach der Formel

$$C = \varepsilon_0 \cdot \varepsilon_r \cdot A/d \tag{81.2}$$

ε_r Dielektrizitätszahl; A Plattenfläche in m^2; d Plattenabstand in m.

Hierin können ε_r, A und d verändert werden. Aus der sich dann ändernden Kapazität kann bei bekanntem rechnerischem Zusammenhang die zu messende Größe bestimmt werden. Dies sind in der Regel: 1. Längen oder Abstände, erfasst über d, 2. Drehwinkel, gemessen über A nach dem Prinzip des Drehkondensators, und 3. Füllstände in Behältern, erfasst über das von Luft verschiedene ε_r des eingefüllten Mediums.

Gemessen wird zunächst die Kapazität C nach Abb. 81.3 entweder durch Erfassen des Spannungsfalls an dieser Kapazität bei Speisung aus einer sinusförmigen Wechselstromquelle bekannter Frequenz oder durch Einfügen in eine Wechselstrommessbrücke. Die entsprechenden Formeln sind in Abb. 81.3 eingetragen. Zur Berechnung der Spannung $\underline{U}_{d0}$ in der Brückendiagonalen wird (80.7) verwendet, wobei für die Widerstände R_1 bis R_4 die Wechselstromwiderstände $\underline{Z}_1 = 1/j\omega C_1$ bis $\underline{Z}_4 = 1/j\omega C_4$ der als ideal angenommenen Kondensatoren eingesetzt werden. C_3 und C_4 sind hier gleich groß gewählt worden; sie können auch unterschiedliche Werte aufweisen und damit dem Messaufbau angepasst werden. Der mathematische Zusammenhang zwischen der Kapazität C und der zu messenden Größe wird in die Skala analog anzeigender Messgeräte eingearbeitet oder bei digital anzeigenden Messgeräten softwaremäßig berücksichtigt.



$$\underline{U} = \underline{I} \cdot \frac{1}{\mathrm{j}\omega c}$$

$$\underline{U} = \underline{I} \cdot \frac{d}{\mathrm{j}\omega e_0 e_r A}$$

C: Plattenkondensator

$$\underline{U}_{d0} = \underline{U}_{sp} \frac{C - C_2}{2(C + C_2)}$$

mit $C_3 = C_4$

$$\underline{U}_{d0} = \underline{U}_{sp} \frac{\frac{e_0 \cdot e_r \cdot A}{d} - C_2}{2\left(\frac{e_0 \cdot e_r \cdot A}{d}\right) + C_2}$$

C: Plattenkondensator

Abb. 81.3 Messschaltungen für kapazitive Aufnehmer

81.1.3 Induktive Aufnehmer

Die Induktivität L einer Spule nach Abb. 81.4a mit der Länge l, der Querschnittsfläche $A = r^2 \cdot \pi$, der Windungszahl N und einem Material innerhalb der Spule mit der Permeabilität $\mu = \mu_0 \cdot \mu_r$ berechnet sich für $r \ll l$ näherungsweise aus

$$L \approx N^2 \mu_0 \mu_r \cdot A/l \tag{81.3}$$

N Windungszahl; μ_r Permeabilitätszahl; A Spulenquerschnittsfläche in m^2; l Spulenlänge in m.

Im Gegensatz zu den kapazitiven Aufnehmern kann in der Praxis nur die Permeabilitätszahl μ_r geändert werden, indem ein Material mit $\mu_r \neq 1$ mehr oder weniger tief in die Spule eingetaucht wird. Die Möglichkeit, die Induktivität durch Abgreifen von Teilwindungen über Schleifkontakte zu ändern, wird kaum angewendet, da hierfür die Potentiometergeber besser geeignet sind. Die Messverfahren zur Messung der Induktivität L nach Abb. 81.4b entsprechen denen der kapazitiven Aufnehmer. Die Ableitung von $\underline{U}_{d0}$ geschieht auf die gleiche Weise wie zu Abb. 81.3b beschrieben. Es gilt $L_3 = L_4$. Induktive Aufnehmer werden bevorzugt zur Messung von Längen bzw. Abständen eingesetzt.

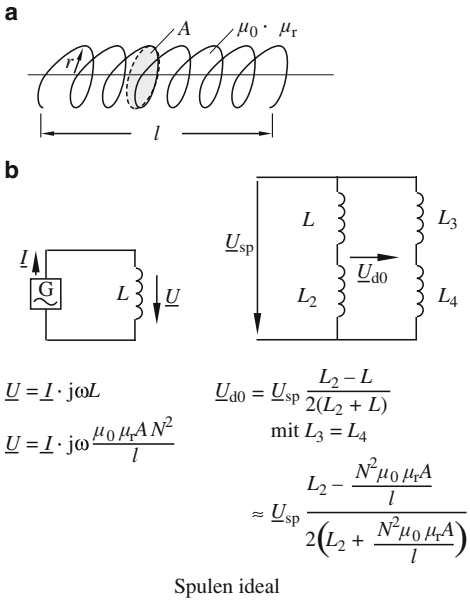
81.1.4 Optische Aufnehmer

Einheiten für die lichttechnischen Größen sind das Lumen (lm) und das Lux (lx).

Ein Lumen ist die Einheit des Lichtstromes Φ, der die gesamte von einer Lichtquelle abgegebene Strahlungsleistung darstellt. Ein Lux ist die Einheit der Beleuchtungsstärke E unter Berücksichtigung der beleuchteten Fläche.

$$\text{Beleuchtungsstärke } E = \frac{\text{Lichtstrom } \Phi}{\text{Fläche } A};$$

$$1\,\mathrm{lx} = 1\,\frac{\mathrm{lm}}{\mathrm{m}^2} \tag{81.4}$$



$$\underline{U} = \underline{I} \cdot \mathrm{j}\omega L$$

$$\underline{U} = \underline{I} \cdot \mathrm{j}\omega \frac{\mu_0 \mu_r A N^2}{l}$$

$$\underline{U}_{d0} = \underline{U}_{sp} \frac{L_2 - L}{2(L_2 + L)}$$

mit $L_3 = L_4$

$$\approx \underline{U}_{sp} \frac{L_2 - \frac{N^2 \mu_0 \mu_r A}{l}}{2\left(L_2 + \frac{N^2 \mu_0 \mu_r A}{l}\right)}$$

Spulen ideal

Abb. 81.4 Induktive Aufnehmer. **a** Zur Berechnung der Induktivität, **b** Messschaltungen

Licht ist der vom Auge wahrnehmbare Teil der Strahlung. Da die von einer Lichtquelle ausgesendete Strahlung meist aus einem Gemisch verschiedener Frequenzen und damit Farben besteht, muss der optische Aufnehmer auf diesen Frequenzbereich abgestimmt werden; er muss dort seine größte *spektrale Empfindlichkeit* aufweisen.

Zu den wichtigen optischen Aufnehmern zählen: Fotowiderstand (wurde bereits bei den ohmschen Aufnehmern behandelt), Fotodiode und Fotovervielfacher.

81.1.4.1 Fotodiode

Sie kann entweder aktiv oder passiv betrieben werden (Abb. 81.5). Die typische Kennlinie ist im Teil a dargestellt. Zum Kennlinienfeld im IV. Quadranten gehört die Schaltung b, in der die Fotodiode als aktives Element geschaltet ist. Der von ihr gelieferte Strom ist der Beleuchtungsstärke E proportional. Durch Einfügen eines Widerstandes R kann eine proportionale Spannung $U_L \sim E$ abgegriffen werden. Nähert sich allerdings die Spannung über der Fotodiode ihrer Schwellspannung, wirkt dies wie eine parallel liegende Diode und ergibt die nach rechts steil ansteigenden Kennlinienabschnitte. Im III. Quadranten wird die Fotodiode mit der Schaltung nach Abb. 81.5c in Sperrrichtung und damit passiv betrieben. Der Sperrstrom steigt mit steigender Beleuchtungsstärke E, wobei der Zusammenhang zwischen Sperrstrom und Beleuchtungsstärke linear ist. Über einen Widerstand R kann eine entsprechende Spannung U_R erzeugt werden.

Beide Betriebsarten sind zur Messung der Beleuchtungsstärke E geeignet, wobei der passive Betrieb zwar eine Hilfsspannungsquelle erfordert, aber durch die in Sperrrichtung vorgespannte

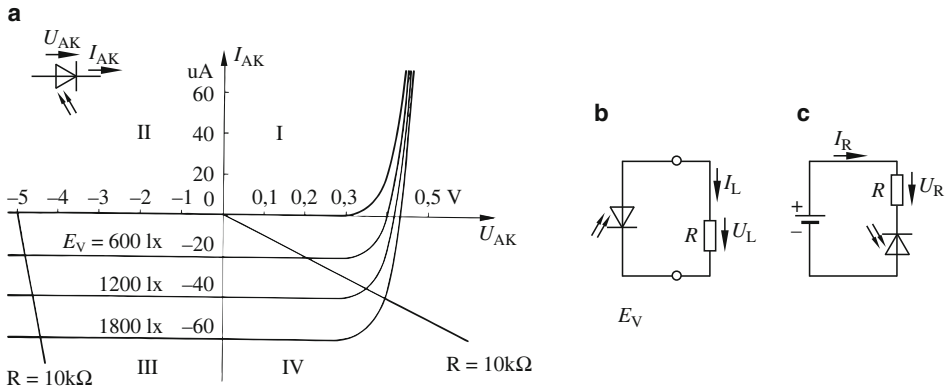


Abb. 81.5 Fotodiode. **a** Kennlinie, **b** aktive Betriebsart, **c** passive Betriebsart

Tab. 81.3 Eigenschaften von optischen Aufnehmern

Aufnehmer	Fotowiderstand	Fotodiode aktiv	Fotodiode passiv	Fotovervielfacher
Grenzfrequenz in Hz	10^2	10^3	10^8	10^2
Stromempfindlichkeit in µA/lx	10^2	10^{-1}	10^{-1}	10^5
Material	CdS	Si	Si	Sb, Cs
Max. spektrale Empfindlichkeit bei der Wellenlänge λ in mm	0,65	0,85	0,85	0,4
Temperatureinfluss auf die Messgröße	Groß	Mittel	Mittel	Sehr klein

Diode eine kleine Sperrschichtkapazität und damit eine hohe Grenzfrequenz für Wechselgrößen von 10 . . . 100 MHz erreicht wird. In der aktiven Betriebsart erhält man nur wenige kHz.

81.1.4.2 Fotovervielfacher

Die Wirkungsweise besteht darin, dass Lichtquanten aus einer fotoempfindlichen Kathode Elektronen (entsprechend einem Strom i_K) auslösen, die auf ihrem Weg zu einer ersten Anode beschleunigt werden, und zwar so, dass jedes Elektron beim Auftreffen mehrere Elektronen (Vervielfachungsfaktor n) auslöst. Diese Elektronen werden wiederum auf ihrem Weg zu einer zweiten Anode beschleunigt, wo sich die Vervielfachung wiederholt. Insgesamt sind bis zu $k = 10$ Anoden vorhanden, so dass der Anodenstrom an der letzten Anode $i = i_K \cdot n^k$ beträgt. Fotovervielfacher sind so empfindlich, dass sich bei einem Verstärkungsfaktor $n^k = 10^8$ einzelne Lichtquanten nachweisen lassen.

Die für die Messtechnik wichtigen Eigenschaften der drei aufgeführten Aufnehmer sind in Tab. 81.3 zusammengestellt. Der Temperatureinfluss wurde nur grob klassifiziert angegeben, da er von der jeweiligen Beleuchtungsstärke abhängt.

81.1.5 Ladungsliefernde Aufnehmer

Bei bestimmten Kristallen (z. B. Quarz oder Bariumtitanat) oder organischen Polymeren (z. B. Polyvinylidendifluorid PVDF) treten zwei Effekte auf.

1. *Piezoelektrischer Effekt*: Durch mechanische Verformung tritt an den gegenüberliegenden Seiten eines Plättchens eine Ladungsverschiebung und damit eine Polarisation auf (Abb. 81.6a). Da der elektrische Widerstand der Materialien sehr hoch ist ($\rho > 10^{13}\,\Omega$ cm), hat eine Ladungsänderung ΔQ eine Spannungsänderung ΔU gemäß $\Delta Q = C \cdot \Delta U$ zur Folge, wobei man diese Spannung über entsprechend angebrachte Elektroden abgreifen kann. Damit ist die Spannung eine Funktion der Verformung oder der die Verformung bewirkenden Größe, z. B. einer Kraft. Die Ladungsänderung ist der Kraftänderung proportional.

$$\Delta Q = k_p \cdot \Delta F \quad \Rightarrow \quad \Delta F = C_p \cdot \Delta U / k_p;$$
$$k_p \approx (2 \ldots 250) \cdot 10^{-12}\,\mathrm{A\,s/N} \qquad (81.5)$$

Darin ist k_p die *piezoelektrische Konstante*. Die durch eine konstante Krafteinwirkung entstandene Spannung steht allerdings nicht beliebig lange zur Verfügung, da sich die Ladungen über den endlichen Widerstand R_i des Piezomaterials ausgleichen. Das elektrische Ersatzschaltbild zeigt Abb. 81.6b.

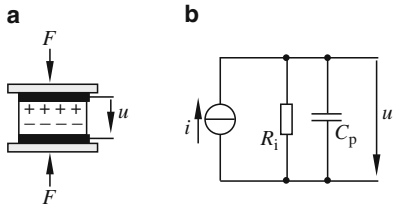


Abb. 81.6 Piezoaufnehmer. **a** Aufbau **b** Ersatzschaltbild

Beispiel 3

Ein Piezomaterial hat folgende Daten:

$k_p = 10^{-11}\,\mathrm{A\,s/N}$, Fläche $A = 10\,\mathrm{cm}^2$, Dicke $d = 1\,\mathrm{mm}$, $\rho = 10^{14}\,\Omega\,\mathrm{cm}$, $\varepsilon_r = 5$, $F = 10\,\mathrm{N}$.

Daraus lassen sich folgende Größen berechnen:

Innenwiderstand $R_i = \rho \cdot d/A = (10^{14} \cdot 0{,}1/10)\,\Omega\,\mathrm{cm} \cdot \mathrm{cm/cm}^2 = 10^{12} \cdot \Omega$.

Kapazität $C_p = \varepsilon_0 \cdot \varepsilon_r \cdot A/d = (8{,}85 \cdot 10^{-12} \cdot 5 \cdot 10^{-3}/10^{-3})\,\mathrm{F} = 44\,\mathrm{pF}$.

Ladung $Q = k_p \cdot F = 10^{-11} \cdot 10\,\mathrm{A\,s} = 10^{-10}\,\mathrm{A\,s}$.

Spannung $U = 10^{-10}\,\mathrm{A\,s}/44\,\mathrm{pF} = 2{,}27\,\mathrm{V}$

Entladezeitkonstante $t = R_i \cdot C_p = 10^{12} \cdot 44 \cdot 10^{-12}\,\mathrm{s} = 44\,\mathrm{s}$

Soll die Entladezeitkonstante durch ein angeschlossenes Spannungsmessgerät nicht wesentlich verringert werden, ist der Innenwiderstand des Spannungsmessgerätes größer als R_i zu wählen. Häufig wird die zu messende Spannung zunächst einem elektronischen Messverstärker mit entsprechend hohem Eingangswiderstand zugeführt und erst anschließend gemessen. ◀

2. *Pyroelektrischer Effekt*: Bei den genannten Materialien wird eine Polarisation auch durch Temperaturänderungen hervorgerufen. Der formelmäßige Zusammenhang lautet:

$$\Delta Q = k_q \cdot A \cdot \Delta T \tag{81.6}$$

k_q pyroelektrische Konstante in $\mathrm{A\,s/(m^2\,K)}$; A Fläche in m^2; ΔT Temperaturänderung in K.
Dieser Effekt wird häufig bei den *Bewegungsmeldern* eingesetzt. Eine sich nähernde Person bewirkt durch ihre Körperwärme eine Änderung der Umgebungstemperatur, die von dem pyroelektrischen Aufnehmer erfasst wird.

Beispiel 4

Eine 50 µm dicke Scheibe liefert mit $k_q = 2 \cdot 10^{-4}\,\mathrm{A\,s/(m^2\,K)}$, $\varepsilon_r = 50$ und $\Delta T = 1\,\mathrm{K}$:

$$\begin{aligned}\Delta U &= \Delta Q/C \\ &= k_q \cdot A \cdot \Delta T/(\varepsilon_0 \cdot \varepsilon_r \cdot A/d) = k_q \cdot \Delta T \cdot d/(\varepsilon_0 \cdot \varepsilon_r) \\ &= 2 \cdot 10^{-4} \cdot 1 \cdot 50 \cdot 10^{-6}/(8{,}85 \cdot 10^{-12} \cdot 50)\,\mathrm{V} = 20\,\mathrm{V}\end{aligned}$$

Das Ersatzschaltbild ist gleich dem des Piezoaufnehmers in Abb. 81.6b. ◀

81.1.6 Thermische Aufnehmer

Zu den wichtigsten thermischen Aufnehmern gehören temperaturabhängige Widerstände (Ohm'sche Aufnehmer), pyroelektrische Aufnehmer (Ladungsliefernde Aufnehmer) und Thermoelemente.

81.1.6.1 Thermoelemente

Verbindet man zwei elektrisch leitende, unterschiedliche Materialien nach Abb. 81.7a punktförmig miteinander und hat diese Verbindungsstelle die Temperatur T_M, so entsteht dort eine Thermospannung (*Seebeck-Effekt*).

$$\textit{Thermospannung} \quad U_{\mathrm{th}} = k_{\mathrm{AB}} \cdot T_M \tag{81.7}$$

k_{AB} Thermoempfindlichkeit der Materialien A und B in V/K oder mV/100 K; T_M Temperatur der Verbindungsstelle.

Sie ist der Temperatur der Verbindungsstelle proportional. Die Konstante k_{AB} ist die *Thermoempfindlichkeit* und wird in der Regel in mV/100 K angegeben. Die Indizierung gibt die Richtung der Spannung, bezogen auf die Materialien A und B, an. Ist k_{AB} positiv, hat Material A das positivere Potenzial gegenüber Material B. In der Praxis bezieht man zur Angabe der Thermoempfindlichkeit die einzelnen Materialien (x) auf das Bezugsmaterial Platin (Pt) und gibt dann $k_{x\,\mathrm{Pt}}$ an. Zur Bestimmung von k_{AB} bildet man die Differenz:

$$k_{\mathrm{AB}} = k_{\mathrm{A\,Pt}} - k_{\mathrm{B\,Pt}} \tag{81.8}$$

In der Praxis treten Thermospannungen grundsätzlich bei Temperatur*differenzen* auf. Insofern entspricht Abb. 81.7a nicht der Wirklichkeit. Vielmehr muss die entstehende Spannung gemessen oder anderweitig ausgewertet werden. Dazu wird an die zwei freien Enden nach Abb. 81.7b ein Messgerät angeschlossen, von dem angenommen werden soll, dass es einen durchgehenden elektrischen Kreis aus *einem* Material besitzt (Beispiel: Drehspulmesswerk mit Cu-Zuleitungen, Cu-Anschlussklemmen, internen Cu-Leitern, Cu-Drehspule). Dann entsteht an jeder Verbindungsstelle zweier unterschiedlicher Materialien eine Thermospannung, deren Wert nach Gleichung (81.7) eingesetzt wird. Die drei Teilspannungen werden zur Gesamt-Leerlaufspannung U_{th0} addiert (Strom durch das Messgerät zu Null angenommen).

$$U_{\mathrm{th0}} = U_1 + U_2 + U_3 = k_{\mathrm{CA}} \cdot T_1 + k_{\mathrm{AB}} \cdot T_2 + k_{\mathrm{BC}} \cdot T_3;$$

mit (81.8) folgt daraus

$$U_{\mathrm{th0}} = (k_{\mathrm{C\,Pt}} - k_{\mathrm{A\,Pt}})T_1 + (k_{\mathrm{A\,Pt}} - k_{\mathrm{B\,Pt}})T_2 + (k_{\mathrm{B\,Pt}} - k_{\mathrm{C\,Pt}})T_3$$

und nach einer Umformung und entsprechender Zusammenfassung

$$U_{\mathrm{th0}} = k_{\mathrm{C\,Pt}}(T_1 - T_3) + k_{\mathrm{A\,Pt}}(T_2 - T_1) + k_{\mathrm{B\,Pt}}(T_3 - T_2).$$

Sorgt man dafür, dass die beiden zum Messgerät führenden Verbindungsstellen die gleiche Temperatur besitzen, also $T_1 = T_3$ und wählt die für die Praxis anschaulichen Bezeichnungen $T_1 = T_3 = T_V$ als Temperatur der *Vergleichsstelle* und $T_2 = T_M$ als Temperatur der *Messstelle*, dann vereinfacht sich die Formel zu:

$$U_{\mathrm{th0}} = k_{\mathrm{A\,Pt}}(T_M - T_V) + k_{\mathrm{B\,Pt}}(T_V - T_M)$$

Schließlich wird unter Verwendung von (81.8) daraus:

$$\begin{aligned}U_{\mathrm{th0}} &= (k_{\mathrm{A\,Pt}} - k_{\mathrm{B\,Pt}}) \cdot (T_M - T_V) \\ &= k_{\mathrm{AB}} \cdot (T_M - T_V)\end{aligned} \tag{81.9}$$

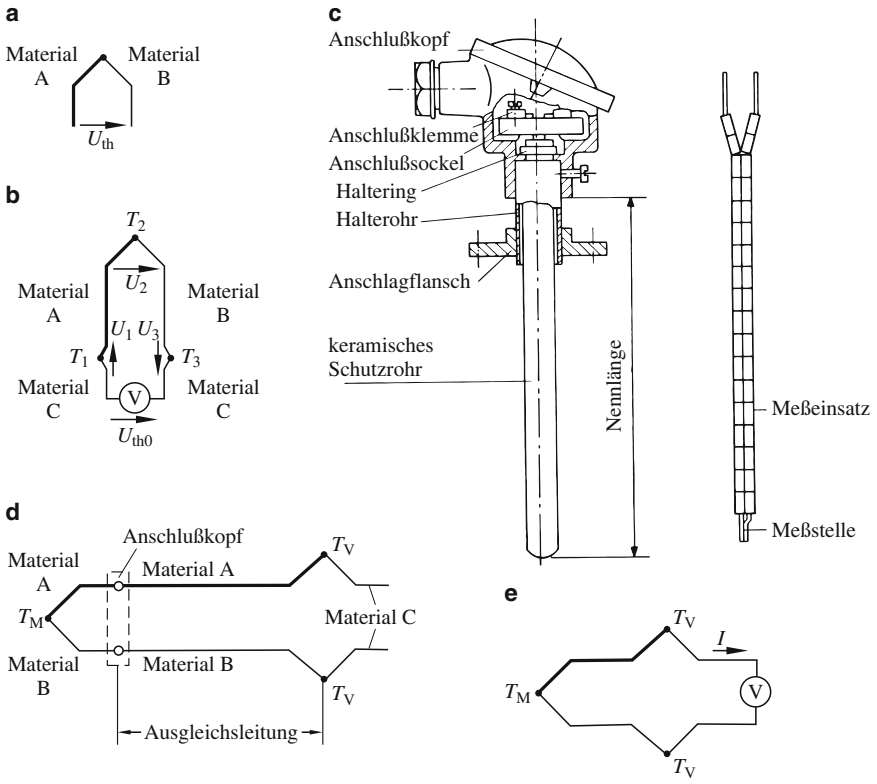


Abb. 81.7 Thermoelemente. **a** Theoretisches Prinzip, **b** praktisches Modell, **c** Ausführungsformen, **d** Ausgleichsleitung, **e** Messschaltung

Teil X

Die Spannung U_{th0} ist der Temperaturdifferenz $T_M - T_V$ proportional. Um die Temperatur T_M messen zu können, muss die Temperatur T_V bekannt sein. In der Praxis wird die Temperatur T_V der Vergleichsstelle durch z. B. „Eiswasser“ mit 0 °C (genau +0,01 °C) hergestellt und die zwei Verbindungsstellen in dieses Eis-Wasser-Gemisch eingetaucht.

In Abb. 81.7c sind zwei gängige Ausführungsformen von Thermoelementen zum Messen auch in Flüssigkeiten dargestellt. Um die Vergleichsstelle aus dem Anschlusskopf an einen geeigneteren Ort zu verlegen, werden *Ausgleichsleitungen* nach Abb. 81.7d verwendet, die aus dem gleichen Material wie das eigentliche Thermoelement bestehen, aber billiger und flexibel sind. Damit entstehen im Anschlusskopf keine Thermospannungen.

Der in der Messschaltung nach Abb. 81.7e fließende Strom berechnet sich aus $I = U_{th0}/(R_i + R_I)$, worin R_i der Innenwiderstand des Thermoelementes und R_I der des Messinstrumentes ist. Der Innenwiderstand des Thermoelementes ergibt sich durch den ohmschen Widerstand der Thermoelementanordnung.

In Tab. 81.4 sind die Thermoempfindlichkeiten einiger Materialien gegenüber Platin eingetragen. Zwar sind theoretisch alle elektrischen Leiter untereinander zu Thermoelementen zu kombinieren, aber nur wenige Paarungen erfüllen die in der Praxis gestellten Forderungen: Die Verbindungsstelle muss schweißbar, die Schweißstelle muss über lange Zeit chemisch und mechanisch stabil und die Thermoempfindlichkeit sollte so groß wie möglich sein.

Tab. 81.4 Thermoempfindlichkeit geeigneter Thermoelement-Materialien

Material x	$k_{x\,\mathrm{Pt}}$ in mV/100 K
Konstantan	−3,47 … −3,04
Nickel	−1,9 … −1,3
Platin	−0,0
Kupfer	+0,75
Eisen	+1,9
Nickel-Chrom	+2,2
Silizium	+42 … +50

Beispiel

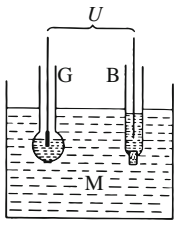
Beispiel 5 Ein Thermoelement aus der Materialkombination Eisen-Konstantan liefert bei einer zu messenden Temperatur $T_M = 800\,°\mathrm{C}$ und einer Temperatur der Vergleichsstelle von $T_V = 0\,°\mathrm{C}$ eine Thermo-Leerlaufspannung von $U_{th0} = (k_{\mathrm{Fe\,Konst}}) \cdot (T_M - T_V) \approx 5{,}2 \cdot (800 - 0)\,\mathrm{K} \cdot \mathrm{mV}/100\,\mathrm{K} = 41{,}6\,\mathrm{mV}$. ◀

Da die entstehenden Thermospannungen relativ klein sind, ist Silizium ein bevorzugtes Thermoelementmaterial. Allerdings gibt es bei der Fertigung zur Zeit noch technologische Probleme, die den Einsatz erschweren.

Tab. 81.5 enthält häufig verwendete Materialkombinationen und deren Eigenschaften.

Tab. 81.5 Eigenschaften häufig verwendeter Thermoelemente

Materialkombination A-B	Typ	Maximal messbare Temperatur in °C	k_{AB} in mV/100 K
Eisen-Konstantan	J	ca. 1800	ca. 5,2
Kupfer-Konstantan	T	ca. 1400	ca. 4,2
Nickelchrom-Nickel	K	ca. 1300	ca. 3,7
Platinrhodium-Platin, verschiedene Legierungen möglich	R, S, B	ca. 1700 … 1800	ca. 0,6 … 1,1



G Glasmeßelektrode enthält Flüssigkeit mit pH_0
B Bezugselektrode
M Meßflüssigkeit mit pH_x
U Meßspannung

Abb. 81.8 Messung des pH-Wertes mit Glaselektroden

81.1.7 Chemische Aufnehmer

Chemische Aufnehmer werden zunehmend für Messungen im Zusammenhang mit Prozesssteuerungen und der Umweltmesstechnik eingesetzt. Hier sollen einige wenige vorgestellt werden: pH-Wert-Messeinrichtung mit Glaselektrode, Aufnehmer zur Messung der Sauerstoffkonzentration und Aufnehmer zur Messung von Gaskonzentrationen allgemein.

81.1.7.1 pH-Wert-Messeinrichtung mit Glaselektrode

Der pH-Wert ist der mit −1 multiplizierte dekadische Logarithmus der Konzentration von H^+-Ionen. Eine Lösung mit dem pH-Wert 7 ist neutral (z. B. Trinkwasser), ein pH-Wert < 7 kennzeichnet saure, ein pH-Wert > 7 alkalische Flüssigkeiten.

Den Aufbau einer Messsonde zeigt Abb. 81.8. Die Glaselektrode G ist mit einer Flüssigkeit vorgegebener Wasserstoffionenkonzentration pH_0 gefüllt und nach unten hin so dünn, dass eine gewisse Leitfähigkeit (Widerstand $10^7 \ldots 10^{10}\,\Omega$) zur Flüssigkeit mit der gesuchten Konzentration pH_x besteht. Die Bezugselektrode B ist mit einer speziellen Salzlösung gefüllt und steht über das poröse Glasgefäß in relativ guter elektrischer Verbindung mit der Messflüssigkeit. Für die sich einstellende Spannung U_{pH} ist folgender Zusammenhang gefunden worden:

$$U_{pH} = 2 \cdot 10^{-4} \cdot T \cdot (pH_0 - pH_x)\,V/K.$$

Mit $T = 273\,K$, entsprechend 0 °C, folgt daraus

$$U_{pH0} = 0{,}0546 \cdot (pH_0 - pH_x) \quad \text{in V} \qquad (81.10)$$

Unterscheiden sich beide pH-Werte um 1 ($pH_0 > pH_x$), beträgt die Spannung 54,6 mV. Da der Innenwiderstand dieser Spannungsquelle im Bereich $10^7 \ldots 10^{10}\,\Omega$ liegt, muss das Messgerät zur Spannungsmessung einen wesentlich höheren Eingangswiderstand aufweisen. Deshalb setzt man häufig spezielle Messverstärker mit entsprechend hohen Eingangswiderständen ein.

Die Glaselektrode hat nur eine begrenzte Lebensdauer und muss jeweils nach einigen Jahren ausgetauscht werden. Deshalb hat man nach weiteren Messmöglichkeiten gesucht. Zurzeit sind pH-Wert-Aufnehmer mit Feldeffekttransistor in der Entwicklung, die eine höhere Lebensdauer und die Zuverlässigkeit der Glaselektrodenausführung haben.

81.1.7.2 Aufnehmer zur Messung der Sauerstoffkonzentration

Erhitzt man Keramik aus Zirkonoxid (ZrO_2) und Yttriumoxid (Y_2O_3) auf mindestens 350 °C, leitet sie Sauerstoffionen. Nach Abb. 81.9a bringt man an den gegenüberliegenden Seiten poröse Platinelektroden an. Stehen die zwei Elektroden unter dem Einfluss unterschiedlicher Sauerstoff-Partialdrücke p_1 und p_2, so entsteht eine Spannung

$$U_0 = 0{,}05 \cdot T \cdot \lg(p_2/p_1) \cdot \frac{\text{mV}}{\text{K}} \quad \text{in mV} \qquad (81.11)$$

T Keramiktemperatur in K; p_1, p_2 Partialdrücke in Pa oder bar.

Bei $T = 800\,K$ und $p_2 = 10 \cdot p_1$ ergibt sich ein Wert von $U_0 = 40\,mV$.

Eine Sonderausführung ist die aus der Automobiltechnik bekannte *Lambda-Sonde* (Abb. 81.9b). Erhält der Motor genau soviel Sauerstoff, wie zur vollständigen Verbrennung des Kraftstoffes erforderlich ist, so ist $\lambda = 1$; bei $\lambda < 1$ fehlt Sauerstoff, und man spricht von „fettem Gemisch“, bei $\lambda > 1$ ist zu viel Sauerstoff und damit ein „mageres Gemisch“ vorhanden. Durch Verwendung des Platin-Katalysators hat man erreicht, dass sich die Spannung an der Sonde im Bereich $0{,}9 \leq \lambda \leq 1{,}1$ sehr stark ändert, nämlich von ca. 1 V auf 40 mV (Leerlaufkennlinie Abb. 81.9b: p_1^* ist der Partialdruck des thermodynamischen Gleichgewichtes).

81.1.8 Aufnehmer zur Messung von Gaskonzentrationen allgemein

Abb. 81.10 zeigt den prinzipiellen Aufbau eines Aufnehmers der Firma Figaro. Durch den Heizdraht wird die Oberfläche soweit erhitzt, dass die im Innern befindlichen Gase entweder oxidieren oder reduzieren und eine Elektronenverarmung oder Anreicherung bewirken. Damit ändert sich die Leitfähigkeit zwischen den Elektroden, und der Widerstand ist ein Maß für die Gaskonzentration. Prinzipiell reagieren diese Aufnehmer auf Konzentrationsänderungen sowohl von reduzierenden

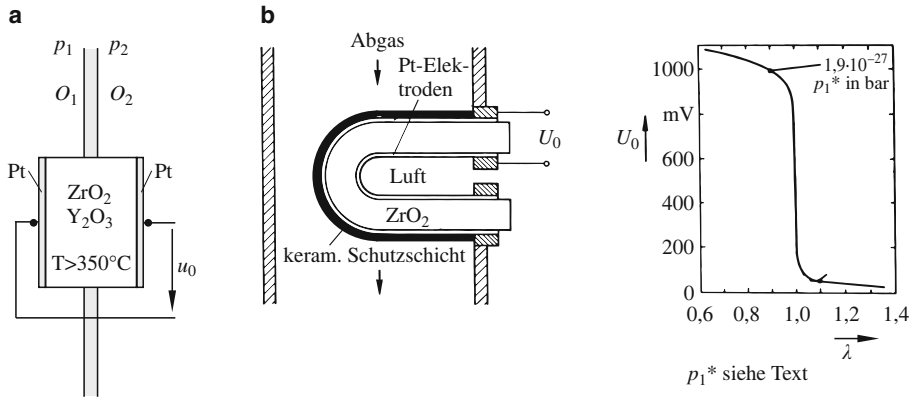


Abb. 81.9 Aufnehmer zur Messung der Sauerstoffkonzentration. **a** Prinzip, **b** Lambda-Sonde (nach Unterlagen Fa. Bosch): Aufbau und Leerlaufspannung

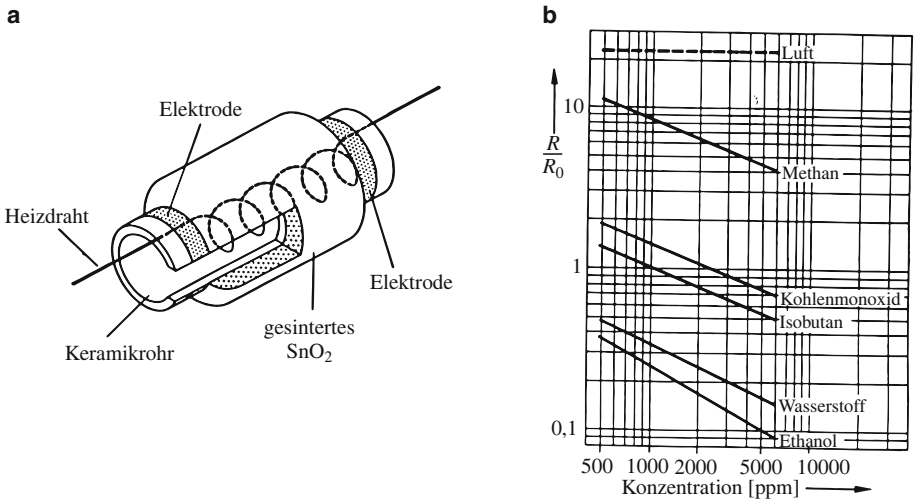


Abb. 81.10 Gassensor (Fa. Figaro). **a** Aufbau, **b** Kennlinie

Gasen wie H_2, CO, CH_4, Kohlenwasserstoffen und Alkoholen als auch von oxidierenden Gasen wie N_2 und O_2, so dass eine Widerstandsänderung nicht einem bestimmten Gas zugeordnet werden kann. Ändert allerdings nur ein Gas seine Konzentration, während die der anderen vorhandenen Gase konstant bleibt, lässt sich die Gaskonzentration messen. Nach diesem Verfahren werden z. B. Räume mit Gasheizanlagen auf ausströmendes Erdgas (CH_4) überwacht.

Durch die erforderliche Heizleistung von einigen Watt sind diese Aufnehmer nicht für Batteriebetrieb geeignet. Wurde die Heizung unterbrochen, stellt sich erst nach einer Betriebszeit von einigen Tagen der stabile Arbeitspunkt wieder ein. Eventuell ist sogar eine Neukalibrierung erforderlich.

81.2 Messverfahren

Ergänzend zu den im Abschn. 80.1 „Messaufnehmer“ beschriebenen Messverfahren sollen hier einige weitere vorgestellt werden, wobei es sich um eine kleine Auswahl unter einer Vielzahl von eingesetzten Verfahren handelt. Dabei wird das der Messung zugrunde liegende Verfahren etwas ausführlicher gezeigt, um damit das Verständnis für hier nicht aufgeführte Verfahren zu erleichtern.

81.2.1 Kraftmessung mit Dehnungsmessstreifen (DMS)

Die Messung von Kräften geschieht in der Mehrzahl der Anwendungen mit Dehnungsmessstreifen, weil dieses Verfahren sehr genau und technisch ausgereift ist. Das Prinzip beruht darauf, dass ein Messkörper (z. B. ein Stab) durch die Einwirkung einer Kraft (Zug oder Druck) eine Längenänderung erfährt. Auf diesen Messkörper ist ein elektrischer Leiter aufgeklebt, der den Längenänderungen zwangsweise folgen muss. Dadurch wird der Leiter länger und dünner oder kürzer und dicker und ändert seinen elektrischen Widerstand. Der mathematische Zusammenhang zwischen der Widerstandsänderung und der Kraft lässt sich für verschiedene Messkörper ableiten, das Ergebnis wird im Folgenden angegeben.

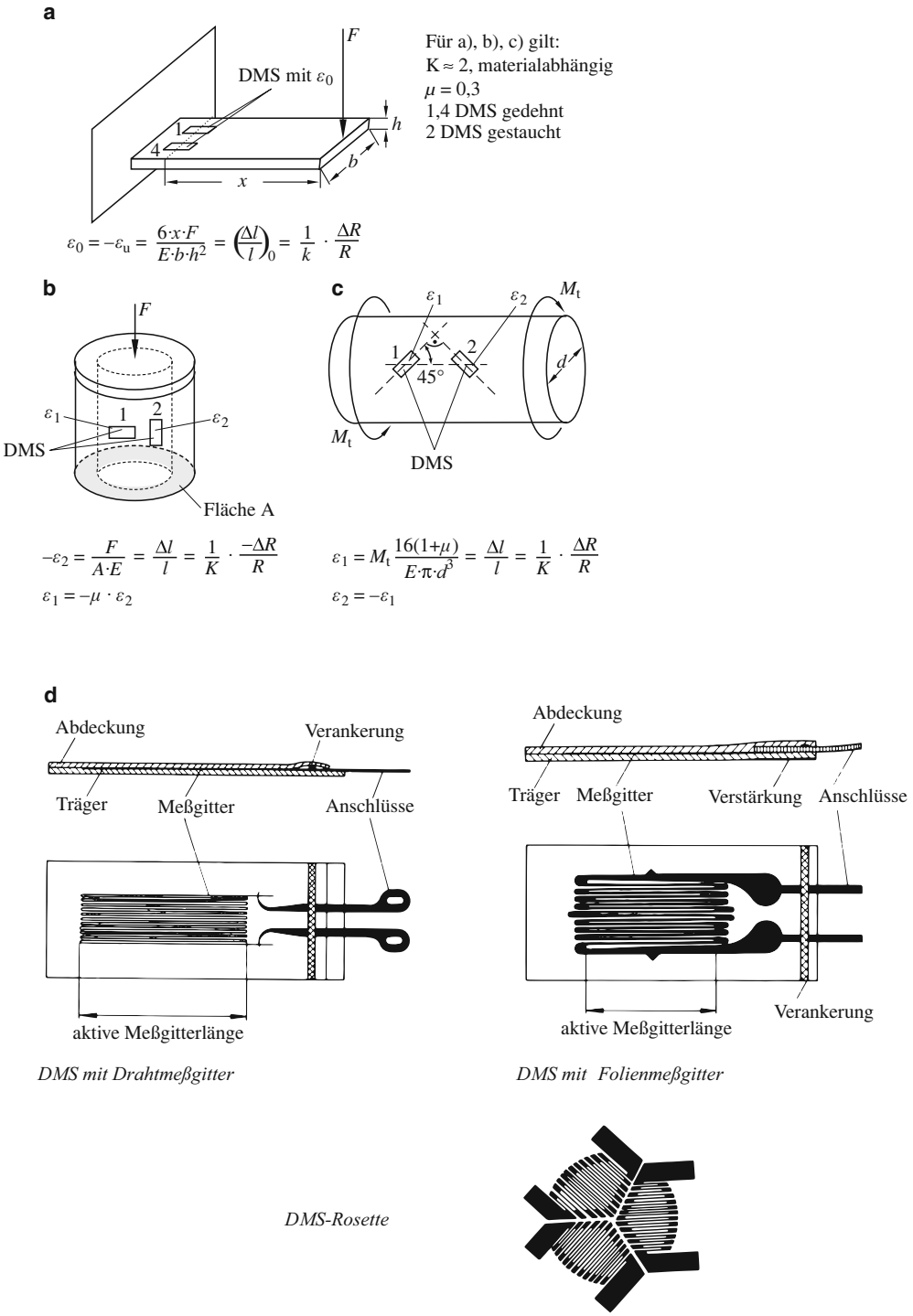


Abb. 81.11 Anordnung von Dehnungsmessstreifen. **a** Einseitig eingespannter Stab, **b** Kraftmessdose, **c** Torsionsstab, **d** technische Ausführungen

Dehnungsmessstreifen (DMS) werden nach Abb. 81.11d entweder als Drahtgitter-DMS oder als Folien-DMS mit aufgedampftem Messdraht in den Vorzugswerten 120, 350 und 600 Ω hergestellt. Bestimmte Anwendungen erfordern einen geeigneten DMS-Aufbau; so dient beispielsweise die DMS-Rosette zum Messen von mehrachsigen mechanischen Spannungsverteilungen. Der Zusammenhang zwischen der relativen Längenänderung $\Delta l/l$ des Messkörpers und damit des DMS und der relativen Widerstandsänderung $\Delta R/R$ ist gegeben durch

$$\Delta R/R = K \cdot \Delta l/l = K \cdot \varepsilon \quad \text{mit } \varepsilon = \Delta l/l \tag{81.12}$$

Formelzeichen siehe Text.

K steht für *K-Faktor*. Sein Wert ist materialabhängig. Für metallische Leiter gilt $K \approx 2$. Die maximale relative Dehnung $\varepsilon_{\max}$ beträgt für Metalle $\approx 10^{-3}$, damit man im elastischen Bereich bleibt. Daraus folgt die maximale relative Widerstandsänderung zu $2 \cdot 10^{-3}$. Um derart kleine Widerstandsänderungen messen zu können, werden DMS in Brückenschaltungen eingesetzt. Gleichung (81.12) gilt auch für den Fall, dass der Messkörper gestaucht wird. Dann sind $\Delta R/R$ und ε negativ einzusetzen, Abb. 81.11b, c.

Ein Problem beim Einsatz von Dehnungsmessstreifen ist die Temperaturabhängigkeit des elektrischen Widerstandes. Bei einer Temperaturänderung von 1 °C errechnet sich die relative Widerstandsänderung von Kupfer mit $\alpha = 4 \cdot 10^{-3}/\mathrm{K}$ zu:

$$\begin{aligned} R &= R_0(1 + \alpha \cdot \Delta T) = R_0 + R_0 \cdot \alpha \cdot \Delta T \\ \Rightarrow \quad (R - R_0)/R_0 &= \Delta R/R_0 = \alpha \cdot \Delta T = (4 \cdot 10^{-3}/\mathrm{K}) \cdot 1\,\mathrm{K} \\ &= 4 \cdot 10^{-3} \end{aligned}$$

Damit bewirkt eine Temperaturänderung von 1/2 °C die gleiche Widerstandsänderung wie die maximal zulässige Dehnung $\varepsilon_{\max}$. Um dennoch mit DMS sinnvoll messen zu können, wird folgendermaßen verfahren:

Erstens wählt man als Material für die DMS Konstantan mit $\alpha < |10^{-5}\,\mathrm{K}^{-1}|$, und zweitens verwendet man in der Brückenschaltung zwei oder vier DMS. Beim Einsatz von zwei DMS spricht man von einer *Halbbrücke*, bei vier DMS von einer *Vollbrücke*. Die zwei DMS in der Halbbrücke werden „untereinander" angeordnet. Da sich deren Widerstände – gleiche Temperatur vorausgesetzt – gleichsinnig mit der Temperatur ändern, wird die Spannung in der Brückendiagonalen von Temperaturänderungen unabhängig. Natürlich muss man dafür sorgen, dass ein DMS gedehnt, der andere gestaucht wird, da sich andernfalls auch die Nutzsignale aufheben würden. Durch die gegensinnigen Widerstandsänderungen der DMS erhält man bei der Halbbrücke weiterhin ein größeres Ausgangssignal gegenüber dem Einsatz von nur einem DMS. Sind außerdem die Widerstandsänderungen bei Dehnung und Stauchung gleich groß ($+\Delta R$ und $-\Delta R$), ergibt sich ein linearer Zusammenhang zwischen der Widerstandsänderung und der Spannung U_{d0} in der Brückendiagonalen. Man wird deshalb die Halbbrücke der Viertelbrücke vorziehen. Kann man Dehnung und Stauchung nicht gleichzeitig mit zwei DMS erfassen, wählt man trotzdem zwei DMS, von denen einer die Dehnung oder Stauchung erfasst, der andere als „Dummy" in innigem Wärmekontakt zum messenden DMS steht, aber nicht aufgeklebt wird. Er besorgt dann lediglich die Temperaturkompensation.

Bei der Verwendung von vier DMS in einer Vollbrücke gelten die gleichen Aussagen wie bei der Halbbrücke mit dem Zusatz, dass die Spannung in der Brückendiagonalen nochmals größer ist, maximal doppelt so groß, wenn $+\Delta R$ und $-\Delta R$ gegeben sind. Die Anordnung der DMS ist in der Brückenschaltung so zu wählen, dass sich die zwei gestauchten diagonal gegenüberliegen, ebenso die zwei gedehnten.

Als Messkörper (Abb. 81.11a–c) werden in der Regel ein einseitig fest eingespannter Stab, ein in Längsrichtung mit einer Kraft beaufschlagter Zylinder, auch *Kraftmessdose* genannt, und ein tordierter Stab zur Messung des Drehmomentes verwendet. In Tab. 81.6 sind die für diese Messkörper möglichen Anordnungen für DMS und die sich daraus ergebenden Spannungen U_{d0} aufgeführt. Dazu wurde in (80.7) für die Widerstände jeweils $R_i \pm \Delta R_i$, $i = 1 \ldots 4$, eingesetzt. Häufig sind die Widerstände R_i gleich groß, d. h. $R_1 = R_2 = R_3 = R_4 = R$, weil die Brückenschaltung dann die größte Empfindlichkeit besitzt und damit maximales U_{d0} liefert. In Tab. 81.6 ist diese Dimensionierung zugrunde gelegt. Für die genannten Messkörper gelten die folgenden Beziehungen:

1. *Einseitig fest eingespannter Stab*: Die Dehnung ist abhängig vom Ort des Aufklebens, die Abhängigkeit ist in (81.13) gegeben und in Abb. 81.11a erläutert.

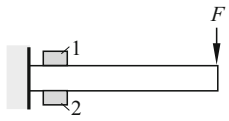
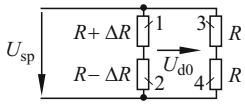
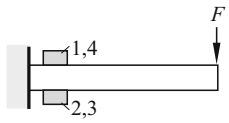
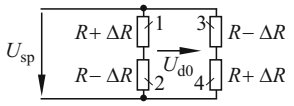
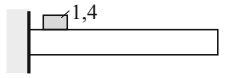
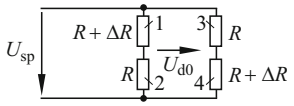
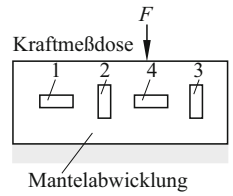
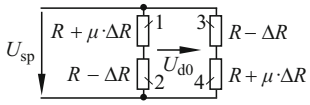
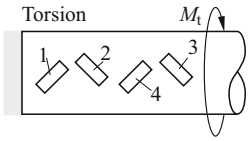
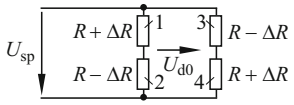
$$\sigma_o = \frac{6 \cdot x \cdot F}{b \cdot h^2} \quad \text{mit } \sigma_u = -\sigma_o; \quad \sigma = e \cdot E \qquad (81.13)$$

   σ_o Spannung auf der Oberseite und σ_u Spannung auf der Unterseite des Balkens in $\mathrm{N/m^2}$; F Kraft in N; b Balkenbreite in m; h Balkenhöhe in m; x Aufklebeort der DMS, vom freien Balkenende aus gemessen; $\varepsilon = \Delta l/l$ relative Dehnung oder Stauchung; E Elastizitätsmodul in $\mathrm{N/m^2}$.
   $\sigma_u = -\sigma_o$ gilt nur für den Fall, dass die DMS oben und unten genau untereinander aufgeklebt sind.
   Ein Sonderfall ergibt sich, wenn der Balken nur von oben oder nur von unten zugänglich ist. Dann kann man zwei DMS aufkleben, die in der Brückenschaltung diagonal angeordnet werden müssen. Diese Anordnung ist weder temperaturkompensiert noch liefert sie eine lineare Abhängigkeit zwischen U_{d0} und $\Delta R/R$. Einziger Vorteil: Die Spannung U_{d0} ist doppelt so groß wie beim Einsatz nur eines DMS. Bei der Vollbrücke müssen die zwei DMS oben genau nebeneinander angeordnet werden, und die zwei DMS unten müssen sich genau unter den zwei oberen befinden, damit gleiche Dehnungen oder Stauchungen erfolgen und die berechnete Abhängigkeit gültig ist.
2. *Kraftmessdose*: Bei Dehnung ($+\varepsilon$) oder Stauchung ($-\varepsilon$) eines Zylinders in Längsrichtung findet gleichzeitig eine Umfangsabnahme ($-\Delta U/U$) oder -zunahme ($+\Delta U/U$) in Querrichtung statt. Beide haben entgegengesetztes Vorzeichen und sind über die *Querkontraktionszahl* $\mu = 0{,}3$ miteinander verknüpft: $\Delta U/U = -0{,}3 \cdot \varepsilon$. Kraftmessdosen werden als Vollbrücke aufgebaut und sind damit temperaturkompensiert. Die Abhängigkeit zwischen $\Delta R/R$ und U_{d0} ist wegen $\mu \neq 1$ nichtlinear.
   Der Zusammenhang zwischen ε und der Kraft F ergibt sich mit dem *Hooke'schen Gesetz* und der Beziehung „Spannung gleich Kraft pro Fläche" zu

$$\varepsilon = \frac{\sigma}{E} = \frac{F}{A \cdot E} \qquad (81.14a)$$

   E Elastizitätsmodul in $\mathrm{N/m^2}$; A Fläche des Zylinders in $\mathrm{m^2}$; F Kraft in N.
   Darin ist A die Querschnittsfläche des Voll- oder Hohlzylinders. In der Regel werden Hohlzylinder verwendet, da sie gegenüber Vollzylindern bei gleicher Querschnittsfläche eine größere Stabilität aufweisen.

Tab. 81.6 Messkörper mit Dehnungsmessstreifen

einseitig eingespannter Stab		
		$U_{d0} = -U_{sp}\frac{\Delta R}{2R}$
		$U_{d0} = -U_{sp}\frac{\Delta R}{R}$
		$U_{d0} = -U_{sp}\frac{\Delta R}{2R+\Delta R}$
		$U_{d0} = -U_{sp}\frac{\Delta R(1+\mu)}{2R+\Delta R(\mu-1)}$
		$U_{d0} = -U_{sp}\frac{\Delta R}{R}$

3. *Tordierter Stab:* Die relative Widerstandsänderung $\Delta R/R$ hängt mit dem Torsionsmoment M_t wie folgt zusammen:

$$\frac{\Delta R}{R} = K \cdot M_t \cdot \frac{16 \cdot (1+\mu)}{E \cdot \pi \cdot d^3} \qquad (81.14b)$$

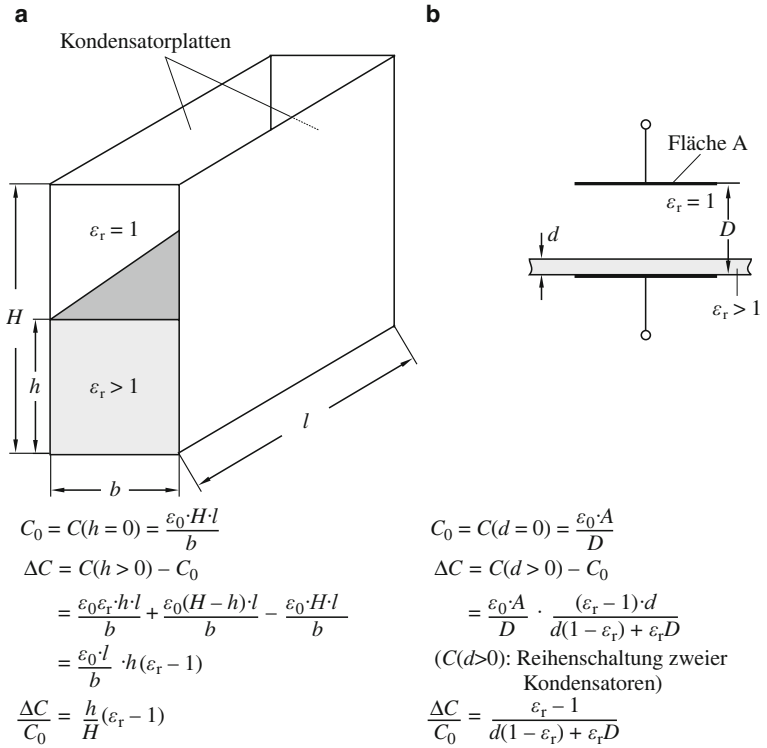
K K-Faktor ($K \approx 2$); μ Querkontraktionszahl ($\mu = 0{,}3$); E Elastizitätsmodul in N/m^2; d Torsionsstabdurchmesser in m; M_t Torsionsmoment in N m.
Bei der gezeigten Anordnung von 4 DMS sind deren Dehnung und Stauchung entgegengesetzt gleich groß, so dass sich ein linearer Zusammenhang zwischen $\Delta R/R$ und U_{d0} ergibt.
Da die Spannung in der Brückendiagonalen bei allen drei aufgeführten Messkörpern maximal einige mV beträgt, verwendet man häufig zur Speisung der Brücke eine sinusförmige Wechselspannung mit 420 Hz oder 10 kHz. Dann kann zur Messung der Brückendiagonalspannung ein selektiver Wechselspannungsverstärker mit phasenempfindlichem Gleichrichter eingesetzt werden, so dass Messfehler durch Störeinflüsse verringert werden.

81.2.2 Füllstandsmessung und Messung der Foliendicke

Das Prinzip beruht darauf, dass das Dielektrikum Luft eines Plattenkondensators durch einen Stoff mit $\varepsilon_r > 1$ verdrängt wird. Damit ändert sich die Kapazität der Gesamtanordnung. Abb. 81.12 zeigt zwei Modelle. Das erste dient zur Messung der Füllstandshöhe, das zweite zur Messung der Foliendicke von Kunststofffolien.

In der Praxis hat sich bewährt, die *relative* Änderung der Kapazität $\Delta C/C_0$ in Abhängigkeit von der gesuchten Füllstandshöhe h oder Foliendicke d anzugeben. Dabei ist C_0 die Kapazität für $h = 0$ oder $d = 0$. Die angegebenen Formeln sind entsprechend dargestellt.

1. *Füllstandsmessung*: Es wird ein rechteckförmiger Behälter verwendet, bei dem zwei gegenüberliegende Seiten mit Folie beklebt oder bedampft sind und bei einem elektrisch leitenden Schüttgut oder einer elektrisch leitenden Flüssigkeit mit einer Kunststoff-Isolierschicht versehen sind. Von Vorteil



$$C_0 = C(h=0) = \frac{\varepsilon_0 \cdot H \cdot l}{b}$$

$$\Delta C = C(h>0) - C_0$$

$$= \frac{\varepsilon_0 \varepsilon_r \cdot h \cdot l}{b} + \frac{\varepsilon_0 (H-h) \cdot l}{b} - \frac{\varepsilon_0 \cdot H \cdot l}{b}$$

$$= \frac{\varepsilon_0 \cdot l}{b} \cdot h(\varepsilon_r - 1)$$

$$\frac{\Delta C}{C_0} = \frac{h}{H}(\varepsilon_r - 1)$$

$$C_0 = C(d=0) = \frac{\varepsilon_0 \cdot A}{D}$$

$$\Delta C = C(d>0) - C_0$$

$$= \frac{\varepsilon_0 \cdot A}{D} \cdot \frac{(\varepsilon_r - 1) \cdot d}{d(1-\varepsilon_r) + \varepsilon_r D}$$

($C(d>0)$: Reihenschaltung zweier Kondensatoren)

$$\frac{\Delta C}{C_0} = \frac{\varepsilon_r - 1}{d(1-\varepsilon_r) + \varepsilon_r D}$$

Abb. 81.12 Kapazitive Verfahren zur Messung von **a** Füllstand, **b** Foliendicke

ist, dass die Messeinrichtung nicht in den Füllraum hineinragt und damit nicht beschädigt werden kann und auch den Füllvorgang nicht behindert. Ist die Oberfläche durch eine Kunststoffschicht abgedeckt, können auch agressive Medien eingefüllt werden. Bei körnigem Schüttgut ist die obere Füllhöhe nicht eindeutig festgelegt, kann aber meist durch Erfahrungswerte oder durch Glätten der Oberfläche ausreichend genau erfasst werden.

2. *Foliendicke*: Man kann zeigen, dass die Kapazität der Anordnung *nicht* davon abhängt, *wo* sich die Folie innerhalb der Kondensatorplatten befindet. Die Dicke d der Folie wird über die Kapazität der Kondensatoranordnung gemessen und bei einer Abweichung vom Sollwert entsprechend eingestellt, wenn sie sich auf dem Weg von der Spritzdüse zu der Aufwickelrolle befindet. Dabei darf die Folie nicht berührt werden, da sie noch heiß ist und sich im Zustand der Vernetzung befindet.

81.2.3 Drehzahlmessung

Die Drehzahl wird in min^{-1} angegeben, die Drehfrequenz in s^{-1}. Zur Messung der Drehzahl kann im einfachsten Fall ein Impuls pro Umdrehung erzeugt und die Zahl der Impulse für genau eine Minute erfasst werden. Dieses Verfahren ist nur sinnvoll, wenn sich die Drehzahl innerhalb der Messzeit 1 min nicht ändert. Außerdem muss eine Minute gewartet werden, bis das Messergebnis zur Verfügung steht. In der Praxis werden deshalb pro Umdrehung mehrere Impulse erzeugt. Dann gilt folgender Zusammenhang zwischen der Zahl der Impulse und der Drehzahl:

$$\textit{Impulszahl} \quad N = n \cdot p \cdot t \tag{81.15}$$

N gezählte Impulse; n Drehzahl in min^{-1}; p Zahl der Impulse pro Umdrehung; t Messzeit in min.

Darin ist N die Zahl der in der Zeit t in einem elektronischen Zähler gezählten Impulse, n die gesuchte Drehzahl in min^{-1}, p die Zahl der Impulse pro Umdrehung und t die Zeit in min, während der die Impulse gezählt werden. Man wird zweckmäßigerweise $p \cdot t$ so wählen, dass sich als Einheit 1 min ergibt, denn dann entspricht die Zahl N der Drehzahl n. Setzt man voraus, dass der Fehler der Messeinrichtung vernachlässigbar klein ist, was in der Regel auch zutrifft, dann ergeben sich für $p > 1$ folgende Vorteile:

1. Die Drehzahl als Zahl N steht in weniger als 1 Minute zur Verfügung und ermöglicht das Erfassen von Drehzahl*änderungen*.
2. Bei unveränderter Messzeit steht die Drehzahl als Zahl N mit einer höheren Stellenzahl und damit höherer Genauigkeit zur Verfügung. Voraussetzung ist, dass sich die Drehzahl während der Messzeit nicht ändert.

Es stehen Messeinrichtungen mit einem maximalen p von etwa 5000 zur Verfügung (Abb. 81.13c).

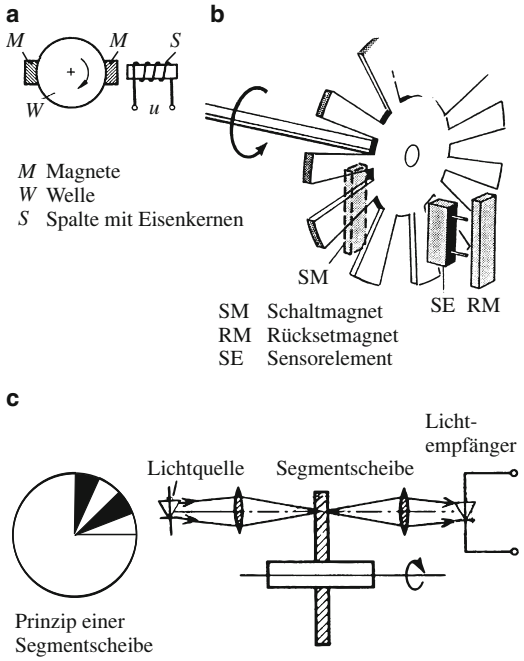


Abb. 81.13 Digitale Verfahren zur Drehzahlmessung. **a** Spannungsinduktion, **b** Fe-Blendenrad (nach Unterlagen Fa. VAC), **c** optische Verfahren

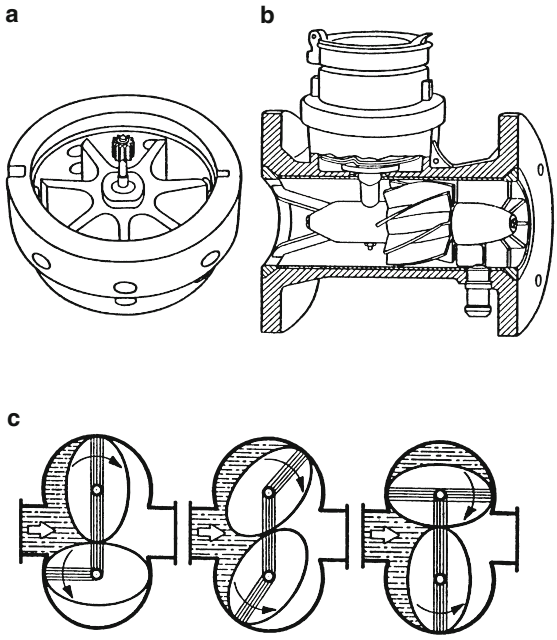


Abb. 81.14 Durchflussmessung (nach Unterlagen Fa. Siemens (**a**, **b**) und Fa. Orlicek (**c**)) **a** Flügelradzähler, **b** Woltmanzähler, **c** Ovalradzähler

Beispiel 6

Eine Drehzahl wird mit einem Geber mit $p = 120$ gemessen. Wählt man $p \cdot t = 1\,\text{min}$, dann verringert sich die Messzeit auf $(1/120)\,\text{min} = 0{,}5\,\text{s}$, und die Zahl N der Impulse entspricht der Drehzahl in min^{-1}. Auf diese Weise ist auch das Erfassen der Drehzahlerhöhung von Motoren vom Einschaltmoment bis zum Erreichen der Nenndrehzahl möglich. Beträgt die Messzeit andererseits 1 min, erhält man bei einer (konstanten) Drehzahl von $3000\,\text{min}^{-1}$ eine Anzeige $N = 3000 \cdot 120 = 360.000$. Hier wird man zweckmäßigerweise $p = 100$ wählen und erhält dann $N = 300.000$. Bei entsprechender Kommasetzung ergibt sich die Drehzahl $n = 3000{,}00\,\text{min}^{-1}$. ◀

In Abb. 81.13 sind einige digitale Verfahren zur Drehzahlmessung dargestellt. Abb. 81.13a zeigt eine Anwendung des Induktionsgesetzes. Nähert sich der Magnet der Spule, nimmt der magnetische Fluss zu, beim Entfernen ab. Durch diese Flussänderungen wird eine Spannung in der Spule induziert, die aber mit sinkender Drehzahl kleiner wird. Bei diesem Verfahren hat der Einsatz mehrerer Magnete auf der Welle noch den zusätzlichen Vorteil, dass die durch *einen* Magneten entstehende Unwucht bei mehreren Magneten verringert werden kann.

Eine Weiterentwicklung zeigt Abb. 81.13b (nach Unterlagen der Fa. Vakuumschmelze). Durch den Einsatz magnetfeldabhängiger Schalter hängt die Signalhöhe nicht mehr von der Drehzahl ab. Das entsprechend geformte Rad muss an oder auf der Welle angebracht werden.

Der Vorteil der Verfahren nach Abb. 81.13a und b liegt darin, dass sie sehr betriebssicher sind. Das gilt auch für das im Abb. 81.13c dargestellte optische Verfahren. Als Lichtsender werden Leuchtdioden (LED) verwendet, und die Lichtempfänger sind Fotodioden. Auf der Welle werden industriell gefertigte Kunststoffscheiben mit Kreissegmenten angebracht, bei denen sich lichtdurchlässige und lichtundurchlässige Segmente abwechseln. Die Anzahl der auf einer Scheibe angeordneten Segmente kann bis zu maximal 10.000 (5000 lichtdurchlässige und 5000 lichtundurchlässige) betragen.

Die analogen Verfahren zur Drehzahlmessung, die u. a. nach dem Prinzip des Wechselstromgenerators aufgebaut sind („Dynamoprinzip"), verlieren an Bedeutung, weil die Auswertung bzw. Anzeige auf digitalem Wege geschieht und der mechanische Verschleiß der Digitalverfahren geringer ist.

81.2.4 Durchflussmessung

Bei den in Abb. 81.14 gezeigten Durchflussmessern ist der Momentanwert der Drehzahl der Flügel- oder Turbinenräder dem Momentanwert der Menge des durchfließenden Mediums und damit seiner Geschwindigkeit proportional. Zur Bestimmung des gesamten in einer gegebenen Zeit durchgeflossenen Mediums müssen die Momentanwerte der Drehzahl addiert werden, d. h., es muss die Gesamtzahl der Umdrehungen erfasst werden. Deshalb ist die Durchflussmessung verbunden mit einer Drehzahlmessung, wobei die erforderliche Summenbildung häufig über ein mechanisches Zählwerk geschieht („Wasserzähler" im Haushalt). Beim Flügelradzähler werden die Flügel tangential angeströmt, beim *Woltmann*-Zähler axial.

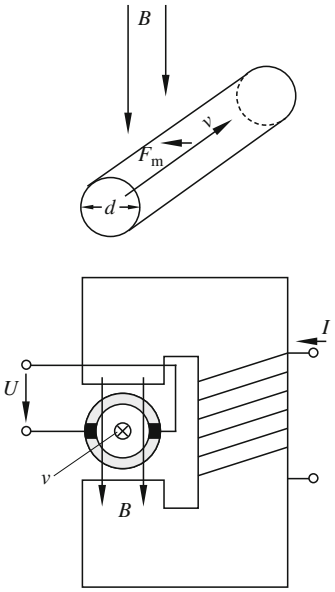


Abb. 81.15 Durchflussmessung nach dem Prinzip der magnetischen Induktion

Bei den Durchflussmessern nach dem Verdrängungsprinzip, wie sie bei den Gasmengenzählern im Haushalt angewendet werden, rollen zwei Ovalräder aufeinander und zur Messkammer hin ab und schließen ein bestimmtes Volumen des Mediums ein (Ovalradzähler Abb. 81.14c). Wird die Zahl der Umdrehungen erfasst, ist auch das insgesamt durchgeströmte Volumen des Mediums bekannt. Die Drehung der Ovalräder wird durch den Druck des anströmenden Mediums auf das untere Rad im Abb. 81.14c links bewirkt.

Beim Prinzip der *magnetischen Induktion* nutzt man die Kraftwirkung auf bewegte elektrische Ladungen Q im Magnetfeld aus (Abb. 81.15). Sie ist durch $F_m = Q \cdot v \cdot B$ gegeben, wobei B und v senkrecht aufeinander stehen. Dann ist F_m wiederum senkrecht zu B und v. Die magnetische Flussdichte B wird durch einen von außen angebrachten Magneten erzeugt. Die erforderlichen elektrischen Ladungen Q stehen bei Flüssigkeiten durch die meist vorhandene Dissoziation in Form von Ionen zur Verfügung. In die Wand des Rohres werden an zwei gegenüberliegenden Stellen elektrische Kontakte in das elektrisch nichtleitende Rohr eingebracht, an denen die Spannung abgegriffen wird. Die elektrische Kraft ist durch $F_e = Q{\cdot}E = Q{\cdot}U/d$ gegeben. Es gilt $F_m = F_e$ und damit $v = U/(B \cdot d)$. Der Durchfluss, also das pro Zeiteinheit durchströmende Volumen V' in m^3/s errechnet sich bei gegebener Strömungsgeschwindigkeit n und der Fläche A aus $V = v{\cdot}A = v{\cdot}d^2 \cdot \pi/4$. Dann folgt für das pro Zeiteinheit durchströmende Volumen V':

$$V' = (d^2 \cdot \pi/4) \cdot (U/(B \cdot d)) = (d \cdot \pi/4) \cdot (U/B) \qquad (81.16)$$

d Rohrdurchmesser in m, U gemessene Spannung an den Elektroden in V, B magnetische Flussdichte in $V\,s/m^2$.

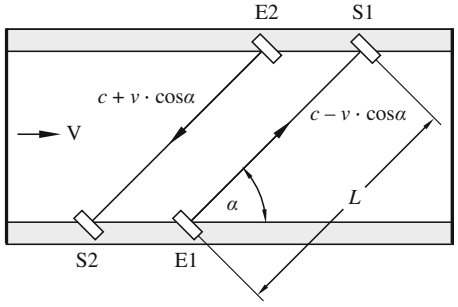


Abb. 81.16 Ultraschall-Durchflussmessung

Beispiel 7

Ein Dauermagnet erzeugt eine Flussdichte von $B = 0{,}1\,V\,s/m^2$; der Rohrdurchmesser beträgt $d = 5\,cm$. Bei einer Strömungsgeschwindigkeit des Mediums von $v = 0{,}1\,m/s$ ergibt sich eine Spannung von $U = 10^{-1}\,(Vs/m^2) \cdot 5 \cdot 10^{-2}\,m \cdot 10^{-1}\,(m/s) = 5 \cdot 10^{-4}\,V = 500\,\mu V$. Für V' folgt daraus $V' = (5 \cdot 10^{-2}\,m \cdot \pi/4) \cdot (5 \cdot 10^{-4}\,V/(10^{-1}\,V\,s/m^2)) = 1{,}96 \cdot 10^{-4}\,m^3/s$. ◀

Die entstehenden Spannungen liegen im µV- oder mV-Bereich. Der Innenwiderstand dieser Spannungsquelle ist sehr groß (Megaohm-Bereich). Deshalb sind geeignete Maßnahmen zur Störunterdrückung erforderlich.

Bei der *Ultraschall-Durchflussmessung* nach Abb. 81.16 erzeugen zwei Ultraschallsender (S1, S2) kurze Ultraschallimpulse, die auf die zugeordneten Empfänger (E1, E2) gerichtet sind und von diesen empfangen werden. c ist die Geschwindigkeit, mit der sich der Ultraschall in dem im Rohr befindlichen Medium ausbreitet, wenn dieses ruht. Strömt das Medium mit der Geschwindigkeit v durch das Rohr, breiten sich die Schallwellen vom Sender S1 langsamer als mit c aus. Der genaue Wert ergibt sich aus $c_1 = c - v \cdot \cos\alpha$. Entsprechend folgt für die Geschwindigkeit der von S2 ausgesendeten Schallimpulse $c_2 = c + v \cdot \cos\alpha$. Mit der Entfernung L der zwei Sender und Empfänger voneinander folgt für die Laufzeiten

$$t_1 = \frac{L}{c - v \cdot \cos\alpha}; \quad t_2 = \frac{L}{c + v \cdot \cos\alpha} \qquad (81.17)$$

Die Laufzeitdifferenz berechnet sich zu

$$t_1 - t_2 = 2 \cdot L \frac{v \cdot \cos\alpha}{c^2 - v^2 \cdot \cos^2\alpha} \qquad (81.18)$$

In Flüssigkeiten kann der Term $v^2 \cdot \cos^2\alpha$ in (81.18) vernachlässigt werden, so dass sich als Näherungslösung ergibt

$$t_1 - t_2 \approx 2 \cdot L \frac{v \cdot \cos\alpha}{c^2} \quad \Rightarrow \quad v \approx \frac{c^2(t_1 - t_2)}{2 \cdot L \cdot \cos\alpha} \qquad (81.19)$$

Teil X

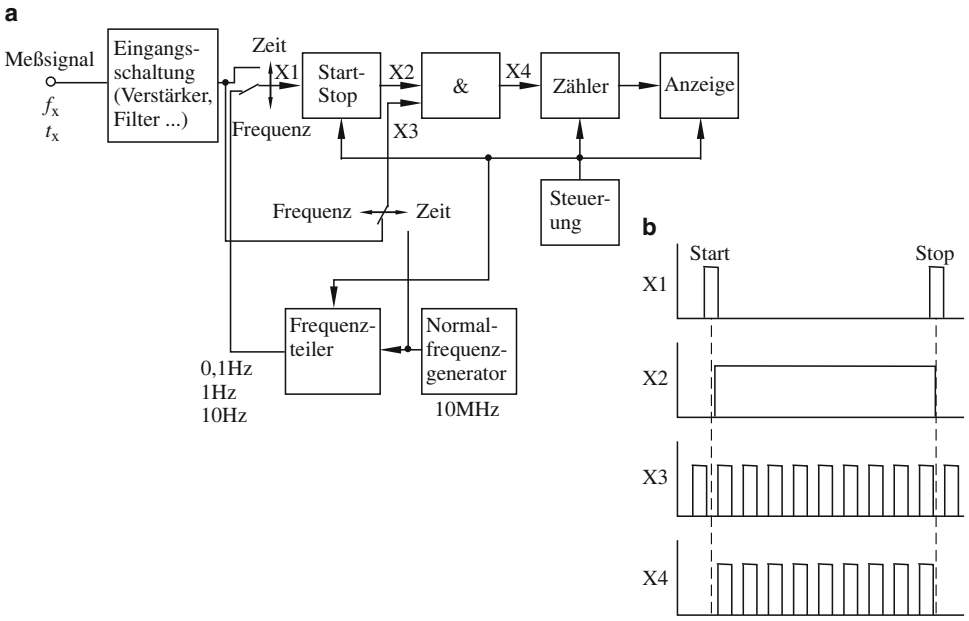


Abb. 81.17 Zeit- und Frequenzmessung. **a** Prinzip, **b** Ablaufdiagramm

Das Messergebnis hängt allerdings noch von der Schallausbreitungsgeschwindigkeit c ab. Um diese Abhängigkeit auszuschalten, hat man z. B. die Ansteuerung der Ultraschallsender so gewählt, dass ein Schallimpuls, von S1 ausgesendet und von E1 empfangen, sofort wieder von S2 ausgesendet wird, von E2 empfangen und unverzögert von S1 ausgegeben wird. Für die Frequenzen f_1 und f_2 der Sender S1 und S2 folgt

$$f_1 = \frac{1}{t_1} = \frac{c - v \cdot \cos\alpha}{L};$$
$$f_2 = \frac{1}{t_2} = \frac{c + v \cdot \cos\alpha}{L}; \qquad (81.20)$$

Aus der Differenz dieser zwei Frequenzen

$$f_2 - f_1 = \frac{2 \cdot v \cdot \cos\alpha}{L} \qquad (81.21)$$

ergibt sich für die Strömungsgeschwindigkeit v unabhängig von der Schallgeschwindigkeit c

$$v = \frac{L}{2 \cdot \cos\alpha}(f_2 - f_1) \qquad (81.22)$$

81.2.5 Zeit- und Frequenzmessung

Die Messung der Zeit und der Frequenz geschieht auf digitalem Wege, wobei die Messeinrichtung die Messung von Zeit und Frequenz ermöglicht. Das Prinzip ist in Abb. 81.17 dargestellt.

Ein Normalfrequenzgenerator stellt die Zeiteinheit in Form einer Sekunde oder dekadischer Teile oder Vielfache davon dar. Das periodische Signal mit der zu messenden Frequenz f_x wird zunächst einer Eingangsschaltung zugeführt, die den Spannungswert an die Erfordernisse der nachfolgenden Torschaltung anpasst. Sie enthält einen Abschwächer für große Signalamplituden, einen Verstärker für kleine Signalamplituden und eine Signalformerstufe. Letztere wandelt das Eingangssignal in ein rechteckförmiges Signal mit gleicher Frequenz um. Weiterhin können frequenzselektive Filter zugeschaltet werden, die unerwünschte Störfrequenzanteile ausfiltern. Außerdem kann häufig ein Gleichanteil im Messsignal durch einen Kondensator unterdrückt werden (entsprechend den Schalterstellungen AC oder DC beim Oszilloskop). Die Torschaltung besteht aus einem UND-Glied und lässt das Signal mit der Frequenz f_x für die von der Zeiteinheit vorgegebene Zeit passieren. Die Impulse werden vom Zähler gezählt. Nach Schließen des Tores wird der Zählerstand in die Anzeige übernommen und sichtbar gemacht. Wählt man als Torzeit nicht eine Sekunde, sondern dekadische Teile oder Vielfache davon, muss in der Anzeige lediglich das Komma entsprechend gesetzt werden. Je kürzer die Torzeit gewählt wird, desto schneller steht das Messergebnis zur Verfügung, allerdings mit verminderter Stellenzahl, da in der vorgegebenen Zeit weniger Zählimpulse in den Zähler gelangen. Für genaue Messungen wird man deshalb Torzeiten von z. B. 10 s wählen.

Die Steuerung hat innerhalb eines Messzyklus hauptsächlich folgende Aufgaben: Setzen des Zählerstandes auf Null; Öffnen der Torschaltung für die eingestellte Zeit (...; 0,1 s; 1 s; 10 s; ...); Speichern des aktuellen Zählerstandes für die Anzeige in einem Anzeigespeicher; Ansteuerung der Anzeigen im meist verwendeten Multiplexverfahren; Ansteuerung eines weiteren Zählers zur Bereitstellung der Torzeit.

Die Torzeit wird dadurch gewonnen, dass ein Quarzoszillator auf einer relativ hohen Frequenz von 1 MHz oder 10 MHz schwingt. Diese Oszillatoren sind sehr frequenzstabil, wenn man die beiden wesentlichen Einflussgrößen „Temperatur“ und

„Alterung“ so weit wie möglich ausschaltet. Die Temperaturabhängigkeit der Frequenz von $\leq |\pm 10^{-5}\,\mathrm{K}^{-1}|$ kann durch Wahl einer optimalen Schnittebene für den Quarz und durch Einbau in einen Thermostaten weiter verringert werden. Zur Herabsetzung der Frequenzänderung infolge Alterung von $\leq |\pm 10^{-5}\,\mathrm{a}^{-1}|$ werden die Quarze dadurch künstlich gealtert, dass man sie etlichen Temperaturzyklen aussetzt. Zahl, Temperaturendwerte und Dauer dieser Zyklen sind Erfahrungswerte und werden von den Firmen i. Allg. nicht veröffentlicht. Ein Zyklus kann so aussehen: Durchfahren des Temperaturbereiches von 20 °C bis 80 °C und zurück mit einer Dauer von 2 bis 4 Stunden, und das insgesamt 20mal.

Die Frequenz des Quarzoszillators wird über einen Zähler aus mehreren Zähldekaden geteilt, wobei z. B. eine Frequenz von 1 Hz einer Periodendauer von 1 s entspricht und diese dann zur Öffnung der Torschaltung verwendet wird. Andere Toröffnungszeiten werden durch Zu- oder Abschalten einzelner Zähldekaden erreicht.

Die Messung von Zeiten oder Periodendauern geschieht dadurch, dass das Eingangssignal mit der zu messenden Periodendauer t_x mit der ersten positiven Flanke die Torschaltung öffnet und mit der folgenden wieder schließt. Während das Tor geöffnet ist, zählt der Zähler die Impulse des Quarzoszillators (1 oder 10 MHz). Damit ist sein Zählerstand der Periodendauer t_x proportional. Es ist deshalb wichtig, dass die Frequenz des Quarzoszillators möglichst hoch ist, damit der Zähler auch bei kleinen Periodendauern im ms-Bereich noch ausreichend Impulse zählen kann und so eine möglichst hohe Auflösung des Messwertes möglich ist.

Frequenzen > 10 kHz und Periodendauern > 1 ms sind mit einem Zähler mit einer Quarzfrequenz von 10 MHz und einer Torzeit von 1 s mit 4 oder mehr Stellen messbar. Hat das zu messende Signal dagegen Frequenzen < 10 kHz oder Periodendauern < 1 ms, ergibt die Anzeige weniger als 4 Stellen. Dann sollte man zur Frequenzmessung auf die Periodendauermessung und zur Periodendauermessung auf die Frequenzmessung übergehen. Die gewünschte Größe ergibt sich dann über eine Rechnung aus der Beziehung $f = 1/T$. Komfortable Zähler wenden dieses Verfahren bereits automatisch an.

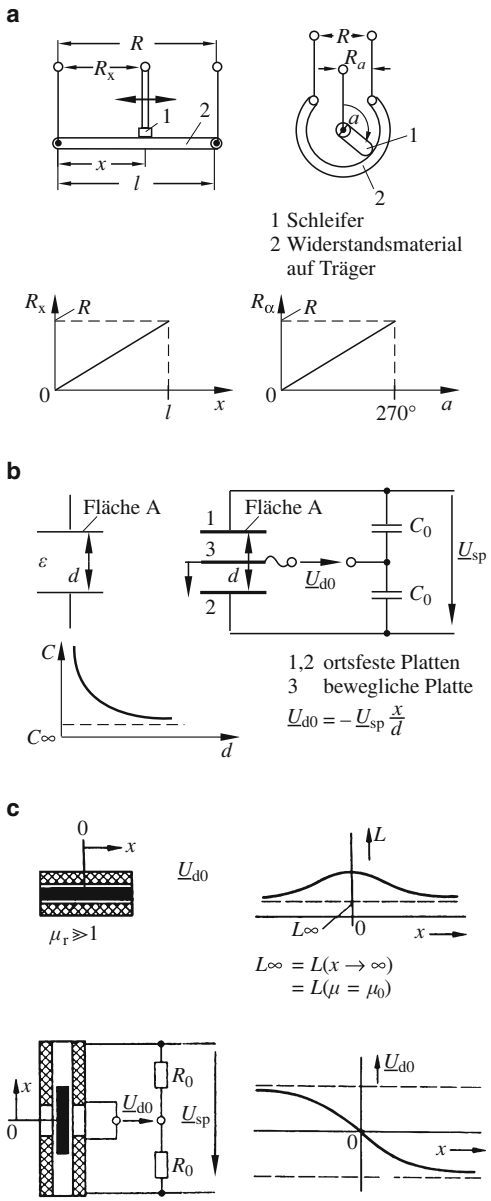


Abb. 81.18 Weg- und Winkelmessung: Prinzip und Kennlinie von **a** ohmschen, **b** kapazitiven, **c** induktiven Aufnehmern

81.2.6 Weg- und Winkelmessung

Mit den im Abschn. 81.1 vorgestellten ohmschen, kapazitiven und induktiven Aufnehmern können auch Wege (Abstände) und Drehwinkel gemessen werden.

In Abb. 81.18 ist das jeweilige Messprinzip gezeigt. Wenn möglich, werden Differentialausführungen verwendet, da die Kennlinie einer angestrebten linearen Abhängigkeit nahe kommt.

Für die resistiven Wegaufnehmer stehen Ausführungen von einigen mm bis zu einem Meter zur Verfügung. Die Winkelaufnehmer werden auch für Winkel bis zu 360° gefertigt. Bei den induktiven Wegaufnehmern lässt sich erkennen, dass die Kennlinie der Differential-Ausführung bereichsweise nahezu linear ist und außerdem positive und negative Wege vom Bezugspunkt aus messbar sind. Bei den kapazitiven Aufnehmern wird die Kapazität sowohl durch eine wegproportionale Abstandsänderung als auch durch eine wegproportionale Flächenänderung variiert. Verwendet man einen Differentialkondensator und misst die Kapazitätsänderung in einer Wechselstrom-Brückenschaltung, erhält man einen linearen Zusammenhang zwischen Weg- und Kapazitätsänderung. Auch die induktiven Differentialausführungen werden zum Messen in Brückenschaltungen eingefügt.

Für die Wechselstrom-Messbrücken werden Frequenzen im kHz-Bereich verwendet, um den Einfluss von Oberschwingungen der 50-Hz-Energieversorgung klein zu halten. Außerdem lassen sich Filterschaltungen zur Messung der Spannung in der Brückendiagonalen mit kleineren Bauelementewerten realisieren.

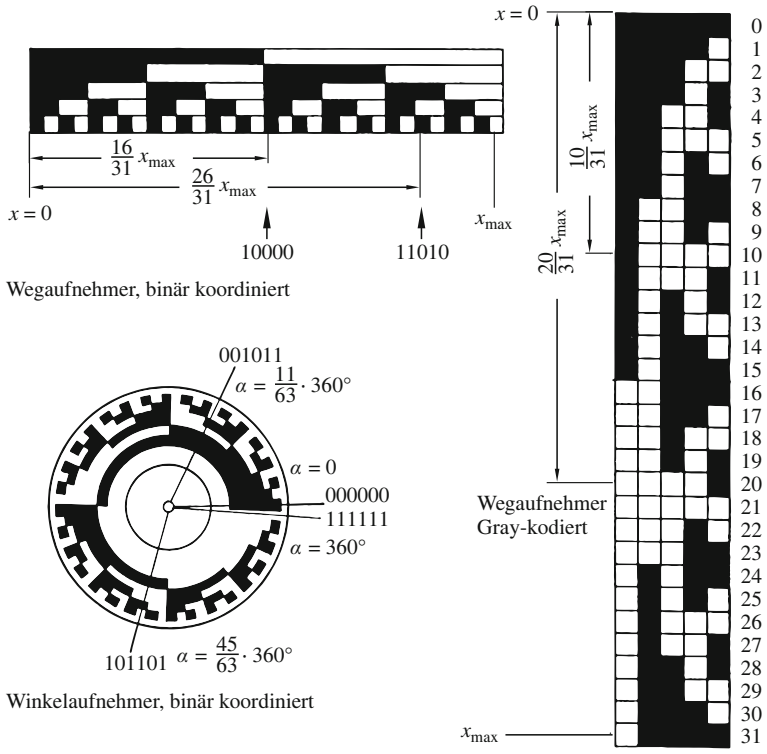


Abb. 81.19 Kodierte Weg- und Winkelaufnehmer

Bei den *kodierten Weg- und Winkelaufnehmern* ist jeder Weglänge oder jedem Drehwinkel ein eindeutig kodiertes digitales Signal zugeordnet (Abb. 81.19).

Diese Signale haben den Vorteil, dass sie auch in einer gestörten Umgebung eine weitgehend sichere Messsignalübertragung gestatten. Außerdem steht das Signal in einer für die Weiterverarbeitung in einem Rechner geeigneten Form zur Verfügung. Auch der Messaufbau ist relativ störunanfällig. Für die Kodierung der Scheiben wird häufig nicht der Binärcode verwendet, der eine eindeutig zugeordnete Stellenwertigkeit aufweist, sondern der Gray-Code, bei dem sich zwei benachbarte Codewörter nur in *einer* Stelle unterscheiden. Da die einzelnen Kodierungen bei Weg- oder Winkeländerungen nacheinander überstrichen werden, sind Übertragungsfehler einfach erkennbar. Allerdings benötigt das Auswertegerät einen Codewandler.

Inkrementale Wegaufnehmer verwenden nach Abb. 81.20a eine Leiste aus Glas oder Kunststoff (1), auf der in Querrichtung abwechselnd lichtdurchlässige und lichtundurchlässige Segmente aufgebracht wurden. Dieser Streifen wird an dem Gegenstand befestigt, dessen Weg gemessen werden soll. Das von einer Lichtquelle ausgesendete Licht wird bei Bewegung des Gegenstandes um eine Rasterlänge Δr durch die ortsfeste Leiste (2) einmal unterbrochen. Zählt man die Zahl der Hell-Dunkel-Wechsel oder der Dunkel-Hell-Wechsel und erhält den Wert n, so ist der zurückgelegte Weg $n \cdot \Delta r$.

Soll auch die *Richtung* des Weges von einem Bezugspunkt aus erfasst werden, bringt man nach Abb. 81.20b auf dem ortsfesten Streifen (2) zwei Hell-Dunkel-Segmente an, die einen Abstand von $\Delta r + \Delta r/4$ zueinander haben. Trifft das Maximum von u_1 mit einer ansteigenden Spannung u_2 zusammen, bewegt sich der Gegenstand nach rechts, bei einer Linksbewegung trifft das Maximum von u_1 mit einer abfallenden Spannung u_2 zusammen.

Zur Festlegung des Bezugswertes für den Weg ($x = 0$) muss der Gegenstand in eine vorgesehene Position gebracht und der Zähler auf Null gesetzt werden. Auf die Streifen lassen sich Raster mit einem Abstand von minimal einigen mm aufbringen. Damit sind auch Wegmessungen bis zu dieser Größenordnung herab möglich.

81.2.7 Beschleunigungsmessung

Die Messung nutzt den Zusammenhang zwischen der Kraft F und der Beschleunigung a ($F = a \cdot m$) aus. In Abb. 81.21a ist ein gedämpftes Feder-Masse-System dargestellt. Wird das Gehäuse beschleunigt, ändert die Masse wegen ihrer Trägheit ihre Lage innerhalb des Gehäuses. Die zugeordnete Differentialgleichung lautet:

$$m \cdot \frac{d^2 y_M}{dt^2} + w \cdot \frac{dy_M}{dt} + c \cdot y_M = -m \frac{d^2 y_G}{dt^2} = -m \cdot a \quad (81.23)$$

Formelzeichen siehe Text.

Darin ist m die Masse des an der Feder aufgehängten Körpers, c die Federsteifigkeit (Federrate), w ein Maß für die Dämpfung und a die gesuchte Beschleunigung. Zu beachten ist, dass Gehäuse und Masse nicht starr miteinander verbunden sind und deshalb y_G- und y_M-Koordinaten eingesetzt worden sind.

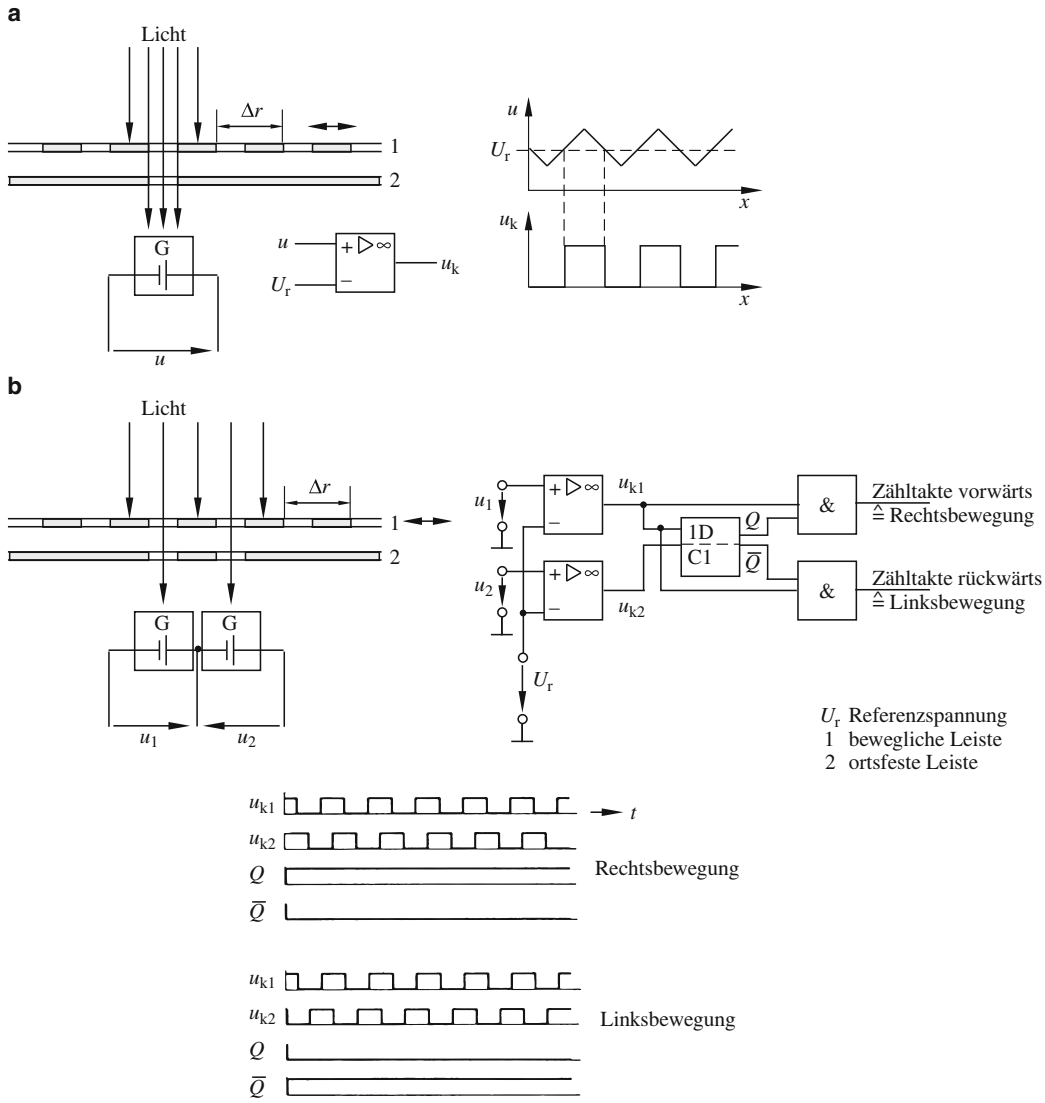


Abb. 81.20 Inkrementale Wegmessung. **a** Prinzip, **b** Erfassung der Richtung

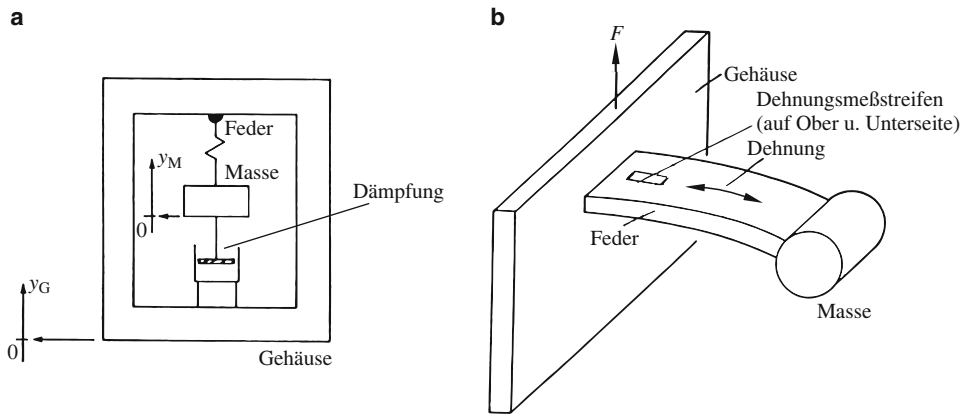


Abb. 81.21 Beschleunigungsaufnehmer. **a** Gedämpftes Feder-Masse-System, **b** Biegebalken mit Dehnungsmessstreifen

Je nach Dimensionierung erhält das System unterschiedliche Eigenschaften.

1. *Steife Feder und kleine Masse (c groß und m klein)*:
   Dann können die zwei ersten Summanden der linken Seite aus (81.23) vernachlässigt werden, und die Differentialgleichung geht näherungsweise über in

$$-a \approx c \cdot y_M/m \tag{81.24}$$

   Diese Dimensionierung ergibt einen Beschleunigungsaufnehmer, da die Auslenkung y_M der Beschleunigung $a$näherungsweise proportional ist. Die Eigenfrequenz $\omega_0^2 = c/m$ dieses schwingungsfähigen Systems ist sehr groß.
2. *Kleine Masse, weiche Feder und große Dämpfung (m und c klein, w groß)*:
   Dann folgt in erster Näherung

$$\frac{w}{m} \cdot \frac{dy_M}{dt} \approx -\frac{d^2y_G}{dt^2} \tag{81.25}$$

   Integriert man beide Seiten, ergibt sich

$$\frac{w}{m} \cdot y_M \approx -\frac{dy_G}{dt} = -v_G \tag{81.26}$$

   Damit hat man einen Geschwindigkeitssensor für die Geschwindigkeit v_G des Gehäuses realisiert.
3. *Große Masse, weiche Feder und geringe Dämpfung (m groß, c und w klein)*:
   In (81.23) bleibt nur noch zu berücksichtigen

$$\frac{d^2y_M}{dt^2} \approx -\frac{d^2y_G}{dt^2} = -a \tag{81.27}$$

   Nach zweimaliger Integration folgt

$$y_M \approx y_G \tag{81.28}$$

   Es handelt sich um einen Wegaufnehmer, der in dieser Form z. B. zur Messung von Erdbeben und zur Überwachung von Schwingtischen eingesetzt wird.

Eine weitere Ausführungsform ist in Abb. 81.21b dargestellt. Gezeigt wird nur der eigentliche Aufnehmer ohne Gehäuse. Die Dimensionierung wird so vorgenommen, dass sich die Abhängigkeit aus (81.24) ergibt. Der Biegebalken wird mit Dehnungsmessstreifen auf Ober- und Unterseite beklebt. Die Widerstandsänderungen sind der Dehnung und diese wiederum der Kraft proportional. Die Kraft ist schließlich über die Masse mit der Beschleunigung verknüpft. Die Widerstände der Dehnungsmessstreifen werden in einer Brückenschaltung gemessen.

Messdatenaufbereitung

82

Wilfried Plaßmann



82.1 Verringerung der Störeinflüsse von außen 944
82.2 Messverstärker 945

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_82

Die von den Messaufnehmern gelieferten elektrischen Größen sind häufig Spannungen im µV- oder mV-Bereich. Zur Weiterverarbeitung in Rechnern müssen sie zunächst auf Werte im V-Bereich verstärkt werden, weil die Analog-Digital-Umsetzer entsprechende Größen benötigen. Außerdem wird eine eventuell erforderliche Übertragung der Messgröße in analoger Form über größere Entfernungen mit geringem Fehler möglich. Zu diesem Zweck muss der Einfluss der stets vorhandenen Störungen auf das Messsignal so klein wie möglich gehalten werden. Theoretische Untersuchungen und praktische Erfahrungen haben zu den im Folgenden dargestellten Schaltungen geführt, die für eine Vielzahl von Messaufgaben anwendbar sind.

82.1 Verringerung der Störeinflüsse von außen

Unter Störungen sollen hier unerwünschte Signale verstanden werden, die von außen in den Messkreis gelangen. Das kann nach Abb. 82.1 induktiv oder kapazitiv geschehen. Störquellen sind beispielsweise das 50-Hz-Wechselstromnetz, Gleichstromsteller für Motoren, Dimmer, Leuchtstofflampen, nahe gelegene Hochfrequenzsender oder Schaltvorgänge an Verbrauchern.

Induktive Einwirkung: Die Störspannung $u_{\text{Stör}}$ hat im Leiter 1 einen Strom $i_{\text{Stör}}$ zur Folge, dessen magnetischer Fluss $\Phi_{\text{Stör}}$ in der Zuleitung zum Messverstärker (Leiter 2) eine Störspannung induziert, die sich dem Messsignal überlagert. Um die induzierte Störspannung klein zu halten, sollte der Abstand d möglichst groß und die vom Leiter 2 eingeschlossene Fläche A möglichst klein gemacht werden. Letzteres erreicht man z. B. durch Verdrillen der Hin- und Rückleitung von Leiter 2. Allerdings hat das den Nachteil, dass die Kapazität zwischen diesen Leitern größer wird und die im Messsignal enthaltenen hohen Frequenzanteile schwächt.

Kapazitive Einwirkung: Zwischen den zwei Leitern 1 und 2 (Abb. 82.1b) besteht die durch den Aufbau bedingte Koppelkapazität C_K, die mit der Parallelschaltung aus der Kapazität des Leiters 2, dem Eingangswiderstand R_e des Messverstärkers und dem Innenwiderstand R_i der Signalquelle u_{Mess} einen Spannungsteiler bildet. Der Einfluss von C_k ist umso kleiner, je größer d ($C_k \sim 1/d$) und C_L und je kleiner R_e und R_i sind.

Die genannten Maßnahmen zur Störspannungsverringerung sind nur in begrenztem Maße anwendbar, sodass man grundsätzlich die Zuleitung vom Messwertaufnehmer zum Verstärker abschirmt (Abb. 82.2a). Dann wird bei induktiver Einwirkung eine Spannung in der Abschirmung induziert, die aber nicht mehr in den Messkreis gelangt. Bei kapazitiver Einwirkung entsteht die Koppelkapazität zur Abschirmung und damit gegen den gemeinsamen Bezugspunkt („Masse“). Allerdings bildet nun die Abschirmung mit dem Innenleiter eine Kapazität C_{LS}, so dass eine frequenzabhängige Spannungsteilung zwischen R_i und C_{LS} stattfindet. Um die Kapazität C_{LS} unwirksam zu machen, wird das Potential der Abschirmung auf dem gleichen Wert gehalten wie das des Innenleiters, und zwar über eine nachgeführte Spannungsquelle (Impedanzwandler mit Operationsverstärker) mit dem Innenwiderstand ≈ 0 (Abb. 82.2b).

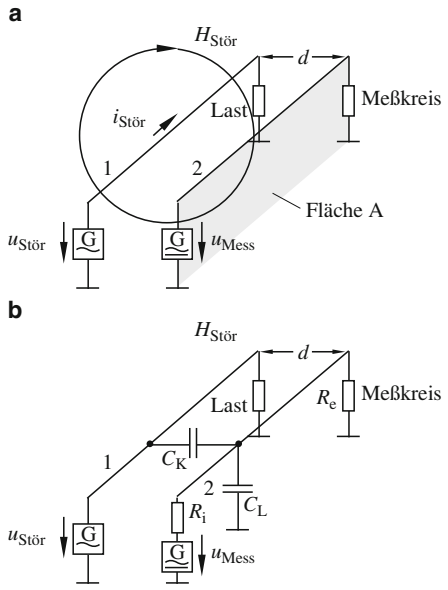


Abb. 82.1 Einwirken von Störungen in einem Messkreis. **a** induktiv, **b** kapazitiv

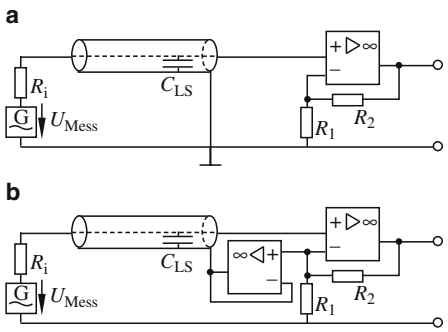


Abb. 82.2 Abschirmung von Signalleitungen. **a** Grundschaltung, **b** Abschirmung mit nachgeführtem Potential

Dann ist die Spannung an C_{LS} ebenfalls ≈ 0 und dieser Kondensator ohne Wirkung.

Eine besondere Bedeutung kommt dem Anschluss mit den gebräuchlichen Bezeichnungen „Erde, Masse, Systemerde oder Gehäuse“ zu. An diesen Anschluss, der in der Regel mit dem Gehäuse verbunden ist und der außerdem mit „Erde“ (meist über den Schutzleiter) in Verbindung steht, wird die Abschirmung der Leitungen gelegt. Dabei werden die Abschirmungen nur an einer Seite, zweckmäßigerweise an der Verstärkereingangsseite, mit diesem Punkt verbunden.

Abb. 82.3a zeigt zunächst den Fall, dass die Abschirmung an beiden Enden mit Masse verbunden ist. Da die beiden Masseanschlüsse in der Regel nicht auf gleichem Potential liegen, sondern einen Potentialunterschied Δu_p aufweisen, überlagert dieser sich dem Nutzsignal u_{Mess}. Wird dagegen die Abschirmung nur einseitig an Masse angeschlossen und der untere Anschluss der Messspannungsquelle getrennt von der Masse zum Eingang des Verstärkers geführt, ist der Potentialunterschied wirkungslos. Deshalb sind auch die Masseanschlüsse einzelner Baugruppen und weiterer Abschirmungen *sternförmig*

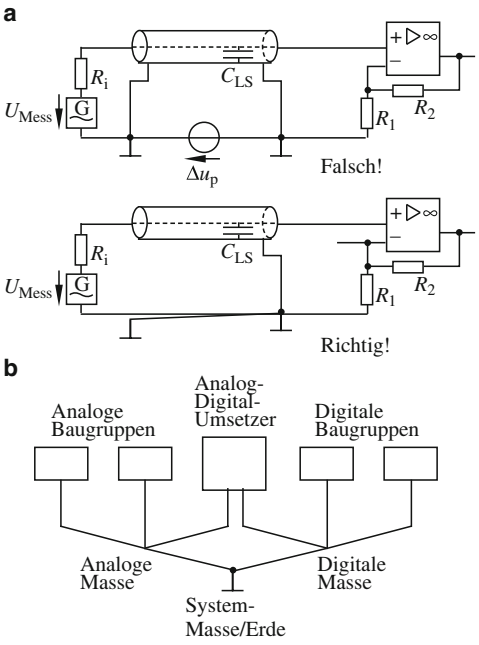


Abb. 82.3 Erdung von Messsystemen. **a** Wirkung einer Erdschleife, **b** sternförmige Erdung, Analog- und Digitalteil getrennt

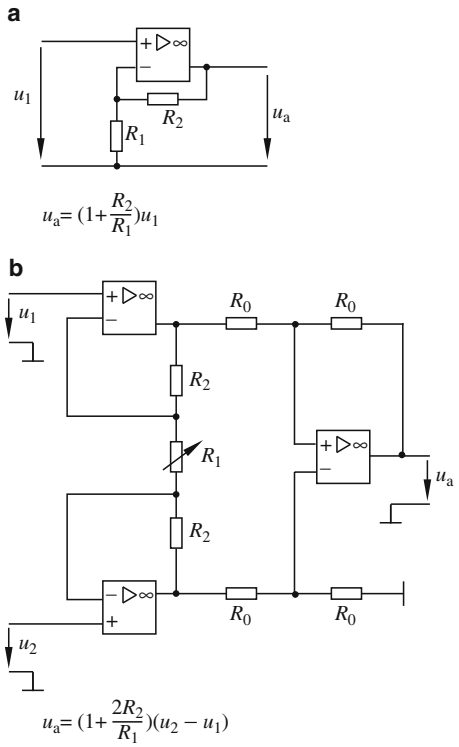


Abb. 82.4 Messverstärker mit Operationsverstärkern. **a** Nichtinvertierender Verstärker, **b** Messverstärker oder Instrumentation Amplifier

in einem Punkt zusammenzuführen (Abb. 82.3b). Bei Baugruppen, die sowohl einen Analog- als auch einen Digitalteil enthalten (A-D- oder D-A-Umsetzer), sind in der Regel die Masseanschlüsse für den Analog- und den Digitalteil getrennt herausgeführt, damit die relativ hohen Ströme im Digitalteil in der Masseleitung des Analogteiles keinen Spannungsfall hervorrufen. Er würde sich sonst dem Analogsignal überlagern.

82.2 Messverstärker

Als Messverstärker werden Operationsverstärker eingesetzt, die geeignet ausgesucht und beschaltet werden. Da die zu verstärkenden Spannungen häufig im mV- oder µV-Bereich liegen, dürfen die durch den Verstärker selbst hervorgerufenen „inneren“ Störungen nicht in die Größenordnung des Nutzsignals kommen, denn sie werden um den gleichen Faktor verstärkt wie das Nutzsignal. Damit bleibt das Verhältnis von Nutz- zu Störsignal am Ein- und Ausgang des Verstärkers gleich. Als wichtigste innere Störquellen sind zu nennen:

Rauschen des Verstärkers: Das Rauschen wird hervorgerufen durch thermische Vorgänge in den Widerständen und Halbleitern, die nach statistischen Gesetzen ablaufen. Es tritt prinzipiell bei Temperaturen oberhalb 0 K auf und lässt sich in zwei Anteile aufteilen: *Weißes Rauschen*, das für alle Frequenzen die gleiche Amplitude aufweist, und $1/f$*-Rauschen*, dessen Amplituden mit steigender Frequenz abnehmen. Durch die Wahl geeigneter technologischer Verfahren bei der Fertigung der Verstärker lässt sich das Rauschen in gewissen Grenzen verringern. Weiterhin kann man bei Verstärkern des gleichen Typs Unterschiede im Rauschen feststellen, so dass sich das Aussuchen rauscharmer Verstärker aus einer großen Stückzahl lohnen kann. Die am Eingang handelsüblicher Verstärker auftretende Rauschspannung ist von der Bandbreite B abhängig und liegt bei $B = 1\,\text{kHz}$ in der Größenordnung 100 nV bis 10 µV (siehe auch Abschn. 92.4.11.1).

Offsetspannung: Durch innere Unsymmetrien, die eine statistische Verteilung haben, ist die Ausgangsspannung nicht Null, wenn die Eingangsspannung Null ist. Dem Nutzsignal ist damit eine Fehlspannung (Gleichspannung) überlagert. Durch Anlegen einer entgegengesetzt gleich großen Spannung (Offsetspannung) an den Eingang des Verstärkers lässt sich diese Fehlspannung kompensieren. Sie ist eine verstärkerspezifische Kenngröße und kann wegen der statistischen Verteilung durch Aussuchen entsprechender Exemplare klein gehalten werden. Je nach Verstärkertyp und Fertigungsverfahren liegt die Offsetspannung im Bereich $\pm 5\,\text{mV} \ldots \pm 0{,}5\,\mu\text{V}$. Am Ausgang des Verstärkers tritt die Offsetspannung aber multipliziert mit der am Operationsverstärker eingestellten Verstärkung auf. Operationsverstärker mit Offsetspannungen im unteren µV-Bereich werden entweder bei der Fertigung durch Lasertrimmung von Widerstandsbahnen abgeglichen oder es werden selbstabgleichende Verstärker eingesetzt, die ihre aktuelle Offsetspannung messen und automatisch kompensieren (siehe Drift) oder sie werden gegen Aufpreis aus einer Vielzahl gefertigter Exemplare ausgesucht.

Drift: Kenngrößen ändern sich unter der Einwirkung von Einflussgrößen. Für die hier betrachteten Messverstärker ist besonders die Abhängigkeit der Offsetspannung von der Temperatur

und der Zeit wichtig. Es wurden deshalb für besonders kritische Anwendungen Baugruppen entwickelt, bei denen der eigentliche Verstärker seine Funktion periodisch für kurze Zeit unterbricht, den Momentanwert seiner Offsetspannung selbsttätig ermittelt und diese an seinen Eingang zur Kompensation anlegt.

Messverstärker mit Operationsverstärkern werden häufig in zwei Schaltungen nach Abb. 82.4 eingesetzt. Im Teil a) handelt es sich um einen nichtinvertierenden Verstärker, dessen Verstärkung v durch das Widerstandsverhältnis bestimmt ist ($v = 1 + R_2/R_1$). Der Eingangswiderstand der Schaltung ist gleich dem des Operationsverstärkers. Geeignet ist diese Schaltung für Messspannungen, deren einer Anschluss mit dem Erdanschluss des Operationsverstärkers verbunden werden kann.

Soll dagegen die Spannung in der Brückendiagonalen einer Rückenschaltung verstärkt werden und muss ein Anschluss der Brückenspeisespannung geerdet werden, ist eine Schaltung nach Abb. 82.4a nicht möglich, weil aus Gründen der Störspannungsverringerung auch der übliche Masseanschluss des Operationsverstärkers mit dem Erdanschluss verbunden werden sollte. Dann wird eine Kombination aus drei Operationsverstärkern nach Abb. 82.4b verwendet. Die Verstärkung ergibt sich zu:

$$u_a = (1 + 2R_2/R_1) \cdot (u_2 - u_1) \tag{82.1}$$

Für eine hohe Gleichtaktunterdrückung ist es wichtig, dass gleich bezeichnete Widerstände nach Abb. 82.4b in ihrem Wert um weniger als $\pm 1\,\%$ voneinander abweichen.

Bussysteme für die Messtechnik

83

Wilfried Plaßmann



83.1 Grundbegriffe 948
83.2 IEC-Bus 948
83.3 DIN-Messbus 949
83.4 Aktuator-Sensor-Interface (ASI) 949
83.5 CAN-BUS 952
83.6 USB-Übertragung 952
83.7 LAN (*L*ocal *A*rea *N*etwork) 952

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_83

Die hier betrachteten Bussysteme werden bevorzugt in der Messtechnik eingesetzt und dienen dazu, den Befehls- und Messdatenverkehr zwischen Messgeräten und der Erfassungs- und Verarbeitungsstation (Rechner) zu organisieren.

Tab. 83.1 Spannungspegel der seriellen Rechnerschnittstelle V.24

Daten		Steuersignale	
Low ≙ (−3 . . . −15) V	typ. −10 V	Low ≙ (+3 . . . +15) V	typ. +10 V
High ≙ (+3 . . . +15) V	typ. +10 V	High ≙ (−3 . . . −15) V	typ. −10 V

83.1 Grundbegriffe

Ergänzend zu den im Abschn. 107.1 genannten Begriffen werden einige weitere dargestellt.

Die Befehls- und Datenworte können seriell oder parallel und synchron oder asynchron übertragen werden. Die Umwandlung von parallel anstehenden Digitalworten in serielle und umgekehrt geschieht über Schieberegister.

Serielle Übertragung: Es genügt eine Zweidrahtleitung zur Übertragung, wenn der Takt aus dem Bitmuster zurückgewonnen werden kann, andernfalls überträgt man den Takt über eine dritte Leitung (Abschn. 96.4.4.7). Beginn und Ende des Wortes werden i. Allg. über ein Start- und ein Stopbit gekennzeichnet. Der Informationsaustausch lässt sich vereinfachen und beschleunigen, wenn noch einige zusätzliche Leitungen zur Verfügung stehen. So hat die serielle Rechnerschnittstelle V.24 (auch mit RS-232 bezeichnet) insgesamt 8 Leitungen. Die serielle Übertragung wird bevorzugt für die Überbrückung größerer Entfernungen eingesetzt.

Parallele Übertragung: Die in der Regel 8 Bit breiten Worte werden über 8 parallele Leiter übertragen. Damit wird die Übertragung beschleunigt, erfordert aber einen höheren Aufwand an Leitungen. Bei der parallelen „Centronics"-Rechnerschnittstelle (LPT) sind 8 Datenleitungen und 9 Steuerleitungen sowie 8 Masseleitungen vorhanden. Diese Schnittstelle wurde zunächst speziell für die Ansteuerung eines Druckers entwickelt und dann universell für Peripheriegeräte, wie Streamer, Scanner und externe Datensicherungsgeräte, eingesetzt. Sie wird nur noch in Ausnahmefällen, z. B. bei bestehenden Anlagen, eingesetzt.

Synchrone Übertragung: Der Datenfluss geschieht kontinuierlich, und die einzelnen Bits lassen sich anhand des parallel übertragenen Taktsignals eindeutig zuordnen. Synchronisationszeichen kennzeichnen Anfang und Ende der Übertragung. Ein Übertragungsprotokoll stellt die erfolgreiche Übertragung sicher.

Asynchrone Übertragung, Handshake-Verfahren: Im einfachsten Fall sind für diese Art der Übertragung zunächst zwei Leitungen erforderlich, von denen die eine die bevorstehende Übertragung eines Datenwortes durch den Sender signalisiert und die andere nach erfolgter Übertragung die erfolgreiche Übernahme durch den Empfänger meldet. Die Daten werden über weitere Leitungen parallel oder seriell übertragen.

Zugriff „Client-Server": Der Client, z. B. der Benutzerrechner, fordert einen Dienst an, z. B. ein Anwenderprogramm oder eine Datenblattsammlung, den der Server zur Verfügung stellt.

Zugriff „Peer-to-Peer": Alle Rechner sind gleichberechtigt, und jeder stellt jedem seine freigegebenen Ressourcen zur Verfügung.

Sind mehr als zwei Geräte (ein Rechner und/oder weitere Rechner/Messgeräte) miteinander verbunden, kann es zu Kollisionen kommen, wenn gleichzeitig mehrere Geräte bzw. Rechner auf die gemeinsame Verbindungsleitung zugreifen. Hier sollen nur einige Prinzipien zum reibungslosen Ablauf des Datenverkehrs kurz dargestellt werden. Beim *Token-Verfahren* wird ein „Zeichen" zentral vergeben und weitergereicht. Nur der momentane Besitzer des Zeichens kann Daten senden. Die Haltezeit des Zeichens ist zeitlich beschränkt. Greifen z. B. beim Peer-to-Peer-Netzwerk mehrere Rechner/Geräte gleichzeitig auf die Leitung zu, wird ein Signalpegel auf der Leitung verändert (Open-Collector-Verknüpfung aller Geräte), so dass alle Geräte den Zugriff beenden und jeder einzelne nach einer statistisch vorgegebenen Zeit einen erneuten Verbindungsaufbau startet. Eine Verbesserung ergibt sich, wenn ein spezieller Übertragungskanal belegt werden kann und dies von den anderen Geräten eindeutig erkannt wird. Ein Übertragungsprotokoll sorgt dafür, dass jedes Gerät innerhalb einer bestimmten Zeit eine Verbindung aufbauen kann.

Serielle Rechnerschnittstelle V.24 oder RS-232: Die Daten werden üblicherweise seriell und asynchron, mit Blocks zu je 7 oder 8 Datenbits plus Start- und Stopbit, übertragen. Am Rechner befindet sich ein 25- oder 9-poliger Sub-D-Stecker. Die Spannungspegel zeigt Tab. 83.1.

Für die Verbindung von nur zwei Geräten (Rechner zum Messgerät) kann eine Verbindung über ein *Nullmodem-Kabel* ausreichend sein.

Parallele oder *Centronics-Rechnerschnittstelle*: Die Signale werden mit TTL-Pegel (0 V, +5 V) asynchron im Handshake-Verfahren übertragen.

83.2 IEC-Bus

Die folgenden Begriffe werden verwendet:

Talker: Das Gerät ist zurzeit nur als Sender (Sprecher) tätig.

Listener: Das Gerät ist zurzeit nur als Empfänger (Hörer) tätig.

Tab. 83.3 stellt die verwendeten Busbefehle dar.

Jedes der angeschlossenen Messgeräte erhält eine nur ihm zugeteilte Identifikationsnummer (möglich sind 1 bis 14, einstellbar über Schalter an jedem Gerät), mit dem es individuell erkannt und angesprochen werden kann. Es wird unterschieden zwischen Geräte- und Schnittstelleninformationen.

Geräteinformationen: Messbereichseinstellungen, Messdaten, Statusabfragen.

Tab. 83.2 Daten IEC-Bus

Daten		Erläuterungen, Ergänzungen
Normen	IEEE-488	24poliger Stecker[a]
	IEC 625	25poliger Stecker[a]
Max. Gerätezahl	15	1 Controller; 14 Messgeräte
Max. Leitungslänge insgesamt	20 m	
Max. Leitungslänge zwischen 2 Geräten	2 m	Verbindung der Geräte über Stecker-Buchse-Kombination
Übertragungsrate	20 … 1000 kByte/s	Evtl. schnelle Treiber erforderlich
Leitungen	8 Datenleitungen: D0 … D7 3 Steuerleitungen 5 Steuerleitungen	Für parallele Übertragung Datenübergabe/Handshake[b] Steuerung Datenaustausch
Steuersignale	TTL-Pegel: (0 … 1,4) V (2,5 … 5) V	Negative Logik; Open-Collector-Verbindung aller Teilnehmer
Daten	TTL-Pegel: (0 … 1,4) V (2,5 … 5) V	Positive Logik

[a] sonst völlig gleichwertig; [b] in Tab. 83.3

Tab. 83.3 Befehle am IEC-Bus

Befehl	Bedeutung	Funktion
REN	Remote enable	Gerätebedienung nur über IEC-BUS
ATN	Attention	Gerätenachrichten oder Schnittstellennachrichten
SRQ	Service request	Dienstanforderung
EOI	End of identify	Letztes Byte der Übertragung
IFC	Interface clear	Rücksetzen der Geräte in Ausgangszustand
DAV[a]	Data valid	Daten sind gültig
NRFD[a]	Not ready for data	Nicht zur Datenaufnahme bereit
NDAC[a]	No data accepted	Daten nicht übernommen

[a] Handshake-Befehle

Tab. 83.4 Befehle innerhalb der Schnittstelleninformationen

Universalbefehle	Bedeutung	Funktion
LLO	Local lockout	Sperren der Bedienelemente
DCL	Device clear	Rücksetzen
PPU	Parallel poll unconfigure	Ende Statusabfrage
SPE	Serial poll enable	Statusabfrage
SPD	Serial poll disable	Sperren Statusabfrage

Adressierte Befehle	Bedeutung	Funktion
GTL	Go to local	manueller Gerätebetrieb
GET	Group execute trigger	Trigger für Messstart
SDC	Select device clear	Rücksetzen Listener
PPC	Parallel poll configure	Listenerkonfiguration

Weitere Befehle	Bedeutung	Funktion
UNL	Unlisten	Adressen aller Listener-Geräte gelöscht
UNT	Untalk	Adressen aller Talker-Geräte gelöscht
LAD *x*	Listener address	Empfängeradresse *x*
TAD *x*	Talker address	Senderadresse *x*

Schnittstelleninformationen: Sie gliedern sich wie folgt auf:

a) Universalbefehle: Sie wirken auf alle angeschlossenen Geräte.
b) Adressierte Befehle: Mit dem Befehl wird die Geräteadresse übertragen, so dass nur das angesprochene Gerät entsprechend reagiert.
c) Weitere Befehle: Sie lassen sich den genannten zwei Befehlsarten nicht eindeutig zuordnen.

Damit könnte die Datenübertragung über den IEC-Bus prinzipiell so aussehen wie in Tab. 83.5 beschrieben.

83.3 DIN-Messbus

Dieser Bus ist relativ einfach aufgebaut und wurde für Anwendungen in der Messtechnik entwickelt. Seine Kenndaten sind in DIN 66348, Teil 2, enthalten. Tab. 83.6 listet einige Eigenschaften auf.

83.4 Aktuator-Sensor-Interface (ASI)

ASI ist von einer größeren Anzahl von Unternehmen als Schnittstelle im Bereich der Prozesssteuerung eingeführt worden. Es werden nur binäre Elemente miteinander vernetzt, so dass es vom Prinzip her Einschränkungen im Funktionsumfang gibt. Sind diese Einschränkungen ohne Bedeutung, ergibt sich ein kostengünstiges und unkompliziertes System. Abb. 83.1 stellt den prinzipiellen Aufbau dar, Tab. 83.7 listet einige Eigenschaften auf. Weitere Einzelheiten können der Literatur entnommen werden (http://www.as-interface.net).

Kernstück des ASI-Systems ist ein Slave-Chip, der die Sensoren S bzw. Aktuatoren A (Abb. 83.1) an die ASI-Leitung ankoppelt. Er kann in den Sensor/Aktuator eingebaut werden, wenn nur ein Element zum Einsatz kommt (Abb. 83.1, unten links), oder es wird ein Modulbaustein in den ASI-Slave integriert, so dass maximal vier S/A-Elemente verwendet werden können (Abb. 83.1, unten rechts). Ein Sensor kann z. B. der Grenzwertschalter für eine maximal zulässige Temperatur sein, ein Aktuator der Ein-Aus-Schalter für eine Heizung. Der

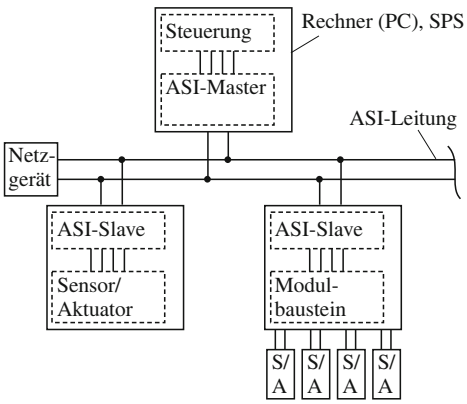


Abb. 83.1 Prinzipieller Aufbau des ASI-Systems

Tab. 83.5 Prinzipieller Ablauf einer Datenübertragung über den IEC-Bus

Steuerbefehle/Steuerleitungen					Datenleitungen D0 . . . D7	Erläuterungen
ATN	REN	SRQ	IFC	EOI		
0	0	0	0	0	–	Ausgangszustand
0	1	0	0	0	–	Gerätebedienung über IEC-Bus aktivieren
1	1	0	1	0	–	Geräte in den Ausgangszustand setzen
1	1	0	0	0	UNL	Adressen aller Listener-Geräte werden gelöscht
1	1	0	0	0	LAD 1 LAD 2 usw.	Gerät 1 als Listener Gerät 2 als Listener usw.
1	1	0	0	0	TAD 3	Gerät 3 als Talker; die vorhergehende Talkeradresse wird automatisch gelöscht
0	1	0	0	0	Datenwort 1	
0	1	0	0	0	Datenwort 2	
. . .	. . .	. . .	. . .	. . .	. . .	
0	1	0	0	1	Letztes Datenwort	
1	1	0	0	0	UNL	Adressen aller Listener-Geräte werden gelöscht

Tab. 83.6 Einzelheiten zum DIN-Messbus (DIN 66348, Teil 2)

Kenngröße	Daten	Erläuterungen
Organisation	1 Master, Slaves	Regelfall: Rechner als Master, Messgeräte als Slaves
Maximale Teilnehmerzahl	32	
Übertragungsart	Seriell, asynchron	
Datenwortlänge	7 bit + Paritätsbit	
Prüfverfahren	Paritätsbit + Blockprüfung (DIN 66022, 66219)	Evtl. Polynomprüfung möglich (DIN 66219)
Betriebsart	Blockübertragung mit Start-Stop-Betrieb, max. 128 Byte je Block	
Übertragungsrate	Typisch 9,6 kbit/s; bis zu 1 Mbit/s möglich	
Maximale Bus-Leitungslänge	500 m	Mit Repeatern einige km; Busleitung an den Enden mit Widerständen abgeschlossen (100 . . . 510 Ω; DIN 66348, Teil 2)
Maximale Leitungslänge Gerät-Bus	ca. 5 m	
Verbindungsleitung (Busleitung und Verbindung Gerät-Bus)	5adrig + Abschirmung: 2 für Senden 2 für Empfangen 1 für Betriebserde 1 für Schirmung	Signalleitungen paarweise verdrillt; Betriebserde: Verbindung der potentialfreien Bezugspunkte untereinander
Spannungspegel	Ähnlich Tab. 83.1; Einzelheiten u. a. in EIA RS-485	Galvanische Trennung zwischen Bus und Geräten vorgeschrieben
Steckverbindung	Sub-D-Stecker, 15-polig, abgeschirmt, am Gerät	Pin-Nr. – Belegung: 1 – Abschirmung, 2 – Sender-Daten, 9 – Sender-Daten, 4 – Empfänger-Daten, 11 – Empfänger-Daten, 8 – Betriebserde

Tab. 83.7 Einzelheiten zum ASI-System

Eigenschaft	Daten	Erläuterungen
Organisation	1 Master; max. 31 Slaves	Je Slave max. 4 binäre Sensoren/Aktuatoren; 5-bit-Adresse erforderlich
Übertragungsart	Seriell	Mit Start- und Stoppbit
Datenwortlänge	Masteraufruf: 14 bit; Slaveantwort: 7 bit	Siehe auch Abb. 83.2
Prüfverfahren (siehe auch Abb. 83.2)	Startbitfehler Alternierungsfehler Pausenfehler Informationsfehler Paritätsfehler Endebitfehler Aufruflängenfehler	1. Impuls stets negativ strenger Wechsel pos./neg. Pause: max. 1 Impulslänge folgt aus Kodeeigenschaft gerade Parität (posit. Imp.) letzter Impuls stets positiv Zeiten definierter Pause
Verbindungsleitung	Energieversorgung (24 V DC) und Datenübertragung über gemeinsame Zweidrahtleitung, nicht abgeschirmt	Entkopplung über Parallelschaltung R mit L je Leiter (39 Ω, 50 μH)
Übertragungsrate	167 kbit/s; davon 53,3 kbit/s für Daten	Systemreaktionszeit max. 5 ms (bei 31 Slaves); verringert sich mit abnehmender Slaveanzahl
Maximale Leitungslänge	100 m	Repeater für größere Leitungslängen
Steckverbindung	M 12, vierpolig, nach IEC 947-5-2 Anhang D	
Netz-Topologie	Linie, Baum und deren Kombinationen	
Energieversorgung der Slaves	24 V DC; max. ca. 100 mA je Slave; max. ca. 2 A insgesamt	Bei höherem Gesamtstrom größere Leitungsquerschnitte erforderlich
Modulationsart	Sensor: Non-Return-to-Zero-Kode (NRZ); daraus Manchester-II-Kode, übertragen in Alternierender Puls-Modulation (APM)	Siehe Abschn. 96.4.4; Impulse in $(\sin^2 x)$-Form

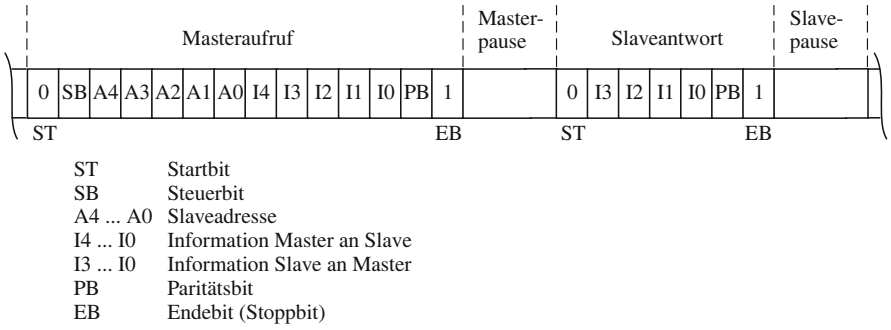


Abb. 83.2 Aufbau einer ASI-Nachricht

Tab. 83.8 Einzelheiten zum CAN-Bus (Auszug)

Topologie: Bus	**Teilnehmerzahl:** maximal 64 (durch Standard-Treiberbausteine begrenzt); durch Repeater zu erhöhen
Takt: NRZ-Codierung; keine Taktübertragung; ständige Flankensynchronisation, auch über Frame-(Rahmen-)Beginn; erzwungener Takt nach 5 aufeinander folgenden Bits ohne Wechsel	**Teilnehmerberechtigung:** alle Teilnehmer gleichberechtigt **Busbelegung:** gemäß festgelegter Prioritätenliste (wichtige Daten haben Vorrang)
Leitungswiderstand: 120 Ω	**Zugriffszeit auf die Daten:** erhöht sich mit wachsender Teilnehmerzahl
Kabellänge: 40 m bei 1 MBit/s; 1 km bei 50 kBit/s	**Ausgänge der Teilnehmer:** Wired-AND-Verknüpfung
Signalübertragung: Differenzspannung; < 0,5 V → High-Pegel > 0,9 V → Low-Pegel	**Datenbits:** maximal 8 Byte (64 Bit)
Schnittstelle: ISO 11898, 11519	**Datenübertragung:** Seriell
2 Ausführungen: Low Speed (max. 125 kBit/s) High Speed (max. 1 MBit/s)	**3adriges Kabel:** 2 Leiter für Differenzspannung, 1 Leiter für Masse (Bezugspunkt)
Frame-Länge: 64 Bit (Daten) + 30 Bit (Steuerdaten) + 10 Bit (Leerzeit) = 104 Bit	**Beispiel Übertragungsrate:** 500 kBit/s → 2 µs/Bit; Verarbeitungszeit im Rechner: 92 µs (geschätzt) Neuer Datenblock aussendbar nach: (104 Bit) (2 µs/Bit) + 92 µs = 300 µs
Fehlererkennung in mehreren Stufen (hier eine Auswahl): – 15 Bits zur Fehlererkennung, Hamming-Distanz 6 – Prüfung der Struktur des übertragenen Frames – Prüfung, ob mindestens *ein* Empfänger das Signal empfangen hat (Empfänger quittieren den erfolgreichen Frame-Empfang oder senden andernfalls eine Fehlermeldung) – Jeder Sender prüft, ob sein momentan geforderter Pegel auch tatsächlich auf der Busleitung vorhanden ist (u. a. zur Kurzschluss- und Kollisionserkennung) – Fehlerzähler trennt Teilnehmer beim Erreichen eines kritischen Wertes vom Bus – Beim Erkennen eines Fehlers wird der Frame erneut angefordert	

ASI-Master verbindet die ASI-Leitung mit dem Steuergerät, das ein Rechner, aber auch eine speicherprogrammierbare Steuerung (SPS) sein kann. Aus dem prinzipiellen Aufbau folgt, dass das ASI-System hierarchisch unterhalb der Feldbussysteme anzusiedeln ist und die Aufgabe hat, den Verdrahtungsaufwand (Stichwort „Kabelbaum“) zu verringern.

Die genannte Modulationsart erfordert eine geringe Kanalbandbreite. Durch die in Abb. 83.2 dargestellte Kodierung ist eine sehr effektive Fehlererkennung möglich: Bis zu 3 Fehler pro Nachricht werden sicher erkannt, 4 und 5 Fehler mit einer Wahrscheinlichkeit von 99,9999 %. Die Korrektur erfolgt dann über eine erneute Datenübermittlung.

83.5 CAN-BUS

Dieser Bus (CAN, *C*ontroller *A*rea *N*etwork) wurde ursprünglich für die Automobiltechnik entwickelt. Die dort zur Übertragung anstehenden Daten sind in nicht unerheblichem Maße Messwerte, so dass der Einsatz für die Messdatenübertragung auch außerhalb des Automobiles naheliegt. Aus dem Einsatz in der Automobiltechnik ergeben sich die folgenden Eigenschaften:

- Niedriger Preis durch die hohen Stückzahlen
- Hohe Zuverlässigkeit und Sicherheit in der Datenübertragung (Datenpaketaufbau)
- Hohe Störsicherheit gegenüber Störungen von außen
- Relativ einfache Verdrahtung durch 3adrige Leitung, auch 2adrig und 1adrig (Übertragung über Lichtleiter) möglich

Tab. 83.8 listet in Auszügen einige Eigenschaften des CAN-Busses auf.

Beispiel 1

Nach dem Beispiel in Tab. 83.8 kann ein neuer Datenblock nach etwa 300 µs ausgesendet werden. Bei einer Gefahrensituation und der höchsten Priorität des Datenframes steht er auch nach 300 µs zur Anwendung bereit. Fährt das Fahrzeug mit der Geschwindigkeit 100 km/h, hat es in 300 µs einen Weg von etwas weniger als 2 cm zurückgelegt. ◄

83.6 USB-Übertragung

Bei dieser Übertragung wird die von den Rechnern eingesetzte USB-Schnittstelle verwendet. Die Vorteile sind:

Sehr schnelle Datenübertragung, USB 2.0 mit maximal 400 MBit/s und USB 3.0 mit maximal 5 GBit/s. Die in der Praxis tatsächlich erreichten Werte sind niedriger und hängen ab von den Komponenten des Gesamtaufbaus der Übertragungsstrecke. Die Systeme 3.0, 2.0 und 1.1 (kaum noch eingesetzt, da technisch überholt) sind in der Regel untereinander abwärtskompatibel.

Vorhandensein von Peripheriebausteinen.

Verfügbarkeit von Literatur, Modellentwicklungen.

Die Geräte können an jeden Rechner angeschlossen und die vorhandenen Daten weiterverarbeitet werden. Weiterhin kann das Messgerät ferngesteuert werden.

83.7 LAN (*L*ocal *A*rea *N*etwork)

Es handelt sich hierbei um eine Entwicklung, die ursprünglich für die Rechnervernetzung einschließlich Peripheriegeräten z. B. in Firmen vorgesehen war. Dieses System bietet sich zur (Mit-)Benutzung an, wenn Messgerätedaten an unterschiedliche Rechner, Server oder Drucker zwecks Weiterverarbeitung oder Archivierung gesendet werden sollen und günstigenfalls schon eine Vernetzung über ein LAN besteht. Durch den Einsatz von Drahtlosverbindungen (**W**ireless **L**ocal **A**rea **N**etwork, WLAN) hat sich der Einsatzbereich weiter vergrößert. Es können auch mobile Messstationen in einem lokal begrenzten Bereich (Firmengelände) ihre Messdaten an Auswertestationen senden. Die übermittelten Daten können verschlüsselt werden. Der Vorteil dieses Systems ist auch, dass Daten in beiden Richtungen versendet werden können. Damit lassen sich Messgeräteparameter einstellen (Messbereich, Abtastrate, ...). Die Datenübertragungsrate liegt momentan bei maximal 1 GBit/s (theoretisch) bzw. (400 ... 600) MBit/s (praktisch).

Wie bei Druckern weit verbreitet werden Messgeräte zunehmend mit den beiden Schnittstellen USB und LAN versehen.

Probleme bei der Digitalisierung analoger Messwerte

84

Wilfried Plaßmann



84.1 Fehler bei der Digitalisierung 954
84.2 Signal-Quantisierungs-Geräuschabstand 954
84.3 Verbesserung des Signal-Rausch-Verhältnisses 955
84.4 Abtast-Halte-Glied 955
84.5 Aliasing 955
84.6 Erfassung von Momentanwerten 956

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_84

Messwerte liegen häufig in analoger Form als Spannungswerte vor. Sie werden in eine digital kodierte Form umgesetzt, wenn

- eine (nahezu) fehlerfreie Übertragung erforderlich ist (siehe dazu auch Abschn. 96.4.4 und 108.1);
- Signal*verläufe* gespeichert werden sollen (Speicheroszilloskop);
- eine Weiterverarbeitung (Rechner, PC) erfolgen soll; Digitalisierungsfehler
- Messungen mit sehr geringem Messfehler notwendig sind (Vorteil Digitalmessgerät gegenüber Zeigermessgerät).

Die zu diesem Zweck eingesetzten Analog-Digital-Umsetzer sind an anderer Stelle dieses Buches beschrieben. Hier soll auf einige Probleme, die durch den Umsetzvorgang entstehen, aus messtechnischer Sicht eingegangen werden. Inwieweit die einzelnen Anteile bei der Anwendung noch berücksichtigt werden müssen, hängt davon ab, ob es sich um ein fertiges Gerät (z. B. Digitalvoltmeter) oder einen zu entwickelnden Messaufbau handelt.

84.1 Fehler bei der Digitalisierung

Der Analog-Digital-Umsetzer hat den Eingangsspannungsbereich $0 \ldots +\hat{u}_{\mathrm{AD}}$ (Unipolarausführung) bzw. $-\hat{u}_{\mathrm{AD}} \ldots +\hat{u}_{\mathrm{AD}}$ (Bipolarausführung). Bei einem N-Bit-Umsetzer beträgt die kleinste Quantisierungsstufe $\hat{u}_{\mathrm{AD}}/(2^N - 1)$ (unipolar) bzw. $2 \cdot \hat{u}_{\mathrm{AD}}/(2^N - 1)$ (bipolar). Daraus ergeben sich die Fehlergrenzen F (absolut) und f (relativ), wenn der Momentanwert der Signalspannung bei einem Unipolarsignal mit u_S bezeichnet wird bzw. bei einem Bipolarsignal mit u_{S+} (größter positiver Signalwert) und $-u_{S-}$ (Bipolarsignal, kleinster negativer Signalwert), zu:

$$F_u = \pm\frac{1}{2} \cdot \hat{u}_{\mathrm{AD}} \frac{1}{2^N - 1} \quad \text{(unipolar)} \quad \text{bzw.}$$

$$F_b = \pm\frac{1}{2} \cdot \hat{u}_{\mathrm{AD}} \frac{2}{2^N - 1} \quad \text{(bipolar)} \tag{84.1}$$

$$f_u = \pm\frac{1}{2} \cdot \frac{\hat{u}_{\mathrm{AD}}}{u_S} \cdot \frac{1}{2^N - 1} \quad \text{(unipolar)}$$

bzw.

$$f_b = \pm\frac{1}{2} \cdot \frac{\hat{u}_{\mathrm{AD}}}{u_{S+} + |-u_{S-}|} \cdot \frac{2}{2^N - 1} \quad (\textit{bipolar}) \tag{84.2}$$

Formelzeichen siehe Text.

Aus (84.2) geht hervor, dass für eine kleine Fehlergrenze die Zahl N möglichst groß und der Wert der Signalspannung nur geringfügig kleiner als der Spannungsbereich des Umsetzers sein sollte. Bei kleinen Signalspannungen u_s kann für $u_S \ll \hat{u}_{\mathrm{AD}}$ durch Zuschalten eines geeigneten Spannungsverstärkers die Fehlergrenze f_u bzw. f_b entsprechend klein gehalten werden.

84.2 Signal-Quantisierungs-Geräuschabstand

Im Abschn. 96.4.4.5, Abb. 96.33, ist das Quantisierungsrauschen (Quantisierungsgeräusch) für den Fall eines sinusförmigen Signalverlaufes dargestellt. Der Verlauf der Quantisierungsrauschspannung u_{NQ} (Abb. 84.1) ist abhängig vom Nutzsignalverlauf u_S, kann aber näherungsweise durch einen sägezahnförmigen Verlauf erfasst werden, der mit x_0 periodisch ist. Ihr Effektivwert U_{NQ} berechnet sich wie folgt:

$$U_{\mathrm{NQ}} \approx \sqrt{\frac{1}{x_0} \cdot \int_0^{x_0} \left[\hat{u}_x \left(1 - \frac{x}{x_0} 2\right)\right]^2 dx}$$

$$= \hat{u}_x \sqrt{\frac{1}{3}} \approx \hat{u}_{\mathrm{AD}} \cdot \frac{1}{2^N} \cdot \sqrt{\frac{1}{3}}$$

mit

$$\hat{u}_x \approx \hat{u}_{\mathrm{AD}} \cdot \frac{1}{2^N} \tag{84.3}$$

Formelzeichen siehe Abb. 84.1, (84.2) und Text.

Für den Signal-Quantisierungs-Geräuschabstand (das Signal-Rausch-Verhältnis) a_{SQ} ergibt sich:

$$a_{\mathrm{SQ}} = 20 \cdot \lg\left[\frac{U_S}{U_{\mathrm{NQ}}}\right] \approx 20 \cdot \lg\left[\frac{U_S}{\hat{u}_{\mathrm{AD}}} \cdot 2^N \cdot \sqrt{3}\right] \tag{84.4}$$

U_S Effektivwert der Signalspannung u_S; U_{NQ} Effektivwert der Quantisierungsrauschspannung; $\hat{u}_{\mathrm{AD}}$ maximale Eingangsspannung des A-D-Umsetzers (siehe Text zu (84.2)).

Für eine sinusförmige Signalspannung u_S mit dem Scheitelwert $\hat{u}_{AD}$ und dem Effektivwert $U_S = \hat{u}_{\mathrm{AD}}/\sqrt{2}$ folgt daraus die im

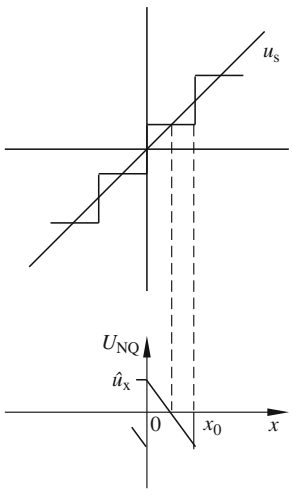


Abb. 84.1 Zur Ableitung des Effektivwertes der Quantisierungsrauschspannung u_{NQ}

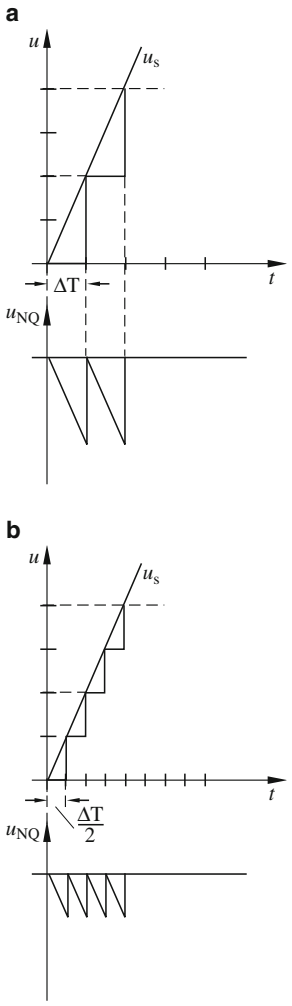


Abb. 84.2 **a** Quantisierungsrauschspannung u_{NQ} des Spannungsverlaufes u_5, **b** wie **a**, aber mit Oversampling, OSR = 2

Abschn. 96.4.4.5 angegebene Gleichung (96.43):

$$a_{SQ_{Sinus}} \approx 20 \cdot \lg\left[\frac{\hat{u}_{AD}}{\hat{u}_{AD} \cdot \sqrt{2}} \cdot 2^N \cdot \sqrt{3}\right] = 20 \cdot \lg\left[2^N \cdot \sqrt{\frac{3}{2}}\right] \approx 20 \cdot \lg\left[2^N\right] \qquad (84.5)$$

Formelzeichen siehe (84.4).

Oversampling: Die Signalabtastfrequenz $f_{ab} = (2{,}2 \ldots 4) \cdot f_{M\,max}$ nach Shannon (siehe Abschn. 96.4.4.2 „Aliasing-Effekt“ und (96.35)) wird um einen Faktor, die *Oversamplingrate OSR*, erhöht. Unter der Voraussetzung, dass sich zwei benachbarte Signalamplitudenwerte ohne Oversampling um mehr als den Wert einer Quantisierungsstufe unterscheiden, wird durch die erhöhte Abtastfrequenz ein „Zwischenwert“ nach Abb. 84.2 eingefügt und damit die Quantisierungsrauschspannung u_{NQ} im Bereich dieses Abtastwertes verringert. Für den Fall, dass die Amplitudenverringerung von u_{NQ} bei *allen* Abtastwerten auftritt, verringert sich der Effektivwert U_{NQ} um den Faktor $1/\sqrt{OSR}$.

84.3 Verbesserung des Signal-Rausch-Verhältnisses

Neben dem Quantisierungsgeräusch sind dem zu messenden Signal noch weitere Rauschanteile überlagert, die sowohl vom Messsignal selber als auch vom Messaufbau herrühren. Für die Wirkung aller Anteile zusammen werden deren Effektivwerte (U_{N1}, U_{N2}) errechnet und, da sie voneinander unabhängig (unkorreliert) sind, in folgender Weise zum resultierenden Signal-Rausch-Verhältnis a_{SN} verknüpft:

$$a_{SN} = 20 \cdot \lg\left[\frac{U_S}{\sqrt{U_{N1}^2 + U_{N2}^2 + \ldots}}\right] \qquad (84.6)$$

U_S Effektivwert der Signalspannung, U_{N1}, U_{N2} Effektivwerte zweier Rausch-(Stör-)Spannungen (unkorreliert).

Handelt es sich bei den zu messenden Größen um Gleichsignale, kann durch n-faches Messen und Mittelwertbildung das Signal-Rausch-Verhältnis $a_{SN(n)}$ verbessert werden [5]:

$$a_{SN(n)} = 20 \cdot \lg\left[\frac{U_S}{U_N} \cdot \sqrt{n}\right] \qquad (84.7)$$

n Zahl der Messungen für die Mittelwertbildung, U_S Effektivwert der Signalspannung, U_N Effektivwert der Rausch-(Stör-)Spannung.

Das Verfahren kann auch bei periodischen Wechselsignalen mit konstanter Kurvenform angewendet werden, sofern durch ein geeignetes Triggersignal die zeitliche Zuordnung und damit die Mittelwertbildung für ein oder mehrere Signalwerte erfolgen kann.

Für die Messung von Gleichgrößen besteht weiterhin die Möglichkeit, den Frequenzbereich von Störungen bzw. Rauschen und damit deren Wirkung auf das Nutzsignal durch Filter auf sehr niedrige Werte einzugrenzen.

84.4 Abtast-Halte-Glied

Über ein Abtast-Halte-Glied muss der analoge Eingangswert für den Umsetzer solange konstant gehalten werden, bis der Umsetzvorgang abgeschlossen ist. Im Abschn. 96.4.4.3, ist dargestellt, inwieweit sich diese Forderung mit nichtidealen Bauteilen erfüllen lässt.

84.5 Aliasing

Der mit Aliasing bezeichnete Effekt bei der Abtastung von periodischen Signalverläufen ist im Abschn. 96.4.4.2 „Aliasing-Effekt“ dargestellt.

84.6 Erfassung von Momentanwerten

Über ein Antialiasing-Filter wird erreicht, dass die höchste im gegebenen Signalverlauf vorkommende Frequenz $f_{M\,\max}$ beträgt. Die größte Steigung ergibt sich für 0° bzw. 180°:

$$u = \hat{u} \cdot \sin(2\pi \cdot f_{M\,\max} \cdot t)$$

$$\Rightarrow \quad \frac{du}{dt} = \hat{u} \cdot 2 \cdot \pi \cdot f_{M\,\max} \cdot \cos(2 \cdot \pi \cdot f_{M\,\max} \cdot t) \qquad (84.8)$$

Zwei benachbarte Abtastwerte unterscheiden sich umso mehr, je kleiner die Abtastfrequenz f_{ab} ist und je näher man sich am Nulldurchgang des Signalanteiles mit der Frequenz $f_{M\,\max}$ befindet. Die größte Differenz ergibt sich für den (theoretischen) Fall $f_{\text{ab}} = 2 \cdot f_{M\,\max}$ zu $2 \cdot \hat{u}$, wenn genau der positive und der negative Scheitelwert erfasst werden (Abschn. 96.4.4.2 „Aliasing-Effekt“). Eine *Abschätzung* des *maximal möglichen* Unterschiedes $\Delta u_{\max}$ zweier benachbarter Abtastwerte erhält man für folgende Annahmen: 1. $f_{\text{ab}} \gg f_{M\,\max}$; 2. Das abzutastende Signal ist sinusförmig mit $f = f_{M\,\max}$ (nur höchstmögliche Signalfrequenz vorhanden). In der Umgebung des Nulldurchganges ergibt sich mit der Näherung $u(t) \approx \hat{u} \cdot 2\pi \cdot t/T$ und der Beziehung $T = 1/f_{M\,\max}$:

$$\Delta u_{\max} = u(t = t_1) - u(t = 0) = u\left(\frac{1}{f_{\text{ab}}}\right) - u(t = 0)$$

$$\approx \hat{u} \cdot 2\pi \cdot \frac{f_{M\,\max}}{f_{\text{ab}}}$$

Für den maximalen relativen Fehler folgt daraus:

$$f_{\max} = \frac{\Delta u_{\max}}{\hat{u}} \approx 2\pi \cdot \frac{f_{M\,\max}}{f_{\text{ab}}} \qquad (84.9)$$

$f_{M\,\max}$ Frequenz des höchsten sinusförmigen Signalanteiles, f_{ab} Abtastfrequenz.

Der Einfluss dieses Fehlers ist dann von besonderer Bedeutung, wenn Momentanwerte, d. h. Signal- bzw. Messwerte in Echtzeit, erfasst werden sollen.

PC-gestützte Messverfahren und Messsignalanalyse

85

Wilfried Plaßmann



85.1 Statistische Verfahren zur Messsignalauswertung 958
85.2 Graphische Darstellung 958
85.3 Ermittlung von Kenngrößen, Klassierung 959
85.4 Messsignalanalyse 959
85.5 Automatisierung von Messabläufen 960
Literatur 960

Teil X

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_85

Der Rechner oder PC wird in der Messtechnik zunehmend eingesetzt, um

1. umfangreichere statistische Berechnungen zur Verringerung der Fehlergrenzen auszuführen,
2. die funktionale Abhängigkeit einer Messreihe in einem Diagramm optimal darzustellen,
3. Kenngrößen des Signalverlaufes zu ermitteln,
4. das Messsignal zu analysieren, damit die gewünschte Größe möglichst fehlerfrei bestimmt wird, und
5. häufig vorkommende Messabläufe zu automatisieren.

Die Messwerte gelangen entweder manuell in den Rechner, z. B. durch die Eingabe einer Messwerttabelle über die Tastatur, oder aber der Rechner kommuniziert über ein Bussystem (siehe Kap. 83) mit der Messeinrichtung und liest die Messwerte automatisch ein. Es hängt vom verwendeten Programm zur Messwerterfassung und Messsignalanalyse ab, welche der im Folgenden aufgeführten Funktionen ausführbar sind bzw. welche zusätzlichen Funktionen noch zur Verfügung stehen.

85.1 Statistische Verfahren zur Messsignalauswertung

In Abschn. 75.3.3 und 75.3.4 sind der arithmetische Mittelwert und die Standardabweichung von Messwerten aufgeführt:

arithmetischer Mittelwert

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{85.1}$$

Standardabweichung

$$s = + \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{85.2}$$

n Anzahl der Messungen; x_i die einzelnen Messwerte.

Sie lassen sich mit einem entsprechenden Rechnerprogramm berechnen.

Sind den Messwerten zufällige Messabweichungen überlagert, ergibt sich (theoretisch, für $n \rightarrow \infty$) die Normal- oder Gauß'sche Verteilung:

Normalverteilung

$$p(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{85.3}$$

$\bar{x}$ siehe (85.1), $\sigma^2 = s^2$ (siehe (75.8a), (75.8b) und (85.2)).

Da die Zahl der Messwerte endlich ist und sie nur bedingt die Normalverteilung erfüllen, ist das folgende Testverfahren entwickelt worden:

- *Chi-Quadrat-Test*: Vereinfacht ausgedrückt werden die Differenzen zwischen den einzelnen Messwerten und den zugeordneten Werten der Normalverteilung gebildet, quadriert und addiert. Überschreitet die Summe einen kritischen Wert, sind die Messwerte nicht normalverteilt.
- *Vertrauensgrenzen*: Weiterhin können die Vertrauensgrenzen errechnet werden, wobei die erforderlichen Zahlenwerte (siehe Abschn. 75.7, Tab. 75.3 und DIN 1319 Teil 3) im Programm abgespeichert sind:

$$\textit{Obere Vertrauensgrenze} \qquad \bar{x} + \frac{t}{\sqrt{n}} \cdot s \tag{85.4a}$$

$$\textit{Untere Vertrauensgrenze} \qquad \bar{x} - \frac{t}{\sqrt{n}} \cdot s \tag{85.4b}$$

- *Quantile*, *Median*, *Zentralwert*: Ein α-Quantil teilt eine nach der Größe der Werte geordnete Wertetabelle (Stichprobe) so in zwei Teile, dass $\alpha \cdot 100\,\%$ der Werte unterhalb und $(1 - \alpha) \cdot 100\,\%$ oberhalb der Trennungslinie liegen. Häufig wird das 0,50-Quantil ($\alpha = 0{,}50$) verwendet, das die Wertetabelle halbiert. Es wird auch mit Median oder Zentralwert bezeichnet. Angewendet werden Quantile, um die Lage von einzelnen Messwerten bezüglich einer Gesamtheit zu erfassen (daher die Bezeichnung „Lagemaß"). Sie werden von Ausreißern relativ wenig beeinflusst (sie sind robust).

85.2 Graphische Darstellung

- *Approximation*: Mit geeigneten Rechnerprogrammen werden die vorliegenden Messwerte (Tabelle oder direkt eingelesen) in Form eines Diagramms dargestellt. Da die funktionale Abhängigkeit in vielen Fällen mathematisch nicht vorgegeben ist, muss eine geeignete Formel gefunden werden. Dazu wählt der Anwender aus einer Vielzahl von möglichen Funktionen eine ihm geeignet erscheinende aus. Beispiele hierfür sind:
  Potenzreihendarstellung:

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 + \ldots$$

Exponentialdarstellung:

$$y = b_0 + b_1 \cdot e^{-\frac{x}{b_2}}$$

Geradendarstellung:

$$y = c_0 + c_1 \cdot x$$

Logarithmische Darstellung:

$$y = d_0 + d_1 \cdot \log\left(\frac{x}{d_2}\right) \tag{85.5}$$

Im allgemeinen Fall geht die gewählte Funktion *nicht* durch alle vorgegebenen Messpunkte, was in der Praxis auch nicht unbedingt wünschenswert ist (sonst beeinflussen „Ausreißer" den Verlauf überproportional). Das Rechnerprogramm ermittelt die Zahlenwerte für die Abweichungen zwischen den einzelnen Messwerten und der gewählten Kurvenannäherung, quadriert und addiert sie. Eine geeignete Kurvenform ist in der Regel diejenige, bei der diese Summe ein

Minimum erreicht. Das Verfahren wird auch mit „kleinste Summe der Abweichungs- bzw. Fehlerquadrate“ bezeichnet. Unter dieser Vorgabe optimiert das Programm anschließend die Koeffizienten $a_0, a_1, a_2, a_3, \ldots$ für die Potenzreihendarstellung, b_0, b_1, b_2 für die Exponentialdarstellung usw.

- *Regression*: Die funktionale Abhängigkeit wird in der Form $y = a_0 + a_1 \cdot x + \ldots + a_m \cdot x^m$ angenommen, und die Koeffizienten a_0 bis a_m werden unter der Zielsetzung „kleinste Summe der Abweichungsquadrate“ bestimmt.
  Für den Fall, dass n Messungen durchgeführt wurden, gilt folgender Ansatz, der sich in Matrizenschreibweise darstellen lässt:

$$\begin{aligned} y_1 &= a_0 + a_1 \cdot x_1 + a_2 \cdot x_1^2 + \ldots + a_m \cdot x_1^m \\ &\vdots \\ y_n &= a_0 + a_1 \cdot x_n + a_2 \cdot x_n^2 + \ldots + a_m \cdot x_n^m \end{aligned}$$

$$\Rightarrow \quad y = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^m \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_n & \ldots & x_n^m \end{pmatrix}}_{X} \cdot \underbrace{\begin{pmatrix} a_0 \\ \cdots \\ \cdots \\ a_m \end{pmatrix}}_{a} = X \cdot a \qquad (85.6)$$

  Die gesuchte Koeffizientenmatrix a ergibt sich zu:

$$a = (X^T \cdot X)^{-1} \cdot X^T \cdot y \qquad (85.7)$$

  X^T transponierte Matrix, $(X^T \cdot X)$ Matrizenprodukt, $(X^T \cdot X)^{-1}$ inverse Matrix des Matrizenproduktes $(X^T \cdot X)$.
  Auf den Beweis, dass es sich bei dem dargestellten Rechengang tatsächlich um einen Kurvenverlauf mit der kleinsten Summe der Abweichungsquadrate handelt, wird hier verzichtet.
- *Lineare Regression*: Es wird vorausgesetzt, dass eine Abhängigkeit in der Form $y = a_0 + a_1 \cdot x$ besteht. Die Koeffizienten a_0 und a_1 werden wiederum unter der Zielsetzung „kleinste Summe der Abweichungsquadrate“ ermittelt. Für die lineare Regression gilt in (85.6) $m = 1$, so dass die Matrix X nur die ersten zwei Spalten besitzt, von denen die erste wiederum nur aus Einsen besteht. Die Koeffizienten der gesuchten Geraden ergeben sich analog zu (85.7):

$$a = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot y \qquad (85.8)$$

  X^T transponierte Matrix, $(X^T \cdot X)$ Matrizenprodukt, $(X^T \cdot X)^{-1}$ inverse Matrix des Matrizenproduktes $(X^T \cdot X)$.
  Die so ermittelte Gerade wird auch als *Ausgleichsgerade* bezeichnet.
- *Abweichungsgrenzen*, Vertrauensbereich: Diese Größen können meist vom Rechnerprogramm berechnet und in das Diagramm eingezeichnet werden.
- *Splines*: Lässt sich keine Kurvenform finden, die alle gegebenen Messwerte ausreichend genau annähert, zerlegt man den Wertebereich der Messwerte in mehrere Teilbereiche und sucht für jeden Teilbereich eine geeignete Kurvenform. Um an der Stoßstelle zweier benachbarter Kurven weder Sprünge noch Knickpunkte zu haben, müssen der Wert und die erste Ableitung beider Kurven im Verbindungspunkt gleich sein. Die so bestimmten Kurvenverläufe bezeichnet man mit Splines.

85.3 Ermittlung von Kenngrößen, Klassierung

Liegt dem Rechner der funktionelle Zusammenhang zweier Größen vor, können wichtige Kenngrößen durch Berechnen der zugeordneten mathematischen Formel ermittelt werden. Dazu gehören:

- Arithmetischer Mittelwert (Berechnung nach (75.15a) und (75.15b));
- Effektivwert (Berechnung nach (75.14a) (75.14b));
- Extremwert (Maximalwert, Minimalwert; durch Berechnung der ersten und zweiten Ableitung und Vorzeichenerfassung);
- Gleichrichtwert (Betragsbildung und Berechnung nach (75.15a) und (75.15b)).

Das Rechnerprogramm kann häufig auch Klassierungen durchführen, d. h. bei einem Kollektiv von gleichartigen Elementen (z. B. gemessenen Spannungswerten) die Häufigkeitsverteilung nach bestimmten Merkmalen (z. B. Abweichungen vom Sollwert) in vorgegebene oder vereinbarte Klassen feststellen (z. B. Abweichung $(0 \ldots +1)\,\%$, $(> +1 \ldots +2)\,\%$) (DIN 1319 Blatt 1). Zur Klassierung gehören auch Begriffe wie Spitzenwert (relativ, absolut), Verweildauer (Zeitdauer, während der ein Wert einen Referenzwert überschritten/unterschritten hat) und Spannenwert (Differenz zwischen einem Maximalwert und dem benachbarten Minimalwert). Weitere Begriffe und Einzelheiten enthält DIN 45667.

85.4 Messsignalanalyse

Bei den nachfolgend dargestellten Verfahren sind die gemessenen Signalwerte zeitabhängig. Dabei kann das Messsignal selber zeitabhängig sein, oder aber ein zeitunabhängiges Messsignal wird von einem zeitabhängigen Störsignal überlagert. Es wird jeweils nur das Prinzip dargestellt, weitere Einzelheiten können den Teilen XIII und XIV in diesem Buch sowie der einschlägigen Literatur entnommen werden. Gemeinsames Kennzeichen der aufgeführten Verfahren ist, dass sie aufwendige mathematische Berechnungen erfordern, womit sich der Einsatz eines Rechners anbietet.

85.4.1 Verfahren

- *Fourierzerlegung*, Fouriertransformation: Für periodische und nichtperiodische Signale werden die im Signalverlauf enthaltenen Frequenzanteile bestimmt. Mit der Fast-Fourier-Transformation (FFT) ist ein für den Rechnereinsatz besonders geeignetes Verfahren entwickelt worden (siehe auch Abschn. 92.4.9.2, 92.4.9.3, und Teil XIV).

- *Autokorrelation*, Kreuzkorrelation: Das gemessene Signal wird bei der Autokorrelation zeitlich verschoben und mit dem nichtverschobenen Signal verglichen. Die Übereinstimmung ist maximal, wenn die Verschiebung Null ist. Bei der Kreuzkorrelation werden zwei Signalverläufe miteinander verglichen und auf Übereinstimmung untersucht (siehe auch Abschn. 92.4.11.2, und Systemtheorie).
- *Digitale Filter*: Sie haben den Vorteil, dass die Filtercharakteristik (Tiefpass, Bandpass, ...) und die Grenz- bzw. Resonanzfrequenzen softwaremäßig programmiert werden können und sich letztere auch auf sehr niedrige Werte einstellen lassen (siehe auch Abschn. 97.2.2).

85.4.2 Anwendungen

85.4.2.1 Messung des Klirrfaktors

Nach der Fourierzerlegung des gemessenen periodischen Signalverlaufes wird der Klirrfaktor nach der Formel

$$Klirrfaktork' = \frac{\sqrt{U_2^2 + U_3^2 + U_4^2 + \ldots}}{U_1} \qquad (85.9)$$

berechnet. Für eine weitere Definition des Klirrfaktors siehe Abschn. 92.4.12.3.

85.4.2.2 Geräuschmessung zur Schadenfrüherkennung

Die Lager von Motoren und Generatoren unterliegen dem Verschleiß. Bei einem notwendigen Austausch des Lagers sind weniger die Kosten für das Lager und den Austausch als vielmehr das entgangene Entgelt während der Ausfallzeit entscheidend. Es wird deshalb angestrebt, Lagerschäden möglichst früh zu erkennen und den Austausch optimal vorzubereiten (auszutauschendes Lager vorhanden, Außerbetriebnahme während einer Zeit geringer Auslastung). Das ist möglich, weil in der Regel zunächst nur geringe Schäden auftreten: Nur eine der im Kugellager vorhandenen 20 Kugeln ist defekt, und das Lager kann noch einige Tage oder Wochen weiter benutzt werden. Erst dann wird die Betriebssicherheit durch den Ausfall weiterer Kugeln verringert, bis schließlich durch das Blockieren des Lagers der Stillstand erzwungen und der Schaden maximal wird. Zur Schadenfrüherkennung nimmt man über ein Mikrofon das Geräusch des intakten Lagers auf und führt eine Frequenzanalyse durch, die im Rechner als Sollmuster gespeichert wird. Die Geräuschmessung mit anschließender Analyse erfolgt jetzt periodisch, und das aktuell aufgenommene Spektrum wird mit dem gespeicherten Sollmuster verglichen. Mit jeder defekten Kugel ändert sich das Spektrum, so dass die oben geschilderten Maßnahmen rechtzeitig ergriffen werden können. Im Laufe der Zeit hat man Erfahrungen sammeln können, wie bestimmte Lagerschäden das Frequenzspektrum verändern, so dass das Messverfahren immer zuverlässiger geworden ist.

85.4.2.3 Abstandsmessung

Ein (bandbegrenztes) Rauschsignal wird von einem Mikrofon aufgenommen (S1), verstärkt und von einem Lautsprecher abgestrahlt (S2). Führt man mit S1 und S2 eine Kreuzkorrelation durch, ergibt sich ein Maximum für eine Zeitverschiebung t beider Signale zueinander, die genau der Laufzeit zwischen Mikrofon und Lautsprecher entspricht (siehe auch Abschn. 92.4.11.2, und Systemtheorie).

85.4.2.4 Erkennung periodischer Signalanteile

Liegt ein durch z. B. Rauschen gestörter Signalverlauf vor, können darin enthaltene periodische Anteile dadurch erkannt werden, dass das Signal autokorreliert wird. Es ergeben sich Maxima im Abstand der Periodendauer des im Signal enthaltenen periodischen Anteils (siehe auch Abschn. 108.3).

85.5 Automatisierung von Messabläufen

Besitzen die Messgeräte eine Schnittstelle zur Verwendung eines Bussystems nach Kap. 83, können Messabläufe automatisiert werden. Der Rechner kann im einfachsten Fall die Messablaufsteuerung (Wahl des Messgerätes, Messbereichseinstellung, Messzeitpunkt) mit anschließender Messwertspeicherung übernehmen. Darüber hinaus können die Messwerte bearbeitet, analysiert und auf Grund der Ergebnisse erforderliche Steueraufgaben ausgeführt werden. Auf diese Weise laufen Industrieprozesse ab.

Literatur

Literatur allgemein

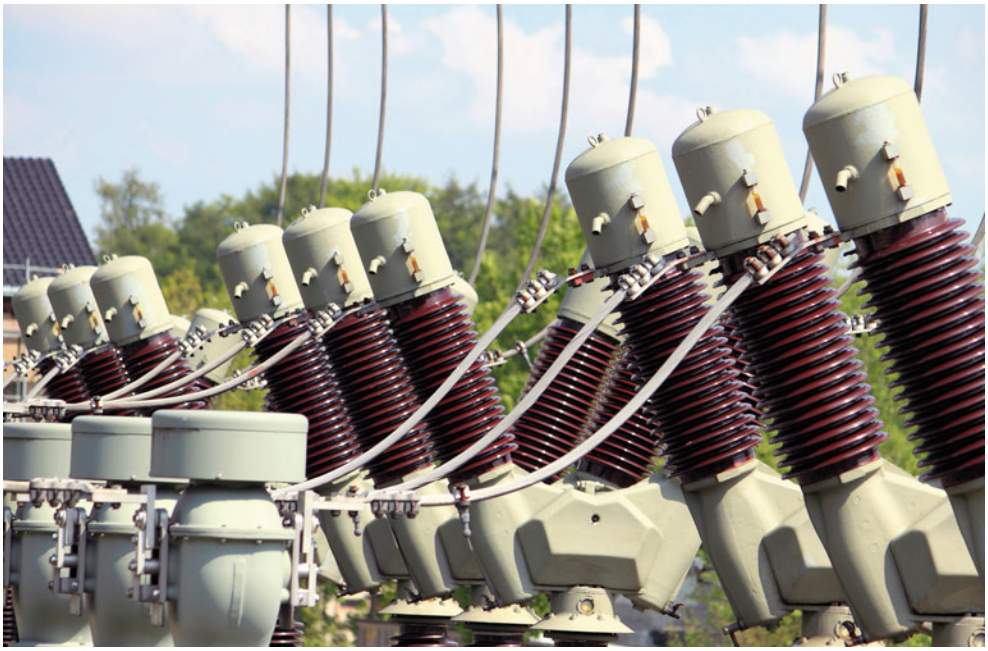
1. *Bergmann, K.:* Elektrische Messtechnik. Verlag Vieweg Braunschweig/Wiesbaden. 6. Auflage 1996
2. *Felderhoff, R.:* Elektrische und elektronische Messtechnik. Verlag Hanser München, Wien. 8. Auflage 2007
3. *Schrüfer, E.:* Elektrische Messtechnik. Verlag Hanser München, Wien. 9. Auflage 2007
4. *Hoffmann, J. (Hrsg.):* Taschenbuch der Messtechnik. Verlag Hanser München. 6. Auflage 2010

Spezielle Literatur

5. *Schrüfer, E.:* Signalverarbeitung. Verlag Hanser München, Wien. 1. Auflage 1990

Energietechnik

XI

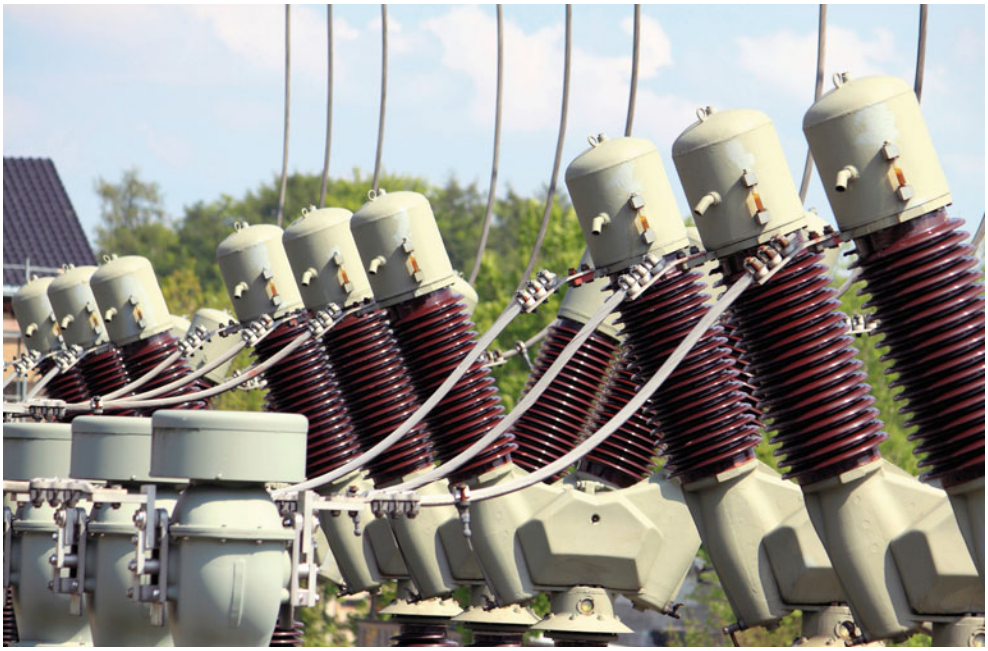


86 Elektrische Maschinen 963
87 Elektrische Energieversorgung 1009
88 Elektrische Energieanwendung 1075

Elektrische Maschinen

86

Reinhard von Liebenstein



86.1 Transformatoren 964
86.2 Drehstrommaschinen 976
86.3 Einphasen-Asynchronmotoren 988
86.4 Drehstrommotor im Einphasenbetrieb 989
86.5 Sonderbauformen 989
86.6 Gleichstrommaschinen 991
86.7 Auswahl von Motoren 1001
Literatur 1007

Teil XI

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_86

Elektrische Maschinen sind Energiewandler, deren Funktion auf den Gesetzen des Elektromagnetismus beruht. Zu den elektrischen Maschinen gehören die Transformatoren (ruhende elektrische Maschinen) und die rotierenden Maschinen, die elektrische Energie in mechanische Energie (Elektromotor) oder mechanische Energie in elektrische Energie (Generator) umformen. Nach der Stromart unterscheidet man Gleichstrom-, Drehstrom- und Einphasenwechselstrommaschinen.

86.1 Transformatoren

86.1.1 Aufgaben eines Transformators

Mit Hilfe eines Transformators wird Wechsel- oder Drehstromleistung gegebener Spannung und Frequenz in solche höherer oder niedrigerer Spannung bei gleichbleibender Frequenz umgewandelt. Ohne die Transformatorentechnik ist der heutige Einsatz der elektrischen Energie in allen Bereichen unseres Lebens nicht denkbar. Die in den Kraftwerken erzeugte elektrische Energie wird durch Transformatoren auf Spannungsebenen umgewandelt, die einen rationellen Transport über weite Entfernungen ermöglichen. Transformatoren großer Leistung werden in den Netzen der Energieversorgungsunternehmen (EVU) eingesetzt. Im unteren Spannungsbereich werden Einphasen-Transformatoren in großer Stückzahl verwendet, um die für die dortigen Anwendungen erforderlichen Spannungen zu erzeugen. Ebenso wandeln Transformatoren Signale der Nachrichtentechnik um (Übertrager), oder sie dienen in der Messtechnik zur Anpassung der zu messenden Signale an den Messbereich des Messgerätes (Wandler). Obwohl Transformatoren kein rotierendes Teil besitzen, werden sie in der Energietechnik zum Bereich der Maschinen gezählt.

86.1.2 Bauteile eines Transformators

Ein einfacher Transformator besteht aus einem geschlossenen Eisenkern und zwei Spulen, die auf diesen Kern gewickelt sind. Der Eisenkern besteht aus einzelnen, gegeneinander elektrisch isolierten speziellen Dynamoblechen, um den Wirbelstrom, der bei Wechselströmen entsteht, zu unterdrücken.

86.1.2.1 Eisenkerne

Die Vielzahl der Einsatzgebiete der Transformatoren erfordert verschiedene Bauformen. Es gibt die *Kern-* oder *Mantelbauform*. In den Bildern Abb. 86.1 und 86.2 sind Einphasen-Transformatoren in verschiedenen Bauformen dargestellt.

Beim *Einphasen-Kerntransformator* sind die beiden äußeren Schenkel mit jeweils der halben Windungszahl der Ober- (OS) und der Unterspannungsseite (US) bewickelt. Beim *Einphasen-Manteltransformator* befinden sich die beiden Wicklungen auf dem mittleren Schenkel. Der mittlere Schenkel wird wie von einem Mantel durch zwei Außenschenkel umgeben.

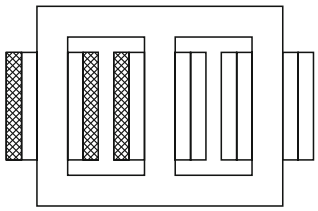


Abb. 86.1 Einphasen-Kerntransformator

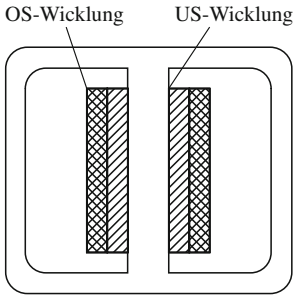


Abb. 86.2 Einphasen-Manteltransformator

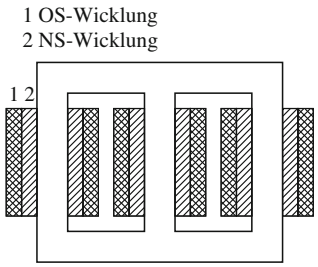


Abb. 86.3 Drehstrom-Dreischenkeltransformator

Beim *Drehstrom-Transformator* erfolgt die Aufteilung der Wicklungen auf die Schenkel des Transformators wie in Abb. 86.3 gezeigt. Die Sonderbauform des *Fünfschenkel-Transformators* findet man bei großen Energietransformatoren, da hierbei die Querschnitte der Joche nur ungefähr 58 % der Säulenquerschnitte betragen. Dadurch verringert sich die Baugröße erheblich.

86.1.2.2 Wicklungen

Im Aufbau der Wicklung unterscheidet man zwei Arten, die *Zylinderwicklung* und die *Scheibenwicklung*.

Bei der Zylinderwicklung wird die Röhre der Oberspannungswicklung über die der Unterspannungswicklung gestülpt (Abb. 86.4). Diese Wicklungsart findet man bei den meisten Transformatoren. Die Scheibenwicklung unterteilt die Ober- und Unterspannungswicklung in einzelne übereinander angeordnete Scheiben (Abb. 86.5).

86.1.2.3 Kühlung

Wie bei allen elektrischen Maschinen erzeugen die Transformatoren Verluste, die Wicklungen und Eisenkerne erwärmen.

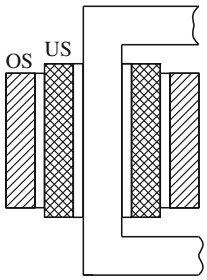


Abb. 86.4 Zylinderwicklung

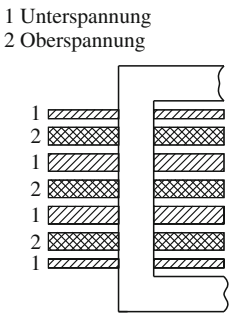


Abb. 86.5 Scheibenwicklung

Diese Verluste müssen vom Transformator an die Umgebung abgeführt werden. Transformatoren werden in Trockentransformatoren und solche mit flüssigem oder gasförmigem Wärmeüberträger eingeteilt. Bei den Trockentransformatoren übernimmt die Umgebungsluft den Wärmetransport über Konvektion oder mittels Strahlung (Selbstkühlung). Neben der Selbstkühlung kann die Wärme durch künstlichen Zug, z. B. durch Lüfter, abgeführt werden (Zwangskühlung). Transformatoren mit flüssigem Wärmeübertrager, in der Hauptsache Öl, weisen gegenüber den Trockentransformatoren einige Vorteile auf. Infolge des größeren spezifischen Gewichts, der größeren spezifischen Wärme und der größeren Wärmeleitfähigkeit kann einerseits die erzeugte Wärmemenge der Wicklungen und des Eisenkerns besser abgeführt, zum anderen eine große Wärmemenge in der Kühlflüssigkeit gespeichert werden, was bei kurzzeitigen Überlastungen bedeutsam ist. Das im Transformator eingesetzte Kühlverfahren erkennt man am Typenschild. Die Bedeutung der Kurzzeichen sind in Tab. 86.1 aufgeführt.

Tab. 86.1 Kühlmittel und Kurzzeichen

Kühlmittel	Kurzzeichen
Mineralöle oder entsprechende synthetische Flüssigkeiten mit einem Brennpunkt $\leq 300\,°C$	O
Andere synthetische Flüssigkeiten	L
Gas, Brennpunkt $> 300\,°C$	G
Luft	A
Wasser	W
Kühlmittelbewegung	
Natürlich	N
Erzwungen, gerichtet	D
Erzwungen, nicht gerichtet	F

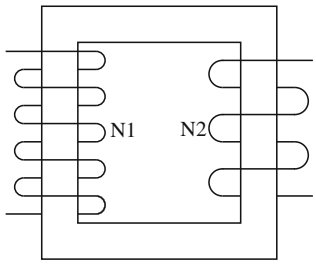


Abb. 86.6 Einphasen-Transformator

Beispiel

Typenschildaufschrift AN Trockentransformator mit natürlicher Luftkühlung; OFAN Öltransformator mit erzwungener Ölkühlung und natürlicher Luftkühlung. ◄

86.1.3 Wirkungsweise eines Einphasen-Transformators

In den Grundlagen der Elektrotechnik werden die Begriffe Selbstinduktion und gegenseitige Induktion anhand zweier gekoppelter Spulen erläutert. Ein Einphasen-Transformator besteht aus einer Wicklung auf dem linken Schenkel (Primärwicklung mit der Windungszahl N_1) und einer Wicklung auf dem rechten Schenkel (Sekundärwicklung mit der Windungszahl N_2).

Beide Wicklungen sind magnetisch über den geschlossenen Eisenkern miteinander verbunden (Abb. 86.6). Bei der weiteren Betrachtung des Transformators wird angenommen, dass die Primärwicklung an eine sinusförmige Wechselspannung U_1 angelegt wird und alle anderen Größen ebenfalls einen sinusförmigen Verlauf aufweisen. Verzerrungen durch Sättigung und Oberschwingungen werden vernachlässigt.

86.1.3.1 Leerlauf

Bei offener Sekundärwicklung ($I_2 = 0$) erregt der Strom I_1, der durch die Primärspule fließt, den magnetischen Fluss $\Phi_1 = \Phi_\eta + \Phi_{1\sigma}$.

Der Hauptfluss Φ_h ist durch den Eisenkern mit beiden Spulen verkettet und induziert in den Spulen die Spannungen (Abb. 86.7):

Leerlauf-Primärspannung

$$U_{10} = j4{,}44 \cdot f \cdot N_1 \cdot \Phi_h \qquad \begin{array}{c|c|c|c} U_{10} & f & N_1 & \Phi_h \\ \hline \mathrm{V} & \mathrm{Hz} & 1 & \mathrm{V\,s} \end{array} \qquad (86.1)$$

Leerlauf-Sekundärspannung

$$U_{20} = j4{,}44 \cdot f \cdot N_2 \cdot \Phi_h \qquad \begin{array}{c|c|c|c} U_{20} & f & N_2 & \Phi_h \\ \hline \mathrm{V} & \mathrm{Hz} & 1 & \mathrm{V\,s} \end{array} \qquad (86.2)$$

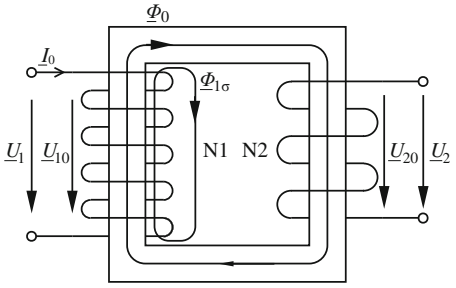


Abb. 86.7 Einphasen-Transformator im Leerlauf

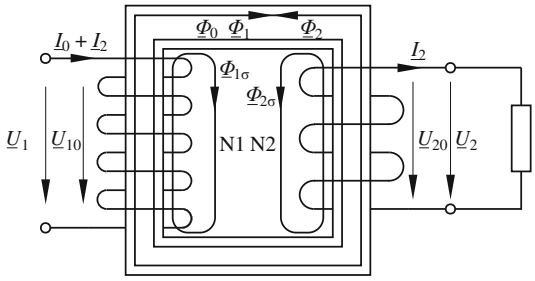


Abb. 86.8 Einphasen-Transformator bei Last

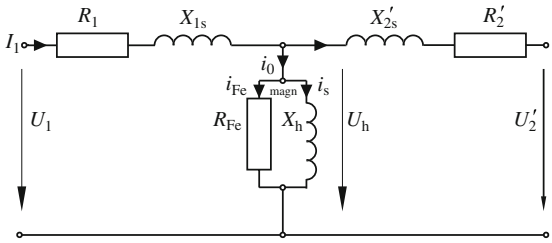


Abb. 86.9 Vollständiges Ersatzschaltbild des Einphasen-Transformators

Übersetzungsverhältnis

$$\ddot{u} = \frac{U_{10}}{U_{20}} = \frac{N_1}{N_2} \tag{86.3}$$

Die beiden Spannungen sind phasengleich; ihre Beträge verhalten sich wie die Windungszahlen. Das Verhältnis der beiden Leerlaufspannungen U_{10} und U_{20} zueinander bezeichnet man als *Übersetzungsverhältnis* $\ddot{u}$ des Transformators. Der Streufluss $\Phi_{1\sigma}$ ist nur mit der Primärspule verkettet und induziert dort die *Selbstinduktionsspannung*.

Selbstinduktionsspannung

$$U_{1\sigma} = j4{,}44 \cdot f \cdot N_1 \cdot \Phi_{1\sigma} \qquad \begin{array}{c|c|c|c} U_{10} & f & N_1 & \Phi_{1\sigma} \\ \hline \mathrm{V} & \mathrm{Hz} & 1 & \mathrm{V\,s} \end{array} \tag{86.4}$$

Hieraus lässt sich die *Streuinduktivität* und die *Streureaktanz* berechnen.

Streuinduktivität

$$L_{1\sigma} = N_1 \cdot \frac{\Phi_{1\sigma}}{I_1} \qquad \begin{array}{c|c|c|c} L_{1\sigma} & N_1 & \Phi_{1\sigma} & I_1 \\ \hline \frac{\mathrm{V\,s}}{\mathrm{A}} & 1 & \mathrm{V\,s} & \mathrm{A} \end{array} \tag{86.5}$$

Streureaktanz

$$X_{1\sigma} = \omega \cdot L_{1\sigma} \qquad \begin{array}{c|c|c} X_{1\sigma} & \omega & L_{1\sigma} \\ \hline \Omega & \mathrm{Hz} & \frac{\mathrm{V\,s}}{\mathrm{A}} \end{array} \tag{86.6}$$

Die Spannungsgleichung für den leerlaufenden Transformator lautet (Abb. 86.9):

Primärspannung

$$\underline{U}_1 = (R_1 + jX_{1\sigma}) \cdot \underline{I}_1 + \underline{U_{1h}} \tag{86.7}$$

Sekundärspannung

$$\underline{U}_2 = \underline{U_{20}} \tag{86.8}$$

R_1 ist der ohmsche Widerstand der Primärwicklung. Durch das ständige Ummagnetisieren des Eisenkerns entstehen dort Verluste. Diese Verluste berücksichtigt man durch einen Ersatzwiderstand R_{Fe}, der parallel zur Hauptinduktivität L_h angeordnet und vom Strom I_{Fe} durchflossen wird. Die in dem Widerstand R_{Fe} erzeugte Verlustleistung P_{Fe} wird auch als *Leerlaufleistung* P_0 bezeichnet, da der Leerlaufstrom sehr gering ist und die Verluste am Widerstand R_1 in diesem Fall vernachlässigbar klein sind.

Leerlaufspannung

$$U_{10} = R_{\mathrm{Fe}} \cdot I_{\mathrm{Fe}} \tag{86.9}$$

Eisenverlustleistung (*Leerlaufleistung*)

$$P_{\mathrm{Fe}} = \frac{U_{10}^2}{R_{\mathrm{Fe}}} \qquad \begin{array}{c|c|c} P_{\mathrm{Fe}} & U_{10} & R_{\mathrm{Fe}} \\ \hline \mathrm{W} & \mathrm{V} & \Omega \end{array} \tag{86.10}$$

86.1.3.2 Belastung

Wird an den Klemmen der Sekundärspule ein Belastungswiderstand angeschlossen (Abb. 86.8), fließt durch die Sekundärspule ein Strom I_2.

Dieser Strom erzeugt am Wirkwiderstand R_2 der Sekundärspule und an der Streuinduktivität X_{2s} einen Spannungsfall. Die Spannungsgleichungen des belasteten Transformators lauten (Abb. 86.9):

Primärspannung

$$\underline{U}_1 = (R_1 + jX_{1\sigma}) \cdot \underline{I}_1 + \underline{U_{1h}} \tag{86.11}$$

Sekundärspannung

$$\underline{U}_2 = -(R_2 + jX_{2\sigma}) \cdot \underline{I}_2 + \underline{U_{20\sigma}} \tag{86.12}$$

Der Strom I_2 verursacht eine magnetische Durchflutung $\Theta_2 = N_2 \cdot I_2$. Hierdurch wird ein magnetischer Fluss Φ_2 erzeugt, der entgegen Φ_1 gerichtet ist. Nach dem Durchflutungsgesetz gilt:

resultierende Durchflutung

$$\Theta_\mu = \Theta_1 - \Theta_2 \qquad \begin{array}{c|c|c} \Theta_\mu & \Theta_1 & \Theta_2 \\ \hline \mathrm{A} & \mathrm{A} & \mathrm{A} \end{array} \qquad (86.13)$$

Durchflutung im Primärkreis

$$\underline{\Theta_1} = N_1(\underline{I_1} - \underline{I_{\mathrm{Fe}}}) \qquad (86.14)$$

Diese Formel gilt für alle Belastungsfälle, muss also ausgehend vom Leerlauf über verschiedene Belastungsfälle bis zum Kurzschluss Gültigkeit haben. Es folgt für den Leerlauf:

Leerlauf Durchflutung

$$\underline{\Theta_\mu} = N_1 \cdot \underline{I_\mu} = \underline{\Theta_{10}} = N_1 \cdot (\underline{I_0} - \underline{I_{\mathrm{Fe}}}) \qquad (86.15)$$

Im Belastungsfall mit I_2 nimmt der Transformator aus dem Netz einen Strom I_1 auf, der die Gegendurchflutung $\underline{\Theta_2} = N_2 \cdot \underline{I_2}$ kompensiert. Für den Primärstrom ergibt sich aus dem Durchflutungsgesetz nach Division durch N_1:

Primärstrom

$$\underline{I_1} = \underline{I_\mu} + \underline{I_{\mathrm{Fe}}} + \frac{N_2}{N_1} \cdot \underline{I_2} \qquad (86.16)$$

Für große Leistungstransformatoren ergibt ein Vergleich der Magnetisierungsdurchflutung Θ_μ mit der Durchflutung Θ_{2N}, dass die durch den sekundären Nennstrom hervorgerufene Durchflutung viel größer ist als die Magnetisierungsdurchflutung. Damit gilt:

$$\underline{\Theta_1} = \underline{\Theta_\mu} + \underline{\Theta_2} \approx \underline{\Theta_2} \qquad (86.17)$$

$$\frac{I_1}{I_2} = \frac{N_2}{N_1} = \frac{1}{\ddot{u}} \qquad (86.18)$$

Die Ströme verhalten sich also umgekehrt zu ihren Windungszahlen. Geht man von einem idealen Transformator ohne Wirkverluste und Streuinduktivitäten aus, so ist die Scheinleistung, die auf der Primärseite vom Transformator aufgenommen wird, gleich der sekundären Scheinleistung.

$$S_1 = U_1 \cdot I_1 = S_2 = U_2 \cdot I_2 \qquad (86.19)$$

Mit dieser Annahme ergeben sich folgende Abhängigkeiten:

$$\frac{U_1}{U_2} = \frac{I_2}{I_1} \qquad (86.20)$$

$$\underline{U_1^2} \cdot Z_1 = \underline{U_2^2} \cdot Z_2 \qquad (86.21)$$

$$\frac{\underline{U_1^2}}{\underline{U_2^2}} = \frac{Z_1}{Z_2} = \frac{N_1^2}{N_2^2} = \ddot{u}^2 \qquad (86.22)$$

Die Widerstände der Sekundärseite verhalten sich mit dem Quadrat der Windungszahl umgekehrt zu denen der Primärseite. Berechnungen am Transformator oder das Zeichnen von Zeigerbildern erfordert einen großen Aufwand, da Primär- und Sekundärseite normalerweise eine andere Spannungsebene haben. Zeigerbilder müssten einen anderen Maßstab erhalten, die

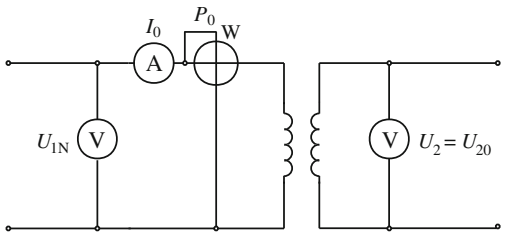


Abb. 86.10 Schaltung im Leerlaufversuch mit Messanordnung

Berechnung muss umständlich ausgeführt werden. Dieses umgeht man, indem beide Seiten des Transformators auf eine Spannungsebene, die der Primärspannung oder die der Sekundärspannung, umgerechnet werden. Die Umrechnung der Sekundärgrößen erfolgt mit dem Übersetzungsverhältnis $\ddot{u}$. Die umgerechneten Größen werden mit einem „′“ gekennzeichnet. Durch die Umrechnung erhält man folgende Größen:

$$U_2' = \ddot{u} \cdot U_2 \qquad (86.23)$$

$$I_2' = \frac{1}{\ddot{u}} \cdot I_2 \qquad (86.24)$$

$$R_2' = \ddot{u}^2 \cdot R_2 \qquad (86.25)$$

$$X_{2\sigma}' = \ddot{u}^2 \cdot X_{2\sigma} \qquad (86.26)$$

Mit Hilfe der umgerechneten Größen kann das Ersatzschaltbild und das Zeigerdiagramm so gezeichnet werden (Abb. 86.9), als handele es sich um einen Stromkreis mit einer Bezugsspannung. Die Größen in dem Ersatzschaltbild können durch zwei Versuche, *Kurzschluss-* und *Leerlaufversuch*, bestimmt werden.

86.1.3.3 Leerlaufversuch

Der Leerlaufversuch wird mit offenen Sekundär- oder Primärklemmen gefahren (Abb. 86.10). Bei offenen Primärklemmen wird die Nennspannung U_{2N} an die Sekundärklemmen angelegt und der sich einstellende Strom I_0, die Leistung P_0 und der Leistungsfaktor $\cos\varphi_0$ gemessen.

Mit den gemessenen Werten können, wenn sie auf Strangwerte umgerechnet sind, die *Hauptinduktivität* L_h und der fiktive *Eisenverlustwiderstand* R_{Fe} bestimmt werden. Dabei werden der Wirkwiderstand der Kupferwicklung und die induktive Streureaktanz vernachlässigt.

Eisenverluststrom $I_{\mathrm{Fe}} = I_{10} \cdot \cos\varphi_0$ (86.27)

Magnetisierungsstrom $I_{\mathrm{magn}} = I_{10} \cdot \sin\varphi_0$ (86.28)

Eisenverlustwiderstand $R_{\mathrm{Fe}} = \dfrac{U_{10}}{I_{\mathrm{Fe}}}$ (86.29)

Blindwiderstand der Hauptinduktivität

$$X_h = \frac{U_{10}}{I_{\mathrm{magn}}} \qquad (86.30)$$

Wird der Leerlaufversuch nicht mit Nennspannung durchgeführt, muss der gemessene Leerlaufstrom I_{10} auf den eigentlichen Nennwert umgerechnet werden.

$$\textit{Nenn-Leerlaufstrom} \quad I_{10N} = I_{10} \cdot \left(\frac{U_{1N}}{U_{10}}\right) \tag{86.31}$$

Da der im Leerlauf fließende Strom im Verhältnis zum Nennstrom sehr gering ist, und um Vergleiche mit anderen Transformatoren herstellen zu können, wird für Transformatoren das *relative Leerlaufstromverhältnis* i_0 angegeben.

relatives Leerlaufstromverhältnis

$$i_0 = \left(\frac{I_{0N}}{I_{1N}}\right) \cdot 100\,\% \tag{86.32}$$

i_0	I_{0N}	I_{1N}
%	A	A

Die im Leerlaufversuch gemessene Wirkleistung P_0 ist identisch mit der am Eisenverlustwiderstand erzeugten Leistung P_{Fe}. Findet der Leerlaufversuch bei Nennbedingungen statt, so gilt:

Eisenverlustleistung

$$P_{\mathrm{Fe}} = P_{10N} = U_{1N} \cdot I_{10N} \cdot \cos\varphi_0 \tag{86.33}$$

P_{Fe}	P_{10N}	U_{1N}	I_{10N}
W	W	V	A

Kann der Versuch nicht unter Nennbedingungen durchgeführt werden, muss die gemessene Leistung auf den Nennwert hochgerechnet werden.

$$P_{0N} = P_0 \cdot \left(\frac{U_{1N}^2}{U_{10}^2}\right) \tag{86.34}$$

Beispiel

Bei einem Einphasen-Transformator sind das relative Leerlaufstromverhältnis $i_0 = 2\,\%$ und der Leistungsfaktor im Leerlauf $\cos\varphi_0 = 0{,}1$ bekannt. Die Nenndaten des Transformators sind $U_N = 60/10\,\mathrm{kV}$; $S_N = 1\,\mathrm{MV\,A}$. Wie groß ist der Eisenverlustwiderstand und die Hauptinduktivität bei $f = 50\,\mathrm{Hz}$?

$$I_{1N} = \frac{S_N}{U_{1N}} = \frac{1\,\mathrm{MV\,A}}{60\,\mathrm{kV}} = 16{,}67\,\mathrm{A}$$

$$I_{10N} = \frac{i_0 \cdot I_{1N}}{100\,\%} = \frac{2\,\% \cdot 16{,}67\,\mathrm{A}}{100\,\%} = 0{,}333\,\mathrm{A}$$

Über den Leistungsfaktor können die Ströme durch den Eisenverlustwiderstand und die Hauptinduktivität berechnet werden.

$$I_{\mathrm{Fe}} = I_0 \cdot \cos\varphi_0 = 16{,}67\,\mathrm{A} \cdot 0{,}1 = 1{,}667\,\mathrm{A}$$

$$I_{\mathrm{magn}} = I_0 \cdot \sin\varphi_0 = 16{,}67\,\mathrm{A} \cdot 0{,}99 = 16{,}58\,\mathrm{A}$$

$$R_{\mathrm{Fe}} = \frac{U_{1N}}{I_{\mathrm{Fe}}} = \frac{60\,\mathrm{kV}}{1{,}667\,\mathrm{A}} = 35{,}99\,\mathrm{k\Omega}$$

$$X_H = \frac{U_{1N}}{I_{\mathrm{magn}}} = \frac{60\,\mathrm{kV}}{16{,}58\,\mathrm{A}} = 3{,}61\,\mathrm{k\Omega}$$

$$L_H = \frac{X_H}{2 \cdot \pi \cdot f} = \frac{3610\,\Omega}{2 \cdot \pi \cdot 50\,\frac{1}{\mathrm{s}}} = 11{,}49\,\frac{\mathrm{V\,s}}{\mathrm{A}}$$ ◄

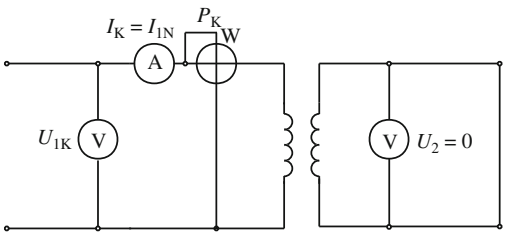


Abb. 86.11 Schaltung im Kurzschlussversuch mit Messanordnung

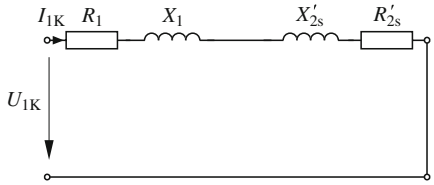


Abb. 86.12 Ersatzschaltbild des Kurzschlussversuchs

86.1.3.4 Kurzschlussversuch

Beim Kurzschlussversuch wird eine Trafoseite kurzgeschlossen und an die andere Trafoseite eine variable Spannung angeschlossen (Abb. 86.11).

Die Spannung wird hochgeregelt, bis im Transformator Nennstrom fließt. Gemessen werden die so eingestellte Spannung U_k, die aufgenommene Wirkleistung P_k und der Leistungsfaktor $\cos\varphi_k$.

Unter Vernachlässigung des Magnetisierungsstroms und der Eisenverluste ($I_0 \ll I_N$) kann das Ersatzschaltbild des Transformators wie im Abb. 86.12 dargestellt werden. Die Größen R_1; R'_2; X_{S1}; X'_{S2} werden durch Rechnung ermittelt.

$$\textit{Kurzschlussimpedanz} \quad Z_k = \frac{U_k}{I_k} \tag{86.35}$$

$$\textit{Kurzschlusswiderstand} \quad R_k = Z_k \cdot \cos\varphi_k \tag{86.36}$$

$$\textit{Kurzschlussblindwiderstand} \quad X_k = Z_k \cdot \sin\varphi_k \tag{86.37}$$

Bei Leistungstransformatoren gilt:

$$R_k = R_1 + R'_2 \tag{86.38}$$

$$R_1 = R'_2 = R_2 \cdot \ddot{u}^2 \tag{86.39}$$

Widerstand der Primärwicklung

$$R_1 = \frac{R_k}{2} \tag{86.40}$$

Widerstand der Sekundärwicklung

$$R_2 = \frac{R_k}{2 \cdot \ddot{u}^2} \tag{86.41}$$

$$X_k = X_{S1} + X'_{S2} \tag{86.42}$$

$$X_{S1} = X'_{S2} = X_{S2} \cdot \ddot{u}^2 \tag{86.43}$$

Streublindwiderstand im Primärkreis

$$X_{S1} = \frac{X_k}{2} \tag{86.44}$$

Streublindwiderstand im Sekundärkreis

$$X_{S2} = \frac{X_k}{2 \cdot \ddot{u}^2} \tag{86.45}$$

Erfolgt die Messung nicht unter Nennbedingungen ($I_k \neq I_N$), müssen die ermittelten Werte auf Nennwerte umgerechnet werden.

Nenn-Kurzschlussspannung
$$U_{kN} = U_k \cdot \left(\frac{I_{kN}}{I_k}\right) \tag{86.46}$$

Nenn-Kurzschlussverluste
$$P_{kN} = P_k \cdot \left(\frac{I_{kN}}{I_k}\right)^2 \tag{86.47}$$

Die im Kurzschlussversuch ermittelte Verlustleistung P_k ist identisch mit der in den Wicklungen erzeugten *Kupferverlustleistung* P_{Cu}. Findet der Kurzschlussversuch mit Nennstrom statt, gilt:

Kupferverluste
$$P_{Cu} = P_{1kN} = U_{1kN} \cdot I_{1N} \cdot \cos\varphi_k \tag{86.48}$$

Ferner wird aus den Messergebnissen die *relative Kurzschlussspannung* u_k bestimmt. Die relative Kurzschlussspannung gibt den Spannungsfall am Transformator bei Nennbetrieb an und liegt zwischen 4 % und 16 %.

relative Kurzschlussspannung

$$u_k = \left(\frac{U_{kN}}{U_N}\right) \cdot 100\,\% \qquad \begin{array}{c|c|c} u_k & U_{kN} & U_N \\ \hline \% & \mathrm{V} & \mathrm{V} \end{array} \tag{86.49}$$

Tritt am Transformator bei Nennbetrieb ein sekundärseitiger Kurzschluss auf, kann der hierbei fließende *Dauerkurzschlussstrom* I_D über die relative Kurzschlussspannung bestimmt werden.

Dauerkurzschlussstrom

$$I_D = \frac{I_N}{u_k} \cdot 100\,\% \qquad \begin{array}{c|c|c} I_D & I_N & u_k \\ \hline \mathrm{A} & \mathrm{A} & \% \end{array} \tag{86.50}$$

Erfolgt der Kurzschlussversuch nicht bei Transformatornenntemperatur von $\vartheta = 75\,°\mathrm{C}$, muss die ermittelte Kupferverlustleistung auf Nenntemperatur umgerechnet werden.

Verlustleistung bei Nenntemperatur

$$P_{Cu\,75} = P_{Cu\,K} \cdot \frac{235 + 75}{235 + \vartheta_K} \qquad \begin{array}{c|c|c} P_{Cu\,75} & P_{Cu\,K} & \vartheta_K \\ \hline \mathrm{W} & \mathrm{W} & °\mathrm{C} \end{array} \tag{86.51}$$

Beispiel

Bei einem Einphasen-Transformator sind die relative Kurzschlussspannung $u_k = 4\,\%$ und der Leistungsfaktor im Kurzschluss $\cos\varphi_k = 0{,}5$ bekannt. Die Nenndaten des Transformators sind $U_N = 60/10\,\mathrm{kV}$; $S_N = 1\,\mathrm{MV\,A}$. Wie groß sind die Widerstände der Wicklungen, welchen Wert haben die Streuinduktivitäten bei $f = 50\,\mathrm{Hz}$? Wie hoch ist der im Kurzschlussfall sekundärseitig maximal fließende Strom?

$$I_{1N} = \frac{S_N}{U_{1N}} = \frac{1\,\mathrm{MV\,A}}{60\,\mathrm{kV}} = 16{,}67\,\mathrm{A}$$
$$U_{kN} = \frac{u_k \cdot U_N}{100\,\%} = \frac{4\,\% \cdot 60\,\mathrm{kV}}{100\,\%} = 2400\,\mathrm{V}$$
$$Z_k = \frac{U_{kN}}{I_N} = \frac{2400\,\mathrm{V}}{16{,}67\,\mathrm{A}} = 144\,\Omega$$

Über den Leistungsfaktor $\cos\varphi_k$ können die Widerstände und Streuinduktivitäten berechnet werden.

$$R_k = Z_k \cdot \cos\varphi_k = 144\,\Omega \cdot 0{,}5 = 72\,\Omega$$
$$X_k = Z_k \cdot \sin\varphi_k = 144\,\Omega \cdot 0{,}866 = 124{,}7\,\Omega$$
$$R_1 = \frac{R_k}{2} = \frac{72}{2}\,\Omega = 36\,\Omega;$$
$$R'_2 = \frac{R_k}{2} = \frac{72}{2}\,\Omega = 36\,\Omega$$
$$R_2 = \frac{R'_2}{\ddot{u}^2} = \frac{36\,\Omega}{(\frac{60\,\mathrm{kV}}{10\,\mathrm{kV}})^2} = 1\,\Omega$$
$$X_{1\sigma} = \frac{X_k}{2} = \frac{124{,}7}{2}\,\Omega = 62{,}35\,\Omega$$
$$X'_{2\sigma} = \frac{X_k}{2} = \frac{124{,}7}{2}\,\Omega = 62{,}35\,\Omega$$
$$X_{2\sigma} = \frac{X'_{2\sigma}}{\ddot{u}^2} = \frac{62{,}35\,\Omega}{(\frac{60\,\mathrm{kV}}{10\,\mathrm{kV}})^2} = 1{,}73\,\Omega$$

Der bei diesem Trafo maximal fließende Strom berechnet sich zu

$$I_{D2} = \frac{I_{2N}}{u_k} \cdot 100\,\% = \frac{I_{1N} \cdot \ddot{u}}{u_k} \cdot 100\,\%$$
$$= \frac{16{,}67\,A \cdot 6}{4\,\%} \cdot 100\,\% = 2500{,}5\,\mathrm{A} \quad \blacktriangleleft$$

86.1.3.5 Wirkungsgrad

Der *Wirkungsgrad* eines Systems ergibt sich aus

Wirkungsgrad
$$\eta = \frac{P_{ab}}{P_{zu}} \tag{86.52}$$

Für den Transformator ergibt sich unter Berücksichtigung der Verluste:

zugeführte Wirkleistung

$$P_{zu} = P_{ab} + P_V = P_{ab} + P_{Fe} + P_{Cu} \tag{86.53}$$
$$\eta = \frac{P_{ab}}{P_{ab} + P_{Fe} + P_{Cu}} \tag{86.54}$$

Die Eisenverluste P_{Fe} werden messtechnisch im Leerlaufversuch bestimmt. Sie sind gleich den Nennverlusten im Leerlauf $P_{Fe} = P_{0N}$. Da die Leerlaufverluste nur von der angelegten Spannung abhängig sind, müssen die Verluste bei Spannungsänderung entsprechend umgerechnet werden.

$$P_{Fe}^{*} = P_{Fe} \cdot \left(\frac{U}{U_N}\right)^2 \tag{86.55}$$

P_{Fe}^{*} Eisenverluste bei beliebiger Spannung U.

Die Kupferverluste P_{Cu} sind identisch mit den Verlusten P_{kN}aus dem Kurzschlussversuch. Sie entstehen an den Wicklungen des Transformators und sind nur abhängig vom Strom. Entspricht der Strom nicht dem Nennstrom, müssen die Kupferverluste entsprechend umgerechnet werden.

$$P_{Cu}^{*} = P_{Cu} \cdot \left(\frac{I}{I_N}\right)^2 \tag{86.56}$$

P_{Cu}^{*} Kupferverluste bei beliebigem Strom I.

Der Wirkungsgrad h des Transformators kann jetzt für jeden Arbeitspunkt bestimmt werden.

$$\eta = \frac{P_{ab}}{P_{ab} + P_{Fe}^{*} + P_{Cu}^{*}} = \frac{P_{ab}}{P_{ab} + P_{Fe} \cdot \left(\frac{U}{U_N}\right)^2 + P_{Cu} \cdot \left(\frac{I}{I_N}\right)^2} \tag{86.57}$$

Der *optimale Wirkungsgrad* des Transformators ergibt sich, wenn die Eisenverluste gleich den Kupferverlusten sind. Für den Normalbetrieb kann davon ausgegangen werden, dass die Spannung U der Nennspannung entspricht und konstant ist ($P_{Fe}^{*} = P_{Fe}$). Der optimale Wirkungsgrad stellt sich ein, wenn gilt:

$$P_{Fe} = P_{Cu} \cdot \left(\frac{I}{I_N}\right)^2 \tag{86.58}$$

$$I = I_N \cdot \sqrt{\frac{P_{Fe}}{P_{Cu}}} = I_N \cdot \sqrt{a} \quad a = \frac{P_{Fe}}{P_{Cu}} \tag{86.59}$$

$$P_{ab} = P_N \cdot \sqrt{a} \tag{86.60}$$

Eine Aussage über den wirtschaftlichen Einsatz eines Transformators kann über den Wirkungsgrad nicht erfolgen, da der Trafo nicht ständig gleichmäßig belastet wird. Eine wirtschaftliche Betrachtung erfolgt mit dem *Jahreswirkungsgrad* η_a.

Jahreswirkungsgrad

$$\eta_a = \frac{W_{ab}}{W_{ab} + W_{Fe} + W_{Cu}} \tag{86.61}$$

elektrische Jahresarbeit

$$W_{ab} = P_{ab} \cdot t_B \qquad \begin{array}{c|c|c} W_{ab} & P_{ab} & t_B \\ \hline \mathrm{W\,h} & \mathrm{W} & \mathrm{h} \end{array} \tag{86.62}$$

Jahres-Leerlaufarbeit

$$W_{Fe} = P_{Fe} \cdot t_E \qquad \begin{array}{c|c|c} W_{Fe} & P_{Fe} & t_E \\ \hline \mathrm{W\,h} & \mathrm{W} & \mathrm{h} \end{array} \tag{86.63}$$

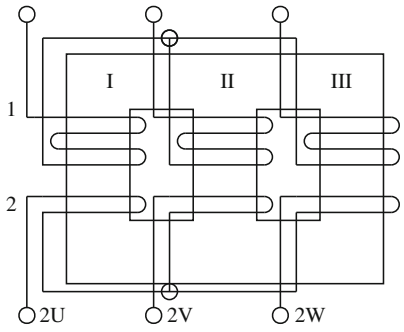


Abb. 86.13 Drehstrom-Transformator

Jahres-Wirkverlustarbeit

$$W_{Cu} = P_{Cu} \cdot t_B \qquad \begin{array}{c|c|c} W_{Cu} & P_{Cu} & t_B \\ \hline \mathrm{W\,h} & \mathrm{W} & \mathrm{h} \end{array} \tag{86.64}$$

t_B Betriebszeit; t_E Einschaltzeit.

Beispiel

Bei einem Einphasen-Transformator sind die Kurzschlussspannung $U_{1k} = 1500\,\mathrm{V}$ und der Leistungsfaktor im Kurzschluss $\cos\varphi_k = 0{,}5$ bekannt. Die Nenndaten des Transformators sind $U_N = 60/10\,\mathrm{kV}$; $S_N = 1\,\mathrm{MV\,A}$. Wie hoch ist der Wirkungsgrad im Nennbetrieb bei $\cos\varphi = 1$, wenn die Leerlaufverluste $P_{FeN} = 10\,\mathrm{kW}$ betragen und die Kurzschlussspannung bei $I_{1k} = 0{,}625 \cdot I_N$ ermittelt wurde?

$$I_{1N} = \frac{S_N}{U_{1N}} = \frac{1\,\mathrm{MV\,A}}{60\,\mathrm{kV}} = 16{,}67\,\mathrm{A}$$

$$P_{kN} = I_N \cdot U_{kN} \cdot \cos\varphi_k$$

$$= 16{,}67\,\mathrm{A} \cdot 1500\,\mathrm{V} \cdot \left(\frac{16{,}67\,\mathrm{A}}{0{,}625 \cdot 16{,}67\,\mathrm{A}}\right) \cdot 0{,}5$$

$$= 20\,\mathrm{kW}$$

$$\eta_N = \frac{P_{ab}}{P_{ab} + P_{FeN} + P_{CuN}}$$

$$= \frac{1\,\mathrm{MW}}{1\,\mathrm{MW} \cdot 10\,\mathrm{kW} \cdot 20\,\mathrm{kW}} = 0{,}97$$ ◀

86.1.4 Aufbau und Schaltung von Drehstrom-Transformatoren

86.1.4.1 Wirkungsweise

In Drehstromnetzen werden zum Transformieren von Spannungen Drehstrom-Transformatoren eingesetzt.

Bei den Drehstrom-Transformatoren werden drei Einphasen-Transformatoren auf einem gemeinsamen Eisenkern mit drei Schenkeln gewickelt (Abb. 86.13). Die Schenkel des Transformators sind, da sie in einer Ebene angeordnet werden,

Tab. 86.2 Kurzzeichen der Wicklungsschaltung

	OS	US
Dreieck	D	d
Stern	Y	y
Zickzack	–	z
Sternpunkt geerdet	N	n

unterschiedlich lang. Diese Anordnung führt zu unterschiedlichen Magnetisierungsströmen. Bei symmetrischer Last sind die Unterschiede so gering, dass der Drehstrom-Transformator in seinem Betriebsverhalten wie ein Einphasen-Transformator betrachtet werden kann. Im Drehstromtransformator können die Wicklungen im Dreieck oder im Stern geschaltet werden. Aus dieser Schaltungsart ergibt sich beim Berechnen der Größen des Transformators, dass sämtliche Rechnungen vorzugsweise im Strang erfolgen sollten.

Für die *Dreiecksschaltung* gilt:

$$S_{N\,\mathrm{Str}} = \frac{1}{3} \cdot S_N \tag{86.65}$$

$$U_{N\,\mathrm{Str}} = U_N \tag{86.66}$$

$$I_{N\,\mathrm{Str}} = \frac{1}{\sqrt{3}} \cdot I_N \tag{86.67}$$

Für die *Sternschaltung* gilt:

$$S_{N\,\mathrm{Str}} = \frac{1}{3} \cdot S_N \tag{86.68}$$

$$U_{N\,\mathrm{Str}} = \frac{1}{\sqrt{3}} \cdot U_N \tag{86.69}$$

$$I_{N\,\mathrm{Str}} = I_N \tag{86.70}$$

S_N Gesamtnennscheinleistung; $S_{N\,\mathrm{Str}}$ Nennscheinleistung eines Strangs (Phase); U_N Außenleiternennspannung; $U_{N\,\mathrm{Str}}$ Strangnennspannung; I_N Leiternennstrom; $I_{N\,\mathrm{Str}}$ Strangnennstrom.

86.1.4.2 Schaltgruppen

Beim Drehstromtransformator können die Wicklungen auf der Oberspannungsseite (primär) und auf der Unterspannungsseite (sekundär) unterschiedlich geschaltet werden. Neben der *Dreiecksschaltung* und der *Sternschaltung* wird auf der Unterspannungsseite die *Zickzackschaltung* eingesetzt. Die eingesetzten Schaltungsarten werden am Transformator durch Buchstaben kenntlich gemacht (Tab. 86.2).

Die gebräuchlichsten Schaltgruppen nach VDE 0532 sind in Abb. 86.14 dargestellt.

Die nach den Buchstaben stehende Zahl gibt die Phasenverschiebung der Unterspannungsseite gegen die Oberspannungsseite an. Die Verschiebung kann durch Multiplikation der nachgestellten Zahl mit dem Winkel 30° berechnet werden.

Beispiel

$$\mathrm{Yy0} \rightarrow 0 \cdot 30^\circ = 0^\circ; \quad \mathrm{Dy5} \rightarrow 5 \cdot 30^\circ = 150^\circ \tag{86.71}$$

◀

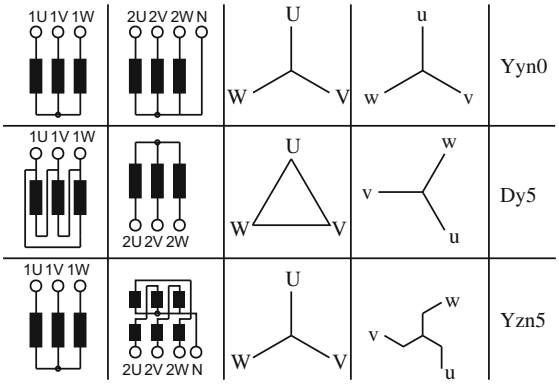


Abb. 86.14 Auswahl von Schaltgruppen

86.1.4.3 Unsymmetrische Belastungen

Jeder Transformator muss sowohl im elektrischen als auch im magnetischen Gleichgewicht sein, d. h., die in den einzelnen Schenkeln des Transformators auftretenden Flüsse müssen sich immer zum Fluss im Leerlauf ergänzen. Unsymmetrische Belastungen führen bei einigen Schaltungsarten zu Streuflüssen, die das Transformatorgehäuse unzulässig erwärmen können. Transformatoren mit den Schaltgruppen Yyn, Dyn, Yzn und Dzn können unsymmetrische Lasten problemlos übertragen.

86.1.5 Parallelschalten von Transformatoren

Sollen mehrere Transformatoren parallelgeschaltet werden, müssen folgende Bedingungen nach VDE 0532 eingehalten werden:

Übersetzungsverhältnis muss gleich sein und Schaltgruppen müssen passen, sonst treten Ausgleichsströme im Leerlauf auf.

Kurzschlussspannungen gleich bei maximaler Abweichung von 10 % ($u_{k1} \cong u_{k2}$), sonst erfolgt eine unterschiedliche Lastaufteilung.

Nennleistungsverhältnis nicht kleiner als 1 : 3.

Die Zusammenschaltung zweier Transformatoren mit unterschiedlichen Schaltgruppen kann nach VDE 0532 Teil 10 erfolgen. Sind zwei Transformatoren mit unterschiedlichen Spannungen zusammengeschaltet (Stufenschalter nicht auf gleicher Stellung), fließt zwischen den Transformatoren ein Ausgleichsstrom folgender Höhe:

Ausgleichsstrom

$$I_a = \frac{\Delta u}{\left(\left(\frac{u_{k1}}{I_{1N}}\right) + \left(\frac{u_{k2}}{I_{2N}}\right)\right)} \tag{86.72}$$

Δu	u_{K1}	I_a	I_{1N}
%	%	A	A

Δu [%] Abweichung der Trafospannungen U_1, U_2.

Dieser Ausgleichsstrom fließt schon im Leerlauf und erzeugt Verluste. Außerdem reduziert er die übertragbare Leistung der

Transformatoren, da er sich ebenfalls zum Laststrom addiert. Die Parallelschaltung von Transformatoren mit unterschiedlicher Kurzschlussspannung u_K führt dazu, dass der Transformator mit der kleineren Kurzschlussspannung einen prozentual höheren Anteil der Last übernehmen muss. Die Lastaufteilung wird wie folgt berechnet:

rel. Ersatzkurzschlussspannung

$$u_{k\,\mathrm{ers}} = \frac{S_{N\,\mathrm{ges}}}{\{(\frac{S_{N1}}{u_{k1}}) + (\frac{S_{N2}}{u_{k2}}) + (\frac{S_{N3}}{u_{k3}})\}} \qquad \begin{array}{c|c|c} u_{k\,\mathrm{ers}} & u_{k1} & S_N \\ \% & \% & \mathrm{V\,A} \end{array} \qquad (86.73)$$

$$S_1 = S_{N1} \cdot \frac{(u_{k\,\mathrm{ers}} \cdot S_{\mathrm{ges}})}{(u_{k1} \cdot S_{N\,\mathrm{ges}})} \qquad (86.74)$$

$u_{k\,\mathrm{ers}}$ relative Ersatzkurzschlussspannung der Parallelschaltung, $S_{N\,\mathrm{ges}}$ Summe der Trafonennleistungen $S_1 + S_2 + \ldots$, S_{ges} tatsächlich abgenommene Scheinleistung der Parallelschaltung.

Beispiel

Die Transformatoren T1 mit $U_N = 10/0{,}4\,\mathrm{kV}$, $S_N = 630\,\mathrm{kV\,A}$, $u_k = 4\,\%$ und T2 mit $U_N = 10/0{,}4\,\mathrm{kV}$, $S_N = 1000\,\mathrm{kV\,A}$ und $u_k = 4{,}5\,\%$ werden parallelgeschaltet. Alle weiteren Voraussetzungen einer Parallelschaltung sind gegeben. Welche maximale Last darf diese Parallelschaltung übertragen, ohne dass einer der Transformatoren überlastet wird?

Ermittlung der relativen Ersatzkurzschlussspannung:

$$u_{k\,\mathrm{ers}} \frac{S_{N\,\mathrm{ges}}}{\{(\frac{S_{N1}}{u_{k1}}) + (\frac{S_{N2}}{u_{k2}})\}} = \frac{1630\,\mathrm{kV\,A}}{(\frac{630\,\mathrm{kV\,A}}{4\,\%} + \frac{1000\,\mathrm{kV\,A}}{4{,}5\,\%})} = 4{,}29\,\%$$

Da keiner der Transformatoren überlastet werden darf, muss das Produkt $\frac{(u_{k\,\mathrm{ers}} \cdot S_{\mathrm{ges}})}{(u_k \cdot S_{N\,\mathrm{ges}})}$ mit der kleinsten relativen Kurzschlussspannung den Wert 1 ergeben.

Scheinleistung

$$S_{\mathrm{ges}} = \frac{(u_{k1} \cdot S_{N\,\mathrm{ges}})}{u_{k\,\mathrm{ers}}} = \frac{4\,\% \cdot 1630\,\mathrm{kV\,A}}{4{,}29\,\%} = 1519{,}8\,\mathrm{kV\,A}$$

◀

86.1.6 Transformatorschutz

Der Schutz von Transformatoren kann über vorgeschaltete Schutzeinrichtungen (Sicherungen, elektronische Schutzrelais) oder über direkt am Transformator eingebaute Schutzgeräte erfolgen. Zu diesen im Trafo eingebauten Schutzgeräten gehört der *Buchholzschutz* ebenso wie eine *Temperaturüberwachung*. Der Buchholzschutz ist eine sehr einfache, aber wirksame Schutzmethode. Treten an den Isolationen der Spulen Schwachstellen auf und kommt es dadurch zu Windungsschlüssen, so entstehen als Folgewirkung dieser Schlüsse Teilentladungen, die Zersetzungsgase produzieren. Diese Gase werden im Buchholzschutz aufgefangen. Ab einer bestimmten Gasmenge erfolgt über Schwimmerkontakte eine Meldung oder die Abschaltung des Transformators. Ein innerer Lichtbogen (Kurzschluss) hat zur Folge, dass eine starke Ölströmung im Transformator auftritt. Diese Ölströmung bewirkt am Stauschieber im Buchholzrelais eine Kontaktgabe. Der Transformator wird sofort abgeschaltet. Die Temperaturmessung verhindert bei Wechselbelastungen eine thermische Zerstörung des Transformators, indem die jeweils zulässige Temperatur über Grenzkontakte eingestellt wird. Bei einer Überschreitung dieser Temperatur wird der Transformator abgeschaltet.

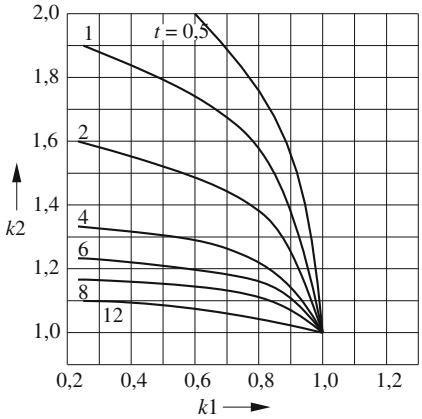


Abb. 86.15 k1 und k2 für Öltransformatoren bei 20 °C Umgebungstemperatur

86.1.7 Überlastung von Transformatoren

Transformatoren sind so ausgelegt, dass sie ihre Nenndaten bei Dauerbetrieb über 24 h und einer bestimmten Kühlmitteltemperatur (normalerweise 20 °C) ohne Schädigung der Isolation und anderer Betriebsgeräte aufrechterhalten können. Für den Notfallbetrieb können Transformatoren überlastet werden, d. h., sie können kurzfristig eine Leistung höher als die Nennleistung abgeben (VDE 0536 oder Herstellerangaben). Erfolgt dieser Notbetrieb nach einer Nennbelastung, muss der Betreiber eine Verkürzung der Lebensdauer des Transformators in Kauf nehmen. Erfolgt eine Überlastung nicht aus dem Nennbetrieb heraus, kann anhand von Diagrammen die mögliche Überlast, die nicht zur Verminderung der Lebensdauer führt, bestimmt werden.

Zur Ermittlung der möglichen Überlast S_2 muss die vorausgegangene Belastung des Transformators S_1 für $(24 - t)$ Stunden und die Nennleistung bekannt sein. Jetzt kann aus (86.75) und (86.76) die mögliche Überlast mit der maximalen Überlastzeit als Parameter aus der Abb. 86.15 ermittelt werden.

Faktor der Vorbelastung $$K_1 = \frac{S_1}{S_N} \qquad (86.75)$$

Faktor der Überlast $$K_2 = \frac{S_2}{S_N} \qquad (86.76)$$

Sollte der Transformator mit Überlast betrieben werden, muss der Strom auf das 1,5fache begrenzt werden (Notfallbetrieb: 2fache des Nennstroms).

Beispiel

Wie lange kann ein Drehstromöltransformator mit $S_N = 1000\,\text{kV A}$ eine Überlast von 200 kVA übernehmen, wenn der Transformator mit einer Dauerlast von $S = 800\,\text{kV A}$ betrieben wird?

$$K_1 = \frac{S_1}{S_N} = \frac{800\,\text{kV A}}{1000\,\text{kV A}} = 0{,}8$$

$$K_2 = \frac{S_2}{S_N} = \frac{1200\,\text{kV A}}{1000\,\text{kV A}} = 1{,}2$$

Nach Abb. 86.15 ergibt sich eine Überlastzeit von ungefähr 5 h. ◀

86.1.8 Aufstellen von Transformatoren

Beim Aufstellen von Transformatoren ist darauf zu achten, dass sie leicht zugänglich sind, die Sicherheitsabstände eingehalten werden, die Austauschbarkeit gewährleistet ist, eine ausreichende Belüftung oder Klimatisierung vorhanden ist und die Geräuschentwicklung berücksichtigt wird.

Bei Transformatoren mit Isolierflüssigkeiten als Kühlmittel sind Auffangwannen oder Gruben vorzusehen, die die gesamte Isolierflüssigkeit und auftretendes Wasser aufnehmen können (Pumpensumpf vorsehen). Bei Transformatoren mit mehr als 1000 l brennbarer Isolierflüssigkeit ist zwischen Transformator und Grube eine Schottervorlage anzuordnen. Nähere Angaben im AGI Blatt J 11 (Arbeitsgemeinschaft Industriebau e. V.).

86.1.9 Sondertransformatoren

Zu den Sondertransformatoren zählen diejenigen Transformatoren, die aufgrund ihrer Bauart besondere Eigenschaften aufweisen oder nicht alle Anforderungen eines Transformators erfüllen.

86.1.9.1 Spartransformatoren

Bei *Spartransformatoren* weisen Ober- und Unterspannungswicklung einen gemeinsamen Teil auf, der als *Parallelwicklung* bezeichnet wird (Abb. 86.16, 86.17).

Der weitere Anteil der Wicklung ist auf die Oberspannung beschränkt und heißt *Reihenwicklung*. Beide Wicklungsteile sind in Reihe geschaltet und sitzen auf einem gemeinsamen Eisenkern. Die elektrische Leistung wird beim Spartransformator nicht nur magnetisch übertragen, sondern je nach Übersetzungsverhältnis auch direkt über die galvanische Verbindung. Hieraus

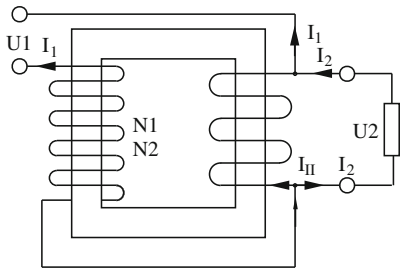


Abb. 86.16 Spartransformator

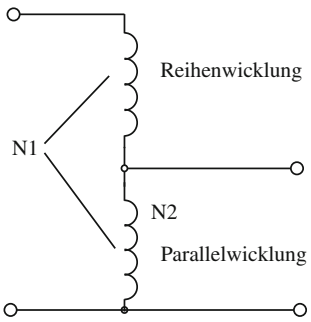


Abb. 86.17 Ersatzschaltbild eines Spartransformators

resultieren zwei Leistungsbegriffe: die *Durchgangsleistung* S_{ND} und die *Typen-* oder *Eigenleistung* S_{NE}. Für den Einphasenspartransformator gilt:

Durchgangsleistung

$$S_{ND} = U_{1N} \cdot I_{1N} \qquad \begin{array}{c|c|c} S_{ND} & U_{1N} & I_{1N} \\ \hline \text{VA} & \text{V} & \text{A} \end{array} \qquad (86.77)$$

Eigenleistung

$$S_{NE} = (U_{1N} - U_{2N}) \cdot I_{1N} \qquad (86.78)$$

Die Eigenleistung ist derjenige Leistungsanteil, der über den magnetischen Kreis übertragen werden muss. Der Vorteil des Spartransformators gegenüber einem Volltransformator ist, dass durch die kleinere Eigenleistung erhebliche Materialeinsparungen erzielt werden. Ein Nachteil ist, dass die Ober- und Unterspannungsseite galvanisch miteinander gekoppelt sind. Eingesetzt werden Spartransformatoren als Kuppeltransformatoren zwischen dem 380 kV- und dem 220-kV-Netz, um Spannungen nach längeren Leitungen hochzutransformieren, als Anlasstransformator für Asynchronmotoren und als Vorschaltgerät für Natrium-Dampflampen.

Beispiel

Ein Spartransformator soll in einem 110-kV-Drehstromnetz den Spannungsfall der Leitung von $U_L = 1\,\text{kV}$ ausgleichen. Die zu übertragende Leistung beträgt 10 MV A. Der Spannungsfall auf der Leitung ist als Strangspannung

zu betrachten. Der Spartransformator wird normalerweise in den Strang als Einphasen-Transformator eingeschaltet. Zur Berechnung der Durchgangsleistung ist deshalb die Strangspannung des Netzes zu berechnen.

Strangspannung

$$U_{\mathrm{Str}} = \frac{U_N}{\sqrt{3}} = \frac{110\,\mathrm{kV}}{\sqrt{3}} = 63{,}51\,\mathrm{kV}$$

Nennstrom

$$I_N = \frac{S_N}{\sqrt{3}\cdot U_N} = \frac{10\,\mathrm{MV\,A}}{\sqrt{3}\cdot 110\,\mathrm{kV}} = 52{,}5\,\mathrm{A}$$

Durchgangsleistung

$$S_{ND} = 63{,}51\,\mathrm{kV}\cdot 52{,}5\,\mathrm{A} = 3{,}34\,\mathrm{MV\,A}$$

Eigenleistung

$$S_{NE} = \Delta U\cdot I_{1N} = 1\,\mathrm{kV}\cdot 52{,}5\,\mathrm{A} = 52{,}5\,\mathrm{kV\,A}$$ ◀

86.1.9.2 Drosselspulen

Drosselspulen werden häufig als Reihendrosselspulen zur Begrenzung von Kurzschlussströmen eingesetzt. Die Drosselspulen werden als kurzschlussstrombegrenzende Spulen stets ohne Eisenkern gebaut, damit die im Kurzschluss auftretenden hohen Feldstärken keine Sättigungserscheinungen hervorrufen. Gekennzeichnet wird die Drosselspule durch ihren Spannungsfall im Nennbetrieb $\Delta U_{N\,\mathrm{Str}}$, den Nennstrom I_N, die Eigenleistung und die Bauleistung.

Eigenleistung $$S_E = 3\cdot\Delta U_{N\,\mathrm{Str}}\cdot I_N \qquad (86.79)$$

Bauleistung $$S_D = \sqrt{3}\cdot U_N\cdot I_N \qquad (86.80)$$

Übliche Baugrößen liegen bei einem prozentualen Spannungsfall Δu_N im Nennbetrieb von 3 bis 10 %.

prozentualer Spannungsfall

$$\Delta u_N = \frac{\Delta U_{N\,\mathrm{Str}}}{U_N}\cdot\sqrt{3}\cdot 100\,\% \qquad (86.81)$$

Der für eine Drosselspule benötigte Spannungsfall ergibt sich aus der Reduzierung der Kurzschlussleistung nach der Drosselspule.

$$\Delta u_N = 1{,}1\cdot 100\,\%\cdot S_D\cdot\frac{S''_{K1}+S''_{K2}}{S''_{K1}\cdot S''_{K2}} \qquad (86.82)$$

S''_{K1} Kurzschlussleistung ohne Drosselspule; S''_{K2} Kurzschlussleistung mit Drosselspule.

Bei der Aufstellung von Drosselspulen ist darauf zu achten, dass die Stromwärmeverluste abgeführt werden können. Die

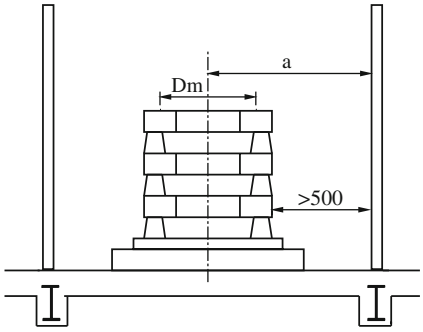


Abb. 86.18 Aufstellung einer Strombegrenzungsspule; *a* Abstand der Drosselspulen-Mittelachse zum Metallteil

Drosselspule muss soweit von Metallteilen entfernt aufgestellt werden, dass keine unzulässige Erwärmung solcher Teile durch Wirbelströme auftreten kann. Der Abstand zu Konstruktionsteilen aus Stahl sowie Eisenarmierungen sollte mehr als 500 mm betragen. Kann nicht auf metallische Konstruktionsteile verzichtet werden, sollten diese durch nichtleitende Materialien getrennt werden. Ferner sollte die magnetische Feldstärke den Wert von $H = 20\,\mathrm{A/cm}$ nicht überschreiten. Zur Ermittlung der Feldstärke kann folgende Formel benutzt werden (Abb. 86.18):

Feldstärke

$$H = 0{,}1\cdot\frac{I_N\cdot N\cdot D_m}{a^2} \qquad (86.83)$$

H	I_N	D_m	a	N
$\frac{\mathrm{A}}{\mathrm{cm}}$	A	cm	cm	–

D_m mittlerer Durchmesser der Drosselspule; a Abstand der Drosselspulenmittelachse zum Metallteil; N Windungszahl der Drosselspule.

Beispiel

Die Kurzschlussleistung an einer Verteilung soll von $S''_{K1} = 600\,\mathrm{MV\,A}$ auf $S''_{K2} = 100\,\mathrm{MV\,A}$ reduziert werden. Die Nennspannung ist $U_N = 10\,\mathrm{kV}$, der Nennstrom $I_N = 600\,\mathrm{A}$. Welche Daten muss die Kurzschlussdrosselspule haben?

Der prozentuale Spannungsfall berechnet sich zu:

$$\begin{aligned}\Delta u_N &= 1{,}1\cdot 100\,\%\cdot S_D\cdot\frac{S''_{K1}+S''_{K2}}{S''_{K1}\cdot S''_{K2}}\\ &= 1{,}1\cdot 100\,\%\cdot\sqrt{3}\cdot 10{,}000\,\mathrm{V}\cdot 600\,\mathrm{A}\\ &\quad\cdot\frac{600\cdot 10^6\,\mathrm{V\,A}-100\cdot 10^6\,\mathrm{V\,A}}{600\cdot 10^6\,\mathrm{V\,A}\cdot 100\cdot 10^6\,\mathrm{V\,A}} = 9{,}53\,\%\end{aligned}$$

Der Spannungsfall an der Drossel sollte mindestens 9,53 % betragen. Die Eigenleistung der Spule hat dann folgenden Wert:

$$\begin{aligned}S_E &= 3\cdot\Delta U_{N\,\mathrm{Str}}\cdot I_N\\ &= 3\cdot\frac{9{,}53\,\%\cdot 10.000\,\mathrm{V}}{100\,\%\cdot\sqrt{3}}\cdot 600\,\mathrm{A} = 0{,}99\,\mathrm{MV\,A}\end{aligned}$$ ◀

Teil XI

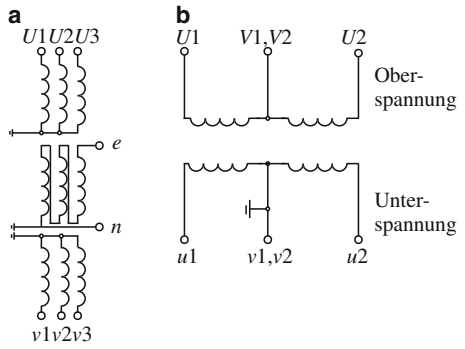


Abb. 86.19 Schaltungen von Spannungswandlern. **a** drei einpolig isolierte Wandler mit Hilfswicklung in offener Dreiecksschaltung, **b** zwei zweipolig isolierte Wandler in V-Schaltung

86.1.9.3 Streufeldtransformatoren

Streufeldtransformatoren haben eine hohe Kurzschlussspannung. Dadurch wird erreicht, dass im Transformator bei auftretendem Kurzschluss kein hoher Strom fließen kann, der den Transformator zerstört. Dieses Verhalten nennt man auch *spannungsweich*. Zu den Streufeldtransformatoren gehören Klingel-, Spielzeug-, Schutz-, Zünd- und Schweißtransformatoren sowie Transformatoren für Leuchtröhrenanlagen.

86.1.10 Messwandler

Messwandler sollen Spannungen und Ströme in elektrischen Anlagen für Mess-, Schutz- und Verrechnungszwecke auf messbare Werte mit genormten Messbereichen umwandeln. Die wesentlichen Anforderungen sind sichere Trennung zwischen Mittelspannungsnetz und Messstromkreis und eine möglichst genaue Abbildung der zu messenden Größen. Diese Forderungen werden bisher überwiegend mit *induktiven Wandlern* gelöst. Induktive Wandler sind Spezialtransformatoren kleiner Leistung entsprechend dem Leistungsbedarf der angeschlossenen Messeinrichtung. Die elektrischen Eigenschaften von Wandlern lassen sich wie bei Transformatoren erklären, wobei Spannungswandler fast im Leerlauf arbeiten ($R_i \ggg$) und Stromwandler fast im Kurzschluss ($R_i \lll$). Messwandler sind entsprechend ihrer Genauigkeit in Klassen eingeteilt. Normwerte sind: Klassen 0,1–0,2–0,5–1,0–3,0–5,0–10. Ferner werden die Wandler noch nach der Art ihres Einsatzes unterschieden. Soll der Wandler für Messungen eingesetzt werden, wird das mit einem „M“ hinter der Klasse verdeutlicht. Wandler für Schutzaufgaben erhalten ein „P“ hinter der Klasse (z. B. die Klasse 1,0 M oder Klasse 1,0 P). Zur Veranschaulichung, bis zu welchem Strom die Klassengenauigkeit eingehalten wird, steht hinter der Einsatzart noch der Faktor, bezogen auf den Nennstrom, bis zu dem die Klassengenauigkeit nicht überschritten wird. Beispiel: Klasse 5P10 bedeutet, dass der Wandler bis $10 \cdot I_N$ einen Messfehler $< 5\,\%$ hat.

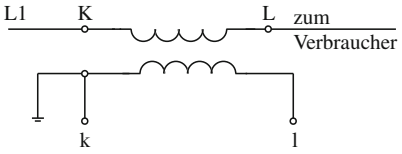


Abb. 86.20 Schaltung eines Stromwandlers

86.1.10.1 Spannungswandler

Spannungswandler werden mit induktiver und kapazitiver Übersetzung gebaut. Die Grundschaltungen von Spannungswandlern in Drehstromnetzen und die Schaltzeichen sind in Abb. 86.19 dargestellt.

Da der Strom im Sekundärkreis des Spannungswandlers sehr klein ist ($R_i \ggg$), folgt, dass bei einem Kurzschluss im Sekundärkreis des Spannungswandlers die Kurzschlussströme zur Zerstörung des Wandlers führen können. Spannungswandler werden zwei- oder einpolig isoliert gefertigt. Zweipolig isolierte Wandler (Anschluss zwischen zwei Außenleitern) besitzen eine Primärwicklung, die mit allen Anschlüssen entsprechend dem Isolationspegel des Hochspannungsnetzes isoliert ist. Bei einpolig isolierten Wandlern (Anschluss zwischen Außenleiter und Erde) ist nur eine Klemme der Primärwicklung für die volle Netzspannung isoliert. Der zweite Anschluss wird geerdet. Als Isoliermittel für Spannungswandler wird Gießharz eingesetzt.

86.1.10.2 Stromwandler

Stromwandler werden hauptsächlich als induktive Wandler gebaut. Die Schaltzeichen und Schaltungen von Stromwandlern sind in Abb. 86.20 dargestellt.

Da der Stromwandler nahezu im Kurzschluss betrieben wird, entsteht beim Öffnen der Sekundärseite an den Sekundärklemmen eventuell eine Überspannung, die zur Gefährdung von Personen und zum Überschlag führen kann. Die genormten Ströme der Sekundärseite sind: Messkern 5 A, 1 A (2 A), 0,5 A; Schutzkern 5 A, 1 A. Stromwandler werden in den verschiedensten Ausführungen hergestellt. Nach Art der Wicklung unterscheidet man zwischen den *Einleiterstromwandlern*, die primär nur eine Windung besitzen, und den *Wickelstromwandlern*, bei denen die Primär- und Sekundärwicklungen aus vielen Windungen bestehen. Der Vorzug der Einleiterstromwandler liegt in ihrer großen Kurzschlussfestigkeit, da auf einen ohne Schleife durch den Kern geführten Leiter praktisch keine mechanischen Beanspruchungen wirksam werden. Der Wickelstromwandler lässt sich dagegen besser bei kleinen primären Strömen einsetzen. Er kann durch Umschaltlaschen in der Primärwicklung an verschiedene Nennströme angepasst werden. In einem Stromwandler lassen sich mehrere Kerne unterbringen, die für verschiedene Aufgaben eingesetzt werden, so z. B. ein Kern für Messungen und ein Kern für Schutzaufgaben.

86.2 Drehstrommaschinen

Werden an einem Ständer drei Spulensysteme in einem Winkel von 120° angeordnet und schließt man diese Spulen an ein Drehspannungssystem wie in Abb. 86.21 dargestellt an, so erzeugt der dann fließende Strom eine resultierende Durchflutung, die sich um den Ständer dreht und im Luftspalt ein umlaufendes Feld (Drehfeld) erzeugt.

Dieses Drehfeld hat, bei drei Spulen im Ständer, einen Winkel von 360° überstrichen, wenn die Spannung eine Periode durchlaufen hat. Die Umlaufgeschwindigkeit des Drehfeldes ist also von der Frequenz der angelegten Spannung abhängig. Ordnet man sechs Spulen, die um 60° verdreht sind, im Ständer an (Abb. 86.22), benötigt das Drehfeld zwei Perioden, bis es einen Winkel von 360° überstrichen hat.

Damit ist die Umlaufgeschwindigkeit des Drehfeldes auch von der Anzahl der im Ständer vorhandenen Pole abhängig.

$$\textit{Drehzahl} \quad n = \frac{f \cdot 60}{p} \qquad \begin{array}{c|c|c} n & f & p \\ \hline \mathrm{min}^{-1} & \mathrm{Hz} & - \end{array} \qquad (86.84)$$

n Drehzahl des Drehfeldes; f Frequenz der Netzspannung; p Anzahl der Polpaare.

Die Anzahl der Polpaare kann aus dem Winkel zwischen den Ständerspulen ermittelt werden.

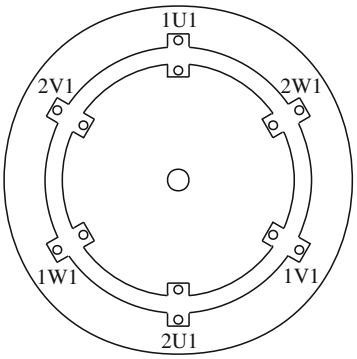


Abb. 86.21 Grundaufbau einer zweipoligen Drehfeldmaschine

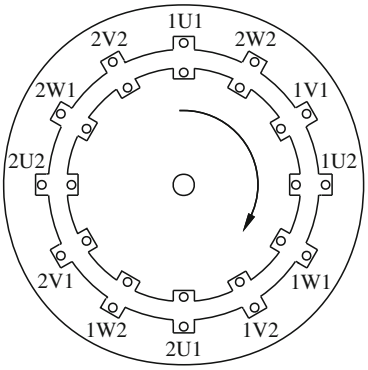


Abb. 86.22 Grundaufbau einer vierpoligen Drehfeldmaschine

Tab. 86.3 Zusammenhang zwischen Polpaarzahl und Drehzahl

Polpaarzahl	Polzahl	Anzahl der Spulen	Winkel zwischen den Spulen in °	Zeit für eine Umdrehung des Drehfeldes
1	2	3	120	$1T$
2	4	6	60	$2T$
3	6	9	40	$3T$
p	$2p$	$3p$	$\frac{360°}{3p}$	pT

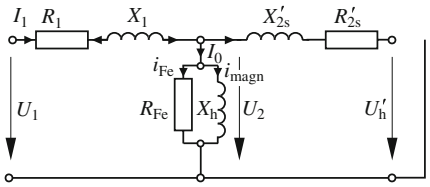


Abb. 86.23 Vollständiges Ersatzschaltbild der Drehstromasynchronmaschine als Käfigläufer (Läuferspulen kurzgeschlossenen) im Stillstand

86.2.1 Die Drehstromasynchronmaschine

Die Drehstromasynchronmaschine spielt in der Antriebstechnik eine überragende Rolle, da sie aufgrund ihres einfachen Aufbaus nahezu wartungsfrei ist und durch die Frequenzumrichtertechnik problemlos gesteuert werden kann. Asynchronmaschinen gibt es von Leistungen $< 1\,\mathrm{kW}$ bis in den MW-Bereich. Die Asynchronmaschine ist eine Drehfeldmaschine und besitzt daher im Ständer eine Drehstromwicklung. Der Läufer ist ebenfalls mit einer Drehstromwicklung ausgestattet, die beim *Schleifringläufer* über Schleifringe und Bürsten herausgeführt wird, beim *Kurzschlussläufer* aus Profilstäben (Kupfer oder Bronze) besteht, die an den Enden verbunden sind. Bei kleineren Maschinen werden die Läuferwicklungen im Spritzgussverfahren aus Aluminium hergestellt.

86.2.1.1 Wirkungsweise der Asynchronmaschine

Die Betrachtung der Asynchronmaschine im Stillstand zeigt ein Verhalten wie ein Transformator. Das Drehfeld im Ständer induziert im Läufer eine Spannung und diese wiederum einen Strom. Die ohmschen Verluste kennzeichnen die Ständer- oder Läuferwicklung, die Streuinduktivitäten entstehen an den Wickelköpfen und an den Nuten.

Das Ersatzschaltbild der Asynchronmaschine ist im Stillstand also identisch mit dem Trafoersatzschaltbild (Abb. 86.23). Bei der Betrachtung der Asynchronmaschine bei sich drehendem Läufer kann das Trafoverhalten nicht mehr ohne weiteres herangezogen werden. Durch die induzierte Spannung im Läufer wird ein Stromfluss hervorgerufen, der mit dem umlaufenden Ständerdrehfeld eine Kraft auf den Läufer ausübt.

$$F = I_L \cdot B_{\mathrm{St}} \cdot l \cdot \sin\alpha \qquad \begin{array}{c|c|c|c} F & I_L & B_{\mathrm{St}} & l \\ \hline \mathrm{N} & \mathrm{A} & \frac{\mathrm{V\,s}}{\mathrm{m}^2} & \mathrm{m} \end{array} \qquad (86.85)$$

I_L Läuferstrom; B_{St} Ständerfluss; l Länge der Spulen im Feld; α Winkel zwischen I und B normal 90°.

Teil XI

Das hieraus resultierende Drehmoment beschleunigt den Läufer in Richtung des Drehfeldes.

$$M = F \cdot \frac{d}{2} \qquad \begin{array}{c|c|c} M & F & d \\ \hline \mathrm{N\,m} & \mathrm{N} & \mathrm{m} \end{array} \qquad (86.86)$$

d Läuferdurchmesser.

Da der Läufer ebenfalls in Richtung des Drehfeldes bewegt wird, entsteht in ihm ein geringerer Stromfluss ($U_i = B \cdot v \cdot l$) und somit ein kleineres resultierendes Moment. Hat der Läufer die gleiche Drehzahl wie das Drehfeld erreicht, ist $I = 0 \Rightarrow M = 0$. Bedingt durch Reibungs- und Widerstandsverluste im Läufer, muss immer ein Drehmoment zur Deckung der Verluste vorhanden sein. Der Läufer kann also niemals synchron mit dem Ständerdrehfeld mitlaufen. Der Unterschied zwischen der synchronen Drehzahl und der tatsächlichen Drehzahl des Läufers wird als *Schlupf s* bezeichnet.

$$\textit{Schlupf} \quad s = \frac{n_0 - n_L}{n_0} = \frac{f - f_L}{f} = \frac{\Delta f}{f} \qquad (86.87)$$

$n_0 = \frac{f \cdot 60}{p}$ synchrone Drehzahl; $n_L = \frac{f_L \cdot 60}{p}$ Läuferdrehzahl.

86.2.1.2 Betriebsverhalten der Asynchronmaschine

Das Betriebsverhalten der Drehstromasynchronmaschine ist nicht mehr mit dem des Transformators zu vergleichen, da diese Maschine eine elektrische Leistung in eine mechanische Leistung umformt. Damit das Verhalten der Asynchronmaschine deutlicher wird, betrachtet man verschiedene Betriebszustände und verallgemeinert hieraus das Verhalten.

86.2.1.2.1 Spannungsgleichung, Ersatzschaltbild

Im Stillstand der Asynchronmaschine gilt das Transformatorersatzschaltbild und die Transformatorgleichung im Kurzschluss. Die Ständerspannungsgleichung lautet:

Ständerspannung

$$\underline{U}'_2 = 0 = -\underline{I}'_2 \cdot (R'_2 + jX'_{s2}) + \underline{I}_0 \cdot \left\{ \frac{(R_{\mathrm{Fe}} \cdot jX_H)}{(R_{\mathrm{Fe}} + jX_H)} \right\} \qquad (86.88)$$

Die Läuferspannungsgleichung lautet:

Läuferspannung

$$\underline{U}_1 = \underline{I}_1 \cdot (R_1 + jX_{s1}) + \underline{I}_0 \cdot \left\{ \frac{(R_{\mathrm{Fe}} \cdot jX_H)}{(R_{\mathrm{Fe}} + jX_H)} \right\} \qquad (86.89)$$

Die induktiven Widerstände X'_{s2} und X_H sowie die Eisenverluste R_{Fe} sind proportional zur Frequenz zwischen Ständer und Läufer. Da sich diese Frequenz mit steigender Drehzahl verringert, verringern sich auch die Widerstände.

Ständer-Läufer-Impedanz

$$X^*_H = 2 \cdot \pi \cdot \Delta f \cdot L_H \qquad (86.90)$$

$$X^*_H = 2 \cdot \pi \cdot \Delta f \cdot L_H \cdot \frac{f}{f} = 2 \cdot \pi \cdot \frac{\Delta f}{f} \cdot f \cdot L_H \qquad (86.91)$$

$$X^*_H = s \cdot X_H \qquad (86.92)$$

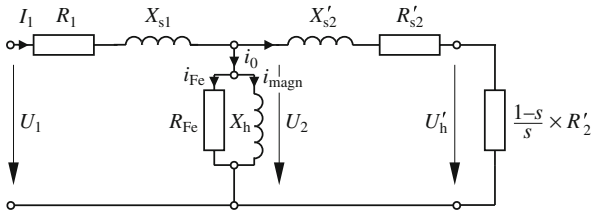


Abb. 86.24 Vollständiges Ersatzschaltbild der Drehstromasynchronmaschine als Käfigläufer

mit $s = \frac{\Delta f}{f}$ und $X_H = 2 \cdot p \cdot f \cdot L_H$.

Diese Beziehung kann gleichermaßen für die Streuinduktivität X_s und den Eisenverlustwiderstand R_{Fe} aufgestellt werden. Für die Ständerspannungsgleichung ergibt sich hieraus:

Ständerspannung

$$\underline{U}'_2 = 0 = -\underline{I}'_2 \cdot (R'_2 + jsX'_{s2}) + \underline{I}_0 \cdot \left\{ \frac{(sR_{\mathrm{Fe}} \cdot jsX_H)}{(sR_{\mathrm{Fe}} + jsX_H)} \right\} \qquad (86.93)$$

$$\underline{U}'_2 = 0 = -\underline{I}'_2 \cdot \left(\frac{R'_2}{s} + jX'_{s2} \right) + \underline{I}_0 \cdot \left\{ \frac{(R_{\mathrm{Fe}} \cdot jX_H)}{(R_{\mathrm{Fe}} + jX_H)} \right\} \qquad (86.94)$$

Gleichung (86.94) ergibt sich aus (86.94) durch Ausklammern des Schlupfes s. Den ohmschen Widerstand im Läuferkreis von (86.95) kann man wie folgt aufteilen:

$$\textit{Ständerwirkwiderstand} \quad \frac{R'_2}{s} = R'_2 + \frac{1-s}{s} \cdot R'_2 \qquad (86.95)$$

Das Ersatzschaltbild der Asynchronmaschine ist identisch mit dem des Transformators (Abb. 86.24). Der auf der Sekundärseite angeordnete Widerstand $R'_2 \cdot \frac{(1-s)}{s}$ entspricht einem Lastwiderstand beim Transformator. Die in diesem Widerstand umgesetzte Leistung ist die an der Welle der Maschine abgegebene oder aufgenommene mechanische Leistung einschließlich der Reibungsverluste.

86.2.1.2.2 Leistungsfluss

Die von einer Asynchronmaschine aufgenommene Wirkleistung berechnet sich wie folgt:

aufgenommene Wirkleistung

$$P_{\mathrm{zu}} = P_{\mathrm{el}} = 3 \cdot U_{\mathrm{Str1}} \cdot I_{\mathrm{Str1}} \cdot \cos\varphi_1 \qquad (86.96)$$

Von dieser Leistung werden die im Ständer auftretenden Verluste P_{Cu1} und P_{Fe} abgezogen. Die verbleibende Leistung wird über den Luftspalt als *Luftspaltleistung* P_δ übertragen. Die an der Welle abgegebene Leistung ist die Luftspaltleistung abzüglich der im Läufer auftretenden Verluste P_{Cu2}.

Luftspaltleistung

$$P_\delta = P_{\mathrm{el}} - (P_{\mathrm{Cu1}} + P_{\mathrm{Fe}}) = 3 \cdot I^2_{2\mathrm{Str}} \cdot \frac{R'_2}{s} = 3 \cdot U_H \cdot I'_2 \cdot \cos\varphi_2 \qquad (86.97)$$

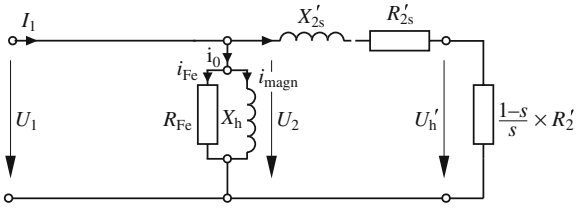


Abb. 86.25 Vereinfachtes Ersatzschaltbild der Drehstromasynchronmaschine als Käfigläufer

Kupferverlustleistung im Läufer

$$P_{Cu2} = 3 \cdot I'^2_{2Str} \cdot R'_2 = s \cdot P_\delta \tag{86.98}$$

mechanisch abgebbare Leistung

$$P_{ab} = P_{mech} = P_\delta - P_{Cu2} = P_\delta - s \cdot P_\delta = (1-s) \cdot P_\delta \tag{86.99}$$

Die mechanische Leistung der Asynchronmaschine ist also stark schlupfabhängig, d. h., für den Schlupf $s = 1$ (Stillstand) kann an der Welle keine Leistung abgegeben werden. Die gesamte Leistung, die über den Luftspalt transportiert wird, setzen die Läuferwicklungen in Verlustwärme um. Beim Schlupf $s = 0$ (Synchronismus) ist die über den Luftspalt transportierte Leistung Null, da der Widerstand im Ständerkreis gegen ∞ geht ($\frac{R'_2}{s} \to \infty$). Zwischen $s = 1$ und $s = 0$ durchwandert die Luftspaltleistung ein Maximum, das beim *Kippschlupf* s_K auftritt. Der Wirkungsgrad η des Asynchronmotors ergibt sich aus dem Verhältnis der abgegebenen zur aufgenommenen Wirkleistung.

Wirkungsgrad

$$\eta = \frac{P_{ab}}{P_{zu}} = \frac{P_{mech}}{P_\delta + P_{Fe} + P_{Cu1}} \cong \frac{P_\delta - P_{Cu2}}{P_\delta} = 1 - s \tag{86.100}$$

Das an der Welle auftretende Drehmoment berechnet sich zu:

Motormoment

$$M = \frac{P_{mech}}{2 \cdot p \cdot n} \qquad \begin{array}{c|c|c} P_{mech} & n & M \\ \hline \text{W} & \text{s}^{-1} & \text{N m} \end{array} \tag{86.101}$$

$$M = \frac{P_\delta}{2 \cdot p \cdot n_0} \tag{86.102}$$

Gleichung (86.102) ergibt sich durch Ersetzen von $n = (1-s) \cdot n_0$ und $P_{mech} = (1-s) \cdot P_\delta$.

86.2.1.2.3 Betriebskennlinien

Mit (86.90) und (86.95) kann die Betriebskennlinie des Asynchronmotors berechnet werden. Diese Berechnung erfordert jedoch einen hohen Rechenaufwand. Deshalb wird die Betriebskennlinie des Asynchronmotors über ein vereinfachtes Ersatzschaltbild (Abb. 86.25) ermittelt.

Hierbei werden die Ständerverluste und Ständerstreuinduktivitäten vernachlässigt, was für Maschinen > 15 kW beim Betrieb

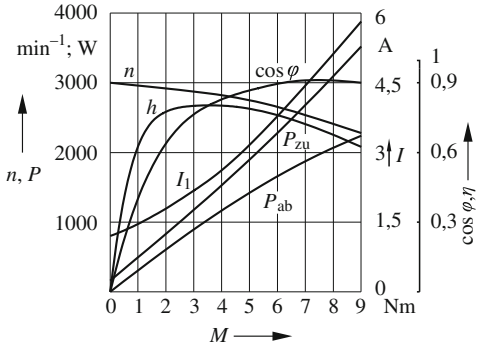


Abb. 86.26 Belastungskennlinie eines 1 kW-Drehstrom-Asynchronmotors

mit Nennfrequenz keinen großen Einfluss auf die Kennlinie hat. Bei kleineren Maschinen werden die Betriebskennlinien für erste Betrachtungen hinreichend genau erfasst. Nach dem vereinfachten Ersatzschaltbild ist die Spannung an der Hauptinduktivität gleich der Netzspannung und somit der Magnetisierungsstrom I_0 nur von der Höhe der angelegten Spannung abhängig. Da der Strom im Ständerkreis auch direkt von der angelegten Spannung abhängig ist, ergibt sich für das Moment folgende Abhängigkeit:

$$M = \frac{d}{2} \cdot F = I'_2 \cdot B_{St} \cdot l \cdot \frac{d}{2} \cdot \sin\alpha;$$
$$B_{St} \approx I_0 \approx U_H; \quad I'_2 \approx U_H \quad \Rightarrow \quad M \approx U_H^2$$

Ebenso wie die Luftspaltleistung hat das Moment sein Maximum beim Kippschlupf. Dieses Moment wird als *Kippmoment* M_K bezeichnet. Die Spannungsgleichung des vereinfachten Ersatzschaltbildes ergibt sich zu:

$$\underline{U}_{Str1} = \underline{U}_H = \underline{I}'_{Str2} \cdot \left(\frac{R'_2}{s} + jX'_s \right) \tag{86.103}$$

Die Stromgleichung zu:

$$\underline{I}'_{Str2} = \frac{\underline{U}_H}{\left(\frac{R'_2}{s} + jX'_s\right)} = \underline{U}_H \cdot \left[\frac{\frac{R'_2}{s}}{\left(\left(\frac{R'_2}{s}\right)^2 + X'^2_s\right)} - j\frac{X'_s}{\left(\left(\frac{R'_2}{s}\right)^2 + X'^2_s\right)} \right] \tag{86.104}$$

Aus der Gleichung erkennt man, dass der Strom für $s = \infty$ einen Maximalwert annimmt und für $s = 0$ Null wird.

Der Leistungsfaktor steigt mit sinkendem Schlupf bis zu einem Maximalwert und fällt dann wieder. Die Abhängigkeiten von I'_2, M, P und $\cos\varphi$ vom Schlupf sind in Abb. 86.26 dargestellt.

86.2.1.3 Kurzschlussläufer

Die Läuferwicklungen beim Kurzschluss- oder Käfigläufer bestehen aus Stäben, die in eine vorgefertigte Nut eingelegt oder gespritzt werden. Die gebräuchlichsten Nutformen sind der Rundstab, Hochstab und Doppelstab (Abb. 86.27).

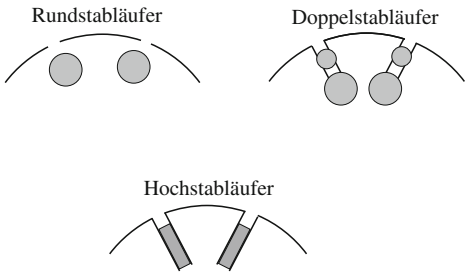


Abb. 86.27 Nutformen des Drehstrom-Asynchronmotors

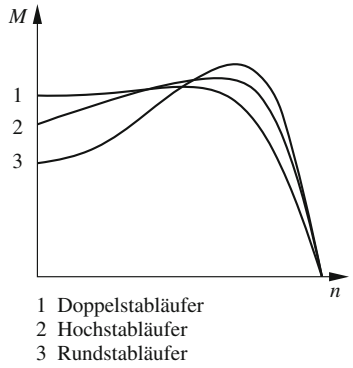


Abb. 86.28 Drehzahl-Drehmomentkennlinien verschiedener Nutformen des Drehstrom-Asynchronmotors

Der Rundstabläufer hat praktisch die gleichen Eigenschaften wie ein Läufer mit gewickelten Spulen. Beim Hochstab- oder Doppelstabläufer bewirkt die Form der Nuten, dass nicht mehr alle Nutabschnitte von der gleichen Anzahl von Feldlinien geschnitten werden. Hieraus resultiert, dass die tiefer liegenden Nutbereiche von mehr Feldlinien, die Bereiche in der Nähe des Luftspaltes dagegen von weniger Feldlinien umschlossen sind. Als Folge dieser ungleichmäßigen Feldlinienverteilung ist die Selbstinduktivität am unteren Nutbereich größer als am Luftspalt. Die Induktionsspannung ist also am Luftspalt am geringsten. Der Strom ist nicht mehr gleichmäßig in der Nut verteilt, sondern wird in Richtung Luftspalt gedrängt. Als Folge dieser *Stromverdrängung* wird der Wirkwiderstand größer, die Streuinduktivität kleiner. Der resultierende Scheinwiderstand wird kleiner, der Strom im Läufer also größer. Ausschlaggebend ist jedoch, dass das Läufermoment wesentlich stärker steigt als der Strom, da das Moment im Anlauf von der im Luftspalt umgesetzten Wirkleistung abhängig ist. Hat der Käfigläufer nicht einen Rundstab, sondern eine andere Stabform, so bezeichnet man den Käfigläufermotor auch als *Stromverdrängungsläufer*.

Der Stromverdrängungsläufer hat jedoch den Nachteil, dass im Bereich der synchronen Drehzahl die Streuinduktivität im Gegensatz zum Normalmotor größer und somit der Leistungsfaktor schlechter ist. Ebenso ist das Kippmoment kleiner als bei stromverdrängungsfreien Motoren. Die realen Drehmoment-Drehzahlkennlinien dieser Maschinen sind im Abb. 86.28 dargestellt. Im Bereich kleiner Drehzahlen fällt auf, dass das Moment von seinem Anlaufwert leicht abfällt und dann bis zum

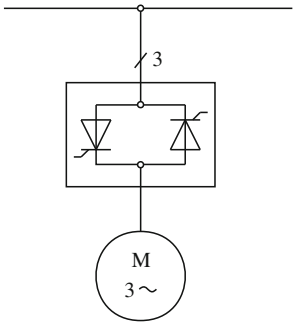


Abb. 86.29 Ständerspannungsänderung mit Thyristorsteller

Kippmoment ansteigt. Diese Verminderung des Moments verursachen Oberwellenmomente, die sich dem Grundmoment überlagern. Die Oberwellen werden bei der theoretischen Betrachtung der Asynchronmaschine vernachlässigt, da ihr Einfluss nur im unteren Drehzahlbereich auftritt. Das durch den Momentenabfall entstehende minimale Moment bezeichnet man als *Sattelmoment*. Beim Anfahren der Maschine ist darauf zu achten, dass das Gegenmoment nicht größer als das Sattelmoment wird (nähere Festlegungen in VDE 0530).

86.2.1.3.1 Anlassverfahren

Befindet sich der Läufer der Asynchronmaschine im Stillstand (Anlauf), wird ein hoher Strom (6–8-facher Nennstrom) aus dem Netz aufgenommen. Dieser Strom verursacht auf den Zuleitungen einen erhöhten Spannungsfall. Deshalb versucht man, diesen Strom zu verringern. Möglichkeiten der Anlaufstrombegrenzung sind die Senkung der Ständerspannung, der Stern-Dreieckanlauf und die Beschaltung mit einem Widerstand im Ständerkreis. Durch Verändern der *Ständerspannung* kann die Höhe des Stroms und des Moments beeinflusst werden. Da der Motor normalerweise für die angelegte Netzspannung gekauft wurde, ist eine Spannungsänderung nur in eine Richtung möglich.

Die Ständerspannung kann durch einen Transformator (Spartrafo) oder durch einen Thyristorschalter (Abb. 86.29) verändert werden. Bei der Spannungverringerung verändert sich der Strom proportional zur Spannung, das Moment jedoch quadratisch ($I \approx U$; $M \approx U^2$). Eine Anlaufschaltung durch Spannungssenkung ist aus diesen Gründen nur für leichte Anlaufbedingungen möglich.

Beim *Stern-Dreieck-Anlauf* (Abb. 86.30) werden die Ständerwicklungen während des Anlaufvorganges im Stern geschaltet. Hierdurch liegt an den Wicklungen eine geringere Spannung (Strangspannung) an, und es fließt ein kleinerer Strom. Mit dem Strom verringert sich jedoch das Anlaufmoment

$$\text{Stromverhältnis} \qquad \frac{I_Y}{I_\Delta} = \frac{\frac{U}{\sqrt{3}\cdot Z}}{\frac{\sqrt{3}\cdot U}{Z}} = \frac{1}{3} \qquad (86.105)$$

$$\text{Momentenverhältnis} \qquad \frac{M_Y}{M_\Delta} = \frac{(\frac{U}{\sqrt{3}})^2}{U^2} = \frac{1}{3} \qquad (86.106)$$

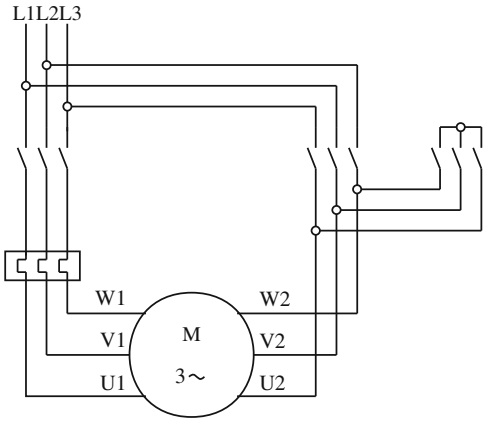


Abb. 86.30 Stern-Dreieck-Anlauf-Stromplan

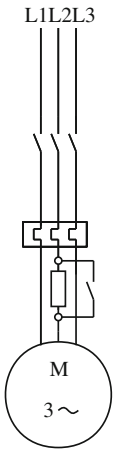


Abb. 86.31 Stromlaufplan der KUSA-Schaltung

Der Stern-Dreieck-Anlauf darf nur bei leichten Anlaufbedingungen eingesetzt werden, wenn der Motor mit verminderter Last anläuft, da beim Umschalten von Stern- auf Dreieckschaltung ein Stromstoß auftritt, der annähernd gleich dem Anlaufstrom sein kann.

Schaltet man einen Widerstand in einen Strang (Abb. 86.31), so verhält sich der Asynchronmotor wie ein Wechselstrommotor mit Hilfswicklung. Diese Schaltung wird auch *KUSA-Schaltung* (Kurzschlussläufer-Sanftanlauf) genannt. Durch den Widerstand in einem Strang wird das Drehfeld in ein elliptisches Feld verwandelt, und das Moment sinkt mit zunehmendem Widerstand. Der Strom in dem mit dem Widerstand behafteten Strang sinkt zwar, dafür steigen jedoch die Ströme in den anderen beiden Strängen. Durch die KUSA-Schaltung erhält man ein kleineres Moment, jedoch keine kleineren Ströme.

86.2.1.3.2 Bremsverfahren

Mögliche Bremsverfahren für den Asynchronmotor sind die Abbremsung über mechanische Bremsen, die Gleichstrombremsung, die Gegenstrombremsung und die übersynchrone Bremsung. Bei der *mechanischen Bremsung* wird der Motor nicht belastet. Nachteil dieser Bremsart ist, dass die Bremsbeläge einem hohen Verschleiß unterliegen und somit ein hoher Wartungs- und Ersatzaufwand notwendig ist. Diese Bremsart sollte nur als Zusatz- oder Notbremse eingesetzt werden.

Tab. 86.4 Faktor zur Bestimmung des Gleichstroms beim Gleichstrombremsen von Asynchronmotoren

Schaltungsart	a	b	c	d
k-Faktor	1,225	1,41	2,21	2,45

Tab. 86.5 Faktor zur Bestimmung der Gleichspannung beim Gleichstrombremsen von Asynchronmotoren

Schaltungsart	a	b	c	d
C-Faktor	2	1,5	0,667	0,5

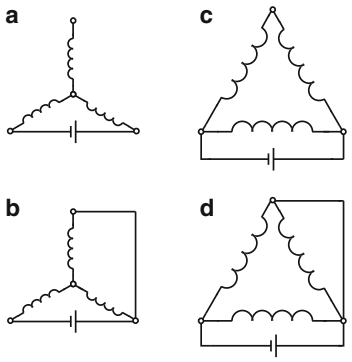


Abb. 86.32 Schaltung zum Gleichstrombremsen von Asynchronmotoren

Die *Gleichstrombremsung* ist die effektivste Bremsmethode. Hierbei wird der Motor vom Netz getrennt und die Ständerwicklungen nach Abb. 86.32 an einen Gleichstrom niedriger Spannung angeschlossen. Das sich dabei einstellende Moment kann aus der Drehzahl-Drehmomentenkennlinie ermittelt werden, indem die Drehzahl n durch die Bremsdrehzahl $n_{BG} = n_0 - n$ ersetzt wird. Der Bremsgleichstrom I_{BG} sollte den 2,5fachen Wert des Nennstromes nicht überschreiten. Bei dieser Bremsart ist kein Anlauf in entgegengesetzter Richtung möglich, und die Wärmeentwicklung in den Spulen ist gering. Anhaltswerte für die Berechnung des Gleichstroms sind in (86.107) und Tab. 86.4 zusammengefasst.

$$\textit{Bremsgleichstrom} \quad I_{BG} \leq k \cdot I_N \cdot 1{,}5 \qquad (86.107)$$

Die Gleichspannung berechnet sich nach (86.109) und Tab. 86.5

$$\textit{Bremsgleichspannung} \quad U_{Gl} = C \cdot I_{BG} \cdot R_{Str_{warm}} \quad \text{oder}$$

$$U_{Gl} = 1{,}3 \cdot C \cdot I_{BG} \cdot R_{Str_{kalt}} \qquad (86.108)$$

$$U_{Gl} = 1{,}3 \cdot C \cdot I_{BG} \cdot R_{Str_{kalt}} \qquad (86.109)$$

Teil XI

Beispiel

Ein Drehstromasynchronmotor mit einem Nennstrom $I_n = 50\,\text{A}$ und einem Strangwiderstand $R_{\text{Str}} = 2\,\Omega$ wird im warmen Zustand über eine Gleichstrombeschaltung nach Schaltungsart b abgebremst. Welcher Gleichstrom darf maximal zugelassen und welche Gleichspannung sollte an die Wicklungen angelegt werden?

Für Schaltungsart b gilt:

$$I_{\text{BG}} \leq k \cdot I_N \cdot 1{,}5 = 1{,}41 \cdot I_N \cdot 1{,}5 = 1{,}41 \cdot 50\,\text{A} \cdot 1{,}5 = 105{,}75\,\text{A}$$

$$U_{\text{GI}} = C \cdot I_{\text{BG}} \cdot R_{\text{Str}_{\text{warm}}} = 1{,}5 \cdot 105{,}75\,\text{A} \cdot 2\,\Omega = 317{,}25\,\text{V}$$ ◀

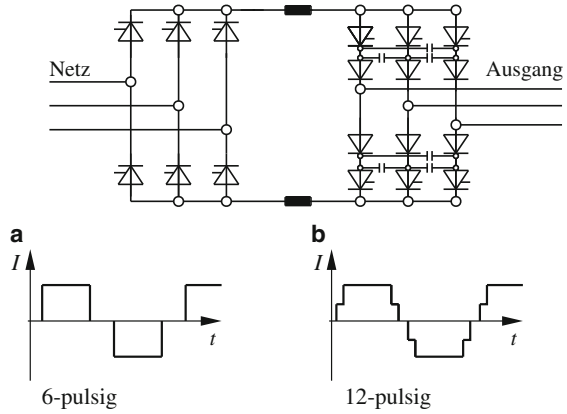


Abb. 86.33 Stromzwischenkreisumrichter

Die *Gegenstrombremsung* hat eine extreme Bremswirkung. Hierbei werden zwei Phasen des Ständers vertauscht. Am Ständer stellt sich ein gegenläufiges Drehfeld ein, das versucht, den Läufer mitzureißen. Der Schlupf wird größer 1. Wird eine Gegenstrombremsung aus einer Nenndrehzahl nahe der Leerlaufdrehzahl ausgeführt, ist die thermische Motorbelastung zweimal so groß wie beim Anlauf. Bei der Gegenstrombremsung wird die mechanische Leistung und zusätzlich die aufgenommene elektrische Leistung in Wärme umgewandelt. Deshalb muss überprüft werden, ob der Bremsvorgang länger als 5 s dauert, da die Maschine sonst thermisch überlastet wird. Damit die Maschine nicht in entgegengesetzte Richtung anläuft, muss ein Drehzahlwächter den Bremsvorgang überprüfen und die Netzspannung kurz vor dem Stillstand abschalten.

Die *übersynchrone Bremsung* kann nur eingesetzt werden, wenn die Last den Motor antreibt. Die Asynchronmaschine arbeitet dann als Generator und hat einen Schlupf $s < 0$; der Läufer dreht also schneller als das Ständerfeld. Dieser Bremsvorgang hat wenig Stromwärmeverluste. Eingesetzt wird er bei Dahlanderantrieben, bei Frequenzumrichterantrieben und bei einer Beschleunigung der Maschine durch die Last, wie z. B. bei Hebezeugen.

86.2.1.3.3 Drehzahlsteuerung

Eine Drehzahländerung des Käfigläufermotors kann durch Verändern der Ständerspannung, Veränderung der Frequenz und geeignetes Umschalten von Spulengruppen in der Ständerwicklung erfolgen.

86.2.1.3.4 Ständerspannungsänderung

Die *Ständerspannungsänderung* lässt eine Drehzahlsteuerung nur bis zum Kippschlupf zu. Einschränkungen: Die Verluste vergrößern sich, der Wirkungsgrad sinkt, und es besteht die Möglichkeit, dass durch Lastwechsel ein Gegenmoment auftritt, das größer als das Kippmoment ist und damit zum Stillstand des Motors führt. Spannungsverminderung wird bei getakteten Motoren eingesetzt. Läuft der Motor im Leerlauf, kann der Leerlaufstrom durch eine Spannungsverminderung gesenkt werden. Dadurch treten weniger Verluste in der Primärwicklung und in den Zuleitungen auf.

86.2.1.3.5 Frequenzänderung

Eine Frequenzänderung hat zur Folge, dass die synchrone Drehzahl der Maschine entsprechend der Beziehung $n_0 = f/p$ beeinflusst wird. Da durch eine Frequenzänderung die Höhe der induktiven Widerstände ebenfalls verändert wird, muss neben der Frequenz auch die Spannung in gleicher Richtung variiert werden. Zur Drehzahlsteuerung von Asynchronmaschinen werden deshalb Frequenzumrichter eingesetzt, bei denen durch geeignete Maßnahmen die Frequenz und die Spannung parallel beeinflusst werden. Zum Einsatz gelangen zwei Arten von Frequenzumrichtern, der *Spannungszwischenkreisumrichter* und der *Stromzwischenkreisumrichter*.

Beim Stromzwischenkreisumrichter (Abb. 86.33) wird die Netzspannung über eine gesteuerte Gleichrichterbrücke in eine Gleichspannung mit variabler Höhe umgewandelt. Auf der Motorseite wird die Gleichspannung in 120°-Rechteckblöcke (mit der entsprechenden Frequenz) auf die Wicklungsstränge des Motors gegeben. Eine Induktivität im Gleichstromkreis bestimmt die Höhe des Gleichstroms. Die Motorspannung stellt sich annähernd sinusförmig mit überlagerten Kommutierungsspitzen ein. Die Oberschwingungen des Stroms weisen ihre höchsten Amplituden bei den niedrigen Ordnungszahlen auf. Bei Teilbelastung gehen die Oberschwingungen anteilig zurück, da sie stromabhängig sind. Um die Oberschwingungen klein zu halten, sollte ein zwölfpulsiger Umrichter eingesetzt werden. Stromzwischenkreisumrichter werden bei mittleren und größeren Leistungen eingesetzt, da sie Bremsenergie in das Netz zurückspeisen können. Im Gegensatz zum U-Umrichter kann der I-Umrichter nicht mehrere Motore gleichzeitig versorgen. Der U-Umrichter nimmt Blindleistung aus dem Netz auf und gibt Oberwellen an das Netz ab.

Beim *Spannungszwischenkreisumrichter* (Abb. 86.34) wird über einen Gleichrichtersatz eine Gleichspannung konstanter Höhe aus dem Drehstromnetz erzeugt. Diese Gleichspannung wird mit elektronischen Schaltern (Transistoren, Thyristoren) in eine Wechselspannung mit veränderbarer Frequenz und Span-

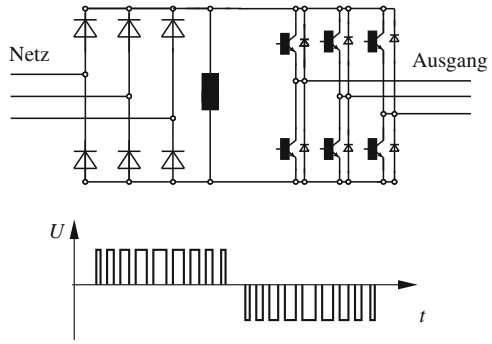


Abb. 86.34 Spannungszwischenkreisumrichter

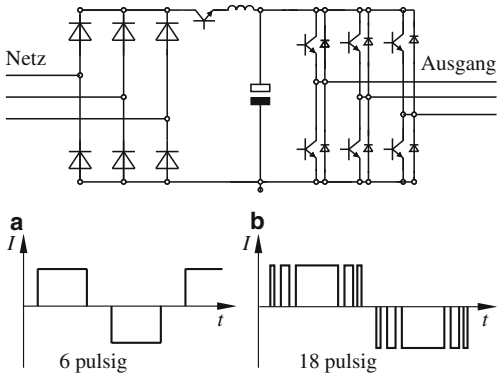


Abb. 86.35 Spannungszwischenkreisumrichter in PAM-(Chopper-)Schaltung

nung umgeformt. Eine veränderbare Spannung wird erzeugt, indem die Pulsbreite der Ausgangsimpulse verändert wird (Pulsweitenmodulation, PWM). Bei dieser Steuerungsart werden Oberschwingungen niedriger Ordnungszahlen weitgehend unterdrückt (5, 7, 11). Die der Pulsfrequenz entsprechenden Oberschwingungen bleiben jedoch unabhängig von der Grundfrequenz in voller Amplitude wirksam. Im Strom bilden sich ebenfalls Oberschwingungen aus, die aus den vorhandenen Spannungsoberschwingungen und den induktiven Einflüssen (Wicklung) resultieren. Moderne Frequenzumrichter arbeiten mit Pulsfrequenzen im kHz-Bereich. Dadurch werden die Stromoberschwingungen stark verringert, da die Induktivitäten einen hohen Widerstand in diesem Frequenzbereich darstellen. Bei der PWM schalten moderne MOS-FET's oder IGBT's Spannungsimpulse von 500 V in 200 ns (Änderungsgeschwindigkeit von 2500 V/ms) auf die Ausgänge des Frequenzumrichters. Wenn diese Potentialsprünge ungefiltert an die Ausgangsklemmen gegeben werden, wirken sie sich direkt auf die Motorzuleitung aus. Mit den Leitungsinduktivitäten und -kapazitäten entstehen Schwingkreise, die von den Rechteckimpulsen des Frequenzumrichters immer wieder angeregt werden und sich unkontrolliert aufschaukeln können. Ebensolche Vorgänge können mit den Streuinduktivitäten und Wicklungskapazitäten des Motors auftreten. Die Folge dieser Schwingungen sind Störstrahlungen, die durch eine abgeschirmte Leitung gemildert werden können. Hierdurch entstehen aber weitere Kapazitäten, die wiederum Schwingkreise aufbauen. Durch die hohe Spannungsänderungs-

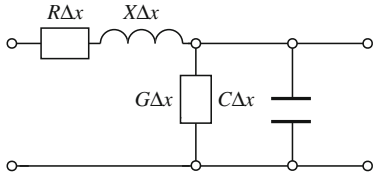


Abb. 86.36 Ersatzschaltbild Leitungsstück

geschwindigkeit können ferner Spannungsspitzen entstehen, die die Isolation des Motors beanspruchen und zerstören können. Durch kapazitive Ableitungen werden Messanordnungen beeinflusst. Filter am Ausgang des Frequenzumrichters können diese Vorgänge vermeiden oder mindern. Sie dürfen die Frequenzen unter 100 Hz jedoch nicht beeinflussen! Je höher die Schaltfrequenz, desto besser der Einfluss der Filter. Je kleiner die Schaltfrequenz, desto teurer die Filter. Optimale Schaltfrequenzen liegen bei 20 kHz. Für Frequenzumrichter ohne Filter kann ein 10 m langes Motorkabel schon als lange Leitung betrachtet werden. Die Spannung am Anfang der Leitung ist dann nicht mehr gleich der Spannung am Ende (in der 50-Hz-Technik liegt die Grenze bei 50 km).

Für einen Leitungsabschnitt gilt folgendes Ersatzschaltbild (Abb. 86.36). Für hohe Frequenzen kann $R \to 0$ und $G \to 0$ gesetzt werden.

Beispiel

Mit der Ausbreitungsgeschwindigkeit v und dem Wellenwiderstand Z ergeben sich für ein Kabel NYY-I $4 \times 2{,}5\,\text{mm}^2$ mit $R = 14\,\Omega/\text{km}$, $L = 0{,}7\,\text{H/km}$ und $C = 100\text{nF/km}$ folgende Werte:

Ausbreitungsgeschwindigkeit

$$v = \frac{1}{\sqrt{L \cdot C}} \qquad \begin{array}{c|c|c} L & C & v \\ \hline \frac{\text{Vs}}{\text{A}} & \frac{\text{As}}{\text{V}} & \frac{m}{s} \end{array} \qquad (86.110)$$

Wellenwiderstand

$$Z = \sqrt{\frac{L}{C}} \qquad \begin{array}{c|c|c} L & C & Z \\ \hline \frac{\text{Vs}}{\text{A}} & \frac{\text{As}}{\text{V}} & \Omega \end{array} \qquad (86.111)$$

$\Rightarrow Z = 84\,\Omega$ und $v = 120\,\text{m/}\mu\text{s}$. ◀

Ist die Leitung nicht mit dem Wellenwiderstand abgeschlossen, liegt eine Fehlanpassung vor, d. h., Teile des Signals werden am Leitungsende reflektiert. Für einen Umrichter mit $R_i \ll Z$ und einem Motor mit $R_i \gg Z$ ist die Verbindungsleitung beidseitig nicht angepasst. Der Reflexionsfaktor berechnet sich zu:

$$\textit{Umrichterreflexionsgrad} \qquad r = \frac{R_i - Z}{R_i + Z} \cong -1 \qquad (86.112)$$

$$\textit{Motorreflexionsgrad} \qquad r = \frac{R_i - Z}{R_i + Z} \cong 1 \qquad (86.113)$$

Am Motor wird die Spannung erhöht (Faktor 2), am Umrichter wird sie umgedreht.

Ein Spannungssprung wird nach einer Laufzeit $t_1 = l/v$ am Kabelende reflektiert.

$$U_2 = \vec{U}_2 + \overleftarrow{U}_2 = 2 \cdot U_2 \tag{86.114}$$

Am Kabelanfang wird die Welle erneut reflektiert und in der Polarität umgekehrt.

Für hohe Frequenzen wird die Motorinduktivität groß und die Motorkapazität klein. Daher wird die Spannungsflanke verkleinert und auf eine Sinuswelle mit einer Wellenlänge der 4fachen Kabellänge heruntergedämpft. Diese Sinusschwingung nimmt exponential ab.

Frequenz der Sinusschwingung

$$f = \frac{v}{\lambda} = \frac{v}{4l} = \frac{1}{4l \cdot \sqrt{L \cdot C}} \tag{86.115}$$

Für eine Länge von 100 m folgt eine Frequenz von $f = 300\,\text{kHz}$. Durch den Frequenzumrichter wird neben der 50 Hz-Sinusschwingung eine Sinusschwingung von 300 kHz auf das Kabel gegeben. Diese hochfrequente Schwingung führt zu Störstrahlungen auf benachbarte Betriebsmittel.

Soll die Ausbildung einer hochfrequenten Sinusschwingung vernachlässigbar klein bleiben, muss die Anstiegszeit der Spannung über dem 10fachen der Signallaufzeit liegen, also unter 10 m Kabellänge (ohne Filter). Nach VDE 0530 T 1 Beiblatt 2 treten keine Motorschäden auf, wenn $U < 1000\,\text{V}$ und $du/dt < 500\,\text{V}/\mu\text{s}$ ist. Schnelle Halbleiterschalter reduzieren zwar die Schaltverluste im Umrichter, erhöhen aber die Verluste im Motor. Deshalb muss auch die Wärmeklasse der Motoren berücksichtigt werden. Y = 90°; A = 105°; E = 120°; B = 130°; F = 155°; H = 180° (Grenztemperatur). Durch hohe oder sich schnell ändernde Spannung wird die Isolation einer schnelleren Alterung unterzogen! Beim Einsatz von Frequenzumrichtern zur Steuerung eines DAsM müssen einige Fakten beachtet werden:

Mit der Drehzahl ändert sich auch die Kühlwirkung des auf der Motorwelle mitlaufenden Lüfters. Bei kleinen Drehzahlen tritt dadurch eine schlechtere Kühlung des Motors ein, die nur zum Teil durch Abnahme der Eisenverluste ausgeglichen wird. Bei Drehzahlerhöhung über die Nenndrehzahl nehmen die Eisenverluste stärker zu als die Kühlwirkung des Lüfters.

Die Umrichterausgangsspannungen und -ströme sind mit Oberwellen behaftet, die im Motor Zusatzverluste verursachen.

Die Motorgeräusche können durch die Oberwellen höher werden. Der Maximaldrehzahl sind durch mechanische Vorgaben Grenzen gesetzt (Kugellager, kritische Drehzahl).

Die Vorteile gegenüber dem Gleichstrommotor sind: Verschleißfrei, hohe Schutzart, höhere Überlastbarkeit, schneller unsteuerbar.

Baugrößen zurzeit bei *U*-Umrichter bis 900 kV A, bei *I*-Umrichtern bis 3,4 MV A.

86.2.1.3.6 Polumschaltung

Die Umschaltung von Spulen im Ständer erfordert getrennte Wicklungen, woraus sich eine geringere Ausnutzung des Wicklungsmaterials ergibt. Die Aufteilung der Wicklung in verschiedene Bereiche (Dahlander) ermöglicht ebenfalls die Veränderung der Drehzahl. Die Drehzahländerungsmöglichkeiten sind bei diesen Maschinen abhängig von der Polpaarzahl und können nur unter $3000\,\text{min}^{-1}$ liegen. Da die Läuferpolpaarzahl entsprechend der des Ständers ausgeführt werden muss, eignen sich hauptsächlich Käfigläufermotoren für die Polumschaltung. Eine Besonderheit der Dahlanderschaltung ist die Ausführung der Wicklungen als *Polamplitudenmodulationswicklung* (PAM-Wicklung). Diese Wicklungsform lässt eine höhere Ausnutzung des eingesetzten Materials zu.

86.2.1.4 Der Schleifringläufer

Beim Schleifringläufer sind die Läuferspulen gewickelt und über Schleifringe herausgeführt. Es ist somit möglich, die Läuferwicklung zu beschalten oder eine Fremdspannung an diese Wicklung anzuschließen. Ein großer Nachteil des Schleifringläufers im Verhältnis zum Käfigläufer ist der erheblich höhere Wartungsaufwand.

86.2.1.4.1 Anlassverfahren

Die für den Käfigläufer eingesetzten Anlassverfahren können überwiegend auch für den Schleifringläufer eingesetzt werden. Durch die Möglichkeit der Beschaltung der Läuferspulen bietet sich beim Schleifringläufer als Anlassverfahren die Beschaltung mit Vorwiderständen an. Für das Motormoment gilt:

$$\textit{Motormoment} \quad M \approx P_\delta = 3 \cdot I_2' \cdot \frac{R_2'}{s} \tag{86.116}$$

Ein gleiches Moment erhält man, wenn bei einer Veränderung von R_2' der Schlupf s gleichermaßen geändert wird. Hieraus ergibt sich:

$$\frac{R_2'}{s} = \frac{(R_2' + R_v')}{s^*} \tag{86.117}$$

$$R_v = R_2 \left(\frac{s^*}{s} - 1 \right) \quad \text{für } s^* > s \tag{86.118}$$

Durch einen zusätzlichen Läuferwiderstand kann der Anlaufstrom begrenzt und gleichzeitig das Anlaufmoment erhöht werden. Durch eine stufenweise Verringerung der Größe des Vorwiderstandes kann der Motor nun angelassen werden.

Der Anlassvorgang erfolgt durch Widerstandskaskaden, die über Schütze abgeschaltet werden, oder durch einen Thyristorschalter (Abb. 86.37), der den Widerstand stufenlos zu- oder abschalten kann. Bei dieser Anlaufvariante werden zusätzliche Verluste im Vorwiderstand $P_{R_v} = 3 \cdot I_2^2 \cdot R_v$ erzeugt, die abgeführt werden müssen.

Das maximale Anlaufmoment wird erreicht, wenn der Vorwiderstand folgenden Wert hat:

$$\textit{Läufervorwiderstand} \quad R_v = R_2 \left(\frac{1}{s_K} - 1 \right) \tag{86.119}$$

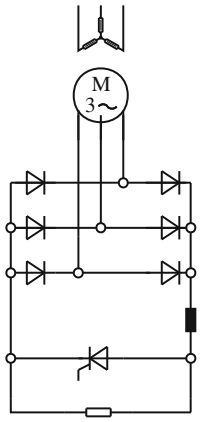


Abb. 86.37 Widerstandsanlassen von Schleifringläufermotoren

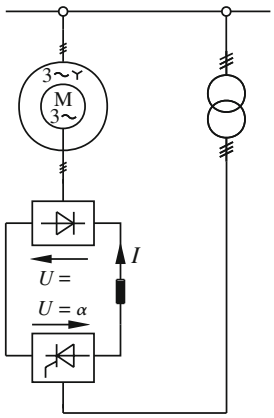


Abb. 86.38 Untersynchrone Kaskade bei Schleifringläufermotoren

Hierbei läuft der Motor mit Kippmoment an. Nachteile dieser Anlaufvariante sind, dass im Vorwiderstand eine erhebliche Verlustleistung erzeugt wird und die Schaltung recht aufwendig ist.

86.2.1.4.2 Bremsverfahren

Zusätzlich zu den bekannten Bremsverfahren des Käfigläufermotors kann beim Schleifringläufer die Kennlinie beim Gleichstrombremsen durch Vorschaltwiderstände so verändert werden, dass der Motor mit Kippmoment abgebremst wird.

86.2.1.4.3 Drehzahlsteuerung

Eine zusätzliche Möglichkeit, die Drehzahl und das Moment des Schleifringläufers zu ändern, bietet die Änderung der Läuferspannung (Untersynchrone Kaskade, USK). Hierbei werden die herausgeführten Läuferspulen an einen Gleichrichter, der eine Gleichspannung erzeugt, angeschlossen (Abb. 86.38).

Diese Gleichspannung ist über einen nachgeschalteten Wechselrichter und Transformator mit dem Netz verbunden. Die erzeugte Gleichspannung verhält sich zur Läuferspannung folgendermaßen:

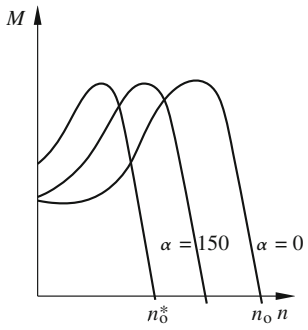


Abb. 86.39 Drehzahl-Drehmomenten Kennlinie für verschiedene Ansteuerwinkel der Untersynchrone Kaskade bei Schleifringläufermotoren

Gleichspannung des Läuferkreises

$$U_{=} = \left(\frac{\sqrt{2}\cdot 3}{\pi}\right)\cdot U_L \tag{86.120}$$

Verändert sich der Schlupf, so verändert sich auch die gleichgerichtete Läuferspannung.

$$U_{=s} = s \cdot U_{=} \tag{86.121}$$

Diese Gleichspannung muss betragsgleich mit der am Wechselrichter erzeugten Spannung sein.

$$\textit{Wechselrichterspannung} \quad U_{=s} = -U_{=\alpha} \tag{86.122}$$

Die am Wechselrichter erzeugte Spannung muss nun wiederum mit der Netzspannung gleich sein.

$$U_{=\alpha} = U_{=T} \cdot \cos\alpha \tag{86.123}$$

α Zündwinkel des Wechselrichters.

$$\textit{Netzspannung} \quad U_{=T} = U_T \cdot \left[\frac{\pi}{(3\cdot\sqrt{2})}\right] \tag{86.124}$$

Da die Läuferspannung am Gleichrichter nicht umgekehrt werden kann, ist mit dieser Schaltung nur untersynchroner Betrieb möglich ($0 \leq s \leq 0{,}5$). Hieraus ergibt sich, dass der Motor mit dieser Beschaltung nicht ohne Hilfsmittel anlaufen kann. Der Motor muss bis zur halben Drehzahl mit einem Anlasswiderstand hochgefahren werden. Aus der Bedingung für die Spannungen im Gleichstromkreis ergibt sich für den Zündwinkel α ein Wert zwischen $90° \leq \alpha \leq 150°$.

Durch diese Steuerung wird die Drehzahl-Drehmomentenkennlinie in ihrer Form erhalten (Abb. 86.39); es stellt sich abhängig vom Ansteuerwinkel eine neue Leerlaufdrehzahl n_0^* ein.

$$n_0^* = n_0 \cdot \left[1 + \left(\frac{U_{=T}}{U_{=}} \cdot \cos\alpha\right)\right] \tag{86.125}$$

Ein Teil der in dem Läufer übertragenen Wirkleistung kann wieder an das Netz abgegeben werden. Da diese Art der Drehzahlsteuerung sehr aufwendig ist, wird sie nur bei größeren Maschinen (20 kW … 20 MW) eingesetzt.

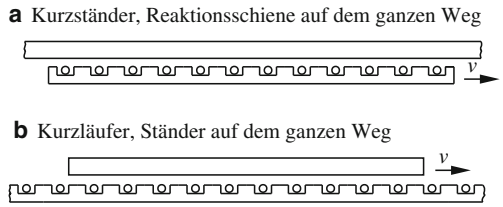


Abb. 86.40 Linearmotor als **a** Kurzständer, **b** Kurzläufermotor

86.2.2 Linearmotor

In der Antriebstechnik wird normalerweise eine geradlinige Bewegung aus einer drehenden Bewegung eines Motors abgeleitet. Der *Linearmotor* formt die elektrische Energie direkt in eine geradlinige Bewegung um, ein Getriebe kann entfallen.

86.2.2.1 Aufbau des Linearmotors

Beim Linearmotor wird der Ständer als flaches, doppeltes oder einseitiges Blechpaket gebaut. Zwischen, oberhalb oder unterhalb des Ständerblechpaketes liegt der gestreckte Läufer (Läuferschiene), die Reaktionsschiene.

Durch die gestreckte Anordnung (Abb. 86.40) wird das Drehfeld in ein *Wanderfeld* (Schubfeld) umgeformt. Neben dem asynchronen Linearmotor kann auch ein Synchron- oder Gleichstromlinearmotor aufgebaut werden. Beide haben in der technischen Anwendung jedoch keine Bedeutung. Beim asynchronen Linearmotor wird im Ständer das Wanderfeld aufgebaut, indem die Spulen durch ein Drehstromnetz gespeist werden. Das Wanderfeld hat eine Synchrongeschwindigkeit v_S. Es wandert während einer Periode um die Strecke eines Polpaares $2{\cdot}\tau_p$ (doppelte Polteilung).

Synchrongeschwindigkeit

$$v_S = 2 \cdot \tau_p \cdot f \qquad \begin{array}{c|c|c} v_S & \tau_p & f \\ \hline \frac{\mathrm{m}}{\mathrm{s}} & \mathrm{m} & \frac{1}{\mathrm{s}} \end{array} \qquad (86.126a)$$

τ_p Polteilung; f Netzfrequenz.

Die Polzahl hat im Gegensatz zum Asynchronmotor keinen Einfluss auf die Synchrongeschwindigkeit. Das Wanderfeld induziert in den Läufer (Reaktionsschiene) eine Spannung, die einen Strom fließen lässt. Strom und Wanderfeld bilden eine magnetische Kraft, die zu einer geradlinigen Bewegung führt. Wie beim Asynchronmotor tritt beim asynchronen Linearmotor ein Schlupf s zwischen Wanderfeld und Bewegungsgeschwindigkeit v auf.

Bewegungsgeschwindigkeit

$$v = (1 - s) \cdot 2 \cdot \tau_p \cdot f \qquad (86.126b)$$

Vorteile eines Linearmotors: Kein Getriebe erforderlich, daher weniger Verluste; Kühlung sehr gut, da der Läufer nur kurz erwärmt wird; Läufer nicht mehr am Gerät; Bremsen wie beim Asynchronmotor.

Nachteile: Luftspalt ist größer als beim Asynchronmotor, dadurch größere Magnetisierungsströme, größere Verluste; Läufer muss den gesamten Weg abdecken, dadurch hohe Kosten; die Speisefrequenz muss variabel sein (Anlauf, Abbremsen). Eine Realisation hat der Linearmotor beim Transrapid gefunden. Hier wird jedoch der Ständer mit dem Läufer vertauscht. In die Schiene sind die Drehstromspulen eingebaut, und im Fahrzeug befindet sich die Läuferschiene. Dieser Aufbau hat den Vorteil, dass dem Fahrzeug die Antriebsenergie nicht über Schleifringe zugeführt werden muss.

86.2.3 Drehstromsynchronmaschinen

Im Unterschied zur Asynchronmaschine (DAsM) wird bei der Synchronmaschine (DSM) das Polrad (Läufer) mit Gleichstrom erregt. Der Ständer besitzt wie bei der DAsM eine Drehstromwicklung. Da der Strom im Polrad nicht über den Luftspalt induziert, sondern von einer Gleichspannung erzeugt wird, ist die Läuferfrequenz $f_2 = 0$. Die Frequenz des Ständers wird aus der Drehzahl und den Polpaarzahlen bestimmt, wenn es sich um einen Generator handelt, sonst bestimmen die Frequenz und die Polpaare die Drehzahl. Ein weiterer Vorteil der DSM ist, dass der zur Erregung des Läufers benötigte Strom nicht aus dem Netz als Blindleistung bezogen werden muss, sondern aus einem Gleichstromnetz stammt. Deshalb kann der Luftspalt der DSM etwas vergrößert werden. Die Größe des Luftspalts der DSM liegt zwischen 0,5 cm und 11,5 cm; im Vergleich dazu bei DASM $<$ 0,3 cm. Beim Einsatz der DSM als Generator wird die Polpaarzahl von der Antriebsmaschine bestimmt. Als Antriebsmaschinen dienen Dieselmaschinen mit $n = 1000 \ldots 1800\,\mathrm{min}^{-1}$, Wasserturbinen mit $n = 100 \ldots 400\,\mathrm{min}^{-1}$ und Dampf- oder Gasturbinen mit $n = 3000\,\mathrm{min}^{-1}$. Die Erregung der DSM kann über eine direkt auf der Welle angebrachte Erregermaschine, meist eine *Außenpolmaschine*, oder über Umrichter direkt aus dem Netz erfolgen. Die DSM kann als *Außen-* oder *Innenpolmaschine* gebaut werden. Bei der Außenpolmaschine liegen die Erregerwicklungen im Ständer, und der Läufer nimmt die Drehstromwicklung auf. Bei der Innenpolmaschine nimmt der Ständer die Drehstromwicklung und der Läufer die Erregerwicklung auf. Ferner wird die DSM als *Schenkelpol* (geringe Drehzahl) oder als *Turboläufer* (hohe Drehzahl) gebaut.

Da die Erregung der DSM aus einem Gleichstromnetz entnommen wird, hängt die Leerlaufspannung nur vom fließenden Erregerstrom ab. Die Abhängigkeit ist im Abb. 86.41 dargestellt.

86.2.3.1 Wirkungsweise der Synchronmaschine

Bei der DSM besteht der Ersatzstromkreis aus dem im Ständer enthaltenen Widerstand, der Streuinduktivität des Ständers und der Spannung U_i (Abb. 86.42).

Die Spannung U_i setzt sich aus dem stromabhängigen Spannungsfall an der Hauptinduktivität und der durch das Polrad induzierten Spannung U_p zusammen. Vernachlässigt man die Abhängigkeit der Hauptinduktivität von der Sättigung und setzt

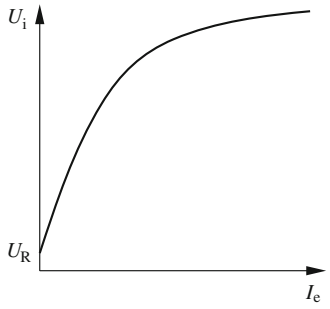


Abb. 86.41 Leerlaufkennlinie U_i-$f(I_e)$ der Synchronmaschine

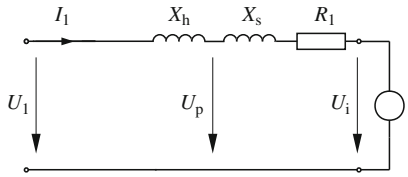


Abb. 86.42 Ersatzschaltbild der Drehstromsynchronmaschine

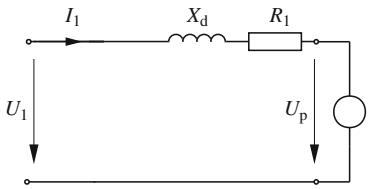


Abb. 86.43 Vereinfachtes Ersatzschaltbild der Drehstromsynchronmaschine

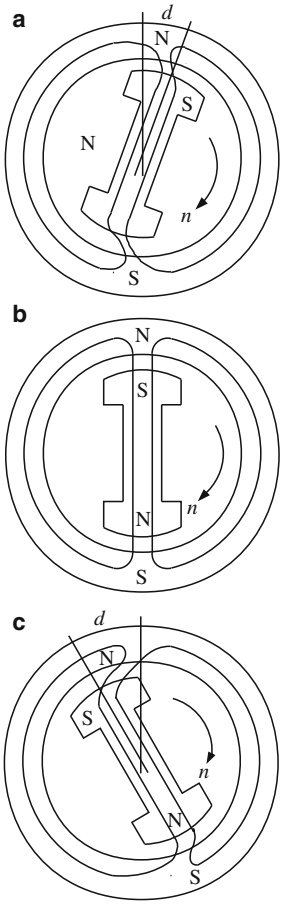


Abb. 86.44 Stellung des Polrades der Synchronmaschine bei **a** Generatorbetrieb, **b** Leerlauf, **c** Motorbetrieb

den Widerstand zu Null, so vereinfacht sich der Ersatzstromkreis der DSM wie im Abb. 86.43 dargestellt.

Die Polradspannung U_p berechnet sich zu:

Polradspannung

$$U_P = B \cdot l \cdot v \cdot N \qquad \begin{array}{c|c|c|c|c} N & v & l & B & U_P \\ - & \frac{m}{s} & \text{m} & \frac{\text{V s}}{m^2} & \text{V} \end{array} \qquad (86.127)$$

Durch entsprechende Ausführung des Polrades erzielt man einen sinusförmigen Verlauf der Polradspannung. Beim DS-Motor hängt die Drehzahl wie bei der DAsM von der Polpaarzahl des Polrades ab ($n = f/p$). Beim DS-Generator hängt die erzeugte Frequenz von der Polpaarzahl und der Drehzahl des Polrades ab ($f = n \cdot p$). Wird die DSM an ein Drehstromnetz angeschlossen, fließt im Ständer ein Drehstrom, der zusammen mit dem induzierten Strom vom Polrad ein resultierendes Drehfeld erzeugt. Dieses Drehfeld rotiert mit synchroner Drehzahl parallel zum Polrad, wenn die DSM nicht belastet wird. Im motorischen Zustand verschiebt sich das Polrad um einen nacheilenden Winkel δ, *dem Polradwinkel*, entsprechend der Last. Hierdurch wird das Polrad vom resultierenden Magnetfeld wie durch ein Gummiband bei synchroner Drehzahl hinterhergezogen.

Im generatorischen Zustand verschiebt sich das Polrad um einen voreilenden Winkel und zieht somit das resultierende Drehfeld hinter sich her (Abb. 86.44). Wird die DSM als Generator eingesetzt, muss die an den Klemmen abgegebene Spannung immer einen konstanten Wert haben. Der Erregerstrom muss ständig der Belastung nachgeführt werden. Das kann eine erhebliche Erhöhung des Erregerstroms gegenüber dem Leerlauf zur Folge haben. Der Erregerstrom kann jedoch nicht immer schnell genug nachgeführt werden. Deshalb können bei plötzlichen Laständerungen Spannungsschwankungen auftreten.

Bei steigender Belastung wird die Klemmenspannung sinken, bei sinkender Belastung wird die Klemmenspannung ansteigen. Im Abb. 86.45 sind verschiedene Betriebszustände einer DSM dargestellt. Aus Abb. 86.45 erkennt man, dass die Spannung U_1 mit der Spannung U_p in Phase liegt, wenn ein reiner Blindstrom fließt. Wird ein Wirkstrom gefordert, verschieben sich die Spannungen zueinander; im motorischen Bereich wandert die Spannung U_p in Richtung der negativen, imaginären Achse (der Winkel δ wird negativ); im generatorischen Bereich wandert die Spannung U_p in Richtung der positiven, imaginären Achse (der Winkel δ wird positiv). Da die Spannung U_p über den Erregerstrom eingestellt wird, kann durch Veränderung des Stromes I_E die Spannung U_p variiert werden. Durch Übererregen ($U_p > U_1$) wird der Strom I_1 so gelegt, dass sich die Maschine wie eine Ka-

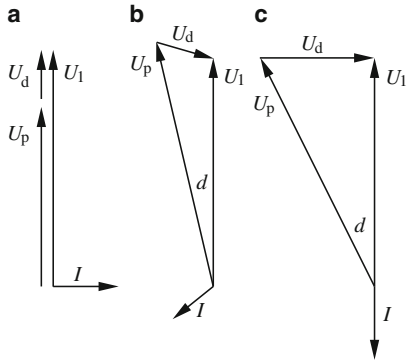


Abb. 86.45 Belastung eines verlustlosen Synchrongenerators **a** im Leerlauf untererregt, **b** bei Belastung kapazitiv, **c** bei rein ohmscher Belastung

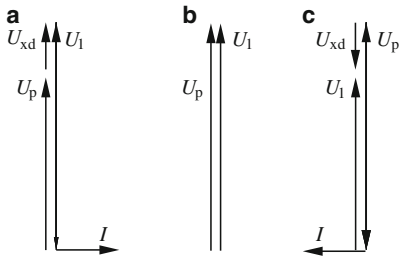


Abb. 86.46 Synchronmaschine im Phasenschieberbetrieb. **a** Untererregt, induktiv, **b** Leerlauf, **c** übererregt kapazitiv

pazität verhält. Durch Untererregen ($U_p < U_1$) verhält sich die Maschine wie eine Induktivität. Man erkennt, dass der Strom I_1 annähernd in Phase oder in Gegenphase mit der Netzspannung U_1 ist. Neben der Wirkleistung wird auch Blindleistung abgegeben oder verbraucht. Dieses ist jedoch nicht beliebig regulierbar.

Regeln für die Synchronmaschine

Blindleistung kann nur über die Erregung gesteuert werden. Wirkleistung kann nur durch Änderung der mechanisch zu- oder abgeführten Leistung an der Welle eingestellt werden. Diese beiden Kriterien können nicht nur einzeln, sondern auch zusammen zum Steuern der Maschine eingesetzt werden.

Hieraus ergeben sich folgende Betriebszustände der Maschine:

In Netzen sind meist induktive Verbraucher vorhanden, somit kann über die DSM ein Ausgleich der Blindleistung erzeugt werden, indem man die Maschine im übererregten Zustand betreibt. Wird im über- oder untererregten Zustand keine Wirkleistung abgenommen, betreibt man den sogenannten *Phasenschieberbetrieb* ($\cos\varphi = 0$). Synchronmaschinen sind generell im Stern geschaltet, da die dritte Oberwelle bei dieser Schaltung nicht auftritt. Bei Dreieckschaltung würde die dritte Oberwelle einen Kreisstrom in den Wicklungen erzeugen, der erhebliche Verluste zur Folge hätte.

86.2.3.2 Spannungsgleichung der Synchronmaschine

Nach vollständigem Ersatzschaltbild lautet die Spannungsgleichung für den Ständer:

Ständerspannung

$$\underline{U}_1 = \underline{I}_1 \cdot (R_1 + jX_{S1}) + [(\underline{I}_1 + \underline{I}_e) \cdot jX_H] \tag{86.128}$$

Polradspannung

$$\underline{U}_P = \underline{I}_e \cdot jX_H \tag{86.129}$$

$$\underline{U}_1 = \underline{I}_1 \cdot (R_1 + j[X_{S1} + X_H]) + \underline{U}_P \tag{86.130}$$

Die Spannung U_p wird als Polradspannung bezeichnet. Sie ist diejenige Spannung, die vom Erregerfeld in die Drehstromspulen induziert wird. Fasst man die Induktivitäten von (86.130) zusammen und vernachlässigt den Widerstand R_1, ergibt sich die vereinfachte Spannungsgleichung der DSM

$$\underline{U}_1 = \underline{I}_1 \cdot jX_d + \underline{U}_P \tag{86.131}$$

86.2.3.3 Anlauf und Synchronisation

Im Leerlaufbetrieb ist der Ständerstrom $I_1 = 0$, die Klemmenspannung gleich der Polradspannung $U_1 = U_p$. Bei offenen Klemmen ist die Klemmenspannung abhängig von der Erregung. Wird die Synchronmaschine an ein Netz konstanter Spannung angeschlossen und die Polradspannung gleich der Klemmspannung eingestellt, so ist der Ständerstrom gleich Null. Eine Erhöhung der Polradspannung hat zur Folge, dass das Spannungsgleichgewicht nicht mehr vorhanden ist. Es muss ein Ständerstrom fließen, der an dem induktiven Widerstand der Maschine einen Spannungsfall U_{Xd} erzeugt. Dieser Spannungsfall stellt das Spannungsgleichgewicht in Phasenlage und Größe wieder her.

Man nennt diesen Betriebszustand *Phasenschieberbetrieb*, da die Maschine bei $U_p > U_1$ einen kapazitiven Strom aus dem Netz aufnimmt (kompensiert) und für $U_P < U_1$ einen induktiven Strom aufnimmt (Abb. 86.46). Die DSM wird als Phasenschieber in Verbrauchernähe eingesetzt, um das Übertragungssystem von der Bereitstellung von Blindleistung zu entlasten. Setzt man die DSM als Motor ein, erfolgt kein Anlauf ohne Hilfseinrichtungen, da das Ständerfeld den Läufer so schnell überstreicht, dass kein resultierendes Moment gebildet werden kann. Bei größeren Maschinen erfolgt der Anlauf mit Anwurfmotoren. Hat der Anwurfmotor die DSM soweit beschleunigt, dass die synchrone Drehzahl erreicht ist, kann die Maschine unter folgenden Bedingungen an das Netz geschaltet werden: gleiche Spannungen; gleiche Phasenlage; gleiche Phasenfolge; gleiche Frequenz.

Eine weitere Hilfseinrichtung ist eine zusätzliche Käfigwicklung im Läufer. Wird die Netzspannung angelegt, läuft die Maschine asynchron hoch und geht nach Erreichen der Leerlaufdrehzahl in den Synchronismus über. Bei diesem Anlaufverfahren ist zu beachten, dass die Erregerwicklungen über einen Widerstand kurzgeschlossen werden müssen, da sonst eine unzulässig hohe Spannung in den Spulen durch das Drehfeld induziert wird. Nach Erreichen der synchronen Drehzahl verhält sich der unbelastete Motor wie ein Phasenschieber. Wird der Motor belastet, muss über den Luftspalt eine Wirkleistung

übertragen werden; das Netz muss Wirkleistung liefern. Die Belastung bremst die Welle ab, es entsteht zwischen Läuferfeld und Ständerfeld ein Winkel. Dieser Winkel erzeugt eine Kraft zwischen den beiden Feldern; mit dem Läuferradius wird ein Moment auf die Welle $M = F \cdot r$ übertragen. Ein Moment bedeutet, dass Wirkleistung übertragen wird. Da die Synchronmaschine „keine" Wirkleistungsverbraucher besitzt, wird die gesamte aus dem Netz aufgenommene Wirkleistung an die Welle weitergegeben.

86.3 Einphasen-Asynchronmotoren

Einphasen-Asynchronmotoren (Wechselstrommotoren) werden wegen des einfachen Einphasenanschlusses dort eingesetzt, wo preiswerte und robuste Motoren benötigt werden. Der Einphasenmotor hat einen ähnlichen Aufbau wie ein Drehstrommotor. Der Läufer ist immer als Kurzschlussläufer ausgebildet. Man unterscheidet den *einsträngigen Motor* und den *zweisträngigen Motor*.

86.3.1 Einsträngiger Motor

Beim einsträngigen Motor besteht die Ständerwicklung im einfachsten Fall nur aus einem Wicklungsstrang, der Arbeitswicklung. Diese Wicklung ist auf die Klemmen U1 und U2 herausgeführt. Sie erzeugt beim Anlegen einer Wechselspannung ein Wechselfeld; der Motor kann nicht anlaufen. Soll der Motor zum Drehen gebracht werden, muss er in eine beliebige Richtung angeworfen werden (Anwurfmotor). Nach diesem Anwurfvorgang bleibt der Motor im Betrieb. Der einsträngige Motor hat sehr schlechte Betriebswerte. Seine abgegebene Leistung beträgt nur etwa 50 bis 60 % der eines gleich großen Drehstrommotors, das Kippmoment liegt bei 40 % eines Drehstrommotors und der Leistungsfaktor ist sehr gering. Aus diesen Gründen wird diese Motorform nur sehr selten eingesetzt.

86.3.2 Zweisträngiger Motor

Beim zweisträngigen Motor wird zusätzlich zur Arbeitswicklung eine um 90° elektrisch versetzte Hilfswicklung angeordnet. Durch diese Hilfswicklung (Klemme Z1 und Z2) sollte ein Strom fließen, der zum Strom durch die Arbeitswicklung eine Phasenverschiebung von 90° hat. Dadurch erreicht man, dass beim Anlegen der Spannungen im Motor ein magnetisches Feld entsteht, das sich ähnlich dem Drehfeld um den Ständer bewegt. Durch die Hilfswicklung kann der Motor selbständig anlaufen; die Betriebseigenschaften verbessern sich. Am häufigsten wird der zweisträngige Motor als *Kondensatormotor* oder als *Spaltpolmotor* ausgeführt.

86.3.3 Kondensatormotor

Wird ein Kondensator vor die Hilfswicklung geschaltet, ist der Strom in der Hilfswicklung voreilend zum Strom in der Arbeitwicklung. Soll bei dieser Anordnung ein Kreisdrehfeld erzeugt werden, müsste die Spannung der Hilfswicklung U_H gegenüber der Spannung der Arbeitswicklung U_A um genau 90° voreilend sein. Die Ströme müssten zu ihren Spannungen die gleiche Phasenlage haben, und die Durchflutung der Wicklungen müsste gleich sein. Diese Bedingungen lassen sich nicht im gesamten Betriebsbereich des Motors verwirklichen, da durch den Schlupf sowohl die Widerstände und Ströme als auch der Phasenwinkel in den beiden Zweigen verändert werden. Man wählt deshalb den Nennbetriebspunkt für die Symmetriebedingung. Als Richtwert für die in den Hilfsstrang einzuschaltende Betriebskapazität C_B gilt: $C_B \approx 25\,\mu F$ bis $35\,\mu F$ je kW Motorleistung an 230 V.

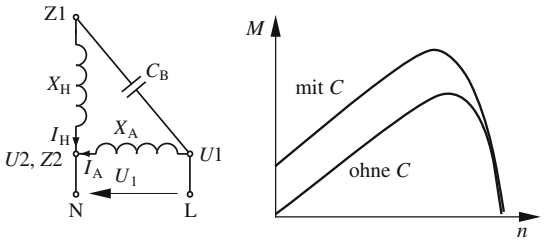


Abb. 86.47 Schaltbild und Drehzahl-, Drehmomentverlauf eines Kondensatormotors mit Betriebskondensator

Tab. 86.6 Kapazität in Abhängigkeit von der Nennspannung

C_B (µF/kW)	U_N (V)
220	127
70	230
25	380

In Abb. 86.47 ist die Schaltung eines Kondensatormotors mit Betriebskondensator dargestellt. Bei der Auswahl der Kondensatoren sind die Vorschriften nach DIN 48501 und VDE 0560 zu beachten, da die Spannung am Kondensator stets größer ist als die Speisespannung des Netzes. In der Praxis kommen Kondensatorspannungen von bis zu 550 V vor, obwohl die Netzspannung nur 230 V beträgt. Kondensatormotore laufen selbständig an und weisen verbesserte Betriebseigenschaften gegenüber dem Anwurfmotor auf. Allerdings ist im Vergleich zum Drehstrommotor das Anlauf- oder Kippmoment noch sehr gering. Eine Erhöhung des Anlaufmomentes kann durch eine Reihenschaltung eines weiteren Kondensators (Anlaufkondensator) zum Betriebskondensator erreicht werden. Dieser Kondensator verschiebt den Symmetriepunkt zum Anlaufpunkt. Der Nachteil dieser Beschaltung ist jedoch, dass durch die hohe Kapazität in Verbindung mit der Spule der Blindwiderstand herabgesetzt wird und damit ein höherer Strom erzeugt wird, der den Motor zu stark erwärmt. Aus diesem Grund wird der Anlaufkondensator über eine geeignete Einrichtung beim Erreichen einer vorgegebenen Drehzahl abgeschaltet.

Der Anlaufkondensator sollte die dreifache Kapazität des Betriebskondensators haben. Im Abb. 86.48 ist die Schaltung eines Kondensatormotors mit Anlauf- und Betriebskondensator dargestellt (Tab. 86.6).

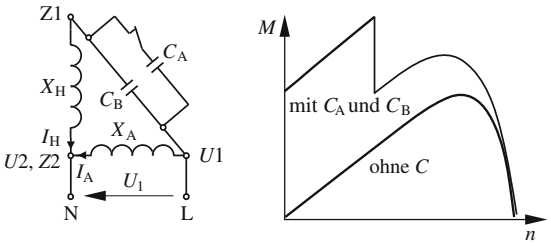


Abb. 86.48 Schaltbild und Drehzahl-, Drehmomentverlauf eines Kondensatormotors mit Betriebs- und Anlaufkondensator

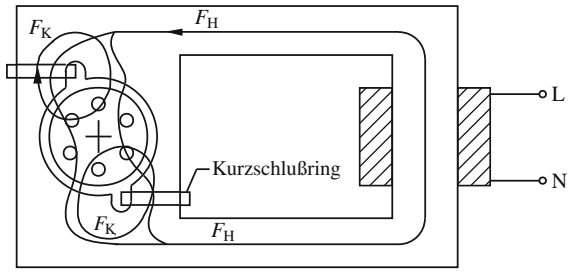


Abb. 86.49 Prinzipdarstellung des Spaltpolmotors mit Kurzschlussring

86.3.4 Spaltpolmotor

Der Spaltpolmotor ist dem Prinzip nach ein Einphasenmotor. Der Läufer hat – wie ein normaler Asynchronmotor – einen Kurzschlusskäfig. Der Ständer ist durch ausgeprägte Pole mit konzentrischen Wicklungen und durch einen Kupfer- oder Aluminiumring, der teilweise vom Hauptfluss durchsetzt wird, gekennzeichnet (Abb. 86.49).

Dieser Ring dient als Hilfsphase. In ihn wird ein Strom induziert, der ein Magnetfeld aufbaut, das dem Hauptfeld entgegengesetzt ist. Durch die Phasenverschiebung zwischen Hauptfluss und Ringfluss entsteht ein sich um den Ständer bewegendes Feld. Dieses Feld ermöglicht ein selbständiges Anlaufen des Spaltpolmotors. Eingesetzt wird der Spaltpolmotor wegen seines geringen Anzugsmomentes für Kleinantriebe.

86.4 Drehstrommotor im Einphasenbetrieb

Der einphasige Betrieb eines Drehstrommotors kann durch eine Störung, aber auch gewollt, herbeigeführt werden. Ohne zusätzliche Beschaltung verhält sich ein Drehstrommotor im Einphasenbetrieb wie ein Einphasen-Asynchronmotor. Er läuft nicht selbständig an, das Motormoment wird kleiner, die aufgenommene Leistung wird bei gleicher Belastung höher als im dreiphasigen Betrieb. Somit steigen die Verluste im Motor, und er erwärmt sich. Soll ein Drehstrommotor am Einphasennetz betrieben werden, muss, wie beim Einphasen-Asynchronmotor, eine Beschaltung mit einem Kondensator erfolgen. Diese Schaltung wird *Steinmetzschaltung* genannt.

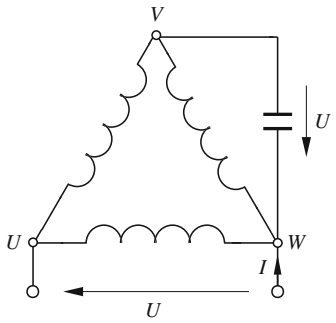


Abb. 86.50 Schaltung eines Asynchronmotors im Einphasenbetrieb (Steinmetzschaltung)

Bei der Steinmetzschaltung wird ein Kondensator entsprechend Abb. 86.50 parallel zu einer Drehstromwicklung geschaltet. Dadurch erzielt man einen selbständigen Anlauf des Motors mit ungefähr 25 bis 35 % des Nennanlaufmomentes. Bei folgender Auslegung der Kapazität kann der Motor mit etwa 75 bis 80 % seiner Nennleistung betrieben werden.

Ein höheres Anlaufmoment erzielt man durch einen Anlaufkondensator, der die zweifache Kapazität des Betriebskondensators aufweist. Dieser Kondensator sollte nach erfolgtem Anlauf abgeschaltet werden.

86.5 Sonderbauformen

Zu den Sonderbauformen von Drehstrommotoren gehören der *Schrittmotor* und der *Servomotor*.

86.5.1 Schrittmotor

Der Schrittmotor entspricht in seinem Aufbau und seiner Wirkungsweise dem Synchronmotor. Mit dem Schrittmotor kann je nach Bauform ein elektrischer Impuls in eine Drehbewegung mit definiertem Winkel umgeformt werden.

Der Schrittmotor ist aus einem Läufer, bestehend aus Permanentmagneten, und einem Ständer mit mehreren, voneinander unabhängigen Spulen, aufgebaut. Im Abb. 86.51 ist ein Schrittmotor in unipolarer Ausführung dargestellt. Werden die Spulen A1 und B1 erregt, nimmt der Läufer die Stellung „I“ ein. Wird anstelle der Spule A1 die Spule A2 erregt, dreht sich der Läufer in die Position „II“. Der Läufer hat seine Position um den Winkel α (Schrittwinkel) verändert. Wird anstelle von B1 die Spule B2 erregt, dreht sich der Läufer wiederum um den Winkel α in die Stellung „III“. Wird als nächster Schaltschritt die Spule A1 wieder erregt, dreht sich der Läufer in die Position „IV“. Der nächste Schritt, Erregung A1 und B1, bringt den Läufer wieder in die Ausgangsposition zurück. Der Schrittmotor hat nach vier Schritten seine Ausgangsposition erreicht. Dieses Beispiel zeigt, dass ein Schrittmotor mit zwei Phasen im Ständer und einem Polpaar im Läufer pro Schritt einen Winkel von 90° zurückle-

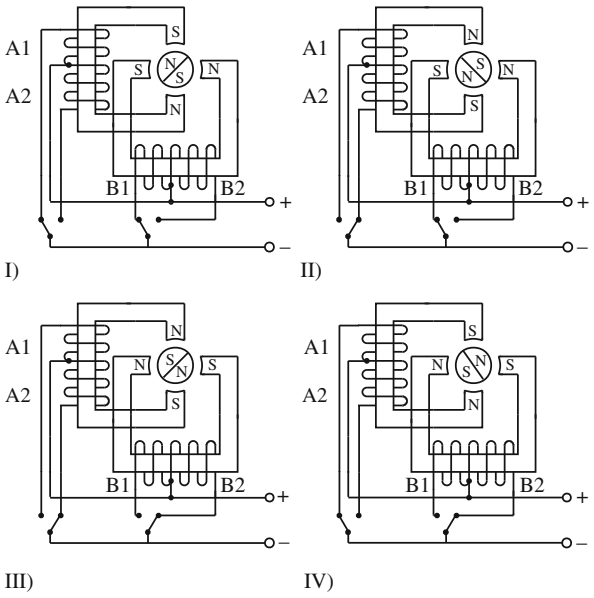


Abb. 86.51 Schrittmotor

gen kann. Der Schrittwinkel α hängt also von der Polpaarzahl und der Phasenzahl des Motors ab.

$$\textit{Schrittwinkel} \quad \alpha = \frac{360°}{2 \cdot m \cdot p} \tag{86.132}$$

p Polpaarzahl des Läufers; m Phasenzahl des Ständers.

Neben dem *Vollschrittbetrieb* (bei einem Schrittmotor mit 90° Schrittwinkel sind vier Schritte für eine Umdrehung notwendig) kann der *Halbschrittbetrieb* eingesetzt werden. Beim Halbschrittbetrieb wird nach jedem zweiten Schritt eine Spule abgeschaltet. Das führt dazu, dass sich der Rotor in eine Zwischenlage (einen halben Schritt) bewegt. Im Halbschrittbetrieb können anstatt der vier Schritte acht Schritte ausgeführt werden. Eine weitere Verfeinerung stellt der *Mini-* oder *Mikroschrittbetrieb* dar. Hierbei wird der Strom in den einzelnen Phasen schrittweise vermindert oder erhöht. Dadurch kann ein Vollschritt in kleine Einzelschritte unterteilt werden. Die genaue Steuerung dieser Schritte benötigt eine umfangreiche Steuerlogik und ist deshalb sehr teuer. Machbar ist heute eine Auflösung eines Vollschritts in bis zu 250 Einzelschritte.

86.5.2 Servomotor

Der Maschinenbau fordert in zunehmendem Maße Motoren, die kurzzeitig die Drehzahl verändern können. Hierzu eignen sich besonders die Servomotoren. Mit ihnen lassen sich Winkel, Wege und Lagen hochdynamisch und beschleunigungsoptimal einstellen. Der Servomotor kann als *Scheibenläufer-* oder *Stabankermotor* gefertigt werden. Mit Ausnahme des Drehstrom-Asynchronmotors sind Servomotoren fast ausschließlich mit Permanentmagneten bestückt.

Das *Impuls-* oder *Spitzendrehmoment* M_{Imp} ist das maximal zulässige Drehmoment, bei dem weder mechanische noch elektrische Grenzwerte überschritten werden. Es kann bis zum zehnfachen Wert des Nennmomentes betragen. Der *Kurzzeit-* oder *Impulsstrom* I_{Imp} ist der Strom, der das Impulsdrehmoment erzeugt. Bei Drehstrommotoren ist die Strombegrenzung auf den Scheitelwert dieses Stroms einzustellen. Der Regelungsfaktor k_N gibt an, wie sich die Motordrehzahl bei konstanter Klemmenspannung ändert, wenn an der Welle eine zusätzliche Belastung auftritt.

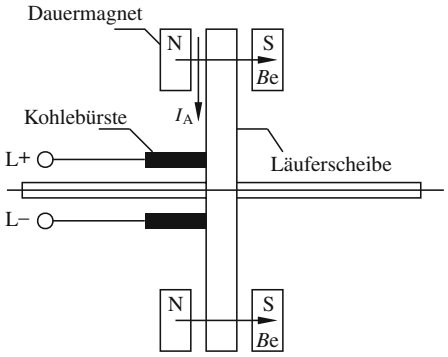


Abb. 86.52 Servoscheibenläufermotor

Regelungsfaktor

$$k_N = \frac{\Delta n}{\Delta M} \qquad \begin{array}{c|c|c} n & M & k_N \\ \hline \text{min}^{-1} & \text{N m} & \frac{1}{\text{min N m}} \end{array} \tag{86.133}$$

Je kleiner der Regelungsfaktor k_N ist, desto weniger Einfluss haben Momentenänderungen auf die Drehzahl.

86.5.2.1 Scheibenläufermotor

Der Scheibenläufermotor ist nach Abb. 86.52 aufgebaut. Das Magnetsystem besteht aus axial angeordneten Dauermagneten, die Ankerscheibe aus Leiterbahnen, die auf die Scheibe aufgeklebt oder durch fotochemische Ätzverfahren aufgebracht werden.

Die Stromzuführung erfolgt über axial angeordnete Kohlebürsten. Im Gegensatz zum herkömmlichen Motor sind die Richtungen von Ankerstrom und Ständerfluss vertauscht, das Funktionsprinzip ist aber erhalten geblieben. Die magnetische Kraft entsteht durch Zusammenwirken des Ständerflusses und des Ankerstroms. Durch die relativ große Scheibenoberfläche kann die erzeugte Verlustwärme sehr gut abgeführt werden. Scheibenläufermotoren sind daher gut geeignet für Anwendungen, die kurze Baulängen, hohe Beschleunigungen, große Wellendurchmesser und geringe Induktivität fordern.

86.5.2.2 Stabankermotoren

Stabankermotoren (Abb. 86.53) gleichen in ihrer Bauweise den herkömmlichen Gleichstrommotoren. Die Erregung wird wie bei den Scheibenläufermotoren durch Dauermagnete erzeugt.

Durch geeignete Formgebung des Ankers und der Kommutatorstege kann der Stabankermotor bis zu Drehzahlen von 0,2 min^{-1} eingesetzt werden. Kleine Ankerdurchmesser ergeben

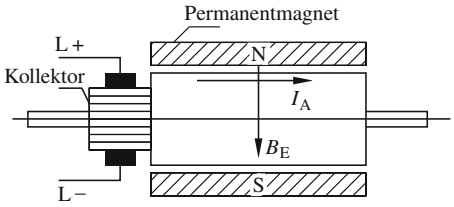


Abb. 86.53 Servo-Stabankermotor

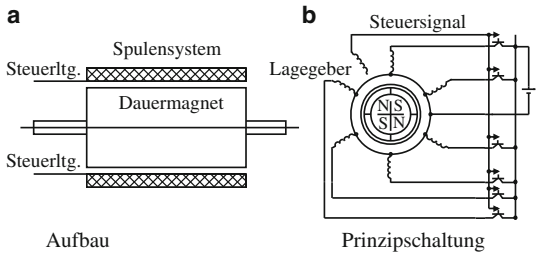


Abb. 86.54 Bürstenloser Servomotor

kleine Massenträgheitsmomente und somit kleine mechanische Zeitkonstanten. Nachteil des Stabankermotors ist, dass das Spitzendrehmoment im Bereich hoher Drehzahlen zu reduzieren ist, da die Kommutierungsgrenze leicht erreicht wird (Strom muss gesenkt werden $\Rightarrow U_i \sim \Delta I/\Delta t$). Ferner treten bei diesem Motortyp Bürsten- und Kommutatorverschleiß auf (wartungsintensiv). Eine Alternative zu dem Stabankermotor mit Kommutator ist der *bürstenlose Gleichstrom-Servomotor* (Abb. 86.54).

Bei diesem Motortyp ist die Erregung in den Rotor und die Wicklungen in den Stator eingebaut. Die Drehbewegung wird durch eine elektronische Kommutierungseinrichtung erzeugt, die die Statorspulen entsprechend der vorgegebenen Drehzahl mit einem getakteten Gleichstrom versorgt. Ein reibungsloser Betrieb dieser Maschine ist nur mit einem Lagegeber möglich. Der Lagegeber meldet der Steuerelektronik die jeweilige Lage des Rotors. Vorteile des bürstenlosen Gleichstrom-Servomotors sind wegfallende Bürstengeräusche und kein Verschleiß oder Abbrand an den Bürsten, stufenlos steuerbares Moment und Drehzahl, schneller Reversierbetrieb und hohe Schutzart.

86.6 Gleichstrommaschinen

Gleichstrommaschinen werden wegen ihrer günstigen Betriebseigenschaften für regelbare Antriebe eingesetzt. Nachteil der Gleichstromantriebe gegenüber den Drehstromantrieben ist, dass für die Rotorwicklung ein Stromwender vorhanden sein muss. Trotz der daraus resultierenden hohen Wartungskosten hat sich der Gleichstrommotor bei großen Leistungen gegenüber den umrichtergesteuerten Drehstrommotoren behauptet.

86.6.1 Aufbau und Wirkungsweise

Der Aufbau einer Gleichstrommaschine ist im Abb. 86.55 dargestellt. Ein drehbar gelagerter Rotor (Anker) ist mit Spulen

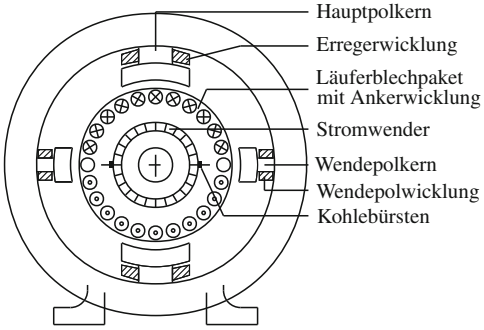


Abb. 86.55 Gleichstrommaschine

bewickelt, die die Spannung über einen Schleifkontakt (Kollektor, Bürsten) zugeführt bekommen.

Der Anker dreht durch ein Magnetfeld, das von feststehenden Spulen (Stator) erzeugt wird. Beim Antrieb der Gleichstrommaschine wird in den Ankerspulen eine Spannung U_i induziert. Bei Beschaltung des Kollektors mit einem Widerstand treibt diese Spannung einen Strom durch die Spule. Die am äußeren Widerstand messbare Spannung errechnet sich aus der Induktionsspannung und dem Spannungsfall der Spule.

$$\textit{Klemmenspannung} \quad U = U_i - I \cdot R_a \tag{86.134}$$

R_a Widerstand im Ankerkreis.

Die Gleichstrommaschine arbeitet also als Generator. Wird dagegen an den Kollektor eine Spannung angelegt, treibt diese Spannung einen Strom durch die Ankerspulen. Der Ankerstrom erzeugt mit dem Statorfeld ein magnetisches Moment, die Maschine fängt an, sich zu drehen. Dreht sich die Maschine, induziert das Statorfeld wiederum eine Spannung in den Ankerwicklungen, die den Strom verringert.

$$\textit{Netzspannung} \quad U = U_i + I \cdot R_a \tag{86.135}$$

Die Gleichstrommaschine arbeitet als Motor. Die induzierte Spannung U_i berechnet sich zu

induzierte Spannung

$$U_i = B \cdot l \cdot v \qquad \begin{array}{c|c|c|c} U_i & B & l & v \\ \hline \mathrm{V} & \frac{\mathrm{Vs}}{\mathrm{m^2}} & \mathrm{m} & \frac{\mathrm{m}}{\mathrm{s}} \end{array} \tag{86.136}$$

$$U_i = -N \cdot \frac{\Delta \Phi}{\Delta t} \qquad \begin{array}{c|c|c|c} U_i & N & \Phi & t \\ \hline \mathrm{V} & - & \mathrm{Vs} & \mathrm{s} \end{array} \tag{86.137}$$

Bei der Gleichstrommaschine sind die mechanischen Daten des Stators und des Ankers konstant, sie verändern sich während des Betriebes nicht. Die Veränderung des magnetischen Flusses nach der Zeit kann über die Drehzahl erfasst werden. Gleichung (86.137) für die in den Ankerspulen induzierte Spannung verändert sich unter diesen Voraussetzungen:

$$U_i = k_1 \cdot \Phi \cdot n \qquad \begin{array}{c|c|c|c} U_i & k_1 & \Phi & n \\ \hline \mathrm{V} & - & \mathrm{Vs} & \mathrm{s^{-1}} \end{array} \tag{86.138}$$

k_1 Maschinenkonstante der Gleichstrommaschine.

Die Leistungsbilanz der Gleichstrommaschine ergibt sich, wenn die Spannungsgleichungen (86.135) und (86.136) mit dem fließenden Strom multipliziert werden.

$$\textit{Generator} \qquad U \cdot I = I \cdot U_i - I^2 \cdot R_a \tag{86.139}$$

$$P_{\text{el}} = P_{\text{mech}} - P_v \tag{86.140}$$

$$\textit{Motor} \qquad U \cdot I = I \cdot U_i + I^2 \cdot R_a \tag{86.141}$$

$$P_{\text{el}} = P_{\text{mech}} + P_v \tag{86.142}$$

Das an der Gleichstrommaschine vorhandene Drehmoment berechnet sich zu:

$$\textit{Drehmoment} \quad M = \frac{P_{\text{mech}}}{2 \cdot \pi \cdot n} \qquad \begin{array}{c|c|c} M & P & n \\ \hline \text{N m} & \text{W} & \text{s}^{-1} \end{array} \tag{86.143}$$

Gleichung (86.142) eingesetzt in (86.143) ergibt:

$$\textit{Drehmoment} \quad M = \frac{U_i \cdot I}{2 \cdot \pi \cdot n} \qquad \begin{array}{c|c|c|c} U_i & I & n & M \\ \hline \text{V} & \text{A} & \text{s}^{-1} & \text{N m} \end{array} \tag{86.144}$$

Wird die induzierte Spannung durch (86.138) ersetzt, ergibt sich für das Drehmoment der Gleichstrommaschine:

$$M = \frac{k_1 \cdot \Phi \cdot n \cdot I}{2 \cdot \pi \cdot n} \tag{86.145}$$

$$M = \frac{k_1 \cdot \Phi \cdot I}{2 \cdot \pi} \tag{86.146}$$

Beispiel

Von einem Gleichstrommotor sind folgende Werte bekannt: $U_N = 200\,\text{V}$; $I_N = 10\,\text{A}$; $P_N = 1{,}9\,\text{kW}$; $n_N = 1800\,\text{min}^{-1}$. Wie groß ist der Ankerwiderstand R_a, die im Nennbetrieb induzierte Spannung U_{iN} und das Nennmoment M_N?

$$M_N = \frac{P_N}{2 \cdot \pi \cdot n_N} = \frac{1900\,\text{V A}}{2 \cdot \pi \cdot 1800\,\text{min}^{-1} \cdot \frac{1\,\text{min}}{60\,\text{s}}} = 10{,}08\,\text{W s}$$

$$U_{iN} = \frac{P_N}{I_N} = \frac{1900\,\text{V A}}{10\,\text{A}} = 190\,\text{V}$$

$$U_N = U_{iN} + I_N \cdot R_a$$

$$\Rightarrow \quad R_a = \frac{U_N - U_{iN}}{I_N} = \frac{200\,\text{V} - 190\,\text{V}}{10\,\text{A}} = 1\,\Omega \quad \blacktriangleleft$$

86.6.1.1 Ankerrückwirkungen

Durchströmt ein Strom die Ankerspule, wird von ihm ein Magnetfeld aufgebaut, das quer zum Feld des Stators liegt. Beide Felder ergeben zusammen ein nicht mehr symmetrisch von einem Statorpol zum anderen verlaufendes, sondern ein verzerrtes Feld (Abb. 86.56).

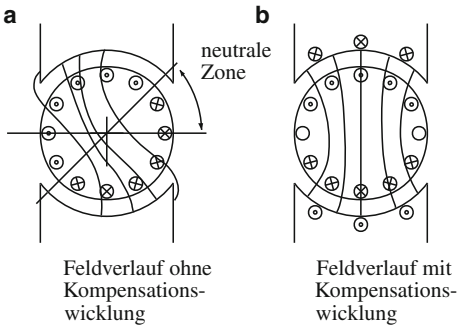


Abb. 86.56 Feldverlauf Gleichstrommaschine

Der Stromwender liegt nicht mehr in der *neutralen Zone*, also nicht mehr im Winkel von 90° zum Gesamtfeld der Maschine. Dadurch wird am Generator nicht mehr die maximale Spannung abgenommen und bei Motorbetrieb nicht mehr das maximale Drehmoment erzeugt. Ferner erfolgt die Stromwendung nicht mehr im spannungslosen Zustand der Wicklung. Durch den bei der Stromwendung auftretenden Kurzschluss kann es zum *Bürstenfeuer* an den Kommutatorstegen kommen. Diese Ankerrückwirkungen lassen sich durch Einbau der *Wendepole* beseitigen. Wendepole sind immer direkt im Bereich der Stromwendung angebracht. In diesen Bereichen wird der Strom in den Spulen von $+I$ nach $-I$ oder umgedreht geändert. Da diese Stromänderung relativ schnell erfolgt, wird in der jeweiligen Spule eine Spannung erzeugt ($U_i \sim di/dt$). Je höher die Drehzahl der Maschine, desto größer die induzierte Spannung. Das hätte eine Funkenbildung zwischen den Bürsten und Kommutatorlamellen zur Folge (Bürstenfeuer). Zur Vermeidung des Bürstenfeuers wird durch die Wendepole ein Magnetfeld aufgebaut, das eine Gegenspannung im Bereich der Stromwender erzeugt, die der induzierten Spannung entgegenwirkt. Der Wendepol wird vom Ankerstrom durchflossen, damit beide induzierten Spannungen bei unterschiedlichen Betriebszuständen gleich sind. Eine weitere Aufgabe der Wendepole ist, die Verschiebung des Hauptfeldes im Bereich der Bürsten zu kompensieren. Bei großen, schnelllaufenden Maschinen können die Ankerrückwirkungen nicht allein durch die Wendepole aufgefangen werden. Deshalb müssen zusätzlich *Kompensationswicklungen* eingebaut werden. Die Kompensationswicklungen sind in den Polschuhen der Maschine untergebracht und werden ebenfalls vom Ankerstrom durchflossen. Durch die Kompensationswicklung werden Feldverzerrungen im Bereich der Polschuhe verhindert.

86.6.2 Betriebsverhalten von Gleichstrommaschinen

Die Gleichstrommaschinen unterscheidet man nach der Zusammenschaltung der Spulensysteme des Ankerkreises und des Stators. Haben Anker und Stator eine separate Spannungsversorgung, handelt es sich um eine *fremderregte Maschine*. Haben beide Systeme dieselbe Spannungsversorgung und sind parallelgeschaltet, handelt es sich um eine *Nebenschlussmaschine*. Sind die Anker- und die Statorspule in Reihe geschaltet, handelt es sich um eine *Reihenschlussmaschine*.

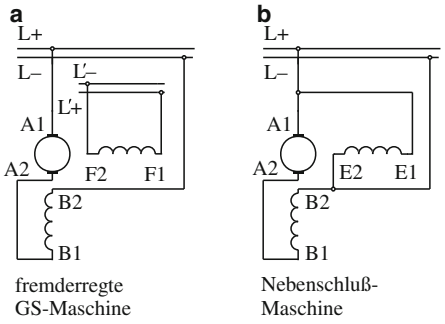


Abb. 86.57 Schaltungen von Gleichstrom-Maschinen

86.6.2.1 Nebenschlussmotor

Für den fremderregten Motor (Nebenschlussmotor) gelten die im Abschn. 86.6.1 angeführten Grundgleichungen. In Abb. 86.57 sind die Schaltungen der Maschinen dargestellt.

Die Erregerwicklung liegt jeweils an einer konstanten Spannung. Damit sind Erregerstrom und Erregerfeld konstant (Ankerrückwirkungen vernachlässigt). Nach (86.138) ergibt sich für die Drehzahl

$$\textit{Drehzahl} \quad n = \frac{U_i}{\Phi \cdot k_1} \tag{86.147}$$

$$n = \frac{U - I \cdot R_a}{\Phi \cdot k_1} \tag{86.148}$$

$$n = \frac{U}{\Phi \cdot k_1} - \frac{I \cdot R_a}{\Phi \cdot k_1} \tag{86.149}$$

Durch Ersetzen von I durch die umgeformte Gleichung (86.147) folgt

$$n = \frac{U}{\Phi \cdot k_1} - \frac{M \cdot R_a}{\Phi^2 \cdot k_1 \cdot k_2} \tag{86.150}$$

Maschinenkonstante

$$k_2 = \frac{k_1}{2 \cdot \pi} \tag{86.151}$$

Der Motorbetrieb stellt sich bei positivem Moment und positiver Drehzahl ein (Abb. 86.58). Der Nennbetrieb (U_N; Φ_N; R_a) liegt im Bereich von $0{,}9 \ldots 0{,}99 \cdot n_0$ (kleine bis große Maschinen). Aus diesem Grund wird nur der Teil der Kennlinie im Bereich der Leerlaufdrehzahl betrachtet. Für den Leerlaufpunkt ($M = 0$) gilt:

$$\textit{Leerlaufdrehzahl} \quad n_0 = \frac{U_N}{k_1 \cdot \Phi_N} \tag{86.152}$$

Für den Anlaufpunkt ($n = 0$) gilt:

$$\textit{Anlaufmoment} \quad M_A = \frac{k_2 \cdot \Phi_N \cdot U_N}{R_a} \tag{86.153}$$

Da die induzierte Spannung im Anlauf gleich Null wird, hängt die Größe des Anlaufstroms nur noch von U_N und R_a ab. Hieraus

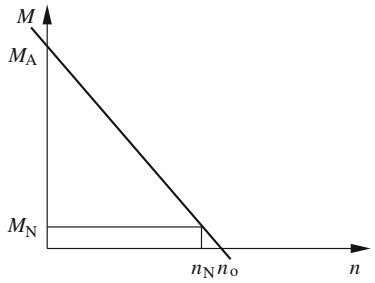


Abb. 86.58 Drehzahl-, Drehmomenten-Kennlinie der Nebenschluss-Maschine

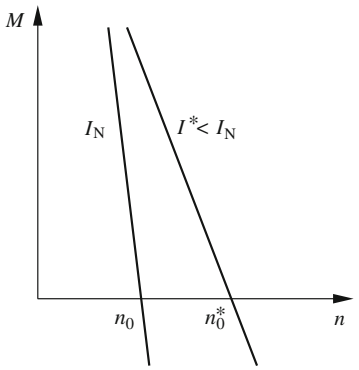


Abb. 86.59 Drehzahl-, Drehmomenten-Kennlinie der Nebenschluss-Maschine in Abhängigkeit von der Erregung

ergibt sich, dass ein fremderregter Gleichstrommotor im Anlauf einen Zusatzwiderstand zur Strombegrenzung erhalten muss. Bereiche oberhalb vom Anlaufmoment auf der Kennlinie sind nicht zugelassen. Hierbei müsste das Lastmoment größer als das Anlaufmoment werden. Der Motor würde entgegen seiner ursprünglichen Richtung gedreht, der entstehende Strom wäre so groß, dass die Motorwicklungen verbrennen würden. Eine *Drehzahlsteuerung* bei der fremderregten Gleichstrommaschine und der Gleichstromnebenschlussmaschine ist durch folgende Maßnahmen möglich:

Änderung der Spannung, des Flusses, des Ankerwiderstandes oder der magnetischen Erregung. Da die Nennerregung meist im Bereich der Sättigung liegt, kann eine Änderung nur durch Schwächung der Erregung erfolgen. Das wird durch einen Vorwiderstand im Erregerkreis möglich. Durch eine Feldschwächung verschiebt sich die Leerlaufdrehzahl zu höheren Werten.

Wenn mit Feldschwächung gearbeitet wird, muss darauf geachtet werden, dass die Drehzahl nicht in Bereiche gelangt, die zur mechanischen Zerstörung der Maschine führt (Abb. 86.59). Das Anlaufmoment wird kleiner, da es proportional von Φ abhängig ist. Soll bei höherer Drehzahl das gleiche Lastmoment abgenommen werden, muss der Ankerstrom erhöht werden. Das geht für den Dauerbetrieb jedoch nur bis zu I_n. Es folgt für das Moment:

$$M^* = k_2 \cdot \Phi^* \cdot I_N \tag{86.154}$$

$$\frac{M^*}{M_N} = \frac{\Phi^*}{\Phi_N} \tag{86.155}$$

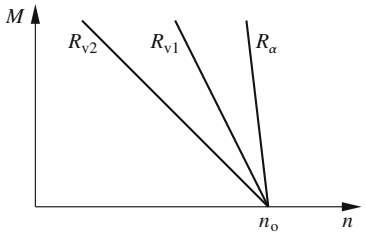


Abb. 86.60 Drehzahl-, Drehmomenten-Kennlinie der Nebenschluss-Maschine in Abhängigkeit vom Ankerkreiswiderstand

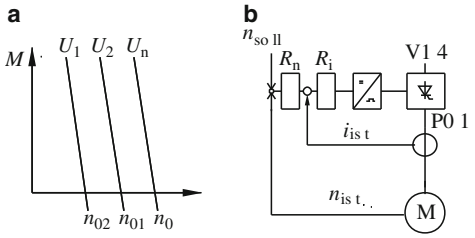


Abb. 86.61 **a** Drehzahl-, Drehmomenten-Kennlinie der Nebenschluss-Maschine in Abhängigkeit von der Spannung, **b** Thyristorschaltung zur Spannungsänderung

Durch die Erhöhung der Drehzahl erfolgt die Kommutierung an den Wendepolen schneller, und die Stromänderungsgeschwindigkeit nimmt zu (dI/dt). Hieraus folgt eine höhere Induktionsspannung, die zu Funkenüberschlägen an den Kommutatorlamellen und zum Bürstenrundfeuer führen kann. Aus diesen Gründen wird die maximale Drehzahl in der Praxis auf $3 \cdot n_0$ beschränkt. Mit einem zusätzlichen Widerstand R_v im Ankerkreis wird der Ankerstrom heruntergesetzt. Da das Moment direkt proportional zum Ankerstrom ist $(M \sim I_a)$, folgt für die Abhängigkeit vom Vorwiderstand:

$$M \sim \frac{1}{R_a + R_v} \tag{86.156}$$

Eine Drehzahlsteuerung der fremderregten Gleichstrommaschine wird in der Praxis kaum durchgeführt, da durch Vorschalten eines Widerstandes zusätzliche Verluste an diesem entstehen $(P_{vz} = I^2 \cdot R_a)$. Ferner ist die Drehzahl im Nennbereich durch Widerstandsregelung kaum variierbar.

Der Zusatzwiderstand dient nur zum Hochfahren der Maschine, um einen hohen Anlaufstrom zu vermeiden (Kennlinie Abb. 86.60). Durch die Änderung der Ankerspannung kann die Drehzahl der Maschine kontinuierlich von Null bis zur Nenndrehzahl gefahren werden (Abb. 86.61).

Durch Umpolen der Ankerspannung ist auch ein Drehzahlbereich von $-n_0 \Leftrightarrow +n_0$ möglich. Die Spannungsregelung wurde früher mit einem Leonard-Umformer realisiert. Heute werden Thyristorschaltungen eingesetzt. Durch den Einsatz der modernen Elektronik ist eine Rückspeisung von Energie in das Netz möglich, wenn der Gleichrichter bei Lastbremsung als Wechselrichter eingesetzt wird. Die Steuerung der Drehzahl durch Spannungsänderung ist ohne Zusatzverluste möglich!

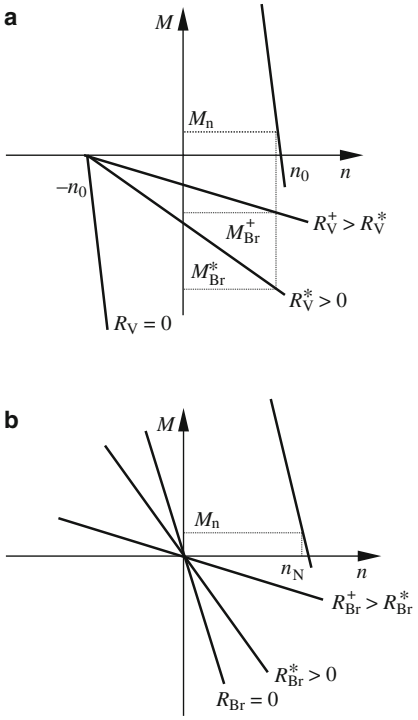


Abb. 86.62 Drehzahl-, Drehmomenten-Kennlinie der Nebenschluss-Maschine **a** beim Gegenstrombremsen **b** beim Widerstandsbremsen

Das Abbremsen von *fremderregten* Gleichstrommaschinen kann auf verschiedene Arten erfolgen. Beim *mechanischen Bremsen* oder *Wirbelstrombremsen* wird Energie auf ein Bremssystem gegeben, das direkt auf die Welle wirkt. Beim *Nutzbremsen* muss die Drehzahl über der Leerlaufdrehzahl liegen. Dann werden Moment und Ankerstrom negativ. Diese Bremsart wird beispielsweise bei Gefällestrecken ausgenutzt (Förderbänder, Triebfahrzeuge). Der Motor geht in den generatorischen Zustand über. Beim *Widerstandsbremsen* wird der Motor vom Netz abgeschaltet, und im Ankerkreis wird ein Widerstand eingeschaltet. Aus der Maschengleichung folgt:

$$U = U_i + I \cdot R_a \quad \Rightarrow \quad U = 0 \quad \Rightarrow \quad U_i = -I \cdot R_a \tag{86.157}$$

Im Bremsbetrieb wird der Strom umgekehrt (Momentenumkehr). Beim *Gegenstrombremsen* wird die Ankerspannung umgekehrt. Hierdurch kehren sich auch Strom und Moment um. Die Gleichung hierfür lautet:

$$\begin{aligned} -U = I^* \cdot R_a + U_i^* \quad &\Rightarrow \quad U_i^* = U_i \\ &\Rightarrow \quad -I^* \cdot R_a = U + U_i \end{aligned} \tag{86.158}$$

Damit der Strom (I^*) im Bremsbetrieb nicht zu groß wird, muss ein Bremswiderstand in den Ankerkreis geschaltet werden.

Beim Gegenstrombremsen muss im Nulldurchgang der Drehzahl abgeschaltet werden, da die Maschine sonst beschleunigt wird und entgegengesetzt anläuft. Im Abb. 86.62 sind verschiedene Bremskennlinien dargestellt.

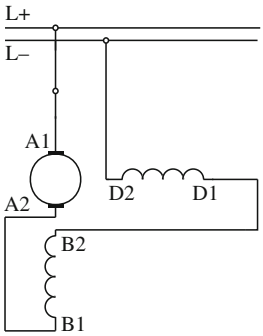


Abb. 86.63 Schaltung einer Reihenschluss-Maschine

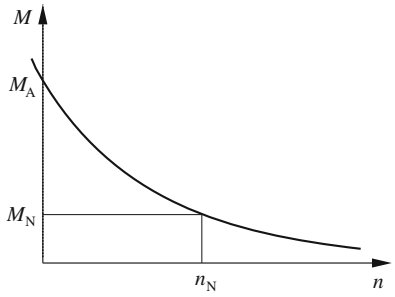


Abb. 86.64 Drehzahl, Drehmomenten-Kennlinie der Reihenschluss-Maschine

Beispiel

Von einem Gleichstromnebenschlussmotor sind bekannt: $U_N = 200\,\text{V}$; $I_N = 10\,\text{A}$; $P_N = 1{,}9\,\text{kW}$; $n_N = 1800\,\text{min}^{-1}$; $R_a = 1\,\Omega$; $U_{iN} = 190\,\text{V}$. Wie groß ist die Leerlaufdrehzahl n_0, der Anlaufstrom I_a, der Strom beim Gegenstrombremsen $I*$ und der Strom beim Widerstandsbremsen ohne Zusatzwiderstand?

$$n_N = \frac{U_{iN}}{\Phi_N \cdot k_1}$$

$$\Rightarrow \quad \Phi_N \cdot k_1 = \frac{U_{iN}}{n_N} = \frac{U_{i0}}{n_0}$$

$$\Rightarrow \quad n_0 = \frac{U_{i0}}{U_{iN}} \cdot n_N = \frac{200\,\text{V}}{190\,\text{V}} \cdot 1800\,\text{min}^{-1}$$

$$= 1894{,}7\,\text{min}^{-1}$$

$$U_N = I_A \cdot R_a + U_{iA}$$

$$\Rightarrow \quad I_A = \frac{U_N - U_{iA}}{R_a} = \frac{200\,\text{V} - 0\,\text{V}}{1\,\Omega} = 200\,\text{A}$$

$$-I^* \cdot R_a = U_N + U_{iN}$$

$$\Rightarrow \quad I^* = -\frac{U_N + U_{iN}}{R_a} = \frac{200\,\text{V} + 190\,\text{V}}{1\,\Omega} = 390\,\text{A}$$

$$U_{iN} = -I \cdot R_a$$

$$\Rightarrow \quad I = \frac{U_{iN}}{R_a} = \frac{190\,\text{V}}{1\,\Omega} = 190\,\text{A}$$ ◀

86.6.2.2 Reihenschlussmotor

Die Schaltung des Reihenschlussmotors ist in Abb. 86.63 dargestellt.

Es gelten die Grundgleichungen der Gleichstrommaschine. Für den *Ankerkreiswiderstand* gilt jedoch:

$$\textit{Ankerkreiswiderstand} \qquad \sum R_a = R_a + R_b + R_d \qquad (86.159)$$

R_b Widerstand der Wendepolwicklung; R_d Widerstand der Erregerwicklung.

Da die Erregerwicklung in Reihe mit dem Ankerkreis geschaltet ist, gilt hier die Beziehung $\Phi = f(I)$. Für den linearen Bereich der Magnetisierungskennlinie kann die Beziehung in eine linearabhängige Form überführt werden:

$$\textit{Erregerfluss} \quad \Phi = c^* \cdot I \qquad (86.160)$$

Gleichung (86.139) und (86.160) ergeben:

Induktionsspannung

$$U_i = k_1 \cdot \Phi \cdot n = k_1 \cdot c^* \cdot I \cdot n = k_3 \cdot I \cdot n \qquad (86.161)$$

Maschinenkonstante

$$k_3 = k_1 \cdot c^* \qquad (86.162)$$

Weiter kann für die mechanische Leistung und das Moment folgende Umformung der Grundgleichung vorgenommen werden:

$$\textit{mechanische Leistung} \qquad P_{\text{mech}} = k_3 \cdot n \cdot I^2 \qquad (86.163)$$

$$\textit{Motormoment} \qquad M = k_4 \cdot I^2 \qquad (86.164)$$

$$\textit{Maschinenkonstante} \qquad k_4 = \frac{k_3}{2 \cdot \pi} \qquad (86.165)$$

Aus (86.163) und (86.164) geht hervor, dass beide vom Quadrat des Stromes abhängig sind. Eine Drehmoment-Drehzahlabhängigkeit kann über die Spannungsgleichung hergestellt werden.

$$\textit{Motordrehzahl} \qquad n = \frac{U}{k_3 \cdot I} - \frac{\sum R_a}{k_3} \qquad (86.166)$$

$$\textit{Motormoment} \qquad M = \frac{k_4 \cdot U^2}{(k_3 \cdot n + \sum R_a)^2} \qquad (86.167)$$

Aus (86.167) und der dazugehörigen Kennlinie (Abb. 86.64) erkennt man, dass die Drehzahl für ein sinkendes Gegenmoment steigt. Für ein Gegenmoment $M_G = 0$ steigt die Drehzahl $n \rightarrow \infty$; der Motor „geht durch“. Der Reihenschlussmotor muss also immer mit einem Gegenmoment betrieben werden, damit er nicht durch zu hohe Drehzahlen zerstört wird. Da das Drehmoment $M \sim I^2$ ist, steht beim Anlauf ein sehr hohes Moment zur Verfügung. Der Reihenschlussmotor wird für Antriebe eingesetzt, die im Anlauf ein hohes Moment erfordern (Elektrofahrzeuge, Hebezeuge). Eine Drehzahlsteuerung

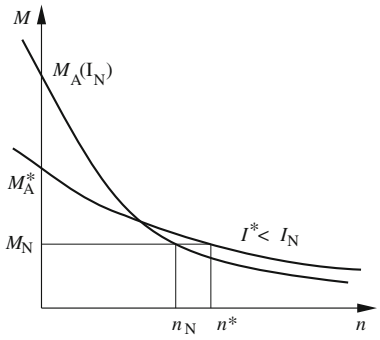


Abb. 86.65 Drehzahl, Drehmomenten-Kennlinie der Reihenschluss-Maschine bei verringerter Erregung

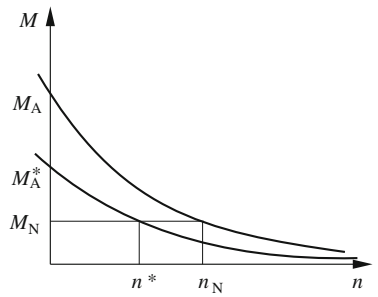


Abb. 86.66 Drehzahl, Drehmomenten-Kennlinie der Reihenschluss-Maschine bei verringerter Spannung

des Reihenschlussmotors erfolgt über eine Feldschwächung, eine Spannungsänderung oder einen Vorwiderstand im Ankerkreis.

Eine Schwächung des Erregerfeldes kann nur erfolgen, wenn ein Widerstand parallel zur Spule geschaltet wird. Dieser Widerstand R_p muss niederohmig ausgeführt werden, da R_D auch niederohmig ist.

Durch eine Feldschwächung wird das Anlaufmoment ebenfalls geschwächt. Im Nennbetrieb erfolgt bei gleichem Moment eine Drehzahlerhöhung (Abb. 86.65). Eine Spannungsänderung (Absenkung) bewirkt, dass weniger Strom durch die Maschine fließt und somit eine Verkleinerung des Anlaufmoments auftritt. Ebenfalls wird bei gleichem Moment die Drehzahl der Maschine reduziert.

Die Drehzahl-Drehmomentenkurve (Abb. 86.66) wird zu kleineren Werten verschoben.

Eine Widerstandsänderung hat die gleichen Auswirkungen auf das Verhalten der Maschine wie eine Spannungsänderung.

Bei der Reihenschlussmaschine ist keine Nutzbremsung möglich (keine konstante Spannung). Das Abbremsen der Maschine erfolgt über die *Widerstands-* oder *Gegenstrombremsung*. Bei der Widerstandsbremsung wird der Motor vom Netz genommen und der Ankerkreis umgepolt. Da der Stromfluss bei der Netztrennung unterbrochen wird, die Maschine aber mit ungefähr gleicher Drehzahl weiter läuft, erzeugt die verbleibende Remanenzspannung einen Stromfluss im Motor. Dieser Strom fließt

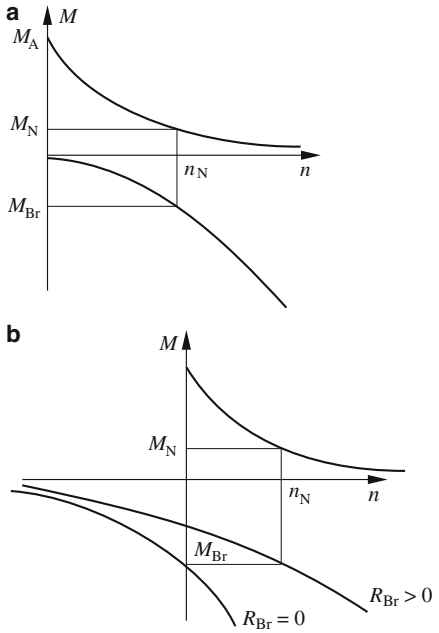


Abb. 86.67 Drehzahl-, Drehmomenten-Kennlinie der Reihenschluss-Maschine im Bremsbetrieb. **a** Mit Vorwiderstand, **b** Gegenstrombremsen

gleichlaufend durch die Erregerwicklung und erzeugt wiederum eine Induktionsspannung im Anker. Durch die Umpolung des Ankers fließt der Strom entgegengesetzt durch den Anker und erzeugt mit dem Erregerfeld ein Moment entgegen der Drehrichtung. Mit sinkender Drehzahl werden U_i und damit auch der Strom kleiner.

Beim Gegenstrombremsen wird die Ankerwicklung umgepolt und ein Bremswiderstand in den Kreis eingeschaltet, der zu hohe Ströme verhindern soll. Im Abb. 86.67 sind die Kennlinienverläufe der Bremsvorgänge dargestellt.

Beispiel

Von einem Gleichstromreihenschlussmotor sind folgende Daten bekannt: $U_N = 200\,\mathrm{V}$; $I_N = 10\,\mathrm{A}$; $P_N = 1{,}9\,\mathrm{kW}$; $n = 1800\,\mathrm{min}^{-1}$; $\sum R_a = 1\,\Omega$.

Das Nennmoment berechnet sich zu:

$$M_N = \frac{P_N}{2 \cdot \pi \cdot n_N} = \frac{1900\,\mathrm{V\,A} \cdot 60\,\frac{\mathrm{s}}{\mathrm{min}}}{2 \cdot \pi \cdot 1800\,\frac{1}{\mathrm{min}}} = 10{,}08\,\mathrm{W\,s}$$

Über das Nennmoment und den Nennstrom können die Motorkonstanten berechnet werden:

$$k_3 = 2 \cdot \pi \cdot k_4 = 2 \cdot \pi \cdot \frac{M_N}{I_N^2}$$
$$= \frac{2 \cdot \pi \cdot 10{,}08\,\mathrm{W\,s}}{10^2\,\mathrm{A}^2} = 0{,}633\,\frac{\mathrm{V\,s}}{\mathrm{A}}$$

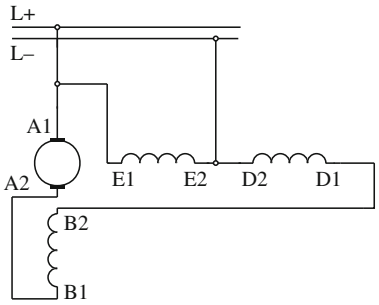


Abb. 86.68 Schaltung einer Doppelschluss-Maschine

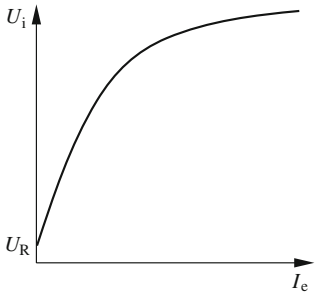


Abb. 86.69 Leerlaufkennlinie $U–f(I)$ des Gleichstromgenerators

Da die Momente quadratisch von den Strömen abhängen, kann der Strom bei halbem Moment bestimmt werden:

$$I^* = I_N \cdot \sqrt{\frac{M_N}{M^*}} = 10\,\text{A} \cdot \sqrt{\frac{10{,}08\,\text{W s}}{0{,}5 \cdot 10{,}08\,\text{W s}}}$$

$$= 10\,\text{A} \cdot \sqrt{\frac{1}{0{,}5}} = 7{,}07\,\text{A}$$

Um die neue Drehzahl berechnen zu können, muss zuerst die Induktionsspannung bestimmt werden:

$$U_i^* = U_N - I^* \cdot \sum R_a$$
$$= 200\,\text{V} - (7{,}07\,\text{A} \cdot 1\,\text{W}) = 192{,}93\,\text{V}$$

Die Drehzahl bei halbem Nennmoment berechnet sich jetzt wie folgt:

$$n^* = \frac{U_i^*}{k_3 \cdot I^*} = \frac{192{,}93\,\text{V} \cdot 60\,\frac{\text{s}}{\text{min}}}{0{,}633\,\frac{\text{V s}}{\text{A}} \cdot 7{,}07\,\text{A}} = 2586{,}6\,\text{min}^{-1} \quad \blacktriangleleft$$

86.6.2.3 Doppelschlussmotor

Der Doppelschlussmotor besitzt im Gegensatz zu den anderen Gleichstrommaschinen zwei Erregerwicklungen (Abb. 86.68).

Diese Erregerwicklungen sind als Nebenschluss- und als Reihenschlusswicklung ausgeführt. Je nachdem, welches Verhalten beim Motor überwiegen soll, wird eine der Wicklungen stärker ausgelegt. Der Doppelschlussmotor hat den Vorteil, dass auch bei einem Reihenschlussverhalten ein „Durchgehen" der Maschine ausgeschlossen ist, da die Nebenschlusswicklung eine definierte Leerlaufdrehzahl vorgibt.

86.6.3 Betriebsverhalten von Gleichstromgeneratoren

Obwohl der Einsatz von Gleichstromgeneratoren durch die Entwicklung der Leistungselektronik stark zurückgegangen ist, soll hier ein Überblick über das Betriebsverhalten der verschiedenen Generatoren gegeben werden.

Ebenso wie bei den Motoren unterscheidet man zwischen fremderregtem Generator, Nebenschlussgenerator und Reihenschlussgenerator. Für die folgenden Betrachtungen wird angenommen, dass die Antriebsdrehzahl konstant ist. Die Spannung, die sich an den Generatorklemmen im unbelasteten Betrieb einstellt (Leerlaufspannung U_0), entspricht der induzierten Spannung U_i nach (86.138).

Im Abb. 86.69 ist die Abhängigkeit der Leerlaufspannung vom Erregerstrom dargestellt. Ausgehend von der Remanenzspannung U_R folgt die Spannungskurve entsprechend der Magnetisierungskennlinie dem Erregerstrom. Wird der Generator belastet, fließt durch die Ankerspule ein Strom, der in den Ankerwicklungen einen Spannungsfall $U_a = I \cdot R_a$ verursacht und ebenfalls ein Ankerquerfeld erzeugt, das zu Ankerrückwirkungen (Feldschwächung) führt. Aus diesen Gründen sinkt die Klemmenspannung bei Belastung am Generator.

86.6.3.1 Fremderregter Generator

Soll der fremderregte Generator mit konstanter Klemmenspannung betrieben werden, muss der Erregerstrom kontinuierlich nachgeregelt werden. Kann die Erregung nicht nachgeführt werden, sinkt die Klemmenspannung. Da die Erregung jedoch nur indirekt von der Belastung abhängig ist (Feldschwächung), liefert der fremderregte Generator bei Belastung eine verhältnismäßig konstante Klemmenspannung. Die Klemmenspannung schwankt je nach Größe der Maschine zwischen 10 und 2 %.

86.6.3.2 Nebenschlussgenerator

Beim Nebenschlussgenerator ist keine Fremdspannungsquelle vorhanden, die den Erregerstrom liefert.

Der Generator muss die Erregerspannung selbst liefern. Wird der Generator mit konstanter Drehzahl angetrieben, erzeugt der vorhandene Restmagnetismus eine Remanenzspannung U_R. Diese Spannung treibt einen Strom durch den Erregerkreis. Dieser Strom erzeugt eine höhere Induktionsspannung, die einen höheren Erregerstrom zur Folge hat. Dieser Vorgang hält so lange an, bis ein Gleichgewicht zwischen der Induktionsspannung und dem Spannungsfall an den Widerständen des Erregerkreises herrscht.

$$\textit{Induktionsspannung} \quad U_i = I_e \cdot (R_a + R_e) \qquad (86.168)$$

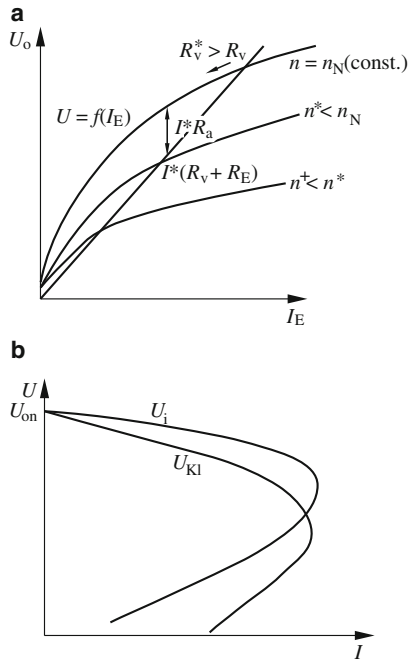


Abb. 86.70 Kennlinie U–$f(I)$ des Gleichstromgenerators. **a** Leerlaufkennlinie für verschiedene Drehzahlen mit Lastwiderstand, **b** Belastungskennlinie für eine Drehzahl

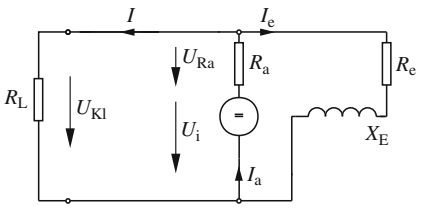


Abb. 86.71 Ersatzschaltbild des Nebenschlussgenerators

Aus der Kennlinie Abb. 86.70 ist abzulesen, dass der stabile Betriebspunkt für die Leerlaufspannung vom Widerstand des Leerlaufkreises und der Drehzahl der Maschine abhängt. Für kleine Drehzahlen kann ein Punkt erreicht werden, an dem keine Selbsterregung der Maschine einsetzt.

Ist die Polung der Erregerwicklung falsch angeschlossen, wird durch den Erregerstrom eine Spannung induziert, die der Remanenzspannung entgegenwirkt. Das Ersatzschaltbild des Nebenschlussgenerators ist in Abb. 86.71 dargestellt. Es gelten folgende Gleichungen für den Erreger- oder Ankerkreis:

$$\textit{Ankerkreisspannung} \qquad U = U_i - R_a \cdot (I + I_e) \qquad (86.169)$$

$$\textit{Erregerkreisspannung} \qquad U = I_e \cdot R_e \qquad (86.170)$$

Unter der Voraussetzung $I \gg I_e$ und $R_e \gg R_a$ ergibt sich:

$$R_a \cdot I = U_i - I_e \cdot R_e \qquad (86.171)$$

Wird ein Widerstand R_L an den leerlaufenden Generator angeschlossen, fließt über den Widerstand R_a nicht nur der Erreger-, sondern auch der Laststrom. Das hat zur Folge, dass Klemmenspannung und treibende Spannung für den Erregerstrom sinken. Je nach Größe von $I \cdot R_a$ stellt sich ein neuer Betriebspunkt am Generator ein. Die Kennlinie zeigt, dass die Klemmenspannung mit steigender Last bis zum Nullpunkt (Kurzschluss) kontinuierlich sinkt. Der Nebenschlussgenerator liefert im Kurzschluss einen Strom, der nicht größer als der Nennstrom wird.

86.6.3.3 Reihenschlussgenerator

Da der Reihenschlussgenerator seine Erregung durch den im Anker fließenden Strom erhält, kann der Generator nur im Kurzschluss angefahren werden. Die vorhandene Remanenzspannung erzeugt einen Strom im kurzgeschlossenen Generatorkreis, der die Generatorspannung erhöht. Bei Belastung des Reihenschlussgenerators schwankt die Klemmenspannung sehr stark mit dem Laststrom, der gleichzeitig Erregerstrom ist. Ein Einsatz dieses Generatortyps ist nicht zu empfehlen.

86.6.3.4 Doppelschlussgenerator

Der Doppelschlussgenerator hat, wie der Doppelschlussmotor, zwei Erregerwicklungen, die auf dem gleichen Polkern übereinandergewickelt sind. Eine der Wicklungen wird mit der Ankerspule in Reihe geschaltet, die andere Wicklung parallel. Je nach Schaltung und Auslegung der Wicklungen spricht man von einem *kompoundierten* Generator. Der Reihenschlussanteil der Wicklung erzeugt ein Feld in Richtung des Nebenschlussfeldes. Die Reihenschlusswicklung ist kleiner als die Nebenschlusswicklung. Dadurch wird die Spannung nahezu konstant gehalten. Der *überkompoundierte* Generator hat einen starken Reihenschlussanteil; die Spannung steigt mit zunehmender Belastung. Der *gegenkompoundierte* Generator hat die Reihenschlusswicklung so geschaltet, dass das Reihenfeld dem Feld der Nebenschlusswicklung entgegenwirkt. Mit zunehmender Belastung sinkt die Spannung.

86.6.4 Gleichstrommaschine am Wechsel- oder Drehstromnetz

Häufig steht kein geeignetes Gleichstromnetz für den Betrieb einer Gleichstrommaschine zur Verfügung. Die benötigte Gleichspannung wird deshalb aus dem vorhandenen Wechsel- oder Drehstromnetz durch Gleichrichterschaltungen erzeugt. In diese Gleichrichterschaltungen können Dioden oder ansteuerbare Ventile (Transistor, Thyristor) eingesetzt werden. Über die ansteuerbaren Ventile besteht die Möglichkeit, die Höhe der Gleichspannung zu verändern. Einen Überblick über diese *stromrichtergespeisten Gleichstromantriebe* gibt Tab. 86.7.

Durch die beliebige Ansteuerbarkeit dieser Stromrichter kann die Drehzahl der Gleichstrommaschinen durch eine Spannungsänderung geregelt werden. Ferner kann durch geeignete Maßnahmen eine Rückspeisung von Energie in das Netz erfolgen (Mehrquadrantenantriebe). Zur optimalen Nutzung dieser Methode sollten die Anker- und Erregerwicklungen durch getrennte Spannungsquellen versorgt werden.

Tab. 86.7 Stromrichter für Gleichstrommaschinen

Bezeichnung	Betriebsbereiche	Zusatzmaßnahmen
Einquadrantenantrieb mit Einfachstromrichter, halbgesteuert	I. oder III. Quadrant, nur Motorbetrieb	Keine
Einquadrantenantrieb mit Einfachstromrichter, vollgesteuert	I., IV. oder III., II. Quadrant, Motor- und Generatorbetrieb	Keine
Mehrquadrantenantrieb, Einfachstromrichter, vollgesteuert	I., IV. und III., II. Quadrant, Motor- und Generatorbetrieb	Umschaltung im Ankerkreis, momentenfreie Pause 0,1 bis 0,2 s
Mehrquadrantenantrieb, Einfachstromrichter, vollgesteuert	I., IV. und III., II. Quadrant, Motor- und Generatorbetrieb	Umschaltung im Erregerkreis, momentenfreie Pause 0,5 bis 2 s
Mehrquadrantenantrieb, Zweifachstromrichter, vollgesteuert	I., IV. und III., II. Quadrant, Motor- und Generatorbetrieb	kreisstromfreie Gegenparallelschaltung, momentenfreie Pause 2 bis 10 ms
Mehrquadrantenantrieb, Zweifachstromrichter, vollgesteuert	I., IV. und III., II. Quadrant, Motor- und Generatorbetrieb	kreisstromführende Kreuzschaltung, keine momentenfreie Pause

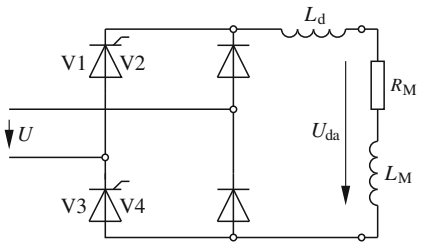


Abb. 86.72 Unsymmetrische halbgesteuerte Einphasen-Brückenschaltung

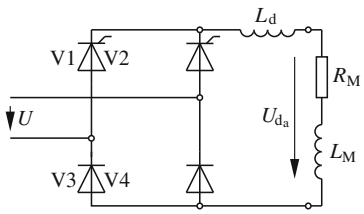


Abb. 86.73 Symmetrische halbgesteuerte Einphasen-Brückenschaltung

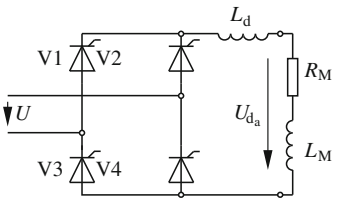


Abb. 86.74 Vollgesteuerte Einphasen-Brückenschaltung

86.6.4.1 Wechselstrombrücken

Wird die Gleichspannung aus dem Einphasenwechselstromnetz erzeugt, stehen die nachfolgend dargestellten Gleichrichterschaltungen zur Verfügung.

86.6.4.1.1 Einquadrantenantrieb (1-Q-Betrieb)

Der Einquadrantenantrieb wird durch eine halbgesteuerte Einphasen-Brückenschaltung erreicht. Hierbei wird die Hälfte der Dioden durch steuerbare Ventile ersetzt. Die Maschine kann nur als Motor im I. oder III. Quadranten betrieben werden. Unterschieden wird bei dieser Brückenschaltung zwischen der symmetrischen und der unsymmetrischen Schaltung.

Bei der *unsymmetrischen Einphasen-Brückenschaltung* (Abb. 86.72) können die Ventile mit einem Zündwinkel von $\alpha = 0°$ bis $\alpha = 180°$ angesteuert werden. Hierdurch ergibt sich eine Gleichspannung von 100 bis 0 %.

Die *symmetrische Brückenschaltung* (Abb. 86.73, 86.74) lässt nur einen Zündwinkel von maximal 150° zu, da sonst Kippgefahr besteht. Der Unterschied der beiden Gleichrichterschaltungen besteht darin, dass sich der Ankerstrom bei der symmetrisch gesteuerten Brücke gleichmäßig auf Thyristoren und Dioden aufteilt, bei der unsymmetrischen Brücke bewirken die zu einem Freilaufzweig angeordneten Dioden eine Entlastung der Thyristoren. Die mit diesen Brücken erzeugte Gleichspannung berechnet sich zu:

Gleichspannung nach der Brücke

$$U_{di\,\alpha} = \frac{\sqrt{2}}{\pi} \cdot U \cdot (1 + \cos\alpha) \qquad (86.172)$$

Die maximale Drehzahl erreicht die Maschine bei $\alpha = 0°$. Die für die Maschine angegebene Nennleistung gilt für Nennspannung und Nennstrom bei vollkommen geglätteter Gleichspannung. Pulsierende Gleichspannungen enthalten Wechselstromanteile, die zusätzliche Verluste in der Maschine erzeugen. Für den Betrieb von Gleichstrommaschinen an Stromrichtern geben die Hersteller deshalb einen *Zuschlagsfaktor F* (der Faktor F ist aus den Herstellerlisten zu ermitteln) an, mit dem die Nennleistung für Stromrichterbetrieb berechnet werden muss.

Maschinenleistung bei Stromrichterbetrieb

$$P_T = M_N \cdot 2 \cdot \pi \cdot \frac{n_N}{60} \cdot F \qquad (86.173)$$

86.6.4.1.2 Zweiquadrantenantrieb (2-Q-Betrieb)

Der Zweiquadrantenantrieb wird durch eine vollgesteuerte Einphasen-Brückenschaltung aufgebaut (Abb. 86.75).

Alle Gleichrichterdioden werden durch steuerbare Ventile ersetzt. Die Zündimpulse für die Ventile werden aus dem Verlauf der Netzspannung abgeleitet. Abwechselnd zündet man zwei diagonal angeordnete Thyristoren, wodurch die beiden anderen gelöscht werden. Im Gleichrichterbetrieb (I. oder III. Quadrant) steht für den Motor eine Spannung $U_{di\,\alpha}$ zur Verfügung, deren Höhe abhängig vom Steuerwinkel der Thyristoren ist.

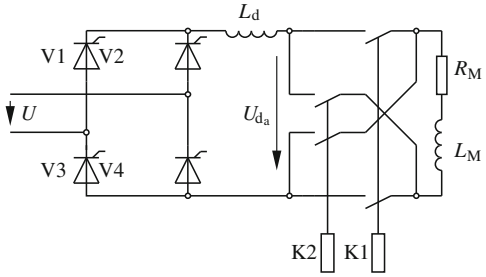


Abb. 86.75 Vollgesteuerte Einphasen-Brückenschaltung mit Wendeschützkombination für Vierquadrantenbetrieb

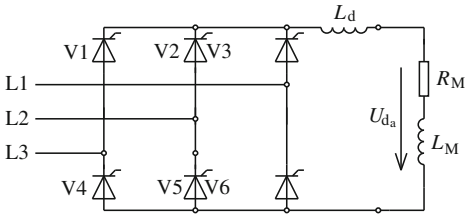


Abb. 86.76 Vollgesteuerte Drehstrom-Brückenschaltung

$$\textit{Brückenspannung} \quad U_{di\alpha} = \frac{2}{\pi} \cdot \sqrt{2} \cdot U \cdot \cos\varphi \qquad (86.174)$$

Bei einem Zündwinkel von $\alpha = 90°$ ist die Gleichspannung $U_{di\alpha} = 0$. Es steht kein Moment mehr zur Verfügung, der Motor steht. Wird der Winkel α über 90° vergrößert, geht die Maschine in den Generatorbetrieb und die Brückenschaltung in den Wechselrichterbetrieb über; die Maschine arbeitet im IV. oder II. Quadranten. Als maximaler Zündwinkel sollte $\alpha = 150°$ nicht überschritten werden, da die Freiwerdezeit der Ventile zu berücksichtigen ist. Eine Momentenumkehr ist mit dieser Schaltung nicht ohne weiteres möglich, daher kein generatorisches Bremsen.

86.6.4.1.3 Vierquadrantenantrieb (4-Q-Betrieb)

Beim Vierquadrantenbetrieb wird ebenfalls eine vollgesteuerte Einphasen-Brückenschaltung eingesetzt.

Über eine Wendeschützkombination im Ankerkreis kann die Spannung am Motor umgepolt und somit die Drehrichtung umgedreht werden (Abb. 86.76). Hierdurch ist ein Betrieb der Maschine in allen vier Quadranten möglich. Eine Umschaltung sollte möglichst stromlos erfolgen, um Schaltüberspannungen und Drehmomentenstöße zu verhindern. Hierzu muss der Strom im Ankerkreis gemessen und der Istwert der Umschalteinrichtung übermittelt werden. Ein Umschaltvorgang läuft wie folgt ab:

Der Steuerwinkel sollte auf 150° gebracht werden. Der dann herrschende Wechselrichterbetrieb lässt den Strom schnell auf Null abklingen. Im Stromnulldurchgang werden die Zündimpulse gesperrt, es wird umgeschaltet. Ist die Umschaltung abgeschlossen, werden die Zündimpulse mit einem Winkel von 150° wieder freigegeben. Durch eine Verringerung des Zündwinkels auf 90° bremst die Maschine bis zum Stillstand ab. Bei einem

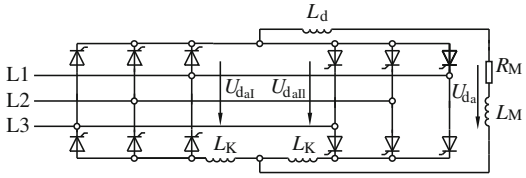


Abb. 86.77 Vollgesteuerte Drehstrom-Brückenschaltung in Gegenparallelschaltung

Winkel $\alpha < 90°$ kann die Maschine in entgegengesetzter Richtung hochfahren.

Der Vierquadrantenbetrieb ist ebenfalls durch eine Umkehr der Erregerspannung möglich. Hierbei ist jedoch eine längere Pause zwischen den Schaltvorgängen einzuhalten!

86.6.4.2 Drehstrombrücken

Steht dem Betreiber ein Drehstromnetz zur Verfügung, bieten sich folgende Brückenschaltungen zur Erzeugung von Gleichstromnetzen an.

86.6.4.2.1 Zweiquadrantenbetrieb (2-Q-Betrieb)

Wird eine vollgesteuerte Drehstrombrückenschaltung nach Abb. 86.77 eingesetzt, so ergibt sich der Wert der Gleichspannung zu:

$$\textit{Brückenspannung} \quad U_{di\alpha} = \frac{3}{\pi} \cdot \sqrt{2} \cdot U \cdot \cos\alpha \qquad (86.175)$$

Diese Brückenschaltung ermöglicht kleine Transformatorbauformen. In der gleichgerichteten Spannung sind nur Oberschwingungen mit Frequenzen vorhanden, die ein Vielfaches der sechsfachen Netzfrequenz betragen. Der Stellbereich ist wesentlich größer als bei der Wechselstrombrücke, und die Stromflussdauer beträgt in allen Brückenzweigen 120° (kein Lückbetrieb vorausgesetzt).

86.6.4.2.2 Vierquadrantenbetrieb (4-Q-Betrieb)

Mit den Wechselstrombrücken ist der Bremsbetrieb nur durch ein Umpolen der Erreger- oder Ankerspannung möglich. Das eignet sich nicht für Antriebe, die eine schnelle Drehrichtungsumkehr verlangen. Wird eine schnelle Drehrichtungsumkehr verlangt, kann die Drehstrombrücke in Gegenparallelschaltung nach Abb. 86.78 eingesetzt werden.

Hierbei wird eine Stromumkehr durch zwei Drehstrombrücken, die mit entgegengesetzten Ventilrichtungen aufgebaut sind, ermöglicht. Die beiden Drehstrombrücken arbeiten parallel, jedoch mit einem Zündwinkel, der gewährleistet, dass sich der Mittelwert der beiden Brückenspannungen in jedem Zeitpunkt vom Betrag her gleicht, jedoch ein entgegengesetztes Vorzeichen hat. Soll die Gleichspannung $U_{d\alpha}$ verändert werden, müssen beide Zündwinkel α_I und $\alpha_{II} = 180° - \alpha_I$ entsprechend verstellt werden. Da die beiden Teilspannungen der Brücken nicht in jedem Zeitpunkt gleich sind, stellt sich über die Teilstromrichter ein Strom ein, der durch diese Wechselspannungskomponente erzeugt wird. Dieser Strom wird durch die

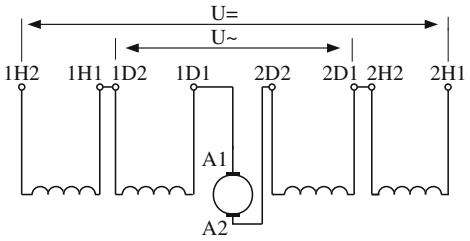


Abb. 86.78 Schaltung eines Universalmotors

Kreisdrossel L_K begrenzt. Bei Antrieben mit größeren Leistungen wird ein Teilstromrichter nicht angesteuert, somit können die Kreisströme vermieden werden. Die Umschaltung erfordert hier jedoch eine gewisse Totzeit.

86.6.5 Universalmotor

Ein Universalmotor ist ein kleiner Reihenschlussmotor, der mit Gleich- und Wechselstrom betrieben werden kann. Im Unterschied zum reinen Gleichstrommotor ist das Ständerblechpaket zur Vermeidung von Wirbelströmen aus Dynamoblech zusammengesetzt. Die Baugröße des Universalmotors geht bis zu einer Leistung von ungefähr 2000 W. Die Drehzahlbereiche sind anders als beim Einphasenwechselstrommotor nicht frequenzabhängig. Universalmotoren werden mit Drehzahlen von 1500 min^{-1} bis 30.000 min^{-1} gebaut. Wird der Motor mit Wechselstrom betrieben, entsteht in der Erregerwicklung neben dem ohmschen Widerstand noch der induktive Widerstand. Dadurch gehen Drehzahl und Leistung beim Übergang von Gleich- auf Wechselstrom zurück.

Soll die Leistung für Gleich- und Wechselstrom konstant gehalten werden, muss die Maschine eine höhere Windungszahl für den Gleichstrombetrieb erhalten.

86.7 Auswahl von Motoren

Wird ein Motor für einen Antrieb ausgewählt, müssen vorher einige Vorgaben geklärt sein. Bekannt sein müssen die für den Antrieb maßgebende Drehzahl/Drehmomentenkennlinie, die geforderte Dauerleistung, die geforderte Drehzahl, die mögliche Spannungsversorgung des Motors, der Aufstellungsort, die Betriebsart des Antriebs, eventuell erforderliche Zusatzmaßnahmen und der notwendige Motorschutz.

86.7.1 Auswahl unter Berücksichtigung der Normen

In den DIN- und VDE-Vorschriften werden dem Betreiber Vorgaben gemacht, die zur Auswahl von geeigneten Motoren herangezogen werden sollten.

Tab. 86.8 Gegenüberstellung der Kurzzeichen für die Bauformen nach DIN IEC 34 Teil 7 und der alten DIN 42950

DIN IEC 34 Teil 7		DIN 42950
IM B 3	IM 1001	B 3
IM V 5	IM 1011	V 5
IM V 6	IM 1031	V 6
IM B 6	IM 1051	B 6
IM B 7	IM 1061	B 7
IM B 8	IM 1071	B 8
IM B 35	IM 2001	B 3/B 5
IM B 34	IM 2101	B 3/B 14
IM B 5	IM 3001	B 5
IM V 1	IM3011	V 1
IM V 3	IM 3031	V 3
IM B 14	IM 3601	B 14
IM V 18	IM 3611	V 18
IM V 19	IM 3631	V 19
IM B 10	IM 4001	B 10
IM V 10	IM 4011	V 10
IM V 14	IM 4031	V 14
IM V 16	IM 4131	V 16
IM B 9	IM 9101	B 9
IM V 8	IM 9111	V 8
IM V 9	IM 9131	V 9

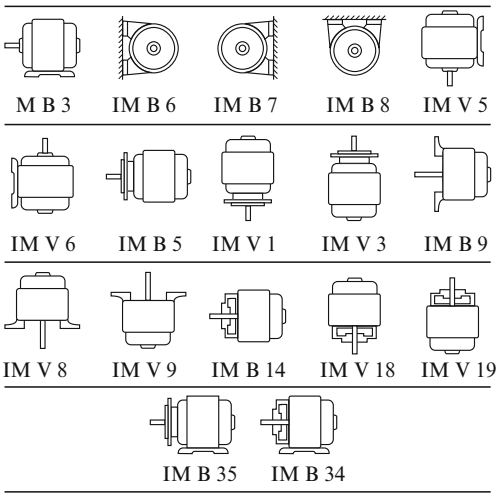


Abb. 86.79 Die gebräuchlichsten Bauformen für Drehstrommotoren

86.7.1.1 Bauform und Baugrößen

Die Bauformen für elektrische Maschinen sind in DIN 42950 (alt) und in der DIN IEC 34 Teil 7 festgelegt. Die DIN IEC 34 unterscheidet zwei Codetabellen für die Bezeichnung der Bauformen. In Tab. 86.8 und Abb. 86.79 sind die gebräuchlichsten Bauformen und Baugrößen aufgeführt.

86.7.1.2 Schutzart

Durch die Wahl einer geeigneten Schutzart wird verhindert, dass schädigende Einwirkung von Wasser, Fremdkörpern und Staub das Betriebsverhalten des Motors verändern. Weiter muss die

Tab. 86.9 Schutzarten

Erste Kennziffer	Erklärung	Zweite Kennziffer	Erklärung	Dritter Buchstabe	Erklärung	Vierter Buchstabe	Erklärung
0	Kein Schutz	0	Kein Schutz				
1	Schutz gegen zufälliges großflächiges Berühren; Schutz gegen Eindringen von Fremdkörpern größer als 50 mm	1	Schutz gegen senkrecht fallendes Wasser	A	Geschützt gegen Zugang mit dem Handrücken (50 mm)	M	Betriebsmittel geprüft auf die schädliche Wirkung durch Eintritt von Wasser, wenn die beweglichen Teile des Betriebsmittels in Betrieb sind
2	Schutz gegen Berühren durch Finger; Schutz gegen Eindringen von Fremdkörpern > 12,5 mm	2	Schutz gegen Tropfwasser bei Schrägstellung des Gerätes bis zu 15°	B	Geschützt gegen Zugang durch Finger (12 mm Durchmesser, 80 mm Länge)	W	Geeignet zur Verwendung unter festgelegten Wetterbedingungen und ausgestattet mit zusätzlichen schützenden Maßnahmen oder Verfahren
3	Schutz gegen Berühren mit Werkzeugen oder gegen Eindringen von Fremdkörpern von einer Dicke > 2,5 mm	3	Schutz gegen Sprühwasser aus einem Winkel bis zu 60°	C	Geschützt gegen Zugang mit Werkzeug (2,5 mm Durchmesser, 100 mm Länge)	H	Hochspannungsbetriebsmittel
4	Schutz gegen Berühren mit Werkzeugen oder gegen Eindringen von Fremdkörpern von einer Dicke > 1 mm	4	Schutz gegen Spritzwasser aus beliebigen Richtungen	D	Geschützt gegen Zugang mit Draht (1,0 mm Durchmesser, 100 mm Länge)	S	Betriebsmittel geprüft auf schädliche Wirkung durch Eindringen von Wasser, wenn die beweglichen Teile im Stillstand sind
5	Vollständiger Schutz gegen Berühren; Schutz gegen schädliche Staubablagerungen	5	Schutz gegen Strahlwasser aus allen Richtungen				
6	Vollständiger Schutz gegen Berühren; Schutz gegen Eindringen von Staub	6	Schutz gegen vorübergehende Überflutung, z. B. schwere See				
7		7	Schutz gegen schädliches Eindringen von Wasser beim Eintauchen				
8		8	Schutz gegen jegliches Eindringen von Wasser				

Berührung von rotierenden Teilen im Inneren eines Motors und das Berühren von spannungsführenden Teilen verhindert werden. Die Schutzarten sind in der DIN 40050 zu EN 60529/VDE 0470 Teil 1 festgelegt. Als Kurzzeichen für die jeweilige Schutzart steht die Abkürzung IP (International Protection), gefolgt von zwei Kennziffern und zwei Buchstaben. Von den beiden Kennziffern gibt die erste Ziffer den Personen- oder Fremdkörperschutz an, die zweite Ziffer den Wasserschutz. Der dritte Buchstabe kann für einen zusätzlichen Schutz stehen, und der vierte Buchstabe stellt eine Ergänzung dar. Die ersten beiden Ziffern müssen nach dem IP-Zeichen vorhanden sein. Ist eine Ziffer nicht ausgeführt, muss sie durch ein X ersetzt werden. Der dritte und vierte Buchstabe muss nicht vorhanden sein.

86.7.1.3 Kühlart

Die Art der Kühlung von Maschinen ist nach DIN IEC 34 Teil 6 festgelegt. Diese DIN sieht zwei Kennzeichnungssysteme vor. Beide beginnen mit den Buchstaben IC (International Cooling). Bei der vollständigen Kennzeichnung folgen auf die Buchstaben IC zwei Blöcke mit je einem Buchstaben und zwei Ziffern, z. B. IC W 337 A 71. Die deutsche Norm empfiehlt dem Anwender bei luftgekühlten Maschinen die vereinfachte Kennzeichnung einzusetzen. Bei der vereinfachten Kennzeichnung beschreibt die erste Kennziffer die Art des Kühlmittelumlaufs, die zweite Ziffer die Art des Antriebs für den Kühlmittelumlauf.

Die für den Antrieb vorgegebene Kühlungsart gewährleistet bei einer Kühlmitteltemperatur (KT) von 40 °C und einer Aufstellungshöhe bis 1000 m, dass der Antrieb mit Nennleistung betrieben werden darf. Bei abweichenden Bedingungen ist die zulässige Leistung entsprechend nachstehender Tabelle zu bestimmen. Diese Angaben sind Richtwerte. Es ist von jedem Hersteller eine genaue Information über die Reduktionsfaktoren einzuholen!

Tab. 86.10 Kühlungsarten von Motoren

1. Kennziffer	Bedeutung	2. Kennziffer	Bedeutung
0	Maschine mit freiem Luftein- und -austritt	0	Selbstkühlung
1	Maschine mit Rohranschluss, ein Einlasskanal	1	Eigenkühlung (Ventilator)
2	Maschine mit Rohranschluss, ein Auslasskanal	2	Eigenkühlung durch eine nicht auf der Welle angebrachten Belüftungseinrichtung
3	Maschine mit Rohranschluss, Ein- und Auslasskanal	3	Fremdkühlung durch eine an die Maschine angebaute Belüftungseinrichtung. Antrieb von der Maschine abhängig
4	Oberflächengekühlte Maschine (Umgebungsluft)	4	
5	Maschine mit eingebautem Wärmetauscher (Kühlmittel Umgebungsluft)	5	Fremdkühlung durch eine eingebaute Belüftungseinrichtung. Antrieb nicht von der Maschine abhängig
6	Maschine mit aufgebautem Wärmetauscher (Kühlmittel Umgebungsluft)	6	Fremdkühlung durch eine an die Maschine angebaute Belüftungseinrichtung. Antrieb nicht von der Maschine abhängig
7	Maschine mit eingebautem Wärmetauscher (Kühlmittel ist nicht Umgebungsluft)	7	Fremdkühlung durch eine nicht auf die Maschine aufgebaute Belüftungseinrichtung. Antrieb nicht von der Maschine abhängig oder durch Druckluft aus dem Versorgungsnetz
8	Maschine mit aufgebautem Wärmetauscher (Kühlmittel ist nicht Umgebungsluft)	8	Verdrängungskühlung (Fahrtwind)
9	Maschine mit getrennt aufgestelltem Wärmetauscher		

Falls Motor und Antriebsmaschine nicht über ein Rohrsystem direkt ins Freie fremdbelüftet sind, muss die von den Maschinen erzeugte Wärmemenge über die Raumluft abgeführt werden. Hierzu ist eine gute Frischluftversorgung zur Kühlung der Maschinennotwendig. Ein oberflächengekühlter 110-kW-Motor mit $1500\,\mathrm{min}^{-1}$ benötigt in einer Stunde ungefähr $2200\,\mathrm{m}^3$ Kühlluft. Der erforderliche Kühlluftstrom berechnet sich nach folgender Formel:

Kühlluftstrom

$$\dot{V}_L = P_V \cdot \frac{0{,}77}{\Delta\vartheta} \qquad \begin{array}{c|c|c} \dot{V}_L & P_V & \Delta\vartheta \\ \hline \frac{\mathrm{m}^3}{\mathrm{s}} & \mathrm{kW} & \mathrm{K} \end{array} \qquad (86.176)$$

P_V gesamte Verlustleistung; $\Delta\vartheta$ zulässige Temperaturdifferenz (Differenz zwischen Lufteintritts- und zulässiger Raumtemperatur).

Tab. 86.11 Reduktionsfaktoren für die Nennleistung abhängig von der Höhe und Temperatur

Aufstellungshöhe über NN in m	Kühlmitteltemperatur (KT) in °C					
	< 30	30–40	45	50	55	60
1000	1,07	1,00	0,96	0,92	0,87	0,82
1500	1,04	0,97	0,93	0,89	0,84	0,79
2000	1,00	0,94	0,90	0,86	0,82	0,77
2500	0,96	0,90	0,86	0,83	0,78	0,74
3000	0,92	0,86	0,82	0,79	0,75	0,70
3500	0,88	0,82	0,79	0,75	0,71	0,67
4000	0,82	0,77	0,74	0,71	0,67	0,63

Tab. 86.12 Grenztemperaturen von Isolierungen

Klasse	Y	A	E	B	F	H	C
Grenztemperatur in °C	90	105	120	130	155	180	> 180

86.7.1.4 Isolierstoffklassen

In der DIN VDE 0530 sind die Isolierstoffe einschließlich der Tränkmittel in *Isolierstoffklassen* eingeteilt, denen genau festgelegte Temperaturwerte zugeordnet sind. Die höchstzulässige Dauertemperatur der Isolierstoffe setzt sich aus der Kühlmitteltemperatur und der Grenzübertemperatur der Wicklung zusammen. Weiter ist noch eine Erwärmungstoleranz zur Sicherheit zu berücksichtigen, da die Messungen nur einen Mittelwert der Erwärmung liefern. Bei den Grenztemperaturen wird davon ausgegangen, dass die Kühlmitteltemperatur den Wert 40 °C nicht übersteigt.

86.7.1.5 Motorschutz

Motoren müssen durch geeignete Maßnahmen gegen zu hohe Erwärmung geschützt werden. Dabei ist zu beachten, dass Sicherungen nur ein Schutz gegen Kurzschlüsse sind und nicht gegen zu hohe Erwärmung, da sie die Vorbelastung des Motors nicht berücksichtigen. Zum Einsatz als Motorschutzgeräte kommen die thermischen Auslöser oder der Motorvollschutz. Bei größeren Motoren auch Relais, die das Wärmeabbild des Motors wiedergeben. Thermische Überlastungen eines Motors sind durch folgende Ursachen möglich:

Erhöhte Verluste durch die Art des Betriebes; Blockieren des Läufers während des Betriebes; Anschluss- und Schaltfehler; erhöhte Verluste durch Netzstörungen; Phasenausfall; Netzspannung; Beeinträchtigung der Kühlung.

86.7.1.5.1 Thermischer Auslöser

Der thermische Auslöser (Motorschutzschalter) befindet sich im Hauptstromkreis und wird vom Motorstrom durchflossen. Der Motorschutzschalter ist für kleinere und mittlere Antriebe geeignet. Die Kosten dieser Schutzeinrichtung sind gering. Die Schutzfunktion ist jedoch nicht optimal, da der Motorschutzschalter nicht die Wicklungstemperatur direkt überwacht,

Tab. 86.13 Betriebsarten

S1 – Dauerbetrieb	Betrieb mit konstantem Belastungszustand, dessen Dauer ausreicht, den thermischen Beharrungszustand zu erreichen. Vorgabe ist die Leistung. Beispiel: S1; 50 kW
S2 – Kurzzeitbetrieb	Betrieb mit konstantem Belastungszustand, der aber nicht so lange dauert, dass der thermische Beharrungszustand erreicht wird. Mit einer nachfolgenden Pause, die so lange besteht, bis die Maschinentemperatur nicht mehr als 2 K von der Temperatur des Kühlmittels abweicht. Vorgabe ist die Leistung und die Betriebsdauer. Beispiel: S2; 20 min; 30 kW
S3 – Aussetzbetrieb ohne Einfluss des Anlaufvorganges	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt, von denen jedes eine Zeit mit konstanter Belastung und eine Pause umfasst. Der Anlaufstrom beeinflusst die Erwärmung nicht merklich. Vorgabe ist die Leistung, die Einschaltzeit t_B und die Spieldauer t_S oder die relative Einschaltdauer t_r. $t_r = t_B/t_S$. Beispiel: S3; 10 %; 50 min; 20 kW
S4 – Aussetzbetrieb mit Einfluss des Anlaufvorganges	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt, von denen jedes eine merkliche Anlaufzeit t_A, eine Zeit mit konstanter Belastung und eine Pause t_{St} umfasst. Vorgabe ist die relative Einschaltdauer, die Zahl der Anläufe pro Stunde und die Leistung. $t_r = (t_A + t_B)/(t_A + t_B + t_{St})$. Beispiel: S4; 35 %; 400 Anläufe; 25 kW
S5 – Aussetzbetrieb mit Einfluss des Anlaufvorganges und der elektrischen Bremsung	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt, von denen jedes eine merkliche Anlaufzeit, eine Zeit mit konstanter Belastung, eine Zeit schneller elektrischer Bremsung t_{Br} und eine Pause umfasst. Vorgabe wie bei der Betriebsart S4, jedoch mit der Angabe der Bremsart. $t_r = (t_A + t_B + t_{Br})/(t_A + t_B + t_{Br} + t_{St})$. Beispiel: S5; 25 %; 250 Spiele/h Gegenstrombremsung; 40 kW
S6 – Durchlaufbetrieb mit Aussetzbelastung	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt, von denen jedes eine Zeit mit konstanter Belastung und eine Leerlaufzeit umfasst. Es tritt keine Pause auf. Vorgabe wie Betriebsart S3. $t_r = t_B/t_S$. Beispiel: S6; 20 %; 45 kW
S7 – Ununterbrochener Betrieb mit Anlauf und elektrischer Bremsung	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt, von denen jedes eine merkliche Anlaufzeit, eine Zeit mit konstanter Belastung und eine Zeit mit schneller elektrischer Bremsung umfasst. Es tritt keine Pause auf. Vorgabe wie Betriebsart S5, jedoch keine relative Einschaltzeit $t_r = 1$. Beispiel: S7; 10 kW; 300 Reversierungen/h
S8 – Ununterbrochener Betrieb mit periodischer Drehzahländerung	Betrieb, der sich aus einer Folge gleichartiger Spiele zusammensetzt. Jedes Spiel umfasst eine Zeit mit konstanter Belastung und bestimmter Drehzahl. Anschließend eine oder mehrere Zeiten mit anderer Belastung, denen unterschiedliche Drehzahlen entsprechen. Vorgaben wie bei Betriebsart S5, jedoch für jede Drehzahl
S9 – Ununterbrochener Betrieb mit nichtperiodischer Last- und Drehzahländerung	Betrieb, bei dem sich Belastung und Drehzahl innerhalb des zulässigen Betriebsbereiches nichtperiodisch ändern. Es treten häufig Belastungsspitzen auf, die weit über der Nennleistung liegen können. Vorgabe ist eine passend gewählte Dauerbelastung, deren Grundlage die Wurzel aus dem quadratischen Mittelwert der Leistung oder des Stromes sein sollte. $P = \sqrt{\frac{P_1^2 \cdot t_1 + P_2^2 \cdot t_2 + P_3^2 \cdot t_3}{t_1 + t_2 + t_3}}$

sondern nur die Stromaufnahme des Motors. Fehlfunktionen des Motorschutzschalters können eintreten, wenn seine Belüftung nicht einwandfrei funktioniert. Dann kann der Motorschutzschalter bei erhöhter Umgebungstemperatur auslösen, ohne dass der Motor überlastet ist. Wird der Motorschutzschalter jedoch besser gekühlt als der Motor, so erfolgt keine Auslösung, obwohl der Motor überlastet wird.

86.7.1.5.2 Thermistor-Motorvollschutz

Von einem *Motorvollschutz* spricht man, wenn in den Wicklungssträngen des Ständers des zu schützenden Motors temperaturabhängige Widerstände eingebaut sind. Diese Widerstände sind meistens Kaltleiter (PTC). Der Kaltleiter wird so ausgelegt, dass ein definierter Strom bei Betriebstemperatur durch den Schutzkreis fließt. Erhöht sich die Wicklungstemperatur auf unzulässige Werte, steigt der Widerstand des Kaltleiters sehr stark an. Im zugehörigen Auslösegerät wird dieser Widerstandsanstieg durch einen Stromabfall registriert; der Motor wird vom Netz geschaltet oder es erfolgt eine Störungsmeldung. Dieser Schutz ist für *ständerkritische* Motoren ausreichend. *Läuferkritische* Motoren benötigen zusätzlich noch ein Überlastrelais. Das Überlastrelais schützt den Motor, wenn der Läufer festgebremst ist. Zusätzlich übernimmt es noch den Leitungsschutz.

86.7.1.6 Abstimmung des Motors auf die Arbeitsmaschine

Für die Auswahl eines geeigneten Motors für einen Antrieb müssen folgende Daten bekannt sein: Betriebsart; Anlauf-, Brems- und Umsteuervorgänge; Fremdträgheitsmomente; Momentenverlauf der Arbeitsmaschine; Drehzahlsteuerung; Netzverhältnisse; Aufstellungshöhe; Kühlmitteltemperatur; Bedarfsleistung der Arbeitsmaschine. Die Nennleistung des Motors soll auf die Bedarfsleistung der Antriebsmaschine abgestimmt sein, da ein zu groß bemessener Motor einen höheren Anzugsstrom verursacht, wodurch größere Sicherungen und stärkere Leitungen erforderlich sind. Auch fährt der Motor in einem unwirtschaftlichen Betrieb, da der Leistungsfaktor und unter Umständen der Wirkungsgrad unter Teillast schlechter sind. Das Drehmoment der Antriebsmaschine ist aus entsprechenden Unterlagen des Herstellers zu entnehmen. In Abb. 86.80 sind einige typische Drehzahl-Drehmomentverläufe dargestellt.

Kurve 1: Drehmoment praktisch gleichbleibend, Leistung proportional der Drehzahl. Gilt z. B. für Hebezeuge, Kolbenpumpen und Verdichter bei Förderung gegen konstanten Druck, Kapselgebläse, Walzwerke, Förderbänder, Mühlen ohne Lüfterwirkung, Werkzeugmaschinen mit gleichbleibender Schnittkraft.

Tab. 86.14 Überlastschutzeinrichtungen im Vergleich (++ sehr gut; + gut; −− gering)

Schutzeinrichtung	Überlastschutz			Kurzschlussschutz		Schalthäufigkeit
	Leitung	Motor (Ständer)	Motor (Läufer)	Motor	Leitung	
Sicherung, Leistungsschalter (Überlast; Kurzschluss)	++	++	++	++	++	−−
Sicherung; Schütz; Überlastschutz	++	++	++	++	++	++
Sicherung; Leistungsschalter (Überlast); Thermistor	+	++	+	++	++	−−
Sicherung; Schütz; Thermistor	+	++	+	++	++	++
Sicherung; Schütz; Überlastschutz; Thermistor	++	++	++	++	++	++

Tab. 86.15 Liste von Wartungsarbeiten in Abhängigkeit vom Motortyp

Zeitraum	Lager		Maschinentyp			Wartungsarbeiten
	Wälzlager	Gleitlager	Käfig	Schleifring	Stromwender	
Alle 1 bis 2 Wochen	×					Lagergeräusch und Lagererwärmung kontrollieren
		×				Ölstandskontrolle
					×	Kontrolle der Oberfläche des Stromwenders und der Bürsten einschließlich Halterung
Alle 1 bis 2 Monate				×	×	Oberflächenkontrolle der Schleifringe und des Stromwenders, der Bürsten, der Haltevorrichtung und der Bürstenabhebevorrichtung
Alle 3 bis 4 Monate				×	×	Überprüfung des Bürstendrucks
			×	×	×	Kontrolle der Anschlussklemmen
Jährliche Wartung	×	×				Kontrolle der Lager und Messung der Lagertemperatur
			×	×	×	Messung des Isolationswiderstandes der Wicklungen
			×	×	×	Reinigung verschmutzter Wicklungen
			×	×	×	Messung der Wicklungstemperatur
				×	×	Kontrolle der Schleifringe und Stromwender mit Zubehör

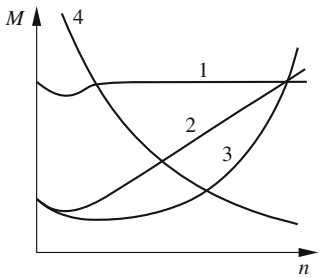


Abb. 86.80 Drehzahl-, Drehmomentkennlinien verschiedener Arbeitsmaschinen

Kurve 2: Drehmoment wächst proportional zur Drehzahl, Leistung proportional zum Quadrat der Drehzahl. Maschinen zum Glätten von Geweben und Papier, Heißmangel und Kalander.

Kurve 3: Drehmoment wächst proportional zum Quadrat der Drehzahl, Leistung proportional zur dritten Potenz der Drehzahl. Gilt für Kreiselpumpen, Ventilatoren, Kolbenmaschinen, die in ein offenes Rohrnetz fördern, Maschinen mit Schleuderwirkung, Schiffsantriebe, Rührwerke.

Kurve 4: Drehmoment nimmt umgekehrt proportional zur Drehzahl ab, die Leistung ist konstant. Dieser Verlauf ist nur für Regelvorgänge interessant. Kommt bei Zerspanungsmaschinen, Aufwickel- und Rundschälmaschinen vor. Für den Anlauf ist das mittlere Lastmoment M_l aus dem Moment M_e nach beendetem Hochlauf zu bestimmen. Eine Überschlagsbestimmung für die angegebenen Kennlinien ergibt: Kurve 1: $M_l = M_e$; Kurve 2: $M_l = M_e/2$; Kurve 3: $M_l = M_e/3$; Kurve 4: M_l. Ist das mittlere Hochlaufmoment M_m bekannt, ergibt sich für die Anlaufzeit t_a des Antriebs folgender Wert:

Anlaufzeit

$$t_a = \frac{\sum J \cdot n_0}{9{,}55 \cdot M_{\mathrm{bmi}}} \qquad \begin{array}{c|c|c|c} t_a & J & n_0 & M_{\mathrm{bmi}} \\ \hline \mathrm{s} & \mathrm{kg\,m^2} & \mathrm{min^{-1}} & \mathrm{N\,m} \end{array} \qquad (86.177)$$

mittleres Beschleunigungsmoment

$$M_{\mathrm{bmi}} = M_m - M_l \qquad (86.178)$$

J Gesamtträgheitsmoment des Antriebes.

Das Gesamtträgheitsmoment setzt sich aus dem Motorträgheitsmoment, dem Trägheitsmoment der Arbeitsmaschine und den Trägheitsmomenten von Kupplungen und Riemenscheiben, umgerechnet auf die Drehzahl der Motorwelle, zusammen. Kann wegen großer Trägheitsmomente/Lastmomente kein einwandfreier Anlauf erreicht werden, muss ein Sonderläufer oder ein

Tab. 86.16 Störungen an Gleichstrommaschinen, ihre Ursache und Behebung

Störungsart	Ursache	Abhilfe
Motor läuft nicht an	Sicherung defekt	Sicherung ersetzen
	Bürsten liegen nicht auf	Bürstensitz überprüfen, Bürstenhalter reinigen, Bürsten auswechseln
	Lager festgefressen	Lager auswechseln
	Anker- oder Feldwicklung unterbrochen	Durchgang überprüfen, Wicklung ersetzen
Motor läuft schwer an	Körperschluss der Wicklung	Auf Körperschluss prüfen (Kurbelinduktor), Wicklung erneuern
	Erregerwicklung unterbrochen	Erregerwicklung auf Durchgang prüfen
	Bürstenbrücke verstellt	Bürstenbrückenstellung prüfen (Markierung beachten)
Unruhiger Lauf	Defekte Kugellager	Kugellager ersetzen
	Verspannte Lagerschilder	Befestigungsschrauben gleichmäßig nachziehen
Bürstenfeuer zu stark	Motor überlastet	Belastung verringern
	Kollektor unrund oder verschmutzt	Kollektor abdrehen, Glimmerisolation auskratzen, Kollektor reinigen
	Wendepole falsch geschaltet	Schaltung überprüfen
Motor läuft zu schnell	Erregerwicklung hat Windungsschluss	Wicklung erneuern
	Klemmenspannung zu hoch	Gleichspannung verringern

Tab. 86.17 Störungen an Asynchronmaschinen, ihre Ursache und Behebung

Störungsart	Ursache	Abhilfe
Motor läuft nicht an	Schutz hat angesprochen	Schutzgerät überprüfen, einschalten
	Lager festgefressen	Lager auswechseln
Motor läuft schwer an	Wicklungsstrang im Läufer unterbrochen	Wicklungsstränge auf Durchgang prüfen
	Ständerwicklungen haben Phasen- oder Körperschluss	Wicklungen überprüfen, auswechseln
	Klemmenspannung zu niedrig	Spannungsfall auf Zuleitung überprüfen
Motor wird im Leerlauf zu warm	Windungsschluss	Strangwiderstände messen, defekten Strang austauschen
	Falsche Ständerschaltung bei zu hoher Betriebsspannung	Ständer von Dreieck- auf Sternschaltung umschalten
Motor wird im Dauerbetrieb zu warm	Belastung zu hoch	Verringerung der Belastung
	Eine Sicherung hat angesprochen (Zweiphasenlauf)	Klemmenspannung prüfen
	Belüftung fehlt	Motor abstellen, bis Belüftung wieder vorhanden
Motor brummt	Wicklungsstrang des Ständers hat Wicklungsschluss	Wicklung durchmessen, austauschen
	Zweiphasenlauf	Klemmenspannung messen
	Am Klemmbrett sind zwei Phasen und der Mittelleiter angeschlossen	Klemmenspannung messen
Drehzahl sinkt bei Belastung stark ab	Maschine überlastet	Verringerung der Belastung
	Läuferstäbe ausgelötet oder abgerissen	Läuferwicklung prüfen (Brandstellen)
	Läuferblechpaket auf der Welle verschoben	Sitz des Läuferblechpaketes prüfen

größerer Motor eingesetzt werden, der dann aber bei Normalbetrieb schlechter ausgenutzt wird. Deshalb sollte ein Schleifringläufermotor, eine Anlaufkupplung oder ein Frequenzumrichter eingesetzt werden.

86.7.1.6.1 Wartung von Maschinen

Elektrische Maschinen haben nur wenige Teile, die einem Verschleiß unterliegen. Diese Verschleißteile bedürfen aber besonderer Beobachtung und regelmäßiger Wartung. Insbesondere Lager sollten mit besonderer Sorgfalt gewartet werden, da sie einem ständigen Verschleiß unterliegen. Welche Wartungsarbeiten in welchen Abständen durchgeführt werden sollten, ist in Tab. 86.15 dargestellt. Die Angaben der Tabelle beziehen sich auf normale Betriebsbedingungen, also Dauerbetrieb über acht Stunden, Nennumgebungstemperatur und keine Verschmutzungen in der Kühlluft. Eine Veränderung der normalen Betriebsbedingungen ergibt andere Wartungsinter-valle. Wartungs- und Kontrollarbeiten sollten in Motorstillstandszeiten durchgeführt werden. Neben den erwähnten Wartungs- und Kontrollarbeiten gehört auch die Überprüfung der Schutzgeräte in den Wartungsplan.

Tab. 86.18 Gegenüberstellung von alten und neuen Bezeichnungen bei Maschinen

Art der Maschine		Neue Bezeichnung	Alte Bezeichnung
Transformator	Oberspannung	1U1, 1V1, 1W1	U, V, W
		1U2, 1V2, 1W2	X, Y, Z
	Unterspannung	2U1, 2V1, 2W1	u, v, w
		2U2, 2V2, 2W2	x, y, z
Gleichstrommaschine	Ankerwicklung	A1–A2	A–B
	Wendepolwicklung	B1–B2	GW–HW
	Geteilte Wendepolwicklung	1B1–1B2 2B1–2B2	
	Kompensationswicklung	C1–C2	GK–HK
	Reihenschlusswicklung	D1–D2	E–F
	Nebenschlusswicklung	E1–E2	C–D
	Fremderregte Wicklung	F1–F2	I–K
Wechselstrommaschine	Dreiphasenmaschine	U1, V1, W1	U, V, W
		U2, V2, W2	X, Y, Z
	Schleifringläufer	K, L, M	u, v, w
	Erregerwicklung	F1–F2	I, K
	Kondensatormotor	U1–U2	U–V
		Z1–Z2	W–Z

86.7.1.7 Störungsbeseitigung

Die Tab. 86.17 soll dem Betreiber einige Hinweise geben, welche Ursache einer eventuell aufgetretenen Störung zugrunde liegt.

86.7.1.8 Anschlusskennzeichnungen von Maschinen

Siehe Tab. 86.18.

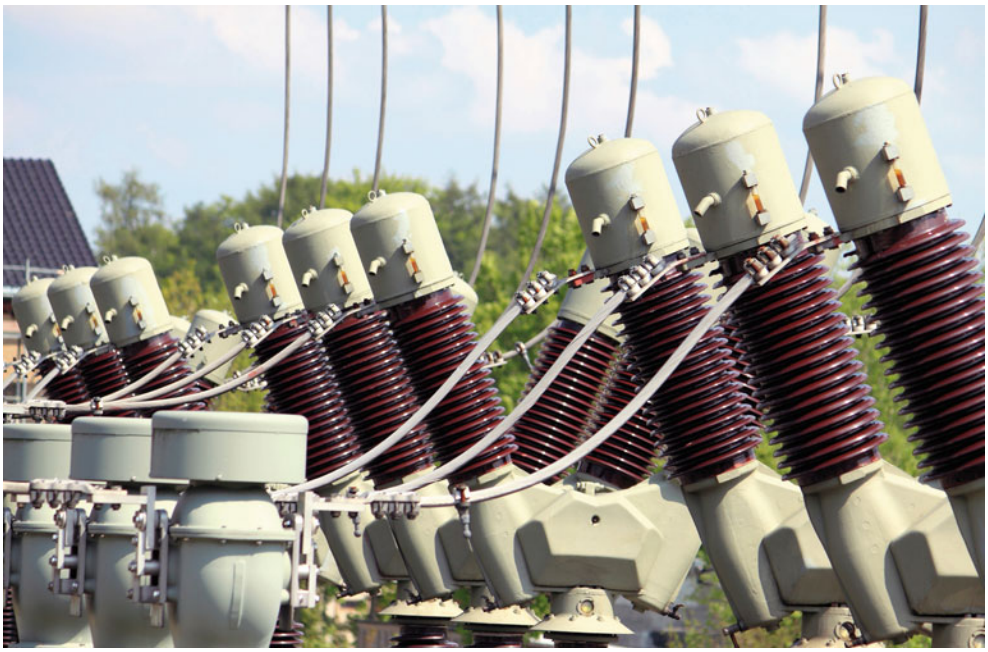
Literatur

1. *Schröder (Hrsg.):* Elektrische Antriebe Bd. 1. Berlin: Springer 2002, ISBN 3-540-66846-2
2. *Giersch/Harthus/Vogelsang:* Elektrische Maschinen. Stuttgart: Teubner, ISBN 3-519,46821,2
3. *Constantinescu/Simon/Liviu (Hrsg.):* Elektrische Maschinen und Antriebssysteme. Wiesbaden: Vieweg 1999, ISBN 3-528-06426-9

Elektrische Energieversorgung

87

Detlef Schulz



87.1 Energiewirtschaftliche Rahmenbedingungen 1010
87.2 Elektrische Energieerzeugung 1013
87.3 Elektrische Energieübertragung und Energieverteilung 1034
87.4 Elektrische Betriebsmittel 1053
Literatur 1074

Teil XI

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_87

87.1 Energiewirtschaftliche Rahmenbedingungen

87.1.1 Einführung

Nur mit einer zuverlässigen und wirtschaftlichen elektrischen Energieversorgung ist eine nachhaltige gesellschaftliche Entwicklung von Industrienationen möglich. Gleichzeitig muss diese selbst so nachhaltig, d. h. umweltverträglich gestaltet werden, dass auch zukünftige Generationen nicht in ihrer Entfaltung behindert werden. Eine versorgungssichere, wirtschaftliche und umweltverträgliche Energiebereitstellung ergibt sich nicht zwingend allein aus einer technisch evolutionären Entwicklung und ist auch bis heute nicht ausschließlich nur mit einer der bekannten Wandlungstechnologien nachweislich realisierbar.

Daher wird die Energieversorgung insgesamt, wegen ihrer gesellschaftlichen Relevanz jedoch insbesondere die elektrische Energieversorgung, durch politische Steuerungsinstrumente und Gesetzgebung gelenkt. Dazu gehören neben dem Energiewirtschaftsgesetz das Stromeinspeisungsgesetz, europäische Richtlinien, das Erneuerbare-Energien-Gesetz sowie verschiedene weitere Instrumente zur Regulierung und Setzung von Investitionsanreizen.

Neben der Gesetzeslage existieren politische Festlegungen. Im Energiekonzept der Bundesregierung ist festgeschrieben, dass der Anteil der Stromerzeugung aus erneuerbaren Energien am Bruttostromverbrauch 35 % bis 2020 betragen soll und danach 50 % bis 2030, 65 % bis 2040 und 80 % bis 2050. Dazu gehört ein Übergangsprozess von der Nutzung fossiler Energieträger hin zu erneuerbaren Energien, der auch als Energiewende umschrieben wird. Dabei stellen der verbesserte Klimaschutz und der bereits 2002 beschlossene Atomausstieg die zwei Oberziele dieses Prozesses dar, denen Maßnahmen in verschiedenen Technologiebereichen zugeordnet werden.

87.1.2 Gesetzliche Rahmenbedingungen

Das von 1935 bis 1998 inhaltlich unverändert bestehende Energiewirtschaftsgesetz (EnWG) verfolgte das Ziel, die erheblichen Investitionen in die Infrastruktur durch Gebietsmonopole abzusichern. Ein Wettbewerb unterschiedlicher Anbieter war daher nicht vorgesehen. Zur Abschaffung der Oligopole (natürliche Monopole) in der Stromerzeugung und zur Sicherstellung eines Wettbewerbs wurde das EnWG im Jahr 1998 neu gefasst. Damit wurde die EG-Richtlinie zum Energiebinnenmarkt umgesetzt, die ein Unbundling (buchhalterische Entflechtung) der Teilbereiche Erzeugung, Übertragung/Verteilung und Vertrieb forderte und damit erstmals den freien Wettbewerb in einem liberalisierten Energiemarkt ermöglichte. Nach vier weiteren Novellierungen ist die aktuelle Fassung des EnWG aus dem Jahr 2011 gültig, die in Teilen letztmalig im Jahr 2014 geändert wurde. In der Tab. 87.1 sind die wesentlichen Inhalte und Änderungen zum EnWG in Kurzform zusammengefasst.

Mit dem Stromeinspeisungsgesetz aus dem Jahr 1990 wurde erstmals die Abnahmepflicht und Vergütung für Strom aus Anlagen mit regenerativer Erzeugung mit Leistungen unterhalb von 5 MW geregelt. Im Jahr 2000 wurde das Stromeinspeisungsgesetz durch das Erneuerbare-Energien-Gesetz abgelöst, das unterschiedliche Vergütungssätze für die regenerativen Wandlungstechnologien vorsah und inzwischen mehrfach bis zur aktuell vorliegenden Form überarbeitet wurde.

87.1.3 Struktur der Elektrizitätswirtschaft

Mit der seit 1998 umgesetzten Liberalisierung der Elektrizitätswirtschaft und dem damit verpflichtenden Unbundling, d. h. der organisatorisch-rechtlichen Entflechtung, hat sich die Struktur der Energieversorgung grundlegend gewandelt. Vor der Liberalisierung existierten zumeist in Großkonzernen organisierte Energieversorgungsunternehmen, die innerhalb der gesamten Kette der Erzeugung, Übertragung und Verteilung eine Monopolstellung hielten und die bei ihnen erzeugten Energiemengen über ihren Vertrieb mittels langfristiger Verträge an die Endkunden verkauften. Heute werden die erzeugten Mengen von den Kraftwerksbetreibern über den Stromhandel an Vertriebe veräußert, die diese Mengen dann an ihre Kunden verkaufen. Unverändert muss dabei jeweils der Grund-, Mittel- und Spitzenlastbedarf sowie der Bedarf an Regelleistung abgedeckt werden. Dazu werden Liefermengen mit unterschiedlichen Vorlaufzeiten und

Tab. 87.1 Entwicklung des Energiewirtschaftsgesetzes

Fassung/Jahr	Jahr	Wesentlicher Inhalt/Änderung
Erstfassung	1935	„die Energieversorgung so sicher und billig wie möglich zu gestalten“ (Präambel), kein „schädlicher“ Wettbewerb
Korrekturen	1978	Streichung überholter Begriffe und Bestimmungen, keine inhaltliche Änderung
Novelle	1998	Unbundling von Erzeugung, Übertragung und Verteilung, Sicherstellung eines Wettbewerbs
Novelle	2003	Äquivalenz Strom/Gas, Verbändevereinbarungen
Novelle	2005	Unbundling für Energieversorger mit mehr als 100.000 Kunden, regulierter anstatt verhandelter Netzzugang, Regulierungsbehörden überwachen die Netzbetreiber, Gasnetzzugangsverordnung, liberalisierte Zähler-Ablesung
Novelle	2008	Öffnung des Messwesens bei Strom und Gas für Wettbewerb
Novelle	2011	weitergehende Entflechtung von Übertragungs-, Fernleitungs- und Verteilnetzbetreibern, Neuregelung der Vorschriften zu Messeinrichtungen und Messsystemen, Unabhängige Regulierungsbehörden
Änderung	2014	Systemwechsel beim Anschluss von Offshore-Windparks, befristete Regelungen zur Gewährleistung der Versorgungssicherheit bei konventioneller Erzeugung, Modernisierungsanreize für bestehende Pumpspeicher

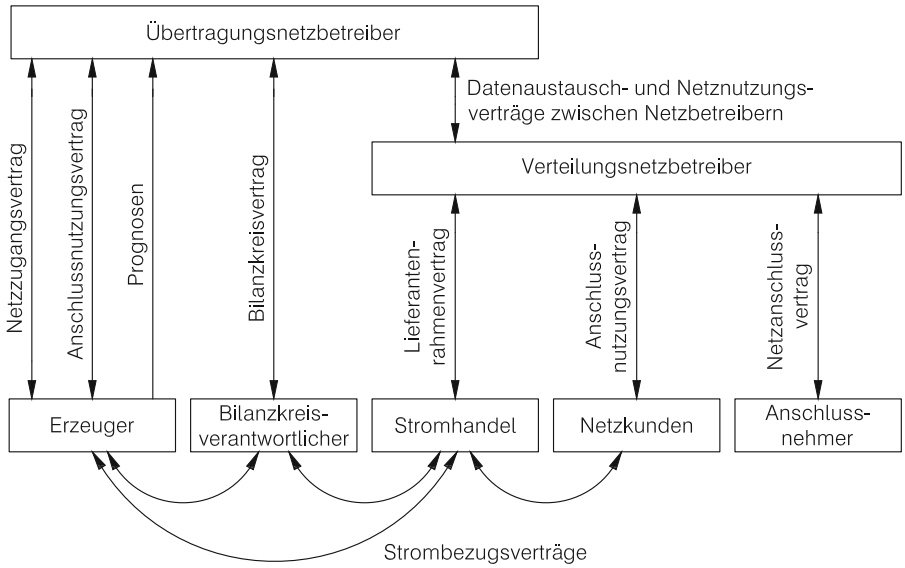


Abb. 87.1 Vertragsbeziehungen in der Energiewirtschaft

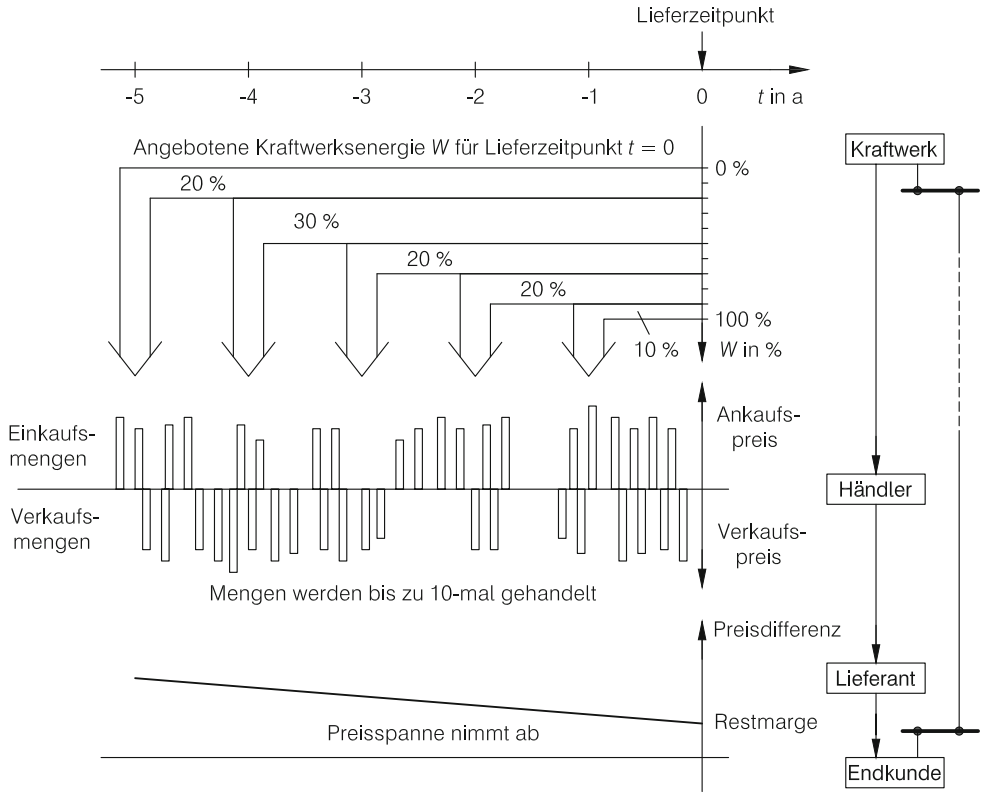


Abb. 87.2 Prinzip des Energiehandels von langfristig vereinbarten Liefermengen

verschiedenen Vertragslaufzeiten gehandelt. Langfristverträge zur Grundlastabdeckung werden bis zu fünf Jahre vor dem Lieferzeitpunkt geschlossen, zur Deckung von Spitzenlasten oder Regelbedarfen findet ein tagesaktueller Einkauf statt. Während langfristige Verträge direkt zwischen den Vertragspartnern gehandelt werden (over the counter), findet der kurzfristige Handel über Strombörsen statt. Dies erfolgt beispielsweise über die Strombörse European Energy Exchange (EEX) in Leipzig, an der Stromprodukte aus Österreich, Frankreich und Deutschland gehandelt werden.

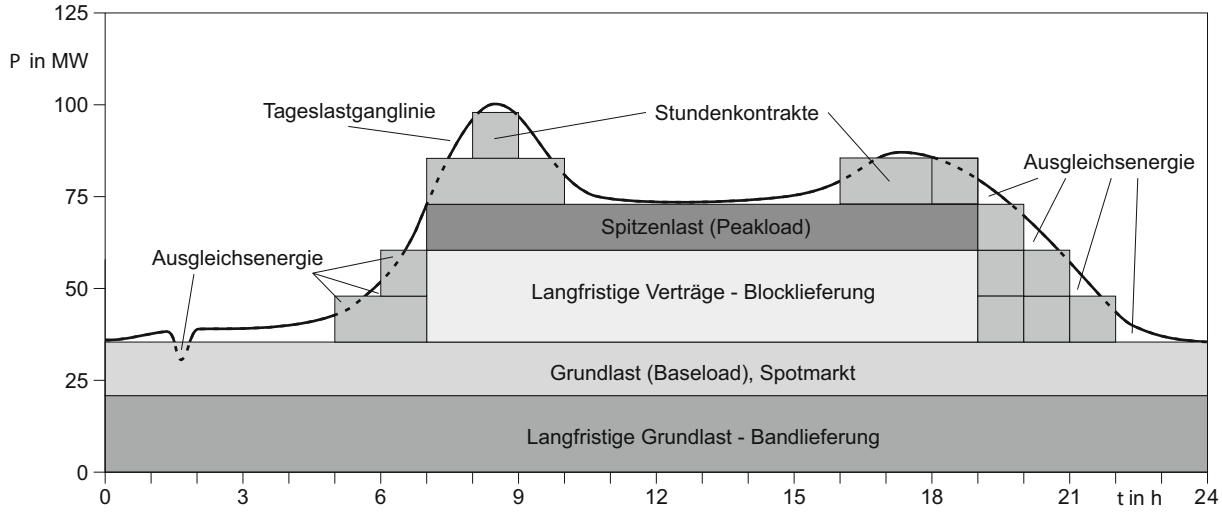


Abb. 87.3 Arten von Stromprodukten zur Bedarfsdeckung

87.1.4 Vertragsbeziehungen und Energiehandel

In Abb. 87.1 sind die Vertragsbeziehungen in der Energiewirtschaft in stark vereinfachter Form dargestellt. Detailliertere Beschreibungen zur Vertragsgestaltung können der Literatur entnommen werden.

Zur Absicherung der Stromproduktion werden die Energiemengen bereits lange vor ihrer Produktion gehandelt, siehe Abb. 87.2. Dadurch entstehen für die Kraftwerke belastbare Informationen für den Brennstoffeinkauf. Auf der Käuferseite werden insbesondere Industriestrompreise gut kalkulierbar, was bei energieintensiven Prozessen wiederum eine marktgerechte Produktkalkulation ermöglicht. Dazu werden von den Kraftwerken fünf Jahre vor dem Lieferzeitpunkt beginnend Liefermengen in Tranchen zwischen 10 % und 30 % angeboten. Bis zum Lieferzeitpunkt erfolgen pro Tranche bis zu zehn weitere Handelsschritte, d. h. Ein- und Verkäufe von/an andere Händler und Lieferanten. Gleichzeitig findet eine Absicherung der Handelsmengen beispielsweise mit anderen Börsengeschäften statt. Im Laufe der Handelsschritte stellt sich ein marktabhängiger Preis ein, d. h. die Preisspannen zwischen An- und Verkauf nehmen mit der Annäherung an den Lieferzeitpunkt ab. Somit trägt der Energiehandel zur marktgerechten Preisbildung der Liefermengen bei. Der Handel findet direkt oder an verschiedenen Strombörsen statt, z. B. auch an der deutschen Strombörse EEX (European Energy Exchange) in Leipzig. Neben direkt (OTC-over the counter) gehandelten Bandlieferungsverträgen mit zeitlich kontinuierlichen Energiemengen und Blocklieferverträgen mit konstanten Mengen über mehrere Stunden werden Stunden- und Spitzenlastmengen entweder OTC oder über die Börse verkauft, Ausgleichs- bzw. Regelenergie wird kurzfristig an der Börse gehandelt, siehe Abb. 87.3.

An der Börse stellt sich der Strompreis als Schnittpunkt zwischen Angebots- und Nachfrageseite ein, siehe Abb. 87.4. Der Angebotspreis für elektrische Energie ergibt sich aus den Grenzkosten der Kraftwerke, die zu deren wirtschaftlichem Betrieb erforderlich sind. Diese werden durch die Kosten für Brennstoffe, Betrieb und Wartung bestimmt. Aufgrund des besseren Umwandlungswirkungsgrads können neuere Kraftwerke ihre Energie günstiger anbieten als ältere Kraftwerke gleichen Typs. Ältere Kraftwerke, die schon abgeschrieben sind, können wiederum vergleichsweise günstig produzieren.

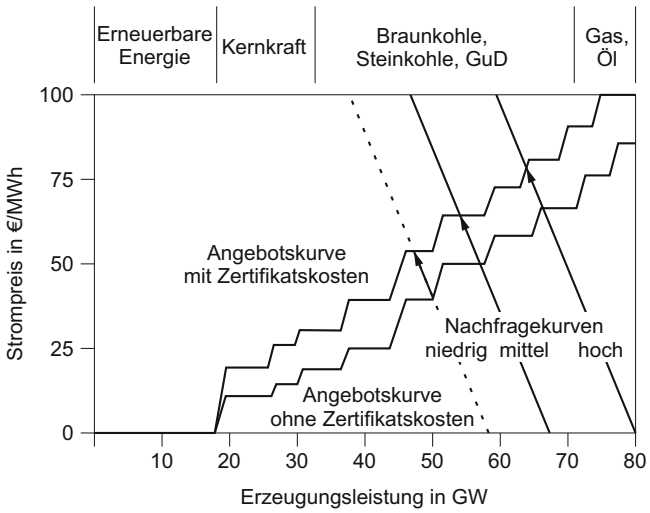


Abb. 87.4 Preisbildung an der Strombörse

Abhängig von den Grenzkosten ergibt sich eine Einsatzreihenfolge (Merit Order) der Kraftwerke. Zuerst kommen erneuerbare Kraftwerke zum Einsatz, da sie Einspeisevorrang besitzen und ihre Grenzkosten am geringsten ausfallen. Dann folgen die konventionellen Kraftwerke in der Reihenfolge ihrer Grenzkosten, bis die verbleibende Last (Residuallast) gedeckt ist. Der Börsenstrompreis ergibt sich aus dem letzten Gebot, das noch einen Zuschlag erhält, dem sog. Market Clearing Price. Somit bestimmt das teuerste noch benötigte Kraftwerk den Preis am

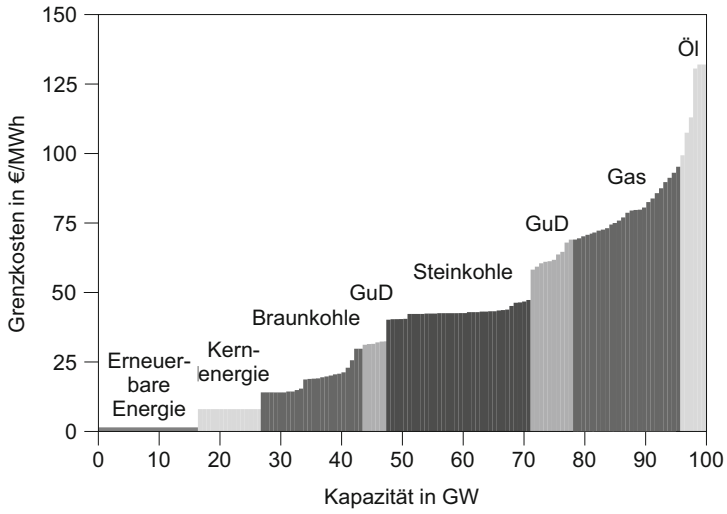


Abb. 87.5 Beispiel für die Grenzkosten-Einsatzreihenfolge (Merit-Order) des deutschen Kraftwerkparks

Strommarkt. In Abb. 87.5 ist ein Beispiel für die Einsatzreihenfolge (Merit-Order) des deutschen Kraftwerkparks gezeigt. Ein höherer Anteil erneuerbarer Energie verschiebt die Angebotskurve nach rechts, dadurch liegt der Schnittpunkt mit der Nachfragekurve bei einem geringeren Preis. Je größer das Angebot aus erneuerbaren Energien wird, umso geringer wird der Spotpreis für die gesamte Handelsmenge. Seit einiger Zeit entstehen dadurch Probleme mit der Wirtschaftlichkeit konventioneller Kraftwerke. Insbesondere die als Spitzenlastkraftwerke notwendigen Gaskraftwerke werden teilweise kurz nach ihrer Fertigstellung außer Betrieb gesetzt, da sie aufgrund ihrer hohen Brennstoffkosten zu teuer anbieten. Ölbefeuerte Kraftwerke weisen die höchsten Kosten auf und kommen nur noch zeitweise bei einem sehr hohem Leistungsbedarf zum Einsatz.

87.2 Elektrische Energieerzeugung

87.2.1 Elektrischer Energieverbrauch

Die elektrische Energieversorgung kann in die Bereiche Erzeugung, Transport sowie Verbrauch unterteilt werden. Entscheidend für den Aufbau des Gesamtsystems ist der Energieverbrauch, d. h. der Bedarf an elektrischer Energie. Wie in Abb. 87.6 gezeigt, sinkt der Primärenergieverbrauch Deutschlands im Mittel kontinuierlich ab. Hingegen steigt der Stromverbrauch weiter an, was durch einen höheren Komfortanspruch insbesondere in den Privathaushalten verursacht wird. Hierbei wirkt sich neben der Anschaffung von mehr Elektrogeräten beispielsweise auch verbrauchserhöhend aus, dass ältere Elektrogeräte trotz ihres Ersatzes durch energieeffizientere neue Geräte weiter in Betrieb bleiben. Der Stromverbrauch erfolgt nach Angaben des Bundesverbandes der Elektrizitäts- und Wasserwirtschaft e.V. (BDEW) im Jahr 2013 zu 47 % in der Industrie, zu 25 % in Haushalten, zu 24 % in Handel, Gewerbe und öffentlichen Einrichtungen, mit 2 % im Verkehr und mit 2 % in der Landwirtschaft.

Für den später in den Abschn. 87.3.3 und 87.3.4 noch detaillierter beschriebenen Netzbetrieb ist zwischen frequenzunabhängigen und -abhängigen Verbrauchern zu unterscheiden, wobei damit die Abhängigkeit der Leistung von der Frequenz beschrieben wird. Vorwiegend finden sich frequenzunabhängige Verbraucher im Netz, wie z. B. Beleuchtung, Elektroheizungen und andere Lasten mit ohmscher Widerstandscharakteristik. Eine Frequenzabhängigkeit liegt bei rotierenden Verbrauchern wie

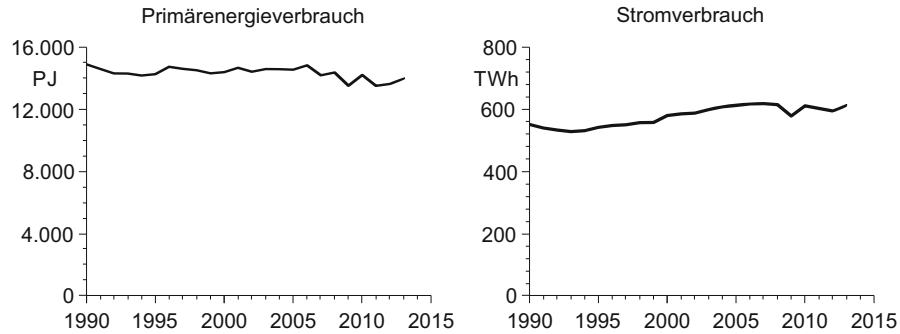


Abb. 87.6 Primärenergie- und Stromverbrauch in Deutschland seit 1990

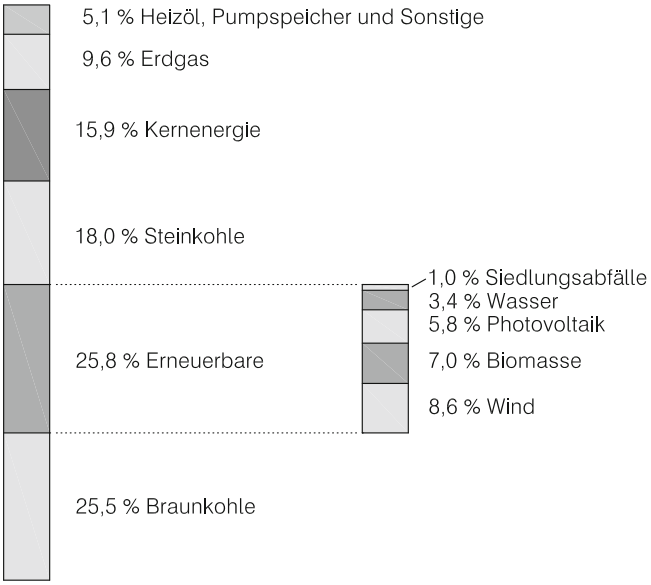


Abb. 87.7 Beiträge der Kraftwerksarten zur deutschen Stromerzeugung im Jahr 2014 nach BDEW

Lüftern, Ventilatoren, Pumpen und anderen Antriebsmaschinen vor. Wesentlich für die Gestaltung des Gesamtsystems ist die Jahreshöchstlast, die in Deutschland bei einem nahezu konstanten Wert von 81 GW liegt. Sie tritt immer im Winter auf, meist in den ersten Januartagen. In den Sommermonaten sinkt der Bedarf bis auf die Mindestlast im Bereich von 25 bis 35 GW ab.

87.2.2 Anteile an der Stromerzeugung und Kraftwerksleistung

Momentan nehmen die fossil befeuerten Kraftwerke noch den größten Anteil bei der Stromerzeugung ein. Doch dies wird sich aufgrund der schon erwähnten Klimaschutzziele und des Atomausstiegs mittel- und langfristig stark verändern. Zur Erreichung dieser Ziele muss ein grundlegender Umbau der Energieversorgung erfolgen. Mit dem Ziel einer klimaverträglichen und risikoarmen Energieversorgung erfolgt der Umbau zu einer CO_2-armen Erzeugung ohne Kernkraft. Langfristig werden daher immer mehr Kohlekraftwerke vom Netz gehen, die letzten Kernkraftwerke werden laut Entscheidung des Deutschen Bundestages vom 30. Juni 2011 im Jahr 2022 abgeschaltet. Im Jahr 2014 lieferten Braunkohlekraftwerke 25,5 %, Steinkohlekraftwerke 18 %, Kernenergiekraftwerke 15,9 % und Gaskraftwerke 9,6 % des Strombedarfs in Deutschland, siehe Abb. 87.7. Die erneuerbaren Energien waren insgesamt bereits mit 25,8 % beteiligt, ihr Anteil wird innerhalb der nächsten Jahre weiter stark ansteigen. Im Jahr 2014 sind bereits 35 GW Windenergie- und 38 GW Photovoltaikleistung installiert. Für den Endausbau allein der Windenergie werden in Deutschland bis zu 160 GW erwartet.

Mit 181,2 GW Kraftwerksleistung im Jahr 2014 liegt der deutsche Kraftwerkspark weit oberhalb der Jahreshöchstlast von 81 GW, siehe Abb. 87.8. Diese Kraftwerksleistung steht nicht in vollem Umfang zur Verfügung, sie wird durch die nicht einsetzbare Leistung, Ausfälle, Revisionen und Reserven für Systemdienstleistungen reduziert. Zur Kategorie der nicht ein-

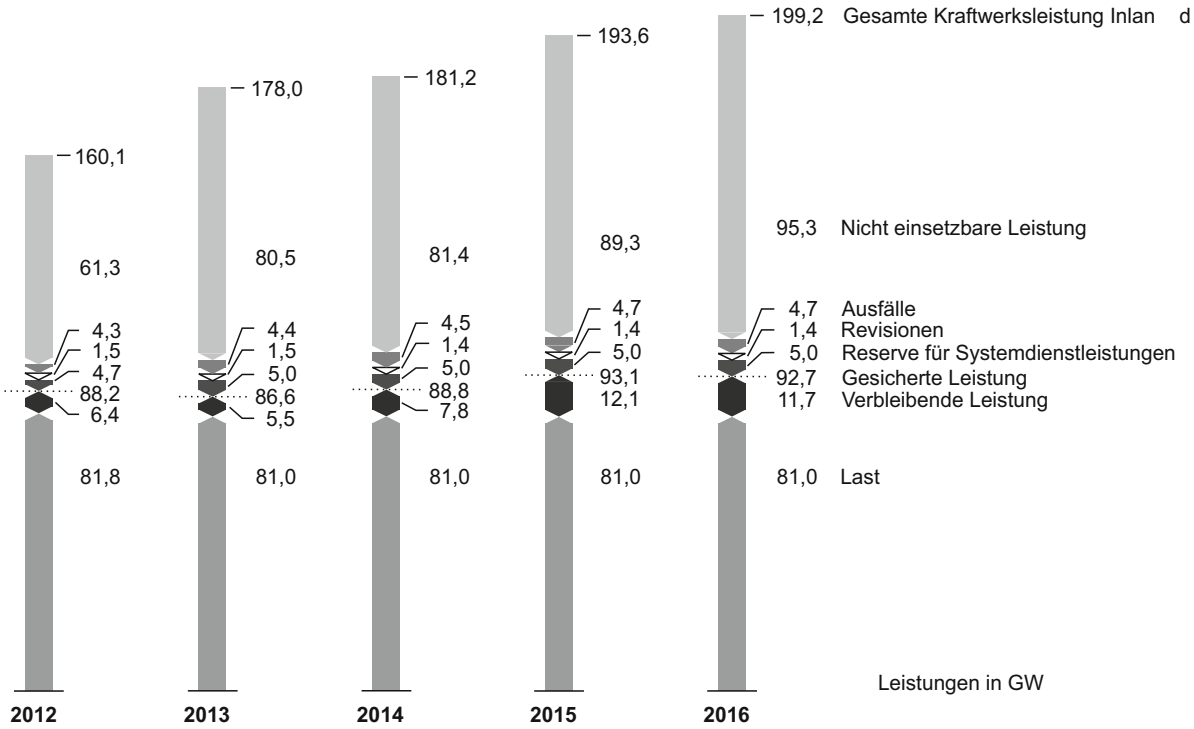


Abb. 87.8 Kumulierte und geschätzte Kraftwerksleistungen in Deutschland in den Jahren 2012 bis 2016, Datenquelle BMWi, Leistungsbilanzbericht 2013

setzbaren Leistung gehören Kraftwerke mit ungünstigen Stromgestehungskosten, wie z. B. ältere Kohlekraftwerke sowie Öl- und Gaskraftwerke. Mit den Systemdienstleistungen wird die Frequenz- und Spannungsregelung beschrieben, für die eine mit den Kraftwerken vertraglich vereinbarte Reserveleistung vorgehalten wird. Die gesicherte Leistung muss einen ausreichenden Abstand zur Jahreshöchstlast aufweisen, dies ist die verbleibende Leistung im Sinne eines Sicherheitspuffers. Im Jahr 2013 betrug dieser Puffer aufgrund des strukturellen Umbaus im deutschen Kraftwerkspark nur 5,5 GW, was zu verstärkten Diskussionen über die Robustheit der Energieversorgung führte. Im Abb. 87.8 ist erkennbar, dass für die nächsten Jahre ein Anstieg der verbleibenden Leistung prognostiziert wird. In den nächsten Abschnitten soll näher auf die Erzeugungstechnologien eingegangen werden.

87.2.3 Überblick und Klassifizierung von Kraftwerksarten

Kraftwerke können auf verschiedene Weise klassifiziert werden, z. B. nach der Art ihrer:

- Feuerung: fossil befeuert (Kohle, Gas, Öl), Kernenergie oder regenerativ versorgt, d. h. mit aus nach menschlichem Ermessen unerschöpflichen Energiequellen versorgt (Wasser, Wind, Biomasse, Sonne, Erdwärme), siehe Abb. 87.9
- Arbeitsmedien: Dampf, Gas, Öl, Wasser, Wind, Sonne, Erdwärme
- CO_2-Beiträge: hoch bei allen fossil befeuerten, gering bei Kernenergie und ohne bei regenerativen und z. B. Brennstoffzellen; Biomasse muss hierbei gesondert betrachtet werden
- Erzeugungscharakteristik: kontinuierlich bei fossil befeuerten Kraftwerken, Kernenergie und Wasser, Biomasse, Erdwärme und Brennstoffzellen und diskontinuierlich bei Wind und Sonne.

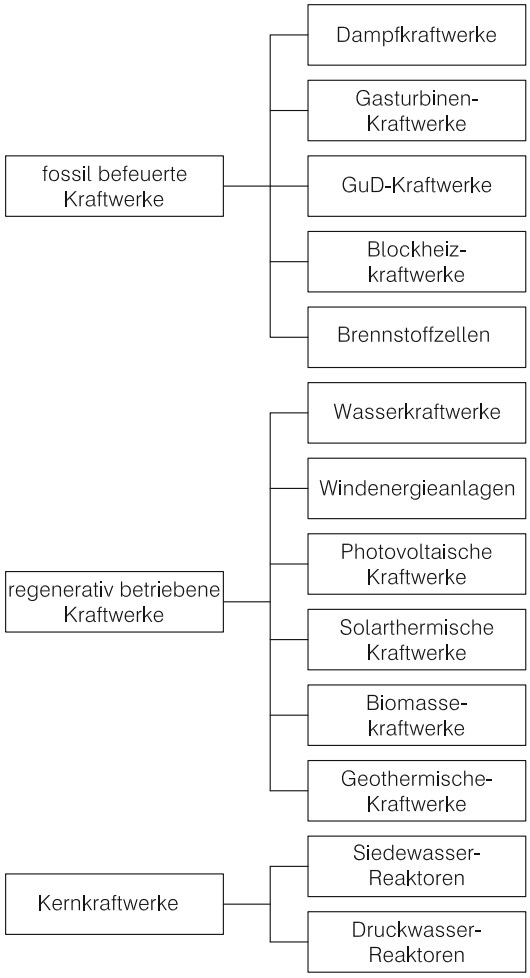


Abb. 87.9 Klassifizierung und Überblick der Kraftwerksarten

Weitere mögliche Bewertungen können nach der Einsatzart, dem Entwicklungsstand, dem Innovationsgehalt, der Regelbarkeit und der Wirtschaftlichkeit erfolgen, wodurch sich z. B. erst ein komplettes Bild der Fusionskraft ergibt. Jedoch ist mit keiner dieser Unterteilungen allein eine eindeutige Aussage über den praktischen Einsatz möglich. In den nächsten Abschnitten sollen die heute relevanten und zukünftig einsetzbaren Kraftwerksarten beschrieben werden. In Abb. 87.10 sind die Umwandlungswirkungsgrade wichtiger Kraftwerksarten abhängig von ihrer Leistung gezeigt.

87.2.4 Fossil befeuerte Kraftwerke

87.2.4.1 Kohlebefeuerte Dampfkraftwerke

87.2.4.1.1 Aufbau und Wirkungsweise

Die ersten kohlebefeuerten Dampfkraftwerke wurden 1882 von Thomas Edison in New York und London in Betrieb genommen. Seit dieser Zeit konnten der Druck und die Temperatur des Prozesses wesentlich erhöht werden, wodurch der Wirkungsgrad auf bis zu 47 % verbessert wurde. Ein Dampfkraftwerk besteht aus dem Kesselhaus mit Brennstoffaufbereitung, Feuerungsraum mit Brennern, Verdampfer und Überhitzer, einem durch das Kesselhaus geführten Wasser-Dampf-Kreislauf mit externem Kühlturm, der Turbine mit angekoppeltem Generator und der Rauchgasführung mit Schornstein, siehe Abb. 87.11. Im Dampfkraftwerk laufen die drei Prozessschritte Verbrennung und Verdampfung sowie Umwandlung der thermischen in mechanische Energie ab. Der Wirkungsgrad wird dabei umso besser, je höher die genutzte Energiedifferenz, d. h. die sog. Enthalpie H ist. Diese hängt von der temperaturabhängigen inneren Energie U sowie den Prozessgrößen Druck p und Volumen V ab:

$$H = U + p \cdot V \tag{87.1}$$

Massebezogen ergibt sich die spezifische Enthalpie h:

$$h = \frac{H}{m} = \frac{U}{m} + \frac{p \cdot V}{m} \tag{87.2}$$

Moderne Dampfkraftwerke mit einem Wirkungsgrad von bis zu 47 % arbeiten mit einem Dampfdruck von ca. 280 bar und einer

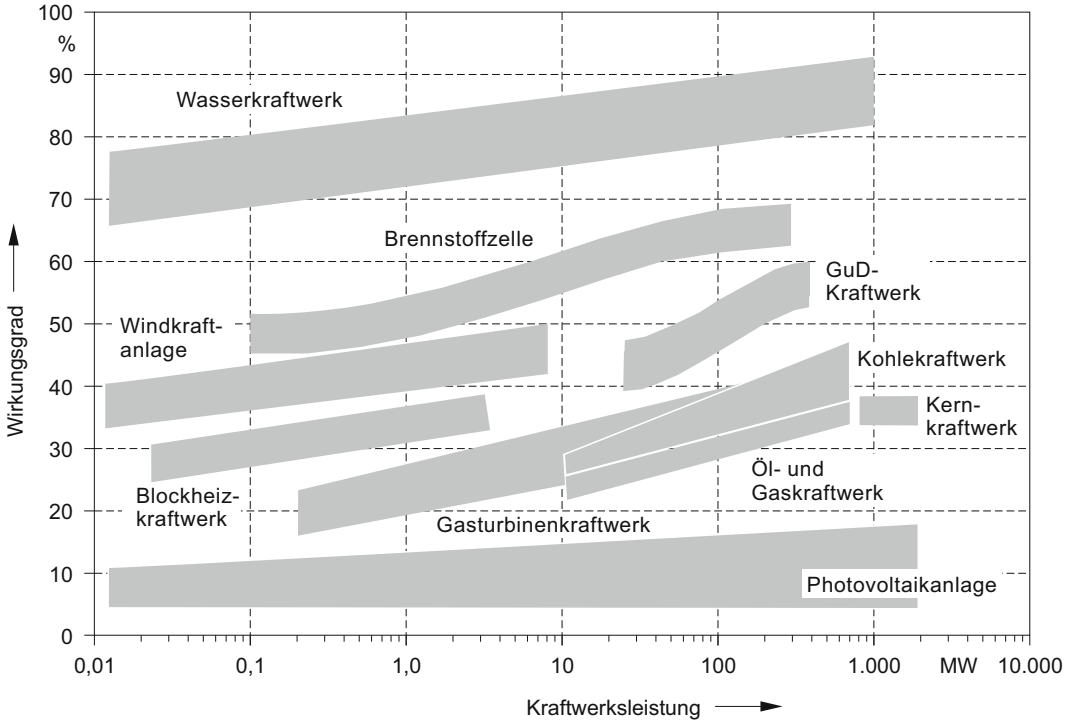


Abb. 87.10 Elektrischer Wirkungsgrad wichtiger Kraftwerksarten

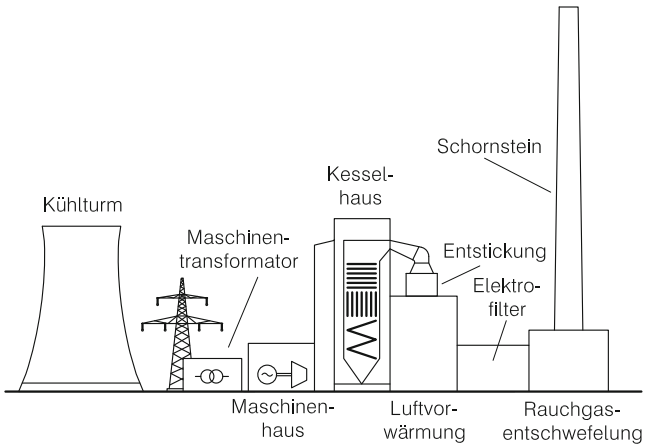


Abb. 87.11 Prinzipieller Aufbau eines Dampfkraftwerks

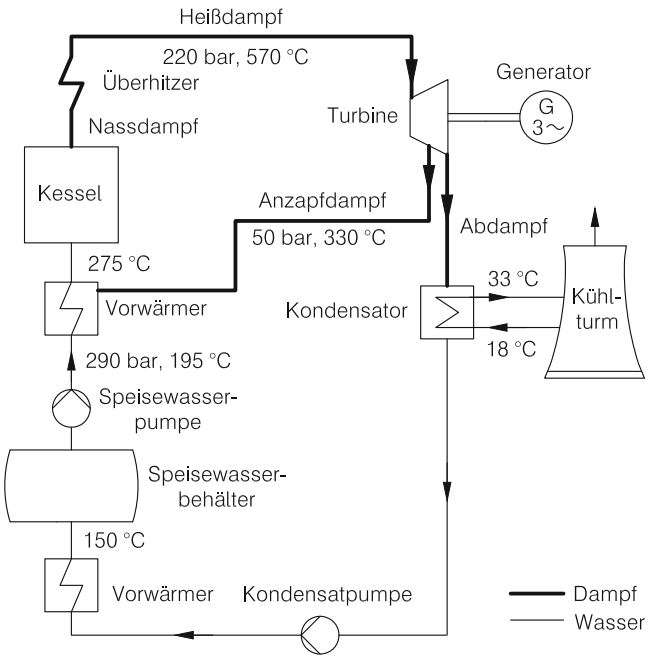


Abb. 87.12 Kreisprozess eines Dampfkraftwerks

Temperatur von bis zu 600 °C. Eine Erhöhung dieser Größen ist nur durch den Einsatz höher belastbarer Werkstoffe möglich. Ältere Kraftwerke arbeiten mit Drücken von 160 bar und Temperaturen um 530 °C; sie erreichen deshalb nur Wirkungsgrade von bis zu 38 %. Wird zusätzlich zur elektrischen Energie auch die Abwärme genutzt, steigt der Gesamtwirkungsgrad wesentlich an. Dies erfordert jedoch große Wärmeverbraucher wie Fernwärmenetze in einem Versorgungsradius von 5 bis 10 km. In diesem Fall spricht man von Heizkraftwerken mit Kraft-Wärme-Kopplung. Heutige Dampfkraftwerke werden in Leistungsstufen, den sog. Blockgrößen, von 500 bis 900 MW gebaut.

87.2.4.1.2 Wasser-Dampf-Kreisprozess

In Dampfkraftwerken läuft ein Kreisprozess ab, mit dessen Beschreibung beim kalten Speisewasser begonnen werden soll, siehe Abb. 87.12. Das Speisewasser wird in mehreren Stufen vorgewärmt und anschließend verdampft. Im Feuerungskessel

wird die Verbrennungswärme der Kohle an Rohre abgegeben. Durch diese Rohre, auch Verdampfer genannt, fließt Wasser mit hohem Druck von bis zu 290 bar, das anschließend in den Dampfzustand übergeht. Dieser Dampf enthält noch kleinste Wassertröpfchen, er wird deshalb auch Sattdampf oder Nassdampf genannt. Im nachfolgenden Überhitzer erfolgt der vollständige Übergang in den Gaszustand. Hier entsteht durch weitere Wärmezufuhr der Heißdampf, auch Frischdampf genannt, mit einer Temperatur von ca. 570 °C. Danach kann die Temperatur noch in einem Zwischenüberhitzer weiter erhöht werden. Anschließend wird der Dampf in eine Turbine geleitet, in der die Umwandlung der thermischen Energie in Bewegungsenergie erfolgt. Die Turbine besteht aus einem Hoch-, Mittel- und Niederdruckteil, die nacheinander vom Dampf durchströmt werden. Bei hohem Druck und hoher Temperatur ist das Dampfvolumen am kleinsten, weshalb die Abmessungen des Niederdruckteils am größten sind. In der Turbine wird die im Dampf gespeicherte Energie in eine Drehbewegung umgesetzt. Mit dem an der Turbinenwelle angekoppelten Generator wird dabei elektrische Energie erzeugt.

Nach der Turbine hat der Dampf nur noch eine geringe Temperatur von z. B. 33 °C und wird in einen Kondensator geleitet, in dem ihm bei Unterdruck so viel Wärme entzogen wird, dass er kondensiert. Dies erfolgt meist durch von Wasser durchflossenen Wärmetauschern, wobei die Wärme in einem Kühlturm abgegeben wird. Das Kondensat wird mit einer Pumpe in den Speisewasserbehälter gefördert, an dem die Beschreibung des Kreisprozesses begann. Der Wirkungsgrad des Wasser-Dampf-Prozesses kann durch die Vorwärmung des Speisewassers mit Anzapfdampf aus der Turbine optimiert werden.

87.2.4.1.3 Feuerung und Abgasbehandlung

Zur Feuerung werden Braun- oder Steinkohle eingesetzt. In Tab. 87.2 sind zum Vergleich die Heizwerte fester Brennstoffe zusammengefasst. Neben der vollständigen Verbrennung des Kohlenstoffs tritt praktisch auch ein unvollständiger Stoffumsatz und wegen des Schwefelgehalts der Kohle eine Schwefeloxidation auf:

$$\begin{aligned} &\text{vollständige Verbrennung:} && \mathrm{C + O_2 \rightarrow CO_2} \\ &\text{unvollständige Verbrennung:} && \mathrm{2\,C + O_2 \rightarrow 2\,CO} \\ &\text{Schwefeloxidation:} && \mathrm{S + O_2 \rightarrow SO_2} \end{aligned}$$

Als Feuerungsarten werden die Rostfeuerung zur Verbrennung fester Brennstoffe, die Staubfeuerung sowie die Wirbelschichtfeuerung eingesetzt. Die Rostfeuerung findet bei neuen Kraftwerksbauten keine Anwendung mehr. Zur Feuerung moderner Dampfkraftwerke wird Kohle genutzt, die zuvor in Kohlemühlen zu Staub gemahlen wurde. Der Kohlestaub wird mit Luft in den Feuerraum geblasen und verbrennt bei einer Temperatur von ca. 1200 °C. Höhere Kesseltemperaturen werden vermieden, da es dann zur Ascheerweichung und zur Bildung von Schlacke im Kessel kommen würde. Das bei der Verbrennung entstehende Abgas muss aufwändig behandelt werden. Dies erfolgt mit einer Rauchgasentschwefelungsanlage (REA), einer Entstickung, d. h. einer Anlage zur Verringerung der Stickoxide (DENOX) und Elektrofiltern zur Entstaubung, siehe Abb. 87.11. In der

Tab. 87.2 Heizwerte fester Brennstoffe bei 25 °C, wf – wasserfrei

Brennstoff	Heizwert H_U (wf) in MJ/kg	Brennwert H_O (wf) in MJ/kg
Nadelholz	18,8	20,2
Laubholz	18,4	19,7
Papier	15	
Getreidepflanze[a]	17,0	19,0
Getreidekörner[a]	17,0	18,4
Getreidestroh[a]	17,3	18,5
Braunkohle	20,6	
Steinkohle	25–32,7	29–32,7
Kohlenstoff	32,8	32,8

[a] Getreide: Weizen/Roggen

REA wird das Rauchgas über ein Gemisch aus gemahlenem Kalkstein und Wasser geführt, als Reaktionsprodukt entsteht dabei Gips. In der DENOX wird das Rauchgas über Katalysatoren geleitet. Der Staub wird in den Elektrofiltern mit einem elektrischen Feld aus dem Abgas entfernt. Bei modernen Kohlekraftwerken kann aufgrund der umfangreichen Abgasbehandlung auf hohe Schonsteine verzichtet werden, das gereinigte Abgas wird in der Mitte des Kühlturms in den warmen Abdampf geleitet und erhält dabei den erforderlichen Auftrieb. Sollen Dampfkraftwerke besonders schadstoffarm betrieben werden, wird bei Blockgrößen bis 300 MW die Wirbelschichtfeuerung eingesetzt. Hierbei wird Kohlestaub mit Ballaststoffen versetzt und mit einer geregelten und vorgewärmten Verbrennungsluft in den Feuerungskessel eingedüst. Dabei wird der Luftstrom so eingestellt, dass sich im Kessel ein Wirbelbett bildet. Dadurch kommt der Kohlestaub intensiv mit der Verbrennungsluft in Berührung und die Verbrennung erfolgt nahezu vollständig. Durch die im Vergleich zur konventionellen Kohlestaubfeuerung geringere Verbrennungstemperatur von 850 °C entstehen keine Stickoxide. Zur Vermeidung von schwefelhaltigem Abgas wird zusätzlich zum Brennstoff gemahlener Kalkstein eingebracht, daher ist keine REA notwendig. Die Entstaubung erfolgt mit Elektrofiltern. Wegen der geringeren Temperatur liegt der Wirkungsgrad bei 35 bis 38 %.

87.2.4.1.4 Einsatz und Entwicklungspotenzial

Dampfkraftwerke benötigen für das Anfahren einige Stunden, da die Bauteile sich temperaturbedingt ausdehnen; der Kessel eines 500-MW-Kraftwerks ist 80 m hoch, er dehnt sich beim Anfahren um bis zu 30 cm in der Höhe aus. Aufgrund ihrer trägen Regelung werden Dampfkraftwerke zur Grundlastversorgung eingesetzt. Teilweise werden in neueren Kraftwerken sog. Regelblöcke mit kleineren Leistungen und geringeren thermischen Zeitkonstanten zur besseren Regelfähigkeit eingesetzt. Aber auch durch die Neuauslegung großer Kraftwerksblöcke mit Leistungen von bis zu 1200 MW wird eine bessere Regelfähigkeit erreicht. Weiteres Entwicklungspotenzial liegt in der Verbesserung des Wirkungsgrades sowie der Verringerung von Emissionen. Soll der Wirkungsgrad weiter erhöht werden, muss der Temperaturbereich über 600 °C erschlossen werden. Dafür müssen geeignete Materialen entwickelt und erprobt werden. Dies ist anspruchsvoll, da sehr

temperaturfeste Materialien oft nur eine unzureichende mechanische Festigkeit aufweisen. Insbesondere für die Turbinen werden daher neue technologische Schritte erforderlich, wie etwa das Gießen von Mischmaterialien mit hoher mechanischer Festigkeit im Kernbereich und hochtemperaturfesten Oberflächen. Eine große Bedeutung kommt zukünftig der erweiterten Abgasbehandlung zu, da Kohlekraftwerke im Vergleich die höchsten Emissionen verursachen. Wesentlich ist hierbei die Abscheidung von Kohlendioxid. Prinzipiell kann die CO_2-Abscheidung auf drei Arten erfolgen: Pre-Combustion-Verfahren, Post-Combustion-Verfahren und Oxyfuel-Prozess. Beim Pre-Combustion-Verfahren werden die kohlenstoffhaltigen Brennstoffanteile vor der Verbrennung entfernt, beim Post-Combustion-Verfahren wird das Kohlenstoffdioxid nach der Verbrennung aus dem Rauchgas entfernt und beim Oxyfuel-Prozess erfolgt die Verbrennung in reiner Sauerstoffatmosphäre mit anschließender Verflüssigung des entstehenden Kohlenstoffdioxids. In Verbindung mit der Abgasbehandlung sind auch die Risiken von Dampfkraftwerken zu sehen. Mit der Abscheidung des Kohlendioxids entstehen zwei wesentliche Probleme. Einerseits sinkt der Wirkungsgrad vorrangig wegen der aufwändigen Aufbereitung der Verbrennungsluft, aber auch durch die Kühlung und Kompression des Abgases um bis zu zehn Wirkungsgradpunkte. Solch eine Verschlechterung kann nur mittel- oder langfristig durch innovative Entwicklungsschritte zur weiteren Prozessoptimierung ausgeglichen werden. Andererseits muss das komprimierte Gas gelagert oder verwertet werden. In Deutschland existiert ein Versuchskraftwerk zur CO_2-Abscheidung im Braunkohlekraftwerk Schwarze Pumpe bei Spremberg, das im Jahr 2014 wegen der geringen Chancen zur industriellen Nutzung dieser Technologie stillgelegt wurde. Projekte zur möglichen Verpressung von flüssigem Kohlendioxid in unterirdische Hohlräume, sog. Kavernen, sind bisher aufgrund widersprüchlicher Bewertungen von Gefahrenpotenzialen am Widerstand der Bevölkerung gescheitert. Es gibt jedoch erste Erfolge mit der stofflichen Nutzung von Kohlendioxid in der chemischen Industrie, z. B. bei der Produktion von Dämmstoffen.

87.2.4.2 Gasturbinen-Kraftwerke

Etwa ab 1935 begann die industrielle Nutzung von Gasturbinen, 1940 setzte ein Kraftwerk in Neuenburg in der Schweiz die erste Gasturbine ein. Ein Gasturbinen-Kraftwerk beinhaltet die Brennstoffzufuhr, die Gasturbine und den angekoppelten elektrischen Generator. Auf der Welle der Gasturbine sind der Verdichter, eine ringförmig aufgebaute Brennkammer, eine Turbine und der Rauchgaskanal hintereinander angeordnet (siehe Abb. 87.13). Im Verdichter wird die Frischluft angesaugt und auf einen Druck von 15 bar bis 20 bar komprimiert, dafür werden ca. 50 % der Wellenleistung benötigt. Rund 70 bis 80 % der verdichteten Luft werden der Brennkammer als Primärluft zugeführt, der Rest wird als Sekundärluft zur Kühlung der Brennkammer verwendet. Die verdichtete Luft wird mit dem Brennstoff Erdgas vermischt und den Brennern in der Ringbrennkammer zugeführt, wo das Gemisch gezündet wird. Bei der Verbrennung werden Temperaturen von bis zu 1500 °C erreicht. Solch hohe Temperaturen sind nur durch

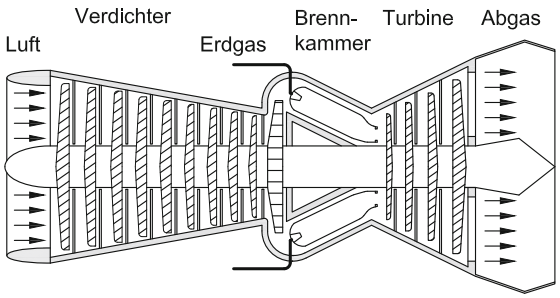


Abb. 87.13 Prinzip einer Gasturbine

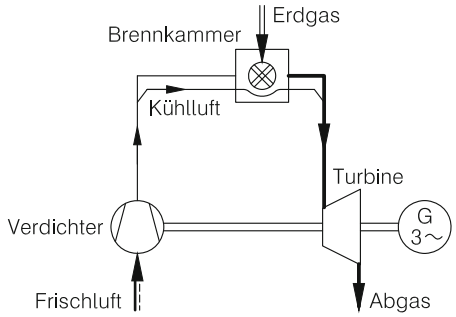


Abb. 87.14 Gasturbinen-Prozess

die Außenkühlung und Keramikauskleidung der Brennkammer sowie wärmestabile Turbinenschaufeln möglich. Die heißen Abgase durchströmen dann die Turbine, dabei wird ihre thermische Energie in mechanische Bewegungsenergie gewandelt. Werden die dann immer noch über 600 °C heißen Abgase anschließend ins Freie geleitet, handelt es sich um einen offenen Gasturbinenprozess, siehe Abb. 87.14. Der Wirkungsgrad moderner Gasturbinen-Kraftwerke beträgt bis zu 40 %, sie werden bis zu einer Leistung von 375 MW gebaut. Im Vergleich zu Kohlekraftwerken weisen sie geringere Investitionskosten und höhere Brennstoffkosten auf. Gasturbinen sind schnell anfahrbar und werden vorrangig als Spitzenlastkraftwerke eingesetzt. Das Entwicklungspotenzial von Gaskraftwerken liegt vorrangig im Einsatz hochtemperaturfester Materialien für die Brennkammer und die Turbinenschaufeln. Neben gasförmigen können auch flüssige Brennstoffe wie Dieselkraftstoff, Kerosin, Heizöl, oder Schweröl genutzt werden. In der Tab. 87.3 sind zum Vergleich die Heiz- und Brennwerte gasförmiger Brennstoffe und in Tab. 87.4 die Kenndaten flüssiger Brennstoffe aufgelistet.

Tab. 87.3 Heizwerte gasförmiger Brennstoffe bei 25 °C

Brennstoff	Heizwert H_U in MJ/kg	Brennwert H_O in MJ/kg
Wasserstoff	120	141,8
Kohlenmonoxid	10,1	10,1
Methan	50	55,5
Erdgas	38	42
Propan	46,3	50,3

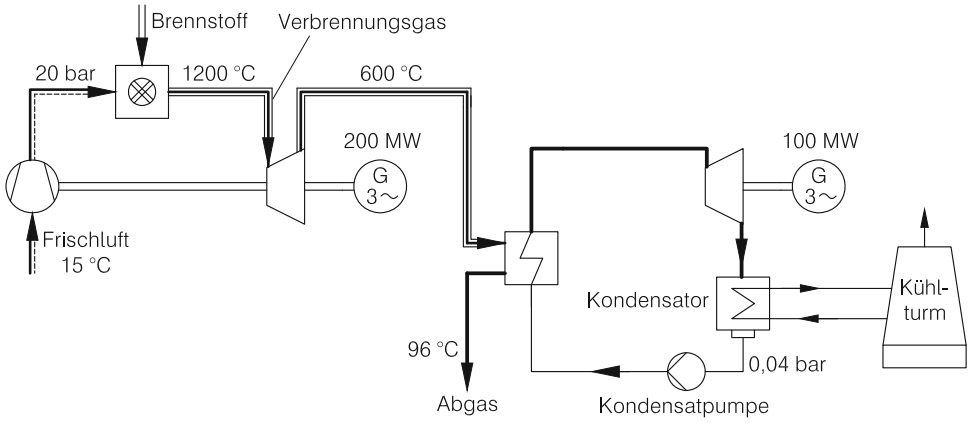


Abb. 87.15 Gas- und Dampfturbinen-Prozess

Tab. 87.4 Heizwerte flüssiger Brennstoffe bei 25 °C

Brennstoff	Heizwert H_U in MJ/kg	Brennwert H_O in MJ/kg
Spiritus	26	
Methanol	19,9	22,7
Ethanol	26,8	29,7
Benzin	41	44
Diesel, Heizöl	42,6	45,4
Kerosin	41	43
Schweröl	39,5	42,5
Biodiesel (RME)	37	40

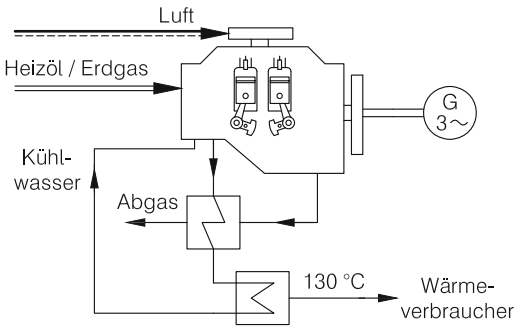


Abb. 87.16 Blockheizkraftwerk mit Kraft-Wärme-Kopplung

87.2.4.3 Gas- und Dampf-Kraftwerke

Beim Gas- und Dampf-Kraftwerk (GuD-Kraftwerk), auch Kombikraftwerk genannt, wird der Gasturbine zur Nutzung der hohen Abgastemperaturen ein Dampfkraftwerksprozess nachgeschaltet. Hierbei erfolgt die Verdampfung des Wassers mit dem Abgas der Gasturbine in einem Wärmetauscher. Die Gesamtanlage ist zweiwellig ausgeführt, d. h. mit je einem eigenen Generator, siehe Abb. 87.15. Der nachgeschaltete Dampfprozess liefert nochmals rund 50 % der Gasturbinennennleistung. Durch die Kopplung kann der Gesamtwirkungsgrad abhängig von der Kraftwerksgröße auf über 60 % gesteigert werden. Auch hier kann der Wirkungsgrad durch Kraft-Wärme-Kopplung weiter verbessert werden.

87.2.4.4 Blockheizkraftwerke

Bei einem Blockheizkraftwerk (BHKW) treibt ein mit Erdgas oder Diesel betriebener Verbrennungsmotor oder mit Wärme versorgter Stirlingmotor einen elektrischen Generator an. Der Wirkungsgrad von Dieselmotoren liegt bei 43 % und von Gasmotoren abhängig von der Motorenart bei 42 bis 44 %. Stirlingmotoren sind Wärmekraftmaschinen, bei denen sich ein fest eingeschlossenes Arbeitsgas durch äußere Erwärmung und Kühlung zwischen zwei Zylindern bewegt und dabei einen Kolben antreibt. Die Kolbenbewegung wird wie bei den Verbrennungsmotoren über eine Pleulstange in eine Drehbewegung umgesetzt. Zur Erhöhung des Wirkungsgrades kann auch bei BHKWs die Kraft-Wärme-Kopplung eingesetzt werden, d. h. die Abwärme wird dann als Heiz- oder Prozesswärme genutzt, siehe Abb. 87.16. Damit können Gesamtwirkungsgrade von bis zu 85 % erreicht werden. Der Leistungsbereich reicht von einigen kW bis in den MW-Bereich, Anlagen mit Leistungen von 1 bis 5 kW elektrisch und 5 bis 10 kW thermisch werden als Mikro-BHKW und von 5 bis 10 kW elektrisch und 10 bis 30 kW thermisch als Mini-BHKW bezeichnet. Mikro- und Mini-BHKW werden für den Einsatz in Ein- und Mehrfamilienhäusern angeboten, für den industriellen Einsatz werden größere Leistungen bis zu 15 MW gefertigt. Die Fertigungszahlen sind in Deutschland in den letzten Jahren stark angestiegen und lagen im Durchschnitt bei über 2 GW pro Jahr, wobei die größten Anteile immer noch in den Export gehen. Es werden in Deutschland bereits einige Anlagenbaureihen vertrieben, in denen anstatt des Verbrennungsmotors Brennstoffzellen eingesetzt werden.

87.2.4.5 Brennstoffzellen

Das Prinzip der Brennstoffzelle wurde bereits 1839 von William Robert Grove entdeckt. Im Jahr 1954 wurde der Prototyp

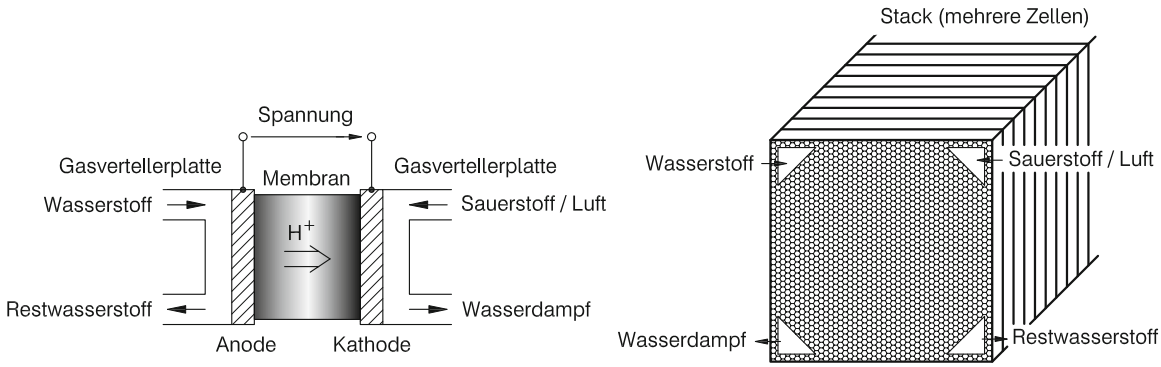


Abb. 87.17 Prinzip einer Brennstoffzelle und eines Brennstoffzellen-Stacks

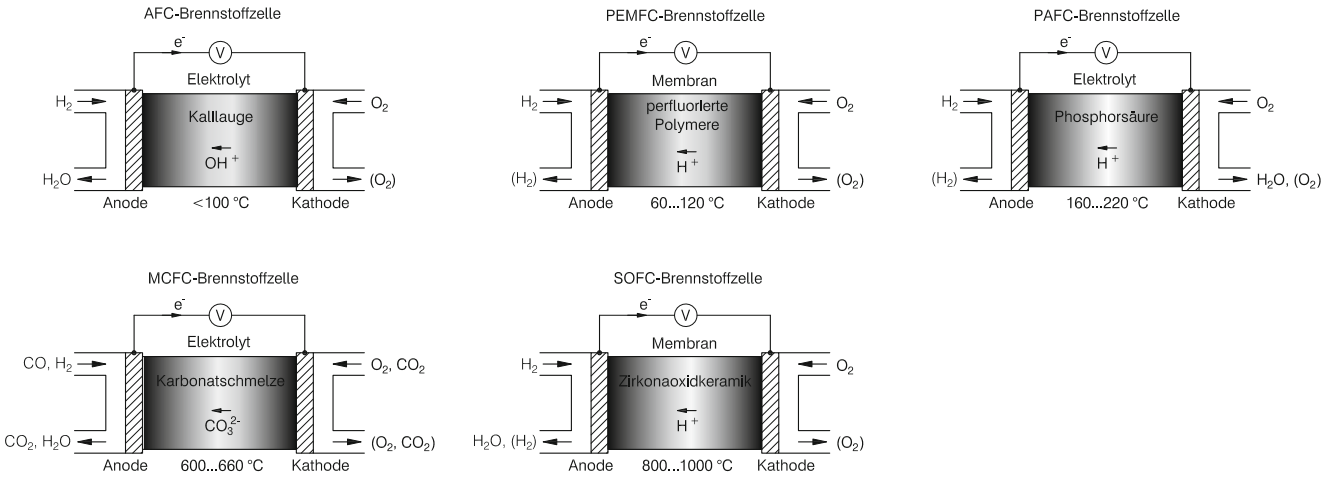


Abb. 87.18 Übersicht über Brennstoffzellen-Typen

einer alkalischen Hochdruckzelle von Francis T. Bacon entwickelt. Ab 1966 begann mit der Anwendung von Niedertemperaturbrennstoffzellen in verschiedenen Raumfahrtprogrammen die industrielle Nutzung, die sich allerdings nur sehr langsam entwickelte. Inzwischen kann diese Technologie als marktnah bezeichnet werden. Der Einsatz von Brennstoffzellen erfolgt in Wasserstoff-Pkws und Bussen, in der Luftfahrt als möglicher Ersatz für die heute genutzte Gasturbine zur Hilfsenergieversorgung, in der Schifffahrt als Energiequelle für den Antrieb von kleineren Schiffen und Unterseebooten sowie als sog. Mikro-, Mini- und Industriekraftwerk mit Kraft-Wärme-Kopplung. In einigen Ländern ist diese Technologie schon weit verbreitet, so sind z. B. in Japan rund 40.000 Brennstoffzellenheizgeräte installiert, in den USA sind einige hundert Brennstoffzellensysteme mit Kraft-Wärme-Kopplung im Leistungsbereich von mehreren hundert MW im Betrieb. Eine Brennstoffzelle besteht aus der Gaszu- und -abführung, dem Elektrolyt sowie der negativen (Anode) und positiven (Kathode) Elektrode zur Stromableitung, siehe Abb. 87.17. Wasserstoff und Sauerstoff reagieren zu Wasserdampf, zusätzlich verbleibt ein Anteil von nicht reagiertem Restgas. Bei dieser Reaktion werden Elektronen freigesetzt, es wird eine Gleichspannung erzeugt, die für bestimmte Nutzungsarten wie z. B. zur Netzeinspeisung, erst noch in Wechselspannung gewandelt werden muss. Die Reaktionsgleichungen der Brennstoffzelle lauten:

$$H_2 \rightarrow 2\,e^- + 2\,H^+$$

Anodenreaktion: Oxidation, Elektronenabgabe

$$2\,H^+ + 2\,e^- + \frac{1}{2}\,O_2 \rightarrow H_2O$$

Kathodenreaktion: Reduktion, Elektronenaufnahme

$$H_2 + \frac{1}{2}\,O_2 \rightarrow H_2O$$

Gesamtreaktion

Pro Brennstoffzelle wird eine Spannung von nur 0,5 bis 1 V erzeugt. Zur Erhöhung der Spannung müssen viele Zellen in Reihe geschaltet werden, praktisch werden sie in sog. Stacks hintereinander verschraubt. Niedertemperaturzellen arbeiten bei 80 bis 220 °C, Hochtemperaturzellen werden bei 600 bis 1000 °C betrieben und benötigen deshalb temperaturfeste Materialien, siehe Abb. 87.18. Dafür können sie direkt mit entschwefeltem Erdgas betrieben werden, während Niedertemperaturtypen für den Betrieb mit reinem Wasserstoff ausgelegt sind und bei der Speisung mit Erdgas zusätzliche Reaktionsstufen zur Gasaufbereitung und -reinigung benötigen. Aus diesem Grund liegt trotz des höheren Temperaturniveaus und der damit verbun-

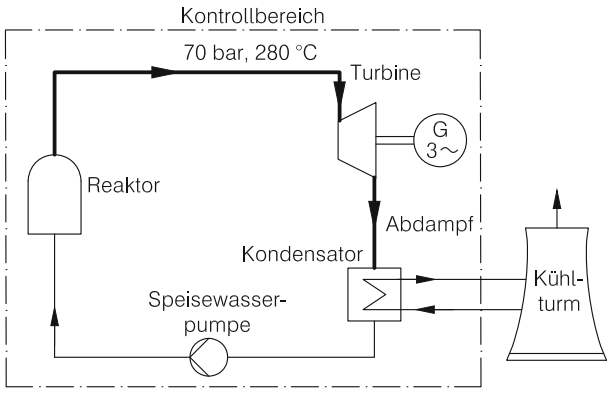


Abb. 87.19 Kernkraftwerk mit Siedewasserreaktor

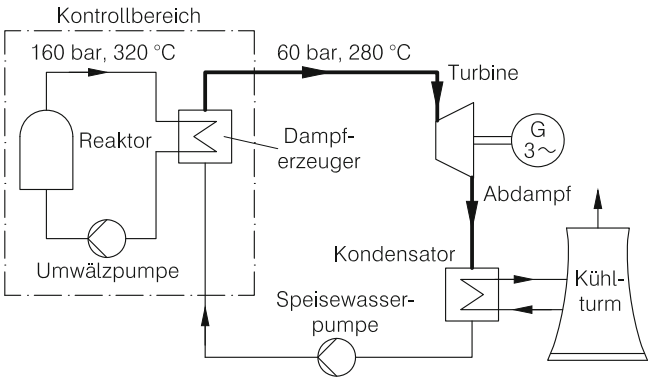


Abb. 87.20 Kernkraftwerk mit Druckwasserreaktor

denen Wärmeverluste der Wirkungsgrad von technisch ausgeführten Hochtemperaturzellen mit 48 % höher als bei den Niedertemperaturtypen mit 34 bis 38 %. Welche Temperaturen erreicht werden, hängt von der Art der Brennstoffzelle ab. Hierbei können folgende Niedertemperaturtypen unterschieden werden: Alkalische Brennstoffzelle (< 100 °C), Polymer Elektrolyt Membran Brennstoffzelle (60 bis 120 °C), Phosphorsaure Brennstoffzelle (160 bis 220 °C); die Direkt Methanol Brennstoffzelle (60 bis 120 °C) nimmt wegen ihrer Speisungsart eine Sonderstellung ein. Zu den Hochtemperaturtypen gehören die Schmelzkarbonat- (600 bis 660 °C) und die oxidkeramische (800 °C bis 1000 °C) Brennstoffzelle. Das Potenzial von Brennstoffzellen ist als sehr hoch einzuschätzen, da ihr theoretisch möglicher Wirkungsgrad im Bereich von 95 % bei geringen Temperaturen und 70 % bei hohen Temperaturen oberhalb des Carnot-Wirkungsgrads für Gasprozesse liegt und technisch noch lange nicht ausgeschöpft ist.

87.2.4.6 Kernkraftwerke

87.2.4.6.1 Kernspaltung

Otto Hahn und Fritz Straßmann entdeckten im Dezember 1938 die Spaltung des Urankerns. Lise Meitner und Otto Frisch lieferten 1939 die erste physikalisch-theoretische Deutung. Das erste Kernkraftwerk der Welt mit einer Leistung von 5 MW ging 1954 im russischen Obninsk in Betrieb. Im Jahr 1955 wurde im englischen Calder Hall das erste kommerzielle Kraftwerk mit 55 MW Leistung eingeschaltet. Der Wasser-Dampf-Kreislauf eines Kernkraftwerks funktioniert prinzipiell wie beim Dampfkraftwerk. Die Verdampfung des Wassers im Reaktor erfolgt durch die Energie einer gesteuerten Kernspaltungs-Kettenreaktion. Beim Siedewasserreaktor befinden sich der Reaktor, die Dampfturbine und der Kondensator in einem gemeinsamen Kreisprozess, das Wasser wird bereits im Reaktorbehälter verdampft, siehe Abb. 87.19. Der Druckwasserreaktor verfügt über zwei getrennte Kreisläufe für den Reaktor und den Turbinenbereich, die Verdampfung erfolgt über einen Wärmetauscher, siehe Abb. 87.20. Somit ist beim Druckwasserreaktor der kontaminierte Kontrollbereich wesentlich kleiner und der Überwachungsaufwand geringer. Während anfangs wegen ihrer einfachen Bauweise Siedewasserreaktoren errichtet wurden, haben sich inzwischen wegen ihres kleineren Kontrollbereichs Druckwasserreaktoren durchgesetzt. Im vom Wasser durchflossenen Reaktorbehälter von Leichtwasserreaktoren befinden sich in bis zu 250 Rohren die Brennstäbe. Als Brennstoff wird angereichertes Uran verwendet, d. h. U 238 angereichert mit 2,5 bis 3,5 % des Isotops U 235. Ein Beschuss des Urans mit Neutronen kann eine Kernspaltungs-Kettenreaktion verursachen, wenn die Geschwindigkeit der Neutronen gering genug ist. Dazu wird Wasser als sog. Moderator eingesetzt, da es die Neutronen abbremst. Durch die Kernspaltung wird das Wasser erhitzt, der Dampf treibt eine Turbine an und erzeugt über den angetriebenen Generator Strom. Prinzipiell kann die Leistung von Kernkraftwerken relativ schnell verändert werden. Wegen des vergleichsweise geringen Druck- und Temperaturniveaus werden Sattdampfturbinen eingesetzt. Der Wirkungsgrad von Kernkraftwerken liegt deshalb nur bei 30 %. Jedoch ist die im Brennstoff gespeicherte Energiemenge sehr hoch, der Heizwert von Uran U 235 beträgt $7{,}9 \cdot 10^7$ MJ/kg. Zum Vergleich beträgt der Heizwert von Braunkohle 20 MJ/kg.

Das Entwicklungspotenzial von Kernkraftwerken ist nahezu ausgeschöpft, da der Druck- und Temperaturbereich nicht ohne weiteres erhöht werden kann. Zu beachten ist, dass die Kühlung der Brennstäbe stets garantiert sein muss, da auch abgeschaltete Brennstäbe Wärme entwickeln. Ein Ausfall der Kühlung kann zu einem starken Temperaturanstieg und damit zur Gefahr einer Kernschmelze führen.

Neben den zwei beschriebenen Reaktorarten existieren folgende weitere Typen:

Schwerwasserreaktor: hier wird als Kühlmittel und Moderator schweres Wasser (Deuterium) genutzt, dessen Herstellung sehr teuer ist. Schwerwasserreaktoren benötigen jedoch kein angereichertes Uran und können somit dann eingesetzt werden, wenn keine Uran-Anreicherung verfügbar ist.

Hochtemperaturreaktor: hier werden die Kernbrennstäbe in Form von Blöcken oder Kugeln mit Graphit als Moderator umhüllt. Es sind theoretisch wesentlich höhere Arbeitstemperaturen möglich. Ein wirtschaftlicher Einsatz dieser Technologie fand noch nicht statt.

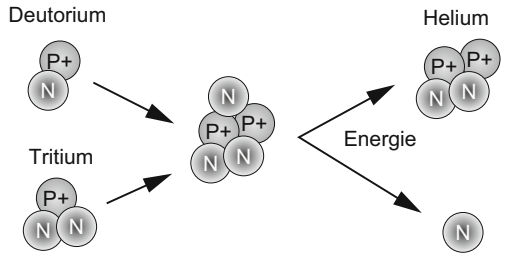


Abb. 87.21 Prinzip der Kernfusion, N – Neutron, P – Positron

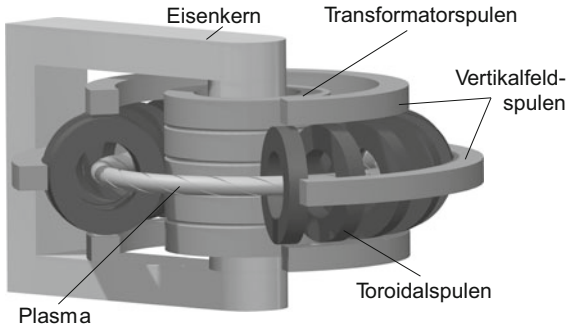


Abb. 87.22 Schematischer Aufbau des Fusionsreaktors ITER

Brutreaktor: hier wird als Brennstoff ein Uran-Plutonium-Mischoxid genutzt. Es kann gleichzeitig mit der Energiewandlung mehr spaltbares Material produziert werden als im Reaktor verbraucht wird. Plutonium ist wegen seiner möglichen militärischen Nutzung wesentlich problematischer als Uran. Zusätzlich erfordert das für diesen Reaktortyp als Kühlmittel verwendete Natrium besondere Brandschutzmaßnahmen.

87.2.4.6.2 Kernfusion

Ernest Rutherford konnte im Jahr 1917 und damit vor der Entdeckung der Kernspaltung erste Fusionsreaktionen im Labor nachweisen, sein Assistent Mark Oliphant führte 1934 die erste gezielte Reaktion im Labor durch. Bei der Kernfusion verschmelzen zwei Atomkerne, zum Beispiel die Isotope des Wasserstoffs Deuterium und Tritium. Da Deuterium auf verschiedenen Wegen erzeugt werden kann, ist kein Uran mehr als Brennstoff notwendig. Bei der Kernfusion haben die Endprodukte der Reaktion nicht die gleiche Masse wie die Anfangsprodukte. Die Massendifferenz wird nach Einsteins Formel $E = m \cdot c^2$ in Energie umgesetzt. Bei der Fusion wird eine im Vergleich zu anderen Reaktionen enorm große Energiemenge freigesetzt, beispielsweise 18,5 MeV bei der Reaktion von Deuterium (D) und Tritium (T) zu Helium (He):

$$^{2}_{1}\mathrm{d} + ^{3}_{1}\mathrm{t} \rightarrow ^{4}_{2}\mathrm{He} + ^{1}_{0}\mathrm{n}$$

(Energieumsatz $\Delta E = 18{,}5\,\mathrm{MeV}$).

Hierbei ist d ein Kern des Deuterium (Deuteron), t ein Kern des Tritium (Triton) und n ein Neutron, siehe Abb. 87.21. Es ergibt sich ein Heizwert von $3{,}6 \cdot 10^8\,\mathrm{MJ/kg}$. Zum Vergleich liegt der Heizwert von Uran U 235 bei der Kernspaltung bei $7{,}9 \cdot 10^7\,\mathrm{MJ/kg}$. Der Vorteil der Kernfusion gegenüber der Kernspaltung liegt somit in der wesentlich höheren Energieausbeute pro Brennstoffmenge. Eine solche Reaktion läuft auf der Sonne und allen anderen leuchtenden Sternen ab. Sie findet in einer Wolke aus Plasma statt. Plasma ist eine Mischung aus allen drei Aggregatzuständen fest, flüssig und gasförmig und wird deshalb oft als vierter Aggregatzustand bezeichnet. Die Probleme bei der technischen Umsetzung ergeben sich aus der hohen Reaktionstemperatur von 100 Mio. °C sowie insbesondere aus einer noch höheren Zündtemperatur des Prozesses von bis zu 400 Mio. °C. Die Kernfusion wird seit den 1930er Jahren intensiv erforscht. Zuerst wurde die unkontrollierte Fusion bei der Entwicklung der Wasserstoffbombe militärisch eingesetzt, 1952 detonierte die erste Wasserstoffbombe im Pazifik. Seitdem wurde intensiv an der kontrollierten Fusion gearbeitet, jedoch gelang erst im Jahr 1991 im englischen Oxfordshire eine nur zwei Sekunden dauernde Kernfusion im Joint European Torus (JET), einer europäischen Forschungseinrichtung, bei der zwei MW erzeugt wurden. Allerdings musste dafür ein Vielfaches der produzierten Energie aufgewendet werden. Im südfranzösischen Forschungszentrum Cadarache wird am „International Thermonuclear Experimental Reactor“ (ITER) gearbeitet, dessen schematischer Aufbau in Abb. 87.22 gezeigt ist. Als Zielstellung soll sich die Fusionsreaktion von Deuterium und Tritium energetisch selbst versorgen und somit ein kontinuierlicher Prozess ablaufen. Hierbei soll eine Leistung von 500 Megawatt über einige Minuten erzeugt werden. Das Plasma soll durch Magnetfelder gehalten werden, die durch gekühlte elektrische Spulen erzeugt werden. Es wird erwartet, dass bis zur ersten industriellen Anwendung noch 50 bis 80 Jahre intensiver Forschung erforderlich sind. Auch bei der Kernfusion entsteht eine erhebliche radioaktive Strahlung, dadurch können hohe gesundheitliche Risiken auftreten. Gleichzeitig gibt es wie bei der Kernspaltung, wenn auch mit vergleichsweise geringerem Umfang, das Problem der atomaren Endlagerung.

87.2.5 Regenerative Kraftwerke

87.2.5.1 Wasserkraftwerke

Wasserkraft wird in Form ihrer bewegten Masse mit einer Dichte von rund 1 kg/dm^3 bei Normaldruck und 3,98 °C für die Energieerzeugung nutzbar. Die Nutzungsmöglichkeiten sind abhängig von den Niederschlagsmengen und den geologischen Bedingungen sehr unterschiedlich. In Norwegen werden z. B. 99 % der elektrischen Energie aus Wasserkraft gewonnen, in Deutschland sind es ca. 3 %. Bereits im Jahr 1880 baute man in England das erste Wasserkraftwerk zur Erzeugung elektrischer Energie, 1896 entstand an den Niagarafällen in den USA das erste Großkraftwerk. Die industrielle Nutzung begann mit der Entwicklung effektiver Turbinen Anfang des 20. Jahrhunderts. Ein Wasserkraftwerk besteht aus einer Wasserturbine mit angekoppeltem elektrischem Generator. Zur Inbetriebnahme der Wasserturbinen müssen nur Schieber geöffnet werden, deshalb beträgt die Anfahrzeit nur ein bis zwei Minuten. Eine Unterscheidung kann nach der Bauart und Wasserspeicherung

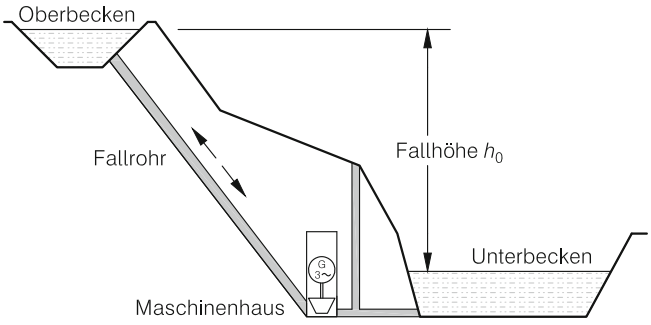


Abb. 87.23 Prinzip eines Pumpspeicherwerks

erfolgen. Laufwasserkraftwerke sind Fließkraftwerke und stellen prinzipiell eine Staustufe in einem Fluss dar. Sie nutzen die anfallende Wassermenge und können abhängig von der Größe bei hohen Leistungen Wirkungsgrade bis 94 % erreichen. Speicherkraftwerke hingegen, auch Talsperrenkraftwerke genannt, nutzen das zufließende Wasser nicht unmittelbar. Es wird in Zeiten mit geringem Stromverbrauch gesammelt und in Zeiten mit hohem Leistungsbedarf genutzt. Pumpspeicherwerke weisen im Gegensatz dazu wie in Abb. 87.23 gezeigt einen oberen und unteren Speichersee auf. In Schwachlastzeiten wird das Wasser mit preiswerter elektrischer Energie in den oberen Speichersee gepumpt. In Zeiten erhöhten Stromverbrauchs wird die potenzielle Energie des Wassers über Turbinen und Generatoren in elektrische Energie gewandelt, indem es in den unteren See abfließt. Bei großen Anlagen sind Wirkungsgrade bis 90 % pro Speicherrichtung erreichbar. Pumpspeicherwerke werden als Spitzenlastkraftwerk und zur Netzregelung eingesetzt. Die in ihnen gespeicherte potenzielle Energie E_p errechnet sich mit der Masse m, der Fallbeschleunigung g und der nutzbaren Fallhöhe h_p:

$$E_p = m \cdot g \cdot h_p \tag{87.3}$$

Für die Masse können das Volumen V und die Dichte ρ eingesetzt werden:

$$E_p = \rho \cdot V \cdot g \cdot h_p$$

Um das Ergebnis in der Einheit kWh anstatt Ws zu erhalten, werden die Umrechnungsfaktoren eingesetzt:

$$E_p = \frac{1}{3{,}6 \cdot 10^6} \cdot \rho \cdot V \cdot g \cdot h_p$$

Die potenzielle Leistung P ergibt sich aus der angenommenen Lade-/Entladezeit t, diese Zeit ist abhängig vom realisierbaren Volumenstrom:

$$P = \frac{E_p}{t} \tag{87.4}$$

In Deutschland werden Pumpspeicherwerke mit einer Gesamtleistung von 6600 MW und einer mittleren Speicherzeit von acht Stunden genutzt.

Neben den beschriebenen Wasserkraftwerken existieren noch weitere Typen:

Gezeitenkraftwerke: Die veränderlichen Anziehungskräfte von Sonne, Mond und Erde verursachen den Gezeiteneffekt, dessen Tidenhub sich zur Energiegewinnung nutzen lässt. Bei Hochwasser wird ein Speicherbecken gefüllt, das bei Niedrigwasser entleert wird. Die dabei auftretenden Wasserströmungen werden über Kaplan-Turbinen geführt, wegen ihrer Mindestfallhöhe ist ein minimaler Tidenhub von 5 m erforderlich. Zur Energiegewinnung wird ein großes Wasserbecken benötigt, wobei erst natürliche Buchten einen wirtschaftlichen Bau ermöglichen. Das größte Gezeitenkraftwerk liegt in der Bucht von St. Malo in Frankreich und weist eine Leistung von 240 MW bei 12 bis 16 m Tidenhub auf.

Wellenkraftwerke: Es wird die Bewegungsenergie von Wellen genutzt, die bei 3 bis 6 m Wellenhöhe bei bis zu 40 kW pro Meter Wellenkamm liegen kann. Als direkte Wandlungsprinzipien sind die Nick-Bewegung von asymmetrischen Schwimmkörpern, die Relativbewegung von Körpern zueinander, z. B. von Schwimmflossen sowie die Bewegung aufrecht stehender Kolben, bekannt. Darüber hinaus kann auch eine Plattform mit Wasser überspült werden, das anschließend über Turbinen zurück ins Meer geleitet wird. Eine Sonderform stellt das Brandungskraftwerk dar, bei dem Wasser in Hohlkammern geleitet wird, aus denen beim Einströmen Luft durch eine weitere Auslassöffnung verdrängt und beim Ausströmen wiederum eingesaugt wird. Dabei wird eine Luftturbine angetrieben. Kommerziell betriebene Wellenkraftwerke existieren bisher als Brandungskraftwerke.

Strömungskraftwerke: Hierbei wird die Wasserströmung im Meer oder in Flüssen genutzt. Dafür lassen sich ähnlich wie bei Windenergieanlagen zwei oder dreiflügelige Horizontalrotoren, aber auch vertikal drehende Rotoren nutzen. Da sich die Meeresströmungen gezeitenabhängig umkehren, müssen die Rotoren um 180° drehbar sein. Es existieren einige Prototypen, wie z. B. die SeaGen-Anlage vor der Küste Nordirlands mit 16 m Rotordurchmesser und 1,2 MW Leistung.

Osmosekraftwerke: Süß- und Salzwasser werden in zwei Behälter gefüllt, die durch eine gemeinsame Membran miteinander verbunden sind. Durch die Membran erfolgt ein Konzentrationsausgleich, es dringt Süßwasser in den mit Salzwasser gefüllten Behälter. Das Süßwasser verdünnt das Salzwasser und erhöht dessen Druck, der z. B. über eine Turbine in Energie umgewandelt werden kann. Das Osmosekraftwerk wird deshalb auch als Salzgradientenkraftwerk bezeichnet. Dieser Kraftwerkstyp wird in Verbindung mit zum Meer ablaufendem Süßwasser verwendet und noch nicht industriell genutzt. Jedoch existiert seit dem Jahr 2009 ein Kleinstkraftwerk im norwegischen Tofte.

87.2.5.2 Windenergieanlagen

Ab 1891 wurden in Dänemark Windenergieanlagen (WEA) zur Versorgung netzferner Regionen mit elektrischer Energie aufgebaut. Die wissenschaftlichen Grundlagen für die systematische Nutzung der Windenergie wurden in den 1920er-Jahren gelegt. Im Jahr 1941 wurde in den USA die erste Großanlage mit 1250 kW errichtet. Ab 1975 wurde die Stromerzeugung aus Windenergie international vorangetrieben. In Deutschland begann die industrielle Nutzung von Windenergieanlagen mit

dem Stromeinspeisungsgesetz von 1990, in dem die Abnahme und Vergütung von elektrischer Energie geregelt wurde. Seit 2000 gilt das Erneuerbare-Energien-Gesetz, dass im Schnitt alle vier Jahre überarbeitet wird. Hierin wurden der vorrangige Netzanschluss regenerativer Energien sowie garantierte Vergütungssätze festgeschrieben, was zu einem starken Zubau geführt hat. Inzwischen wird Windenergie auch direkt vermarktet, wobei ein höheres wirtschaftliches Risiko für den Anlagenbetreiber entsteht. Eine Nutzung der Windenergie ist sowohl an Land (Onshore) als auch auf dem Meer (Offshore) möglich. Offshore-Installationen sind wesentlich kostenintensiver, dafür ist ihre Energieausbeute höher. Heutige Leistungen ausgeführter Anlagen reichen bis zu acht MW, WEA mit zehn MW sind in der Entwicklung, es besteht noch ein Entwicklungspotenzial bis weit oberhalb von zehn MW.

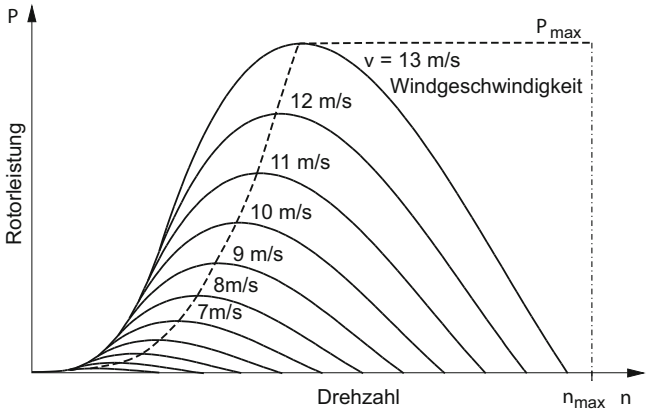


Abb. 87.24 Leistungs-Drehzahl-Kennlinienfeld von Windkraftanlagen und Kennlinie des drehzahlvariablen Betriebs

87.2.5.2.1 Energieentnahme aus dem Wind

Wind entsteht aufgrund von Temperatur- und Druckunterschieden von Luftmassen. Die theoretische Leistung des Windes P_0 ist von der Windgeschwindigkeit v^3 zur dritten Potenz und damit vom Standort sowie der Turmhöhe und der angeströmten Rotorkreisfläche A abhängig. Die Luftdichte ρ hat nur einen geringen Einfluss, sie ändert sich in Abhängigkeit von Druck und Temperatur. Somit erhält man für die theoretische Leistung des Windes:

$$P_0 = \frac{1}{2}\rho \cdot A \cdot v^3 \tag{87.5}$$

Die Energieentnahme aus dem Wind ist physikalisch begrenzt, die Umwandlung der kinetischen Energie bewegter Luft in mechanische Arbeit erfolgt nach (87.6):

$$P = \frac{1}{2}\rho \cdot A \cdot v^3 \cdot c_p \tag{87.6}$$

Der Leistungsbeiwert c_p gibt an, wie viel der im Wind enthaltenen Energie mit der Windturbine entnommen wird. Die theoretisch erreichbare Energieausbeute liegt nach BETZ bei $c_p = 0{,}593$, bei höherer Energieentnahme würde sich der nachfolgende Wind aufstauen. In der Praxis ist der Leistungsbeiwert kleiner, er wird wesentlich durch das Rotorblattprofil bestimmt. Typische Werte moderner WEA liegen bei $c_p = 0{,}43$ bis $0{,}5$. Der Verlauf des Leistungsbeiwerts verursacht den Abfall der Leistungs-Drehzahl-Kennlinie, siehe Abb. 87.24. Das Maximum der Leistung liegt abhängig von der Windgeschwindigkeit bei unterschiedlichen Drehzahlen, weshalb WEA zur Optimierung der Energiewandlung drehzahlvariabel betrieben werden.

87.2.5.2.2 Prinzipieller Aufbau und Netzanbindung

Der Aufbau und die Netzkopplung von Windenergieanlagen sind in Abb. 87.25 gezeigt. In der Praxis finden sich fast ausschließlich Dreiblattanlagen mit Windanströmung von vorn, sog. Luvläufer. Es wird zwischen Konzepten mit Getriebe und Standardgeneratoren mit kleiner Polpaarzahl, wegen der hintereinander liegenden Komponenten aufgelöster Triebstrang genannt, und getriebelosen Anlagen mit hochpoligen Generatoren (konzentrierter Triebstrang) unterschieden. Abhängig von der Windgeschwindigkeit erreicht der Rotor von WEA Drehzahlen im Bereich von 8 bis 22 Umdrehungen pro Minute. Ob für die Drehzahlanpassung zwischen WEA-Rotor und Generator ein Getriebe erforderlich ist, hängt von der Polpaarzahl des Generators ab. Die Summe der magnetischen Nord- und Südpole ist die Polzahl der elektrischen Maschine. Die Polpaarzahl ergibt sich als die Menge der Paare von Magnetpolen, sie entspricht der halben Polzahl. Aus der vom Generatorgleichrichter bestimmten Statorfrequenz des Generators f_1 und der Polpaarzahl p errechnet sich die Synchrondrehzahl n_S des Generators in Umdrehungen pro Minute:

$$n_S = \frac{f_1 \cdot 60\ \frac{\text{s}}{\text{min}}}{p} \tag{87.7}$$

Die Synchrondrehzahl einer elektrischen Maschine beschreibt den Punkt, an dem der Rotor synchron zum Drehfeld des Stators rotiert. Werden im Generator viele Stator-Wicklungsspulen realisiert, resultieren daraus eine hohe Polpaarzahl und nach (87.7) eine geringe Synchrondrehzahl. Liegt die Synchrondrehzahl des Generators im Bereich der WEA-Rotordrehzahl, kann auf ein Getriebe verzichtet werden. Neben der Anlage befindet sich der Transformator, der bei Einzelanlagen den Anschlusspunkt zum öffentlichen Netz der Mittelspannung bei beispielsweise 20 kV bildet. Oft werden viele Anlagen in einem Windpark konzentriert, dann erfolgt der Netzanschluss über eine größere Schaltanlage am Hochspannungsnetz bei 110 kV oder bei sehr großen Windparks auch auf der Höchstspannungsebene 220 oder 380 kV.

87.2.5.2.3 Funktionsprinzip von Windenergieanlagen

Das Funktionsprinzip von Windenergieanlagen besteht in der Umwandlung der Windenergie über Rotorblätter in mechanische Bewegungsenergie. Für die Erzeugung elektrischer Energie wird der Windgeschwindigkeitsbereich von 4 bis 25 m/s (14 bis 90 km/h) genutzt, bei höheren Geschwindigkeiten schalten die Anlagen zum Schutz der Mechanik ab. Anlagen auf See werden für höhere Windgeschwindigkeiten bis 30 m/s (108 km/h) dimensioniert. Zur aerodynamischen Regelung stehen die stall-, active-stall- und pitch-Regelung zur Verfügung,

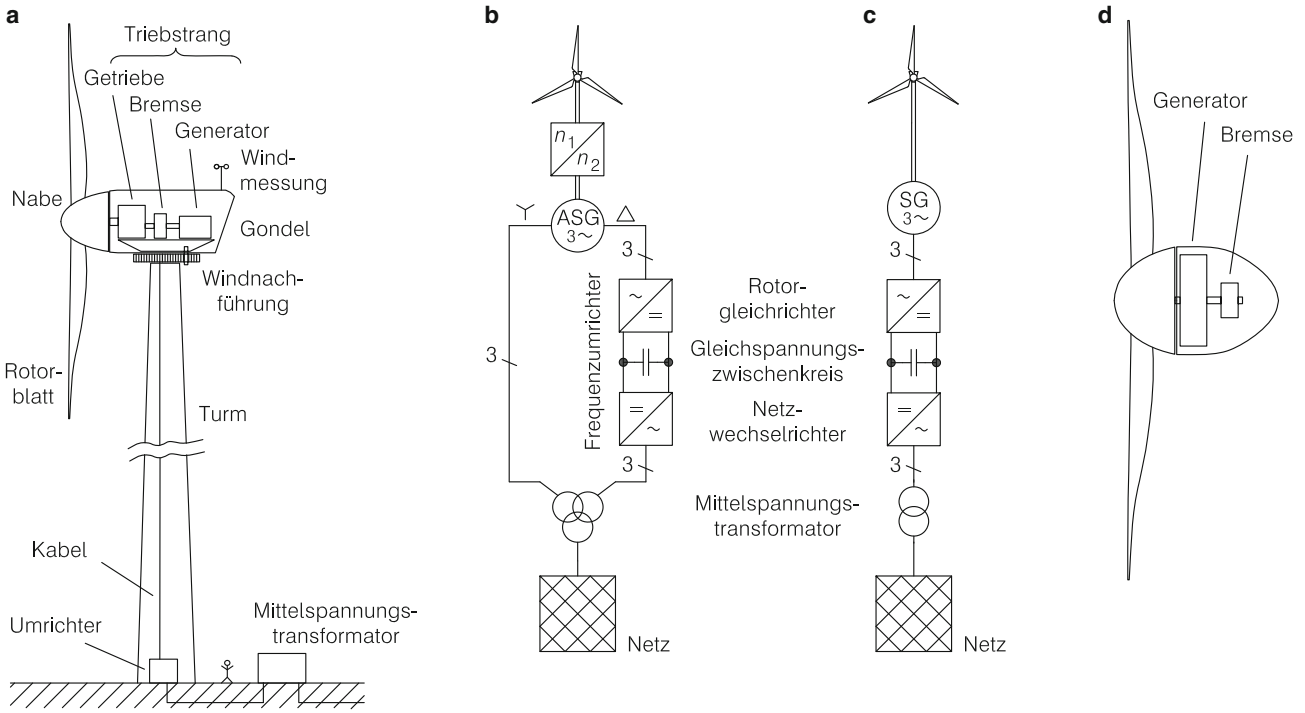


Abb. 87.25 Aufbau von Windenergieanlagen: **a** Prinzipdarstellung einer WEA mit Getriebe und vierpoligem Generator (aufgelöster Triebstrang), **b** Übersichtsschaltbild für den Netzanschluss eines doppelt gespeisten vierpoligen Generators, **c** Übersichtsschaltbild für den Netzanschluss einer getriebelosen Anlage mit vielpoligem Synchrongenerator, **d** Turmkopf einer getriebelosen WEA mit vielpoligem Generator (konzentrierter Triebstrang)

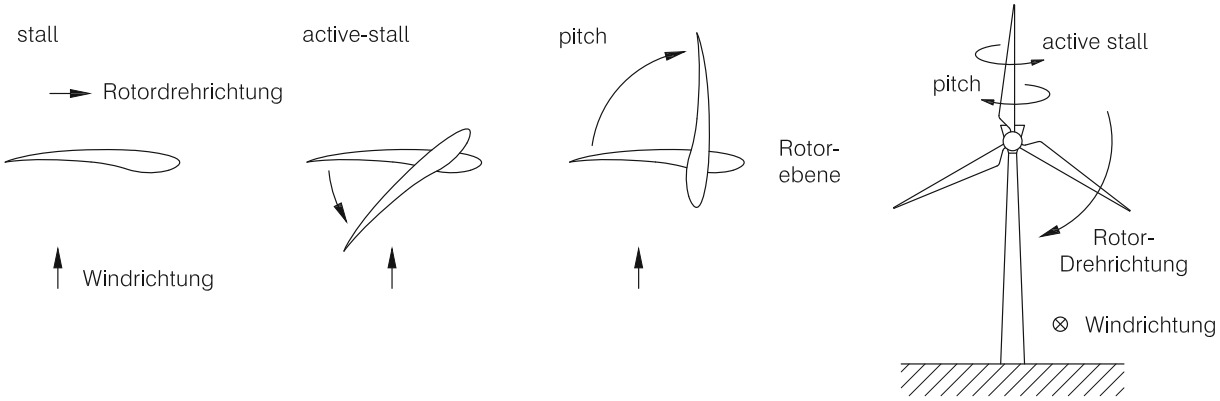


Abb. 87.26 Aerodynamische Leistungsbegrenzung am Rotorblatt von WEA, von *links*: stall, active-stall, pitch, Übersicht des Gesamtaufbaus

siehe Abb. 87.26. Ältere Anlagen mit Leistungen unterhalb 1 MW werden stall-geregelt, d. h. abhängig von der Windgeschwindigkeit findet ein Strömungsabriss an der Hinterkante des Rotorblatts statt. Bei diesen Anlagen wird ein von den Windverhältnissen am Standort abhängiger Rotorblattwinkel fest eingestellt. An den Enden der Rotorblätter befinden sich mehrere Meter lange Bremsspitzen, sog. tip-Spitzen, die mechanisch oder hydraulisch betätigt werden. Diese unterstützen den mechanischen Bremsvorgang zur Stillsetzung der WEA. Auch bei der active-stall-Regelung findet ein Strömungsabriss statt, der jedoch über eine Bewegung des Rotorblatts nach vorn in den Wind eingestellt werden kann. Aufgrund der Gegenkräfte benötigt der Einstellvorgang etwas Zeit. Bei der pitch-Regelung können die Anströmungseigenschaften der Rotorblätter durch drehbare Achsen an die Windgeschwindigkeit angepasst werden. Steigt die Windgeschwindigkeit und damit die mechanische Belastung der WEA an, können die Rotorblätter unterstützt von der Windkraft schnell verstellt werden. Bei modernen Anlagen im Leistungsbereich ab 1 MW wird inzwischen ausschließlich die pitch-Regelung eingesetzt. Zu beachten ist, dass die pitch-Antriebe zur Sicherstellung einer hohen Verfügbarkeit doppelt ausgeführt werden müssen, da keine Bremsspitzen mehr an den Rotorblättern vorhanden sind. Seit einigen Jahren werden die Rotorblätter von WEA abhängig von ihrer Position einzeln pitch-geregelt, um auf die Windscherung reagieren zu können und damit zusätzliche Kraftwirkungen auf die Anlage

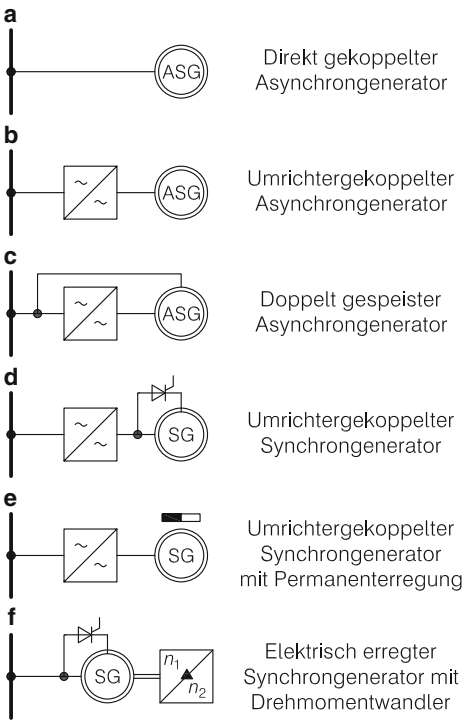


Abb. 87.27 Mögliche Generator-Leistungselektronik-Varianten zur Netzkopplung von Windenergieanlagen

durch höhenabhängig unterschiedliche Windgeschwindigkeiten oder auch den Turmschatteneffekt zu vermeiden.

87.2.5.2.4 Generatoren und Konzepte zur Netzkopplung von Windenergieanlagen

An den Rotor ist ein elektrischer Generator gekoppelt, der die Energie direkt oder über leistungselektronische Wandler in das Netz einspeist. Dafür werden Asynchron- oder Synchrongeneratoren mit meist leistungselektronischer Netzkopplung eingesetzt, siehe Abb. 87.27. Bei den verwendeten Symbolen für die leistungselektronischen Frequenzumrichter ist zu beachten, dass sie jeweils wie in Abb. 87.25 gezeigt aus einem generator- und netzseitigen Wechselrichter bestehen. Die in WEA genutzten Drehstrom-Generatoren unterscheiden sich in der Wirkungsweise nach dem Synchron- und Asynchronprinzip. Damit wird das Verhältnis der elektrischen Größen von Stator und Rotor zueinander beschrieben. Synchron arbeitende Maschinen weisen keine Geschwindigkeitsdifferenz der Drehfelder im Stator- und Rotorkreis auf, beide Felder rotieren synchron zueinander. Synchronmaschinen rotieren stets mit der Synchrondrehzahl nach (87.7). Asynchron arbeitende Maschinen weisen unterschiedliche Geschwindigkeiten von Stator- und Rotorfeld auf. Es tritt eine Drehzahldifferenz zur Synchrondrehzahl auf. Die Differenz wird als Schlupf s bezeichnet. Der Schlupf errechnet sich aus der Synchrondrehzahl n_S und der aktuellen Drehzahl n:

$$s = \frac{\Delta n}{n_S} = \frac{n_S - n}{n_S} \tag{87.8}$$

Für die Netzkopplung von WEA-Generatoren können unterschiedliche Konzepte genutzt werden, die mit unterschiedlichem technischem Aufwand verbunden sind.

Direkt gekoppelter Asynchrongenerator: Bei diesem Typ werden die Statorklemmen des Generators direkt mit dem Netz verbunden, siehe Abb. 87.27a. Für eine netzverträgliche Zuschaltung wird ein Thyristorsteller verwendet, der im Normalbetrieb mit einem Schalter überbrückt wird. Diese Ausführung wird als „Dänisches Konzept“ mit stall-Regelung und drehzahlstarrer Ausführung in den Anfangsjahren der netzeinspeisenden Windenergienutzung eingesetzt. Es ist keine Drehzahlregelung vorhanden und die Einspeisung von Wirk- und Blindleistung ist nicht regelbar. Hier werden alle Windschwankungen unverzögert als Leistungsschwankungen in das Netz übertragen. Der variable Stromfluss verursacht variable Spannungsfälle über der Netzimpedanz und trägt somit zu Spannungsschwankungen bei. Diese wirken sich negativ auf das Drehmomentverhalten von Drehfeldmaschinen aus und verursachen Leuchtdichteschwankungen an Beleuchtungseinrichtungen. Aufgrund seiner starken Netzbeeinflussung ist dieses Konzept inzwischen nicht mehr zulässig.

Umrichtergekoppelter Asynchrongenerator: Hier werden durch den Frequenzumrichter zwischen Generator und Netz deren unterschiedliche elektrische Frequenzen über den Gleichspannungszwischenkreis voneinander entkoppelt, siehe Abb. 87.27b. Somit wird ein drehzahlvariabler Betrieb des Generators möglich. Im Generator wird ein Drehstromsystem erzeugt, dessen Amplitude und Frequenz proportional zur Drehzahl sind. Die Drehzahl des Generators ergibt sich aus der mechanischen Antriebsleistung des Windes. Mit dem Generatorwechselrichter, der als gesteuerter Gleichrichter arbeitet, wird dieses Drehstromsystem in eine Gleichspannung umgeformt. Durch die Gleichspannung können die unterschiedlichen Frequenzen auf der Generator- und Netzseite voneinander entkoppelt werden. Über den Gleichspannungszwischenkreis wird nur Wirkleistung übertragen. Zur Netzeinspeisung wird aus der Gleichspannung des Zwischenkreises mit dem Netzwechselrichter durch die im Abschn. 87.2.5.2.5 beschriebene Pulsweitenmodulation ein Drehstromsystem mit drei Sinusspannungen erzeugt, wobei sich abhängig von der Pulsfrequenz Abweichungen von der Sinusform ergeben. Somit existieren auf der Generator- und Netzseite Drehstromsysteme mit unterschiedlichen Frequenzen, die durch eine Gleichspannung voneinander getrennt und damit entkoppelt sind. Seit ca. 15 Jahren werden nur noch drehzahlvariable WEA mit leistungselektronischer Netzkopplung in Kombination mit pitch-Regelung zur Rotorblattverstellung eingesetzt. Beim drehzahlvariablen Betrieb kann durch den generatorseitigen Wechselrichter der Arbeitspunkt des Generators geregelt werden; zur ökonomischen Optimierung wird bei der Netzeinspeisung ein maximaler Wirk- und minimaler Blindleistungsanteil eingestellt. Damit wird der höchste Wirkungsgrad möglich, gleichzeitig wird auch die Erwärmung des Generators begrenzt. Da bei diesem Konzept die gesamte Nennscheinleistung des Generators über den dadurch kostenintensiven Frequenzumrichter geführt wird, wird dieser auch als Vollumrichter bezeichnet.

Tab. 87.5 Flussrichtung der Leistungen im Generatorbetrieb des DGASG: − Abgabe, + Aufnahme

Betriebsart	Drehzahlbereich	Schlupf s	P_s	P_r
Untersynchron	$n < n_s$	> 0	−	+
Synchronbetrieb	$n = n_s$	0	0	0
Übersynchron	$n > n_s$	< 0	−	−

Doppelt gespeister Asynchrongenerator (DGASG): Bei diesem Konzept wird der Stator des Generators direkt und der Rotor über einen bidirektional arbeitenden Frequenzumrichter mit dem Netz verbunden, siehe Abb. 87.27c. Im untersynchronen Generatorbetrieb, d. h. unterhalb der Synchrondrehzahl des Generators nach (87.7) nimmt der Rotor Leistung auf und der Stator gibt Leistung ab, im übersynchronen Bereich geben Rotor und Stator Leistung ab. Tab. 87.5 zeigt die Leistungsflussrichtungen des DGASG im Generatorbetrieb. Die Gesamtleistung P wird aus der Stator- (P_S) und Rotorleistung (P_R) bestimmt, das Vorzeichen der Summation ergibt sich abhängig vom Drehzahlbereich nach Tab. 87.5:

$$P = P_S \pm P_R \tag{87.9}$$

Die Rotorleistung P_R ist bei vereinfachter Betrachtung unter Vernachlässigung der Verluste direkt von der Statorleistung P_S und dem Schlupf s abhängig:

$$P_R = P_S \cdot s \tag{87.10}$$

Der Betriebsbereich von DGASG wird so gewählt, dass der Umrichter kostensparend dimensioniert werden kann. In der Praxis wird der Drehzahlbereich des Generators um maximal ±30 % um die Synchrondrehzahl variiert. Im Drehzahlbereich $n = \pm 0{,}3 \cdot n_s$ nimmt der Schlupf s einer vierpoligen Drehstromasynchronmaschine mit der Polpaarzahl $p = 2$ nach (87.8) folgende Werte an:

$$s_{-30\,\%} = \frac{n_s - 0{,}7 n_s}{n_s} = 0{,}3$$

$$s_{+30\,\%} = \frac{n_s - 1{,}3 n_s}{n_s} = -0{,}3$$

Die dabei über den Rotorumrichter übertragene Leistung P_R errechnet sich nach (87.10) aus der Statorleistung P_S und dem Schlupf s:

$$P_R = P_S \cdot \pm 0{,}3$$

Somit muss der Rotorumrichter bei einer in der Praxis anzutreffenden maximalen Drehzahlvariation des Generators von ±30 % um die Synchrondrehzahl nur für 30 % der Gesamtleistung dimensioniert werden, siehe Abb. 87.28.

Die Drehzahl des DGASG ergibt sich wegen der doppelten Speisung aus der Frequenzdifferenz der elektrischen Größen im Rotor und Stator:

$$n = \frac{f_R \pm f_S}{p} \tag{87.11}$$

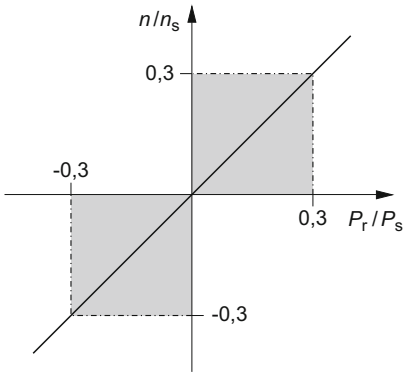


Abb. 87.28 Drehzahl-Leistungsbereich des Rotorkreises des DGASG beim Einsatz in WEA

Auch wenn nur ein Teil der Gesamtleistung über den Rotorumrichter geregelt wird, können die gesamte Wirk- und Blindleistung unter Anwendung der hier nicht näher beschriebenen Zweiachsentheorie voneinander entkoppelt und damit unabhängig voneinander eingestellt werden, d. h. die Verstellung der Wirkleistung kann ohne Veränderung der aktuellen Blindleistung erfolgen.

Umrichtergekoppelter Synchrongenerator mit elektrischer Erregung: Bei diesem Konzept werden Standardgeneratoren mit der Polpaarzahl zwei in Kombination mit Getriebe oder hochpolige Generatoren mit Polpaarzahlen im Bereich von $p = 40$ als getriebelose Anlagen eingesetzt, siehe Abb. 87.27d. Der Rotor des Generators wird mit Gleichspannung versorgt und baut einen Magnetpol auf, der durch die Rotorbewegung der WEA bei anliegender Statorspannungserregung einen Stromfluss in den Statorwicklungen erzeugt. Obwohl der Synchrongenerator gut regelbar ist, muss auch bei diesem Konzept wegen der variablen mechanischen Drehzahl und der damit verbundenen variablen Ausgangsfrequenz des Generators eine Frequenzentkopplung über den Gleichspannungszwischenkreis erfolgen. Der Frequenzumrichter muss wie beim umrichtergekoppelten Asynchrongenerator für die gesamte Nennscheinleistung dimensioniert werden. Grundsätzlich kann der elektrisch erregte Synchrongenerator in Kombination mit einer Hilfsspannungsquelle für den Netzaufbau genutzt werden, d. h. er ist schwarzstartfähig. Dieser Generatortyp weist bei hohen Polpaarzahlen große Abmessungen und ein hohes Gewicht auf, die elektrische Erregung verursacht Verluste. Der Vorteil dieser Erregungsart liegt darin, dass das Verhalten des Generators jederzeit gut regelbar ist.

Umrichtergekoppelter Synchrongenerator mit Permanentmagneterregung: Hierbei wird ein Synchrongenerator mit Permanentmagnetrotor mit einem Vollumrichter kombiniert, siehe Abb. 87.27e. Zur Realisierung einer kompakten Erregung werden auf dem Rotor Permanentmagnete befestigt. Diese Magnete aus einer Neodym-Eisen-Bor-Legierung weisen hohe magnetische Feldstärken im Bereich von $H = 880\,\mathrm{kA/m}$ und volumetrische Energiedichten von $w = 460\,\mathrm{kJ/m^3}$ auf, dadurch können die Abmessungen und das Gewicht dieses Generatortyps im Vergleich zum elektrisch erregten Typ um 30 bis 50 % reduziert

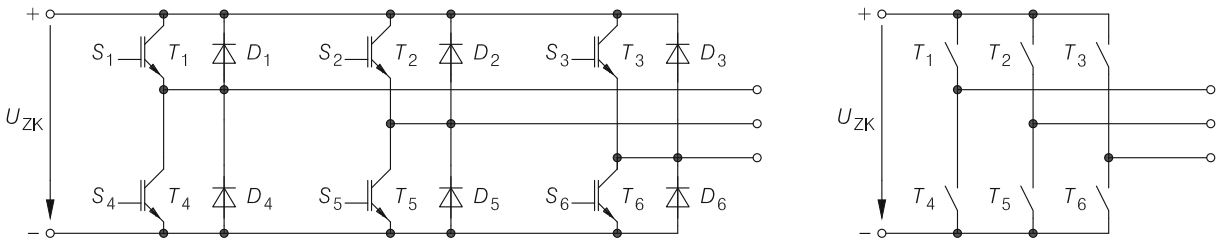


Abb. 87.29 Dreiphasiger Wechselrichter (*links*) und vereinfachte Ersatzschaltung (*rechts*)

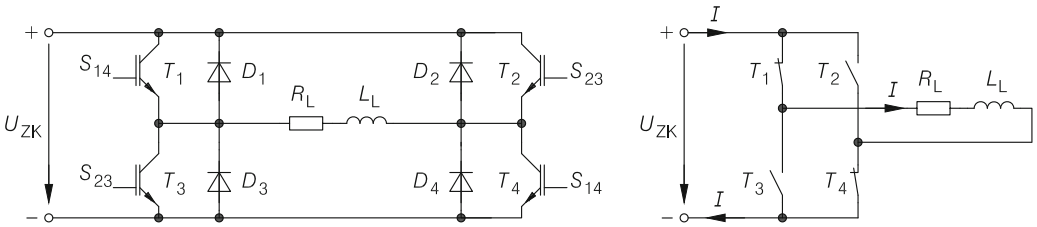


Abb. 87.30 Einphasiger Wechselrichter (*links*) und vereinfachte Ersatzschaltung (*rechts*)

werden. Da die magnetische Feldstärke durch das Material bestimmt wird, kann die Erregung nicht geregelt werden und auch keine Verluste verursachen. Die Ausgangsspannung des Generators U_G hängt wegen der fest eingebauten Magnetisierung nur noch von der Drehzahl n ab:

$$U_G \sim n = k \cdot n \tag{87.12}$$

k-Konstante.

Zum Schutz gegen Korrosion und die damit verbundene Degradation muss das Magnetmaterial versiegelt werden. Durch die Kombination mit einem Vollumrichter bleibt das netzseitige Verhalten der WEA gut über die Leistungselektronik regelbar, Wirk- und Blindleistung werden auch hier unabhängig voneinander eingestellt.

Direkt gekoppelter Synchrongenerator mit Drehzahlwandler: Bei diesem relativ neuen Konzept wird ein elektrisch erregter Synchrongenerator mit einer kleinen Polpaarzahl von meist zwei direkt mit dem Netz verbunden, siehe Abb. 87.27f. Da der Synchrongenerator in direkter Netzkopplung immer bei Netzfrequenz einspeist, muss seine Drehzahl genauso wie die herkömmlicher Kraftwerksgeneratoren konstant bleiben. Dazu ist er mit einem Drehmomentwandler mit dem Rotor der WEA gekoppelt, der in mechanischer Bauart als Planetengetriebe oder hydrodynamisch als Hydraulikgetriebe ausgeführt sein kann. Der Wandler wird so eingestellt, dass er unabhängig von der Eingangsdrehzahl des WEA-Rotors immer eine konstante Ausgangsdrehzahl für den Antrieb des Synchrongenerators liefert. Da der Synchrongenerator mit seiner elektrischen Erregung netzseitig gut regelbar ist, kann die Leistungselektronik bei diesem Konzept entfallen.

87.2.5.2.5 Pulsweitenmodulation (PWM) bei selbstgeführten Wechselrichtern

Da sowohl bei den Windenergieanlagen als auch bei den später behandelten Photovoltaikanlagen die Wandlung von Gleich- zu Wechselgrößen mit der Pulsweitenmodulation (PWM) erfolgt, soll dieses Prinzip hier vertieft werden. In WEA werden ausschließlich dreiphasige Wechselrichter eingesetzt, siehe Abb. 87.29. Diese bestehen aus drei Brückenzweigen mit je zwei Halbleiterschaltern, weshalb diese Schaltungstopologie auch als B6-Brückenschaltung bezeichnet wird. Als Halbleiterschalter werden meist IGBTs (Insulated Gate Bipolar Transistor-Bipolartransistoren mit isoliertem Steueranschluss) verwendet. Antiparallel zu den eigentlichen Schaltern sind Dioden angeordnet, die den Stromfluss bei ausgeschalteten Schaltern übernehmen, um diese gegen Überspannungen und Zerstörung durch Durchschlag zu schützen. Vereinfacht kann ein dreiphasiger Wechselrichter durch Schalter nachgebildet werden, siehe Abb. 87.29 rechts. Die Wechselrichter werden selbstgeführt betrieben, d. h. sie erzeugen ihre Schaltsignale unabhängig von der Netzspannung selbst. Dazu wird das Verfahren der Pulsweitenmodulation (PWM) genutzt.

Deren grundsätzliche Funktionsweise soll zur besseren Veranschaulichung anhand eines einphasigen Wechselrichters erläutert werden, siehe Abb. 87.30, da die Ergebnisse unmittelbar auf dreiphasige Systeme übertragbar sind. Der einphasige Wechselrichter besteht aus zwei Brückenzweigen mit je zwei Schaltern, zu denen jeweils eine Diode antiparallel (parallel, jedoch in Gegenrichtung leitend) verschaltet ist. Der Wechselrichter wird aus dem Gleichspannungszwischenkreis des Frequenzumrichters gespeist. Zwischen den Brückenzweigen befindet sich eine ohmsch-induktive Last. Zur vereinfachten Betrachtung soll die Ersatzschaltung aus Abb. 87.30 rechts verwendet werden, in der die Transistoren durch Schalter ersetzt wurden.

Zur Umwandlung von Gleichspannung in Wechselspannung wird die Periodendauer der gewünschten Sinusspannung in eine Vielzahl von Zeitabschnitten mit der Pulsperiodendauer T_P unterteilt, siehe Abb. 87.31. Innerhalb dieser Pulsperiodendauer kann eine konstante Gleichspannung ein- und ausgeschaltet werden:

$$T_P = t_{\text{ein}} + t_{\text{aus}} \tag{87.13}$$

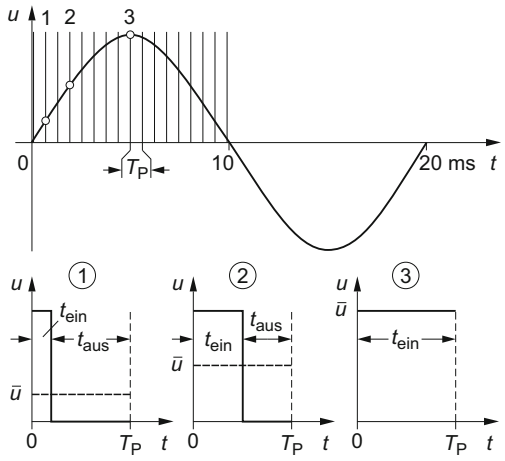


Abb. 87.31 Prinzip der Pulsweitenmodulation, *oben*: Spannungs- und Pulsverlauf, *unten*: Variation der Einschaltdauer innerhalb einer Pulsperiode

Damit ergibt sich im Zeitabschnitt T_P das Verhältnis aus der Einschaltzeit t_{ein} zur Gesamtperiodendauer, das mit dem Tastverhältnis TV beschrieben wird:

$$TV = \frac{t_{\text{ein}}}{t_{\text{ein}} + t_{\text{aus}}} = \frac{t_{\text{ein}}}{T_P} \tag{87.14}$$

Wird die Gleichspannung über die volle Periodendauer T_P eingeschaltet, ist $TV = 1$ und der Mittelwert der Spannung im Zeitintervall entspricht dem vollen Gleichspannungswert. Wird eine volle Periodendauer nicht eingeschaltet, ist $TV = 0$ und der Mittelwert ebenfalls null. Dazwischen liegende Werte der Ausgangsspannung u_a lassen sich durch die Variation der Einschaltzeit t_{ein} proportional einstellen:

$$u_a \sim TV \sim \frac{t_{\text{ein}}}{t_{\text{ein}} + t_{\text{aus}}} \sim \frac{t_{\text{ein}}}{T_P} \tag{87.15}$$

Somit kann im Zeitverlauf durch Variation der Einschaltzeit t_{ein} jeder beliebige Amplitudenwert einer Sinuskurve nachgebildet werden, siehe Abb. 87.31.

Der Spannungswert errechnet sich aus dem Tastverhältnis und der Gleichspannung, die bei Wechselrichtern Zwischenkreisspannung U_{ZK} genannt wird:

$$u_{TV} = \frac{t_{\text{ein}}}{t_{\text{ein}} + t_{\text{aus}}} \cdot U_{\text{ZK}} \tag{87.16}$$

$$u_{TV} = TV \cdot U_{\text{ZK}} \tag{87.17}$$

Das mittlere Tastverhältnis $\overline{TV}$ kann durch Mittelwertbildung über die Anzahl der Pulsperioden m bestimmt werden:

$$\overline{TV} = \frac{\sum_{n=1}^{m} TV_n}{m} \tag{87.18}$$

Daraus lässt sich bei bekannten Einzelwerten in den betrachteten Zeitintervallen der Mittelwert der Spannung $\bar{u}_a$ in einem beliebigen Zeitraum ermitteln:

$$\bar{u}_a = \frac{\sum_{n=1}^{m} u_{TVn}}{m} = \overline{TV} \cdot U_{\text{ZK}} \tag{87.19}$$

Unter der Bedingung eines konstanten Stromes ist auch die Ausgangsleistung des Wechselrichters P_{WR} proportional zum Tastverhältnis:

$$P_{\text{WR}} \sim \overline{TV} \tag{87.20}$$

Damit lässt sich (87.15) um die Leistung erweitern:

$$u_a \sim TV \sim P$$

Der Ausgangsstrom $\underline{i}_a$ des Wechselrichters errechnet sich aus der Ausgangs-spannung $\underline{u}_a$ und den Impedanzen des Wechselrichters $\underline{Z}_{\text{WR}}$ sowie des Netzes $\underline{Z}_N$:

$$\underline{i}_a = \frac{\bar{u}_a}{\overline{\underline{Z}}_{\text{WR}} + \overline{\underline{Z}}_N} = \frac{TV \cdot \bar{u}_{a,\max}}{\overline{\underline{Z}}_{\text{WR}} + \overline{\underline{Z}}_N} \tag{87.21}$$

In der oberen Halbschwingung nach Abb. 87.31, d. h. den ersten 10 ms, werden dabei jeweils die Schalter bzw. Transistoren T1 und T4 innerhalb der Pulsperiodendauer T_P gemäß (87.14) gepulst, es stellt sich der in der Ersatzschaltung in Abb. 87.30 rechts markierte Stromverlauf über der Last ein. Für die untere Halbschwingung erfolgt von 10 ms bis 20 ms die Pulsung der Schalter bzw. Transistoren T2 und T3. Dabei stellt sich ein zur in Abb. 87.30 rechts gezeigten Stromrichtung entgegengesetzter Stromfluss über die Last ein. Zur Erzeugung einer Wechselspannung benötigt der Wechselrichter Steuersignale, die z. B. durch einen Sinus-Dreieck-Vergleich erzeugt werden können, siehe Abb. 87.32. Beim Sinus-Dreieck-Vergleich entstehen in den oberen und unteren Halbschwingungen Schnittpunkte der zwei Funktionen.

Diese liefern die Schaltsignale zur Ansteuerung der vier Schalter des einphasigen Wechselrichters. Soll ein dreiphasiger Wechselrichter angesteuert werden, verwendet man an Stelle einer Sinusfunktion drei um je 120° zeitlich zueinander verschobene Sinuskurven als Vergleichsfunktionen zur Dreiecksfunktion und erhält damit die erforderlichen sechs Schaltsignale zur Ansteuerung der Transistoren.

Es muss sichergestellt werden, dass in den Brückenzweigen des Wechselrichters kein Kurzschluss auftritt. Deswegen dürfen die Transistoren eines Brückenzweigs nie gleichzeitig schalten. Die Sinusspannung wird dadurch erzeugt, dass die Transistoren T1/T4 in der positiven und die Transistoren T2/T3 in der negativen Halbwelle angesteuert werden. Die antiparallelen Dioden führen den Strom in den Zeiten, in denen die Transistoren ausgeschaltet sind, sie ermöglichen somit einen kontinuierlichen Stromfluss. Bei der Sinus-Dreieck-Ansteuerung entsteht ein PWM-Signal nach Abb. 87.32. Die Nachbildung der gewünschten Sinus-Ausgangsspannung ist umso besser, je höher die Frequenz der Dreiecksspannung gewählt wird. Wenn das Verhältnis der Frequenzen von der Dreiecksfunktion und der Sinusfunktion kleiner als 15 ist, sollten die Funktionen synchronisiert sein, d. h. sie sollten identische Nulldurchgänge aufweisen. Wird diese Bedingung nicht erfüllt, verläuft die Ausgangsspannung halbwellenunsymmetrisch. Wird die Frequenz der Dreiecksspannung größer als 750 Hz gewählt, ist dieser Effekt vernachlässigbar. Bei der Wahl einer Frequenz der Dreieckspannung von z. B. 1 kHz entsteht eine gepulste Spannung nach Abb. 87.33. Der Stromverlauf kann ebenfalls dem Bild entnommen werden, seine Form wird durch die Art der Last und

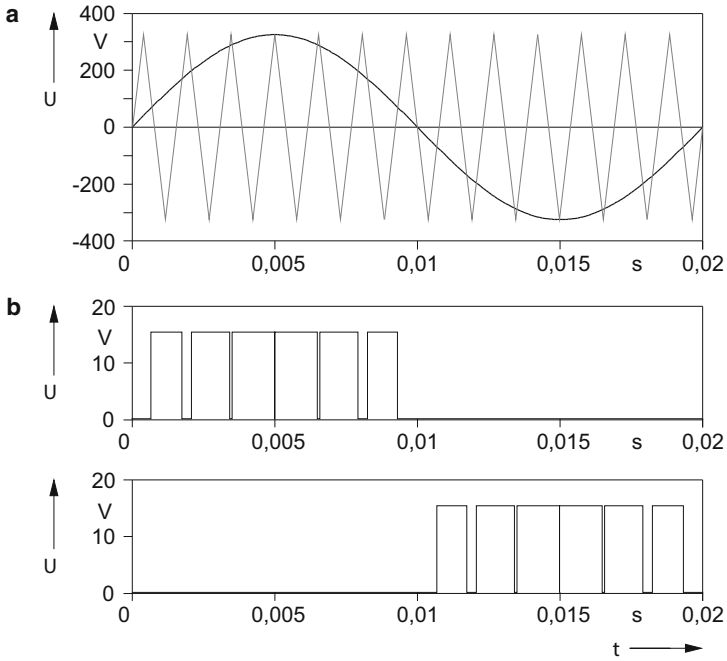


Abb. 87.32 Prinzip des Sinus-Dreieck-Vergleichs. **a** Funktionsverläufe, **b** Schaltsignale für die obere und untere Halbschwingung

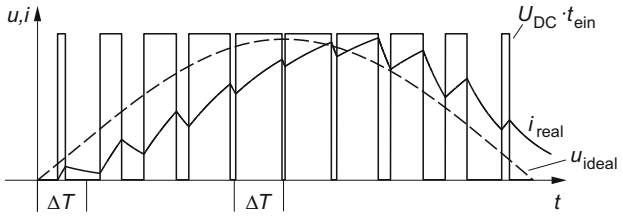


Abb. 87.33 Gepulste Spannung und Strom bei der PWM mit Sinus-Dreieck-Vergleich bei einer Pulsfrequenz von 1 kHz

die Pulsfrequenz der Spannung bestimmt. Innerhalb der Pulsperiodendauern T_P verlaufen die an- und absteigenden Stromverläufe in Form von Exponentialfunktionen. Der Zusammenhang zwischen der gepulsten Spannung und der Sinusspannung wird deutlich, wenn die gepulste Spannung gefiltert wird. Dazu könnte z. B. ein Widerstand mit einem Parallelkondensator genutzt werden. Da der Filter eine Zeitkonstante aufweist, wäre die gefilterte Spannung zeitlich zur Sinusspannung verschoben.

Netzwechselrichter müssen zur Synchronisation vor der Netzaufschaltung immer die vier Zuschaltbedingungen gleiche Spannungsamplitude, gleiche Frequenz, gleiche Phasenfolge und gleiche Phasenlage erfüllen, damit stromlos zugeschaltet wird. Wird asynchron zugeschaltet, kann es zu hohen Ausgleichsströmen und damit verbundenen Schäden kommen.

87.2.5.3 Photovoltaische Kraftwerke

Außerhalb der Erdatmosphäre wirkt die sogenannte extraterrestrische Sonnenstrahlung E_0 mit $1367\,\mathrm{W/m^2}$. Die Einstrahlung E auf der Erdoberfläche wird durch die Atmosphäre gedämpft und liegt bei rund 1000 Watt pro Quadratmeter. Solare Strahlung lässt sich thermisch zur Aufheizung und in Kopplung mit anderen Prozessen zur Stromerzeugung bzw. mit dem photovoltaischen Effekt, das heißt mit der Umwandlung von solarer Strahlung auf dotierten Halbleitern in Elektronen, auch direkt zur Stromerzeugung nutzen.

Obwohl der photoelektrische Effekt bereits 1839 entdeckt wurde und ab den 1950er-Jahren an photovoltaischen Zellen geforscht wurde, begann wegen der im Vergleich hohen Kosten und geringen Wirkungsgrade die industrielle Nutzung der Photovoltaik in Deutschland erst nach der Einführung des Stromeinspeisungsgesetzes im Jahr 1990. Zur direkten Wandlung von solarer Strahlungsenergie in elektrische Energie wird der photoelektrische Effekt genutzt: dotierte, das heißt gezielt mit anderen Materialien verunreinigte Halbleiter, setzen bei der Bestrahlung mit Licht Ladungsträger frei. Meist wird dafür dotiertes Silizium genutzt. Solarzellen können monokristallin, polykristallin oder in Dünnschichttechnologie aufgebaut sein. Typische Wirkungsgrade sind 14 % bis 18 % für monokristalline, 13 % bis 16 % für polykristalline und 10 % bis 12 % für Dünnschichtzellen. Daraus ergibt sich ein Platzbedarf in m pro kW_p, wobei der Indizes p für Peak-(Spitzen)-Leistung steht, die wie auch die Wirkungsgrade unter Standard-Testbedingungen bei einer Einstrahlung E von $1000\,\mathrm{W/m^2}$, einer Temperatur T von 25 °C und einem Sonnenspektrum AM (Air Mass – Einheit für die zu durchdringende Schichtdicke der Atmosphäre) von 1,5 ermittelt wird. Der prinzipielle Aufbau einer Solarzelle ist in Abb. 87.34 gezeigt. Sie besteht aus zwei unterschiedlich dotierten Silizium-Schichten. Silizium als Element der IV. Hauptgruppe wird entweder mit Elementen der III. Hauptgruppe, z. B. Bor, zu einem p-Halbleiter oder mit Elementen der V. Hauptgruppe, z. B. Phosphor, zu einem n-Halbleiter dotiert. Werden n- und p-dotierte

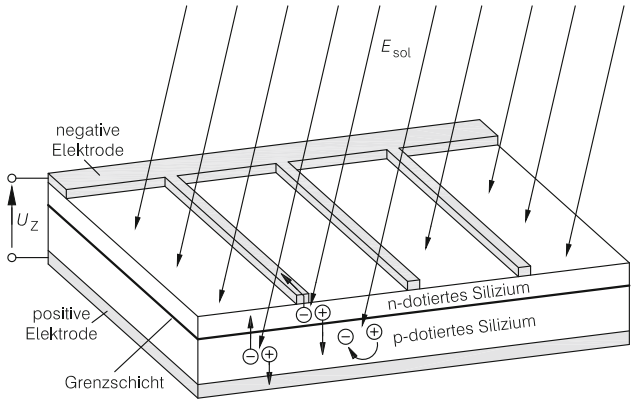


Abb. 87.34 Aufbau einer Solarzelle

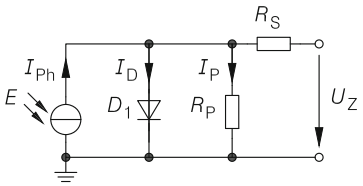


Abb. 87.35 Eindiodenmodell der Solarzelle

Schichten kombiniert, entsteht eine Grenzschicht, die bei Bestrahlung Ladungsträger freisetzt, die an Elektroden abgeleitet werden können. Solarzellen werden heute in Standardabmessungen von 5 Zoll mit 12,7 cm x 12,7 cm oder 6 Zoll mit 15,2 cm x 15,2 cm gefertigt. Das früher übliche Format von 4 Zoll mit 10 cm x 10 cm wird inzwischen nicht mehr gefertigt.

Als Ersatzschaltung zur Berechnung einer Solarzelle kann das Eindiodenmodell verwendet werden, siehe Abb. 87.35.

Für den Photostrom I_{Ph} gilt, das er sich bei konstanter Temperatur T mit dem Faktor c proportional zur Einstrahlung E verhält:

$$I_{Ph} \sim E = c \cdot E \tag{87.22}$$

Der Photostrom wird durch eine gesteuerte Stromquelle nachgebildet. Die Spannung der Solarzelle U_Z erhält man aus der Differenz der Diodenspannung U_D und dem Spannungsabfall am Serienwiderstand R_S:

$$U_Z = U_D - I_Z \cdot R_S \tag{87.23}$$

Der Zellenstrom I_Z ist der um den Diodenstrom I_D und den Parallelstrom I_P verminderte Photostrom I_{Ph}:

$$I_Z = I_{Ph} - I_D - I_P \tag{87.24}$$

Für einen großen Parallelwiderstand R_P wird der Strom I_P sehr klein: $I_p \to 0$ für $R_p \to \infty$. Der Parallelwiderstand entspricht dem fertigungsbedingten Leckwiderstand an den Rändern des p-n-Übergangs, der in der Größenordnung von 1 kΩ liegt. Bei Dünnschichtzellen bzw. bei in Modulen vergossenen Zellen ist dieser Leckwiderstand kleiner, so dass die Leckströme größer werden. Für den Zellenstrom I_Z ergibt sich mit $R_P \neq \infty$ und

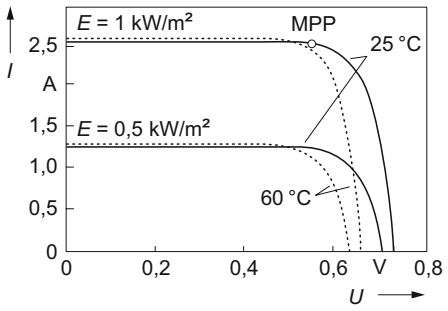


Abb. 87.36 Kennlinie einer Solarzelle

damit $I_P \neq 0$:

$$I_Z = I_{Ph} - I_S \cdot \left\{ \exp\left(\frac{U_D}{m \cdot U_T}\right) - 1 \right\} - \frac{U_D}{R_P} \tag{87.25}$$

I_S ist dabei der Sättigungsstrom der Diode. Der Serienwiderstand R_S bildet die Kontakt- und Übergangswiderstände nach, die so klein wie möglich sein sollen und in der Praxis mit $R_S = 20\,\text{m}\Omega$ angesetzt werden. Bei der Berechnung des Zellstromes könnte vereinfachend der Spannungsabfall über R_S vernachlässigt werden ($R_S \to 0$), hier wurde er jedoch berücksichtigt. Für die Diodenspannung U_D gilt: $U_D = U_Z + I_Z \cdot R_S$, damit errechnet sich der Zellenstrom I_Z zu:

$$I_Z = I_{Ph}(E) - I_S \cdot \left\{ \exp\left(\frac{U_Z + I_Z \cdot R_S}{m \cdot U_T}\right) - 1 \right\} - \frac{U_Z + I_Z \cdot R_S}{R_P} \tag{87.26}$$

Die Temperaturspannung U_T der Diode wird mit 25 mV angesetzt. Der Diodenfaktor m liegt zwischen 1 und 2. Dabei wird unter der Annahme einer idealen Sperrschicht der Diode $m = 1$, bei Annahme von vollständiger Rekombination $m = 2$ gesetzt. Mit dem Eindiodenmodell lässt sich die Charakteristik der Solarzelle gut nachbilden. In Abb. 87.36 ist die Strom-Spannungs-Kennlinie einer Solarzelle gezeigt. Charakteristische Werte sind dabei der Kurzschlussstrom I_K im Bereich von 2,5 bis 3 A und die Leerlaufspannung U_L im Bereich von 0,5 bis 0,7 V. Neben der Abhängigkeit von der Einstrahlung tritt noch eine Temperaturabhängigkeit auf, d. h. die Leerlaufspannung nimmt mit steigender Temperatur ab.

Die einzelnen Zellen werden in Modulen zusammengeschaltet, da jede einzelne nur eine geringe Spannung von 0,5 V bis 0,7 V liefert. Eine Photovoltaikanlage besteht aus dem aus Modulen aufgebauten Solargenerator, dem nachgeschalteten Wechselrichter sowie einer bidirektionalen Sicherheitsschnittstelle zum Netz (BISI), siehe Abb. 87.37. Die Module liefern abhängig

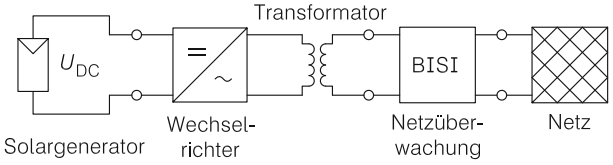


Abb. 87.37 Prinzip einer netzgekoppelten Photovoltaikanlage

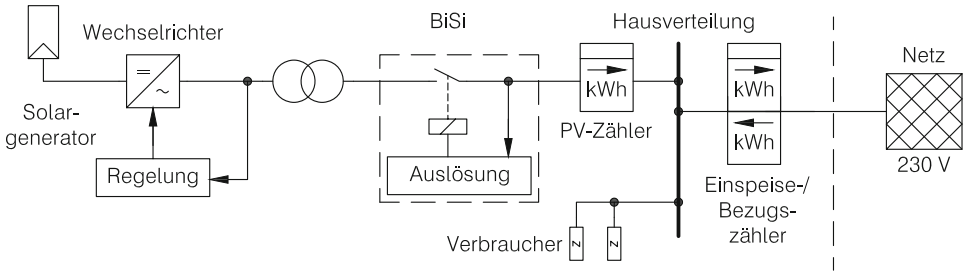


Abb. 87.38 Photovoltaikanlage in einer Hausinstallation

von der Einstrahlungsleistung eine elektrische Leistung. Da es sich um Gleichspannung handelt, muss noch eine Anpassung an den Wechselstrom im Netz erfolgen. Diese Aufgabe übernimmt der Wechselrichter, der dafür das im Abschn. 87.2.5.2.5 beschriebene Prinzip der Pulsweitenmodulation nutzt. Mit der bidirektionalen Sicherheitsschnittstelle wird gewährleistet, dass die Anlage bei einem Netzausfall abschaltet und nicht bei einem zufällig vorhandenen Leistungsgleichgewicht von erzeugter und lokal verbrauchter Wirk- und Blindleistung unkontrolliert in den Inselbetrieb übergeht. Sonst würde bei abgeschaltetem Netz weiter Spannung in der Hausinstallation anliegen. Bei der Wiederzuschaltung des Netzes könnten zudem hohe Ausgleichsströme auftreten, wenn die Phasenlage der Spannung im Inselnetz zur Netzspannung verschoben ist. Durch die Verschaltung vieler Module sind nahezu unbegrenzte Anlagenleistungen realisierbar; hierbei tritt jedoch ein hoher Flächenverbrauch auf. Abhängig vom Wirkungsgrad der Zellen kann hierbei von einem Flächenverbrauch von 10 m^2 pro kW_p installierter Leistung ausgegangen werden. Photovoltaische Anlagen in Mitteleuropa weisen abhängig von der Einstrahlung und dem Material am jeweiligen Standort eine energetische Amortisationszeit von zwei bis drei (Dünnschicht) bzw. drei bis fünf Jahren (polykristallin) auf. Das Entwicklungspotenzial dieser Kraftwerkstechnologie ist noch lange nicht ausgeschöpft, insbesondere der Erhöhung der Wirkungsgrade kommt eine große Bedeutung zu.

In Abb. 87.38 ist die Einbindung einer Photovoltaikanlage in die Hausinstallation gezeigt. Die eingespeiste und bezogene Leistung wird in einem Kombigerät getrennt erfasst und verrechnet. Zukünftig wird der seit einiger Zeit ansteigende Eigenverbrauch von elektrischer Energie aus Photovoltaikanlagen weiter zunehmen, da die Kosten der Eigenerzeugung bereits unterhalb der Strombezugskosten liegen.

87.2.5.4 Solarthermische Kraftwerke

Industriell eingesetzte konzentrierende solarthermische Anlagen nutzen nur den direkten Strahlungsanteil, bei der die Sonnenstrahlung in einem Brennpunkt konzentriert wird. Diese Technologie spielt in Deutschland keine Rolle, da sie erst ab einer direkten Sonneneinstrahlung von 1700 $kW\,h/m^2$ und Jahr sinnvoll einsetzbar ist. Bereits 1912 wurden Parabolrinnen zur Dampferzeugung für eine 45-kW-Dampfmotorpumpe in Ägypten eingesetzt. Die industrielle Nutzung zur Stromerzeugung begann 1984 in den USA, seither wurden weltweit rund 15 Projekte realisiert. Das solare Parabolrinnenkraftwerk besteht aus Reihen von Parabolrinnen, die der Sonne nachgeführt werden, siehe Abb. 87.39 links. Ein nachgeschaltetes Dampfkraftwerk erzeugt über den angetriebenen Generator Strom. Es sind Kraftwerksleistungen im Bereich von 10 MW bis 300 MW realisierbar. Bei solaren Turmkraftwerken wird die Sonnenstrahlung über zweiachsig nachgeführte Spiegel auf das Absorberfenster einer Turmkonstruktion fokussiert, siehe Abb. 87.39 Mitte. Hinter dem Absorber wird Luft erhitzt, die eine nachgeschaltete Gasturbine mit Generator antreibt und Strom erzeugt. Realisiert wurden bisher zwei Versuchsanlagen in Kalifornien und Südspanien sowie ein kommerziell nutzbares Kraftwerk in Südspanien, in den USA sind mehrere Anlagen im Bau. Mit dieser Bauart sind Kraftwerksleistungen zwischen 10 und 1000 MW erreichbar. Eine andere Ausführung solarthermischer Kraftwerke sind die Dish-Stirling-Systeme. Sie bestehen wie in Abb. 87.39 rechts gezeigt aus einem zweiachsig nachgeführtem Parabolspiegel (Dish), in dessen Brennpunkt sich als Wärmekraftmaschine der im Abschn. 87.2.4.4 beschriebene Stirlingmotor befindet. Der Stirlingmotor setzt die Wärme in Bewegungsenergie um und treibt einen elektrischen Generator an. Es sind Leistungen bis zu 250 kW möglich. Diese Kraftwerksart ist vorrangig für netzferne Versorgungen konzipiert.

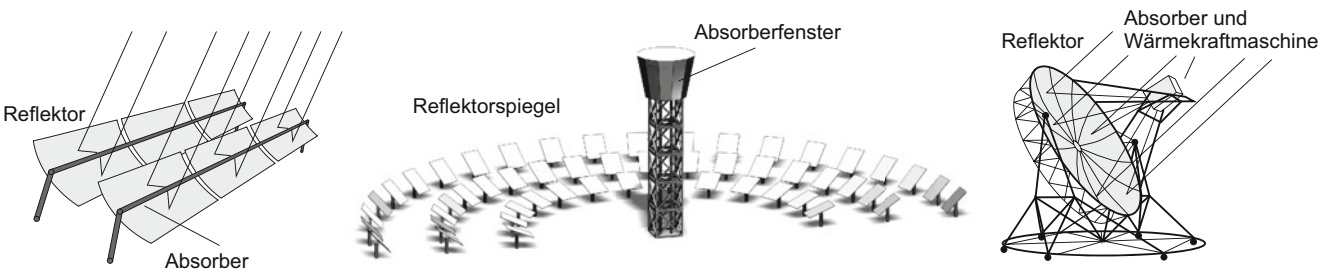


Abb. 87.39 Solarthermische Kraftwerksprinzipien, *links*: Parabolrinnenkraftwerk, *Mitte*: Turmkraftwerk, *rechts*: Parabolspiegel-Kraftwerk mit Stirling-Maschine

87.2.5.5 Biomassekraftwerke

Biomasse kann auf drei Arten zur Strom-, Wärme- und Treibstofferzeugung genutzt werden. Die erste Art, die thermochemische Energiewandlung, erfolgt durch direkte Verbrennung mit Hilfe eines Sauerstoffüberschusses, durch Vergasung unter Sauerstoffmangel oder Verflüssigung bzw. Pyrolyse unter Sauerstoffabschluss. Beispielsweise ist der Betrieb von Wasser-Dampf-Kraftwerken auch mit Biomassefeuerung möglich. Gasturbinen und BHKWs können auch mit Biogas betrieben werden. Bei der zweiten Art, den physikalisch-chemischen Verfahren, wird aus ölhaltiger Biomasse ein Flüssigenergieträger gewonnen. So kann ölhaltige Saat durch Pressung zum Beispiel in einer Ölmühle und Extraktion mit Lösemitteln zu einem Pflanzenöl umgewandelt werden. Dieses ist ein Ausgangsstoff für Biodiesel und kann direkt in umgerüsteten Dieselmotoren verwendet werden. Biodiesel ist ein Ester, der aus Pflanzenölen hergestellt wird. Ester entsteht aus der Reaktion einer Säure mit Alkohol unter Abspaltung von Wasser. Zur dritten Art der Energieumwandlung zählen die biochemischen Verfahren. Sie werden sowohl zur Alkoholerzeugung aus zucker-, stärke- und zellulosehaltiger Biomasse als auch zur Biogasbereitstellung eingesetzt. Um höhere Gesamtwirkungsgrade zu erzielen, werden Biomassekraftwerke meist als Blockheizkraftwerk mit Kraft-Wärme-Kopplung betrieben. In Deutschland hat die Nutzung von Biomasse noch ein hohes Potenzial. Vor allem die Biokraftstoffe der "zweiten Generation" werden dabei eine große Rolle spielen, da hierbei die Nutzungskonkurrenz zu anderen Verwertungsarten schwächer ausgeprägt ist. Während bei der Erzeugung von Biokraftstoffen der ersten Generation nur die Frucht selbst für die Kraftstoffproduktion genutzt wird, findet bei Kraftstoffen der zweiten Generation fast die gesamte Pflanze Verwendung. Darüber hinaus spricht man bei der Nutzung von Algen zur Kraftstofferzeugung von der „dritten Generation" von Biokraftstoffen. Beispielsweise planen große Flugzeughersteller bereits Testflüge mit Algenkraftstoff. Ein großer Vorteil der Stromerzeugung aus Biomasse ist, dass man damit teilweise Schwankungen bei der Windenergieeinspeisung ausgleichen kann.

87.2.5.6 Geothermische Kraftwerke

Eine weitere mögliche Form der Energiegewinnung ist die Nutzung der natürlichen Erdwärme (Geothermie). Vorrangig speist die Erdwärme sich aus radioaktiven Zerfallsprozessen im Erdinneren. Der natürliche tiefenabhängige Temperaturgradient $\Delta T/\Delta L$ liegt bei 3 K pro 100 m:

$$\frac{\Delta T}{\Delta L} = \frac{3\,\mathrm{K}}{100\,\mathrm{m}} \tag{87.27}$$

Die einfachste Form der Nutzung ist bei natürlichen Gegebenheiten mit höheren Temperaturgradienten, sog. Temperaturanomalien, in Form von Heißwasser- und Heißdampfreservoirs möglich. Diese können direkt für Heizzwecke eingesetzt werden. Das vorhandene Temperaturniveau wird über einen Wärmetauscher zur Dampferzeugung genutzt, um Dampfturbinen anzutreiben. Wegen der hohen Erschließungskosten muss für den wirtschaftlichen Betrieb die Kraft-Wärme-Kopplung

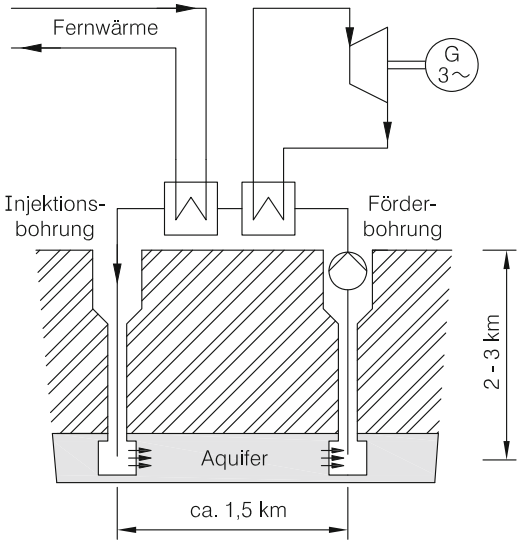


Abb. 87.40 Prinzip einer Geothermieanlage

eingesetzt werden. Deshalb werden nutzbare Standorte neben ihrem Temperaturpotenzial vorrangig so ausgewählt, dass über Heiznetze eine ausreichende Anzahl von Wärmeverbrauchern angeschlossen werden kann. Es wird zwischen oberflächennaher Nutzung und der in Abb. 87.40 gezeigten Tiefengeothermie unterschieden. Die oberflächennahe Nutzung erfolgt mit Erdwärmepumpen; eine direkte Wärmenutzung flüssiger Wärme- bzw. Kältespeicher (Aquiferen) erfordert starke Temperaturanomalien.

Im italienischen Larderello wurde bereits im Jahr 1904 ein Geothermiekraftwerk errichtet. Inzwischen sind dort 21 geothermische Kraftwerke mit einer elektrischen Gesamtleistung von über 560 MW installiert. Bei einer Bohrtiefe von vier Kilometern werden 350 °C Dampftemperatur erreicht. Solche hydrothermalen Lagerstätten können bis zu großen Tiefen von einigen Kilometern genutzt werden. Wenn keine natürlichen Heißwasserquellen vorhanden sind, kann Wasser zum Aufheizen in eine Tiefenbohrung verpresst und wieder an die Erdoberfläche gepumpt werden. Das Wasser erhitzt sich beim Durchströmen des heißen Tiefengesteins (hot dry rock). Mit geothermischen Kraftwerken ist eine kontinuierliche Stromerzeugung möglich, weshalb diese Technologie wie auch die Biomassenutzung gut geeignet ist, den wetterabhängig schwankenden Windstrom auszugleichen. Weltweit sind mehr als 470 geothermische Kraftwerke mit einer Gesamtleistung von über 9000 MW installiert, der Hauptteil davon in den USA. In Deutschland ging im Jahr 2003 das erste Kraftwerk mit einer elektrischen Leistung von 230 kW in Neustadt-Glewe in Mecklenburg-Vorpommern in Betrieb; es wurde bereits seit 1994 als Heizkraftwerk betrieben. Dieses Kraftwerk wird mit 98 °C heißem Wasser aus einer 2250 Meter tiefen Bohrung gespeist. Im Jahr 2007 hat in Landau in der Pfalz ein Geothermiekraftwerk mit einer Bohrtiefe von 3400 Metern und einer elektrischen Leistung von 3,8 MW den Betrieb aufgenommen. Neue Anlagen wurden auch 2009 im bayerischen Unterhaching mit 3580 Metern und 3,4 MW sowie in Bruchsal mit 2500 Metern und 550 kW fertiggestellt.

87.3 Elektrische Energieübertragung und Energieverteilung

87.3.1 Spannungssysteme

87.3.1.1 Gleichstromnetze

Im Gleichstromnetz sind Strom und Spannung immer phasengleich, es tritt nur Wirkleistung auf, siehe Abb. 87.41:

$$P = U \cdot I \tag{87.28}$$

Im Vergleich zur Nutzung von Wechsel- oder Drehstrom tritt keine Blindleistung auf, weshalb Gleichstromnetze zur Übertragung von elektrischer Energie über weite Entfernungen genutzt werden. Die Übertragungsverluste errechnen sich hierbei nur aus den ohmschen Verlusten:

$$P_V = I^2 \cdot R \tag{87.29}$$

Gleichstromnetze können als Zwei- oder Einleiternetze betrieben werden. Zweiphasige Netze werden zur Hochspannungs-Gleichstromübertragung (HGÜ) mit Spannungen von bis zu ±800 kV eingesetzt, siehe Abb. 87.42. Auch im schienengebundenen Verkehr werden Gleichspannungen genutzt, z. B. nutzen U-Bahnen einphasige Netze mit einer Spannung von 750 V.

87.3.1.2 Wechselstromnetze

Bei Wechselstrom kann eine Phasenverschiebung zwischen Strom und Spannung um den Phasenwinkel φ auftreten, siehe Abb. 87.43.

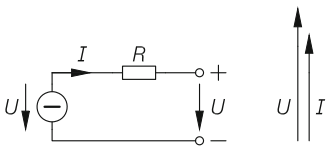


Abb. 87.41 Gleichstromnetz mit Zeigerbild

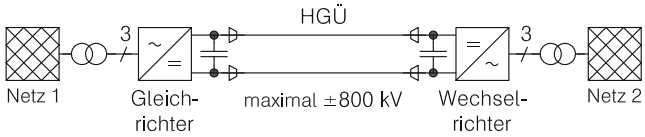


Abb. 87.42 Hochspannungs-Gleichstromübertragung

Die Wirkleistung P errechnet sich aus den Effektivwerten von Spannung und Strom:

$$P = U \cdot I \cdot \cos\varphi \tag{87.30}$$

Für die Blindleistung Q erhält man analog:

$$Q = U \cdot I \cdot \sin\varphi \tag{87.31}$$

Die Scheinleistung S berechnet sich wie folgt:

$$S = \sqrt{3} \cdot U \cdot I \tag{87.32}$$

Da die Leistung im einphasigen 50-Hz-Wechselstromnetz mit 100 Hz pulsiert, kann damit kein zeitlich konstantes Drehmoment an Motoren eingestellt werden. Darüber hinaus ist ein Einphasennetz in seiner Übertragungsleistung durch den Leitungsdurchmesser begrenzt. Bahnnetze werden einphasig versorgt, dann ist nur ein Stromabnehmer erforderlich. Hierbei kommt aus historischen Gründen überwiegend eine Spannung mit einer Frequenz von $16\frac{2}{3}$ Hz zum Einsatz, es existieren jedoch auch einphasige 50-Hz-Bahnnetze.

87.3.1.3 Drehstromnetze

Drehstromnetze können, wie z. B. meist auf der Mittelspannungsebene eingesetzt, entweder nur mit den Leitern L1, L2 und L3 als Dreileiternetz oder mit zusätzlichem Neutralleiter N als Vierleiternetz betrieben werden, siehe Abb. 87.44. Verbraucher können in Stern- oder Dreieckschaltung angeschlossen werden. Zu deren Versorgung werden die Außenleiterspannungen zwischen den Leitern L1, L2 und L3 genutzt, die Ströme werden dann als Leiterströme bezeichnet. Verbraucheranschlüsse werden üblicherweise mit U, V und W bezeichnet. Stranggrößen treten bei Verbrauchern bei der Sternschaltung zwischen Außenleiter und Sternpunkt, bei der Dreieckschaltung zwischen den Außenleitern auf. Die Außenleiterspannungen sind um den Verkettungsfaktor größer als die Strangspannungen:

$$\frac{U_{LL}}{U_{LN}} = \sqrt{3} \tag{87.33}$$

Bei symmetrischer Belastung aller drei Phasenleiter ist die zeitliche Summe der Außenleiterströme null und der Neutralleiter daher stromlos. Bei Dreieckschaltung gilt für die Verhältnisse von Leiter- (ohne Indizes) und Stranggrößen (Index Str):

$$U = U_{Str} \tag{87.34}$$

$$I = \sqrt{3} \cdot I_{Str} \tag{87.35}$$

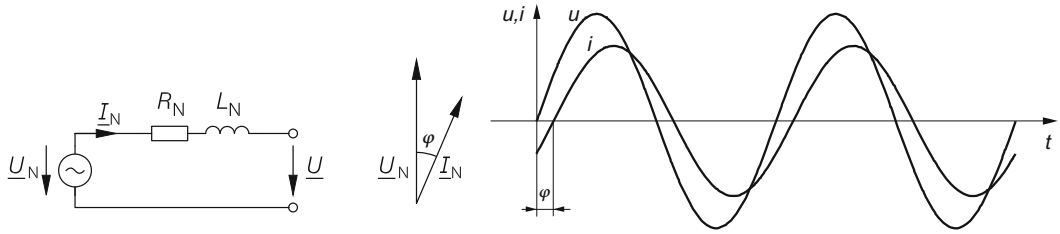


Abb. 87.43 Verlauf von Spannung und Strom im Wechselstromnetz und Zeigerbild

Teil XI

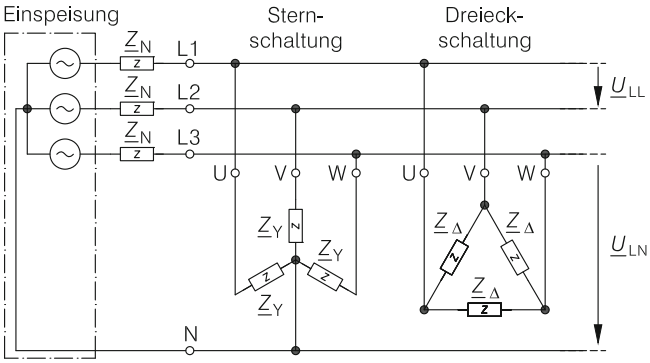


Abb. 87.44 Drehstromnetz mit Verbrauchern in Stern- und Dreieckschaltung

Bei Sternschaltung gilt:

$$U = \sqrt{3} \cdot U_{\mathrm{Str}} \tag{87.36}$$

$$I = I_{\mathrm{Str}} \tag{87.37}$$

Dreiphasige Systeme weisen gegenüber einphasigen Systemen einige Vorteile auf. Es werden bei gleicher Leistungsübertragung im Vergleich zu drei einphasigen Systemen drei Übertragungsleitungen eingespart. In dreiphasigen Systemen ist der Mittelwert der Leistung zeitlich konstant, während er in einphasigen Systemen zeitlich schwankt. Mit dreiphasigen Systemen kann ein magnetisches Drehfeld bereitgestellt werden, wie es für alle rotierenden Drehstrommaschinen erforderlich ist.

Die dreiphasige Wirkleistung P errechnet sich mit der Außenleiterspannung (Leiter-Leiter-Spannung) U:

$$P = \sqrt{3} \cdot U \cdot I \cdot \cos\varphi \tag{87.38}$$

Für die dreiphasige Blindleistung Q erhält man analog:

$$Q = \sqrt{3} \cdot U \cdot I \cdot \sin\varphi \tag{87.39}$$

Die dreiphasige Scheinleistung S berechnet sich wie folgt:

$$S = \sqrt{3} \cdot U \cdot I \tag{87.40}$$

Bei der Anwendung der klassischen Leistungstheorien mit Bezug auf sinusförmige Größen wird stets von einem symmetrischen Drehstromsystem ausgegangen, bei dem die Summe aller Größen zu jedem Zeitpunkt Null ist. Im Bereich der Energieversorgungsnetze können jedoch Asymmetrien im Netzaufbau oder bei der Belastung zu abweichenden Verhältnissen führen. In diesem Fall kann die Methode der symmetrischen Komponenten zur Zerlegung der zeitlichen Größen in ein Mit-, Gegen- und Nullsystem genutzt werden, um beispielsweise Fehlerberechnungen in Netzen durchzuführen. Für den praxisgerechten Einsatz von leistungselektronischen Betriebsmitteln kann es darüber hinaus erforderlich sein, Leistungen bei nicht sinusförmigen Vorgängen oder bei unsymmetrischen Verhältnissen zu berechnen. Für diesen Fall sind die umfassender anwendbaren Leistungsgrößen nach DIN VDE 40110 zu nutzen.

87.3.2 Aufbau und Betrieb von Energieversorgungsnetzen

87.3.2.1 Spannungsebenen

Spannungen bis 1 kV werden als Niederspannung, darüber bis 52 kV als Mittelspannung, bis 150 kV als Hochspannung und über 150 kV als Höchstspannung klassifiziert. Bei der Übertragung und Verteilung elektrischer Energie werden unterschiedliche Spannungsebenen genutzt, siehe Abb. 87.45 links. Zur Vermeidung hoher Transportverluste gilt bei der Wahl der Spannungshöhe, dass die Spannung in kV immer größer als die Entfernung in km sein muss. Mit höherer Spannung lassen sich größere Leistungen bei geringen Verlusten übertragen, deshalb erfolgt die Einbindung in das europäische Verbundnetz und der weiträumige Stromtransport im deutschen Übertragungsnetz mit Spannungen von 380 und 220 kV. An dieser Spannungsebene werden große Kraftwerke und Verbraucher mit Leistungen oberhalb von 300 MW angeschlossen. Mit der 110 kV Spannungsebene werden die Verteilnetze betrieben, z. B. erfolgt die Einspeisung von Großstädten meist über mehrere dieser Verteilnetzleitungen. Innerhalb der Städte wird die Mittelspannung im Bereich von 20 oder 10 kV eingesetzt. Auf der 400-V-Ebene (dreiphasig) bzw. 230-V-Ebene (einphasig) werden die Privatverbraucher versorgt. Deutschland verfügt über ein über lange Zeiträume gut ausgebautes Übertragungs- und Verteilungsnetz, das hervorragend an die Erfordernisse der Versorgung aus großtechnischen Kraftwerksstrukturen angepasst ist. Dazu gehören laut Bundesnetzagentur im Jahr 2013 rund 35.000 km Höchstspannungsleitungen, 96.000 km Hochspannungs-leitungen, 510.000 km Mittelspannungsleitungen und 1.156.000 km Niederspannungsleitungen. Darüber hinaus existieren zwei Hochspannungs-Gleichstrom-Übertragungsleitungen (HGÜ) nach Dänemark und Schweden. Das deutsche Übertragungsnutz ist in vier Regelzonen aufgeteilt, siehe Abb. 87.45 rechts.

87.3.2.2 Netzformen

Als Netzformen werden Strahlennetze, Ringnetze und Maschennetze eingesetzt, siehe Abb. 87.46. Das Strahlennetz ist einfach aufgebaut und daher kostengünstig. Nachteilig sind der Spannungsabfall zum Leitungsende hin und die geringe Versorgungssicherheit. Diese Netzform wird bei geringen Lastdichten und bei kleineren Anschlussleistungen eingesetzt. Ringnetze sind etwas kostenaufwändiger, bieten jedoch eine höhere Versorgungssicherheit. Im Normalbetrieb werden sie offen betrieben, die Lasten werden zur einfacheren Netzführung also immer einseitig über ein Strahlennetz gespeist. Im Fehlerfall kann der verbleibende fehlerfreie Netzabschnitt über die dann zu schließende Trennstelle mit versorgt werden. Dazu müssen die Leitungen so dimensioniert werden, dass sie auch den Laststrom des Parallelzweigs führen können. Das Maschennetz weist viele Querverzweigungen auf. Es bietet eine hohe Versorgungssicherheit und ermöglicht wegen seines geringen Innenwiderstands die beste Spannungshaltung. Aus Kostengründen wird es nur bei hohen Lastdichten in Industrienetzen eingesetzt. Nachteilig sind die hohen Kurzschlussströme, die daraus resultierende Geräteauslegung, ein aufwändiger Netzschutz sowie eine schwierige

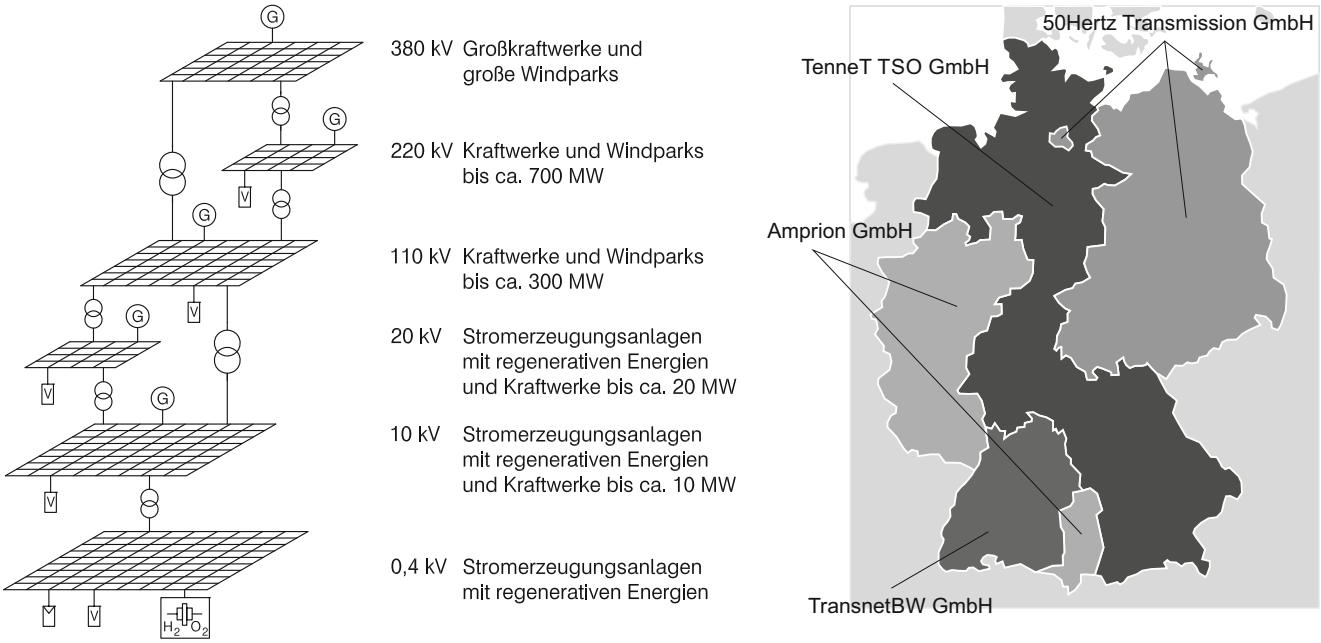


Abb. 87.45 Aufbau der Netzebenen (*links*) und Regelzonen (*rechts*)

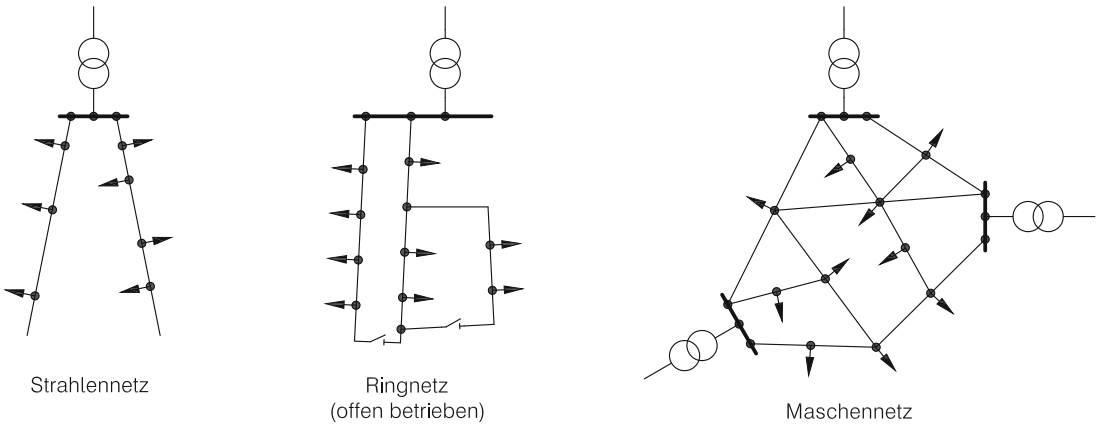


Abb. 87.46 Netzformen: Strahlennetz, Ringnetz mit Verzweigung, Maschennetz mit Mehrfacheinspeisung

Wiederinbetriebnahme nach einem Netzausfall. Daher werden Maschennetze nur im Niederspannungsbereich genutzt, während Strahlen- und Ringnetze sowohl für die Nieder- als auch für die Mittelspannung eingesetzt werden. Für die Mittelspannung werden bei weiter ausgedehnten Netzen wegen der besseren Übersichtlichkeit und der einfacheren Betriebsführung verzweigte Ringnetze anstatt Maschennetze gebaut.

87.3.3 Frequenzregelung

Im Energieversorgungsnetz werden sowohl die Frequenz als auch die Amplitude der Spannung geregelt, dies wird unter dem Begriff der Systemdienstleistungen zusammengefasst. Die Frequenz ist eine netzglobale Größe und weist in jedem Netzzweig den gleichen Wert von 50 Hz auf. Grundlage der Frequenzregelung ist das Gleichgewicht zwischen erzeugter und verbrauchter elektrischer Wirkleistung zu jedem Zeitpunkt. Übersteigt der Verbrauch die erzeugte Wirkleistung, werden die Kraftwerksgeneratoren gebremst und die Frequenz sinkt ab. Bei einem Überschuss an erzeugter Wirkleistung steigt die Frequenz an. Die grundlegende Bedingung für den störungsfreien und zuverlässigen Netzbetrieb ist das Gleichgewicht von erzeugter und verbrauchter Wirk- und Blindleistung zu jedem Zeitpunkt und damit der Momentanleistungen $p(t)$:

$$(P + jQ)_E(t) = (P + jQ)_V(t) \tag{87.41}$$

$$p(t)_E = p(t)_V \tag{87.42}$$

Der Zusammenhang zwischen Frequenz und Wirkleistung lässt sich mit (87.43) verdeutlichen. Aus der Leistung P kann das

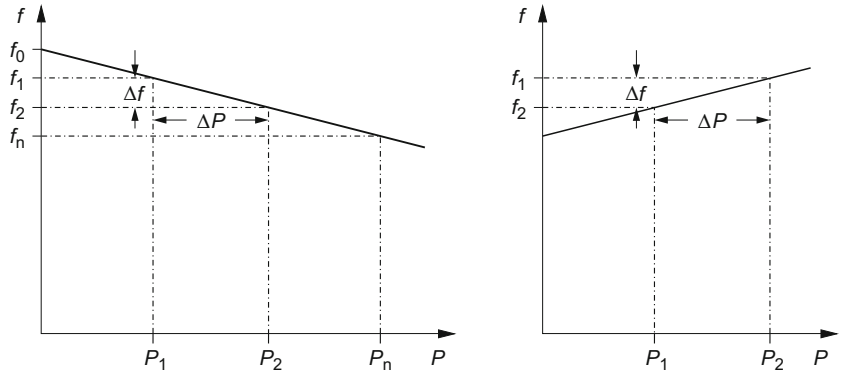


Abb. 87.47 Frequenzkennlinien, *links*: Erzeugerkennlinie, *rechts*: Kennlinie frequenzabhängiger Verbraucher

an der Rotorwelle anliegende mechanische Moment M mit der Kreisfrequenz ω wie folgt ermittelt werden:

$$M = \frac{P}{\omega} \quad \text{mit } \omega = \frac{2 \cdot \pi \cdot n}{60} \tag{87.43}$$

Mit $\omega = 2\pi f$ nimmt bei konstanter Generatorleistung P bei steigendem Lastmoment M die Frequenz f ab:

$$\omega = 2 \cdot \pi \cdot f = \frac{P}{M} \tag{87.44}$$

Auf Grund der Bedingung nach (87.42) müssen bei vernachlässigten Verlusten mechanisches Turbinenmoment M_{mech} und elektrisches Generatormoment M_{el} gleich groß sein:

$$M_{\text{mech}} = M_{\text{el}} \tag{87.45}$$

Wegen des Zusammenhangs nach (87.44) steigt bei $M_{\text{mech}} =$ const die Generatordrehzahl bei abnehmender elektrischer Leistung, sie sinkt bei erhöhter Leistungsabnahme. Steigt die Belastung des Generators durch zusätzliche Verbraucher im Netz um ΔP_{el} an, tritt ein um ΔM höheres elektrisches (M_{el}) und mechanisches (M_{mech}) Generatormoment auf, J ist das Massenträgheitsmoment:

$$M_{\text{mech}} = M_{\text{el}} + \Delta M_{\text{el}} + J \cdot \frac{d\omega}{dt} \tag{87.46}$$

Bei einer Leistungszunahme im Netz mit $\Delta M_{\text{el}} > 0$ gilt $d\omega/dt < 0$, die Generatordrehzahl und -frequenz sinken. Nimmt die Netzleistung ab, gilt $\Delta M_{\text{el}} < 0$ und $d\omega/dt > 0$, die Generatordrehzahl und -frequenz steigen. Dieser Zusammenhang tritt bei allen Kraftwerksgeneratoren auf. Die resultierende Erzeugerkennlinie zeigt einen Frequenzabfall bei Belastung zwischen der oberen Leerlauf-Frequenzgrenze f_o und der unteren Frequenzgrenze f_u bei Nennleistung, siehe Abb. 87.47 links Die Leistung frequenzabhängiger Verbraucher steigt mit der Frequenz an, siehe Abb. 87.47 rechts. Die Steigung der Kennlinien ist die Statik st, der normierte Kehrwert die Netzleistungszahl λ. Für die in Abb. 87.47 links gezeigte Kennlinie wird die Statik mit der oberen und unteren Frequenzgrenze ermittelt:

$$\text{st} = \frac{f_o - f_u}{f_u} \tag{87.47}$$

Die Netzleistungszahl λ hat die Einheit MW/Hz und errechnet sich nach:

$$\lambda = \frac{\Delta P}{\Delta f} \tag{87.48}$$

Eine Umrechnung der Leistungszahl λ in die dimensionslose Statik st erfordert die Normierung auf die Nennleistung und -frequenz:

$$\text{st} = \frac{1}{\lambda} \cdot \frac{P_n}{f_0} \tag{87.49}$$

Die Anpassung der Momente nach (87.46) erfolgt durch die Primärregelung: der Drehzahlregler der Turbine stellt die Leistung entsprechend einer Netzkennlinie (Statik-Kennlinie), deren Steigung die Netzleistungszahl ist, die im ENTSO-E-Netz (european network of transmission system operators for electricity) bei 40.000 MW/Hz liegt:

$$\lambda = \frac{\partial P}{\partial f} = \frac{40.000\,\text{MW}}{\text{Hz}} \tag{87.50}$$

Fällt im ENTSO-E-Netz eine Kraftwerkleistung von $P_{\text{KW}} =$ 4000 MW Leistung aus, ergibt sich eine Frequenzänderung von 0,1 Hz:

$$\Delta f = \frac{P_{\text{KW}}}{\lambda} = \frac{4000\,\text{MW}}{40.000\,\text{MW/Hz}} = 0{,}1\,\text{Hz} \tag{87.51}$$

Erst bei einem Ausfall von 100.000 MW würde sich eine Frequenzänderung von 2,5 Hz einstellen, bei der alle Kraftwerke vom Netz getrennt werden müssten:

$$\Delta f = \frac{P_{\text{KW}}}{\lambda} = \frac{100.000\,\text{MW}}{40.000\,\text{MW/Hz}} = 2{,}5\,\text{Hz}$$

Durch die Primärregelung, d. h. die lokale Drehzahlregelung an Kraftwerksgeneratoren sowie die Sekundärregelung, d. h. pro Regelzone zeitlich nachgeordnete Frequenzregelung, wird die Frequenz in engen Grenzen geregelt, siehe Abb. 87.48. Bereits geringe Frequenzabweichungen würden hohe Qualitätsverluste in Produktionsanlagen verursachen, da die Drehzahl

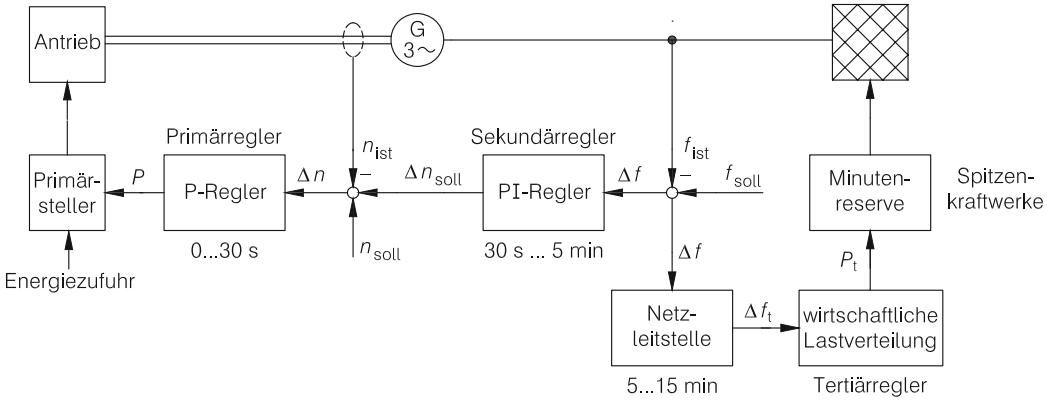


Abb. 87.48 Prinzip der Frequenzregelung

von Drehfeldmaschinen direkt von der Speisefrequenz abhängig ist. Daher werden die Kraftwerke bei einer Unterfrequenz von 47,5 Hz bzw. Überfrequenz von 51,5 Hz vom Energieversorgungsnetz getrennt. Die Frequenzregelung im Netz wird auf die einzelnen Generatoren verteilt. Ab Frequenzabweichungen von 10 mHz reagieren die Primärregler von Kraftwerken mit Leistungen ab 100 MW und regeln die Frequenzabweichung grob aus. Die Primärregelung wirkt als Proportionalregelung, sie arbeitet im Zeitbereich bis maximal 30 s und kann Abweichungen von der Sollfrequenz nicht vollständig ausregeln. Proportionalregler weisen folgendes Stellverhalten zwischen Ausgang y und Eingang x auf:

$$y(t) = K \cdot x(t) \tag{87.52}$$

Abhängig vom Verstärkungsfaktor K wird ohne Zeitverzug ein Ausgangssignal erzeugt, dass jedoch im Regelfall noch vom gewünschten Frequenzwert abweicht. Verbleibende Frequenzabweichungen werden mit einem Leistungsabgabefaktor gewichtet und mit der Sekundärregelung in Form von Spitzenlastkraftwerken ausgeglichen. Dafür wird ein Regler mit Integralcharakteristik genutzt, der folgendes Stellverhalten zwischen Ausgang y und Eingang x aufweist:

$$y(t) = \frac{1}{T_I} \int x(t)dt \tag{87.53}$$

Die Sekundärregelung wirkt im Zeitbereich ab 30 s bis 5 min, aufgrund des Integralverhaltens arbeitet dieser Regler langsamer, es verbleibt nach Abschluss des Regelvorgangs jedoch keine Frequenzabweichung mehr. Zu beachten ist, dass pro Netzregelzone jeweils nur ein Sekundärregler eingesetzt werden darf, da parallel geschaltete Integralregler zum Schwingen neigen, wodurch die Netzregelung gefährdet wäre. Zur Sekundärregelung werden bei Teillast betriebene Kraftwerke sowie Pumpspeicherkraftwerke eingesetzt. Dabei erfolgt eine kostenoptimale Kraftwerksauswahl. Zusätzlich wirkt der Selbstregelungseffekt. Da Verbrauch und Erzeugung im Netz immer gleich groß sein müssen, passt sich der Verbrauch der möglichen Erzeugungsleistung an, wenn die Generatorleistung nicht mehr erhöht werden kann. Dies trifft z. B. auf Lüfter und Pumpenantriebe zu, bei denen eine kubische Abhängigkeit zwischen Leistung und Frequenz besteht:

$$P \sim f^3 = K \cdot f^3 \tag{87.54}$$

Im Zeitraum von 5 bis 15 min wirkt die als Tertiärregelung bezeichnete Minutenreserve, bei der von den Netzleitstellen schnell hochfahrbare Kraftwerke, wie z. B. Gasturbinenkraftwerke oder Kraftwerke benachbarter Regelzonen abgerufen werden. Die Tertiärregelung ist keine Regelung im eigentlichen Sinn, es handelt sich hierbei um eine kostenoptimale Verteilung der Regelleistung auf den Kraftwerkspark. Kleinere Frequenzänderungen innerhalb eines Totbandes von ±10 mHz werden zur Schonung der Kraftwerksturbinen ausschließlich durch die Sekundärregelung ausgeregelt.

Die zuvor beschriebenen Abläufe der Frequenzregelung werden in der Praxis nach einem bei ENTSO-E vereinbarten Organisationsschema ausgeführt, siehe Abb. 87.49. Hierbei findet zur Sicherstellung einer koordinierten Netzregelung auch eine Zeitüberwachung der Regelvorgänge statt.

Zur Vermeidung einer Frequenzabsenkung durch Überlast wurde bei ENTSO-E im Frequenzbereich zwischen 49,2 und 48 Hz ein minimal verpflichtender Unterfrequenz-Lastabwurf über sechs Stufen vereinbart, siehe Abb. 87.50. Diese Vereinbarung wurde vom FNN (Forum Netztechnik/Netzbetrieb im VDE) übernommen.

87.3.4 Spannungsregelung

Im Gegensatz zur Frequenz ist die Netzspannung eine lokale Größe mit ortsabhängig unterschiedlichen Amplituden. Daher wird sie lokal über den Stromfluss geregelt. Je nach Phasenlage des Stroms zur Netzspannung entsteht an der Netzimpedanz ein Spannungsfall mit gleicher oder entgegengesetzter Orientierung des Spannungszeigers. Zu beachten ist, dass neben der nachfolgend zuerst beschriebenen statischen Spannungshaltung im Normalbetrieb auch eine dynamische Spannungsregelung bei Netzfehlern erforderlich sein kann, die im Abschn. 87.3.4.5 erläutert wird.

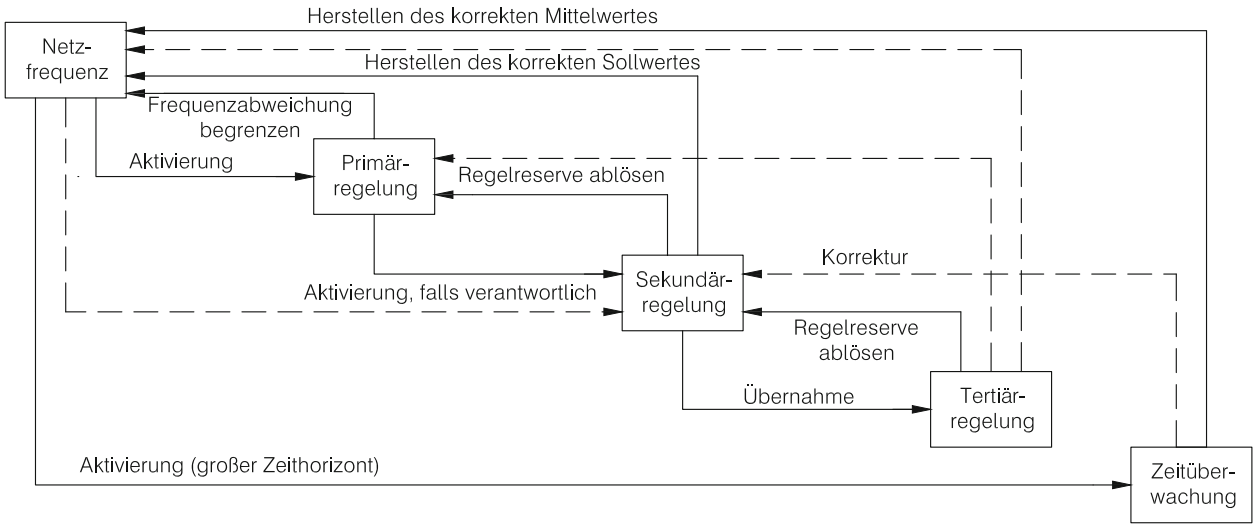


Abb. 87.49 Regelungsschema der Frequenzregelung ENTSO-E

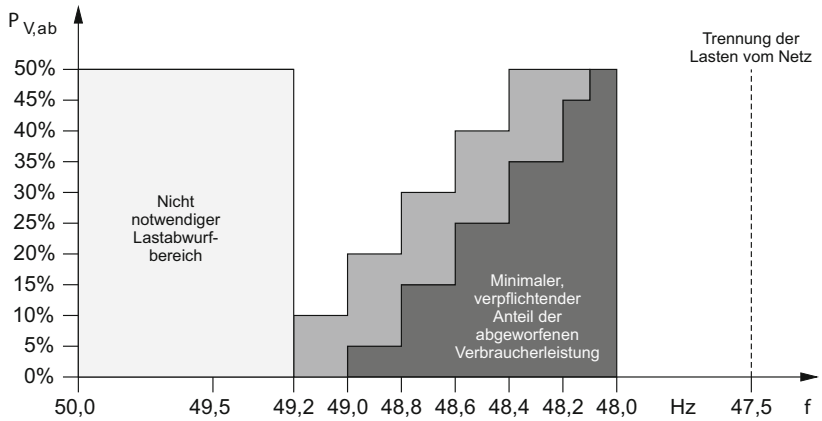


Abb. 87.50 Minimal verpflichtender Unterfrequenz-Lastabwurf in sechs Stufen nach FNN und ENTSO-E

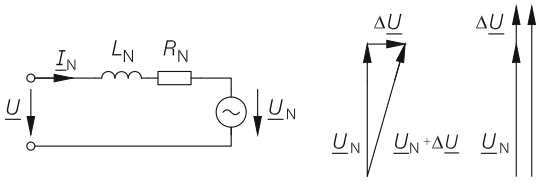


Abb. 87.51 Prinzip der Spannungsregelung, *links*: Ersatzschaltbild, *Mitte*: Zeigerdiagramm bei Wirkstromeinspeisung, *rechts*: Zeigerdiagramm bei Blindstromeinspeisung

87.3.4.1 Allgemeiner Fall

Grundlage der Betrachtung ist das Ersatzschaltbild der Netzeinspeisung in Abb. 87.51 links. Dieses enthält die Netzspannungsquelle sowie die Ersatzgrößen der Leitung R und L. Ist der ohmsche Widerstand R_N des Netzes nicht vernachlässigbar, entsteht bei der Einspeisung von Wirkstrom ein stationärer ohmscher Spannungsfall Δu_R:

$$\Delta u_R = i \cdot R \tag{87.55}$$

Zusätzlich muss der transiente Spannungsfall an der Netzinduktivität L_N berücksichtigt werden. Hierfür muss der Einfluss der Einspeisung von Wirk- und Blindstrom betrachtet werden. Da der Wirkstrom i_P in Phase zur sinusförmigen Netzspannung liegt, wird hierfür die Sinusfunktion verwendet:

$$i_P = -I_P \cdot \sin(\omega \cdot t + \varphi_{U,N}) \tag{87.56}$$

Der transiente Spannungsfall über die Netzinduktivität L_N errechnet sich durch die Ableitung der Sinus-Stromfunktion:

$$\begin{aligned} \Delta u_P = L_N \frac{di_P}{dt} &= -L_N \cdot I_P \cdot \omega \cdot \cos(\omega \cdot t + \varphi_{U,N}) \\ &= -U_P \cdot \cos(\omega \cdot t + \varphi_{U,N}) \end{aligned} \tag{87.57}$$

Als Ergebnis erhält man eine cosinusförmige Spannung, die um 90° zur Netzspannung verschoben ist, siehe Abb. 87.51 Mitte. Somit kann durch eine Wirkstromeinspeisung an der Netzinduktivität kein gutes Stellergebnis für die Spannung erreicht werden. Nun soll die Blindstromeinspeisung betrachtet werden. Da der Blindstrom i_Q zeitlich um 90° zur sinusförmigen

Tab. 87.6 Zulässige Spannungsänderungen durch Schalthandlungen

Spannungsebene	Zulässige Spannungsänderung durch Schalthandlungen
Niederspannung	3 %
Mittelspannung	2 %
Hoch- und Höchstspannung, einzelne Erzeugungseinheiten	0,5 %
Hoch- und Höchstspannung, gesamte Anlage	2 %
Hoch- und Höchstspannung, gestörter Betrieb	5 %

Netzspannung verschoben ist, wird hierfür die Cosinusfunktion verwendet:

$$i_Q = -I_Q \cdot \cos(\omega \cdot t + \varphi_{U,N}) \tag{87.58}$$

Der transiente Spannungsfall über die Netzinduktivität L_N errechnet sich durch die Ableitung der Kosinus-Stromfunktion:

$$\begin{aligned}\Delta u_Q = L_N \frac{di_Q}{dt} &= L_N \cdot I_Q \cdot \omega \cdot \sin(\omega \cdot t + \varphi_{U,N}) \\ &= U_Q \cdot \sin(\omega \cdot t + \varphi_{U,N})\end{aligned} \tag{87.59}$$

Als Ergebnis erhält man eine sinusförmige Spannung, die in Phase zur Netzspannung liegt, siehe Abb. 87.51 rechts. Somit kann durch eine Blindstromeinspeisung an der Netzinduktivität ein linear wirkendes Stellergebnis für die Spannung erzielt werden.

87.3.4.2 Allgemeine Beschreibung der Spannungsregelung auf den Netzebenen

Wendet man den zuvor beschriebenen allgemeinen Fall auf die Spannungsebenen an, kommt man zu folgendem Ergebnis:

1. Ist der ohmsche Widerstand nicht vernachlässigbar klein, beeinflusst der Wirkstrom die Spannungsamplitude.
2. Betrachtet man nur die Netzinduktivität, liefert die Regelung des Blindstromes eine gute Stellmöglichkeit für die Spannungsamplitude.

Für die Niederspannungsebene wirken beide Effekte, weshalb die Spannungsamplitude hier sowohl durch den Wirkstrom als auch durch den Blindstrom beeinflusst wird. Hingegen ist der ohmsche Widerstand des Netzes auf der Mittel-, Hoch- und Höchstspannungsebene vernachlässigbar. Deshalb wird die Spannung auf diesen Netzebenen nur über den Blindstrom geregelt.

87.3.4.3 Direkte und indirekte Spannungsstellung

Auf allen Spannungsebenen müssen die zulässigen Spannungstoleranzen von jeweils ±10 % der jeweiligen Nennspannung eingehalten werden. Dies schließt auch die schaltbedingten Spannungsänderungen ein, die nach Tab. 87.6 nur geringe Prozentwerte annehmen dürfen. Zur Stellung der Spannung kann deren Amplitude direkt über Spannungsabgriffe an der Feinstufenwicklung (F) des Transformators verändert werden, siehe

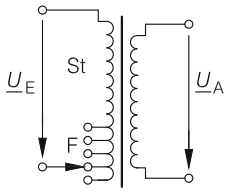


Abb. 87.52 Prinzip eines Transformators mit direkter Spannungsstellung, St – Stammwicklung, F – Feinstufenwicklung

Abb. 87.52. Die Stufenstellung des Schalters erfolgt immer auf der Oberspannungsseite, da er dort einer geringeren Strombelastung unterliegt. Es werden maximal ±13 Wicklungsanzapfungen ausgeführt, die für einen maximalen Stellbereich von ±22 % ausgelegt werden. Bei der indirekten Spannungsstellung wird neben der Amplitude auch die Phasenlage eingestellt. Dazu werden Teilspannungen aus unterschiedlichen Wicklungen addiert, die Gesamtspannung ergibt sich somit indirekt durch die Überlagerung von Amplitude und Phasenlage der Einzelspannungen. In Abb. 87.53 ist das Prinzip der indirekten Spannungsstellung gezeigt. Der Transformator weist neben dem Hauptteil aus Primär- und Sekundärwicklung noch einen Zusatztransformator auf, der aus der Reihenwicklung RW und der Erregerwicklung EW besteht. Der lastabhängige Stromfluss durch die Reihenwicklung erzeugt eine Spannung in der Erregerwicklung, mit der die Zusatzwicklung ZW gespeist wird. Die Zusatzwicklung speist einen zusätzlichen Spannungsanteil in der Sekundärwicklung. Mit dem Umschalter S kann die Stromflussrichtung über die Reihenwicklung RW festgelegt werden, um sowohl den Spannungsabfall bei steigender Belastung als auch den Spannungsanstieg bei zunehmender Einspeisung kompensieren zu können. Die Ausgangsspannung $\underline{U}_A$ ergibt sich aus der Überlagerung der komplexen Größen $\underline{U}_S$ und $\underline{U}_{ZW}$ aus der Sekundär- und der Zusatzwicklung:

$$\underline{U}_A = \underline{U}_S + \underline{U}_{ZW} \tag{87.60}$$

87.3.4.4 Spannungsstellung auf Niederspannungsebene mit regelbaren Ortsnetztransformatoren

Auf der Niederspannungsebene werden die Leitungsquerschnitte so dimensioniert, dass im Belastungsfall durch den Stromfluss

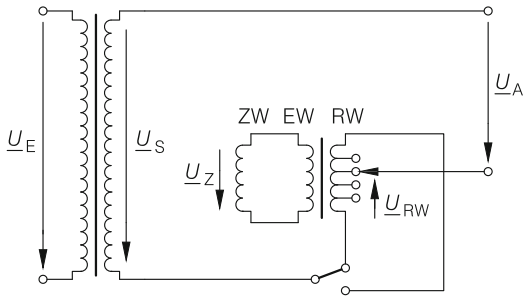


Abb. 87.53 Prinzip eines Transformators mit indirekter Spannungsstellung, ZW – Zusatzwicklung, EW – Erregerwicklung, RW – Reihenwicklung

Tab. 87.7 Maximal zulässiger Spannungsfall im Hauptstromversorgungssystem auf der Niederspannungsebene nach den Technischen Anschlussbedingungen (TAB)

Leistungsbedarf	Zulässiger Spannungsfall
Bis 100 kVA	0,50 %
Über 100 bis 250 kVA	1,00 %
Über 250 bis 400 kVA	1,25 %
Über 400 kVA	1,50 %

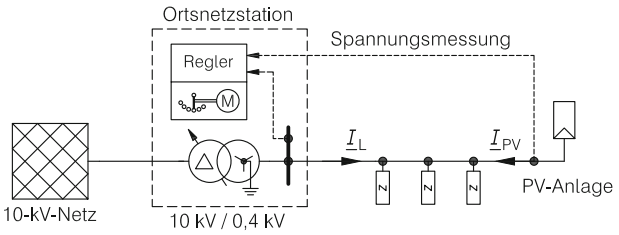


Abb. 87.54 Prinzip des regelbaren Ortsnetztransformators

über der Netzimpedanz ein maximal zulässiger Spannungsfall von Δu nicht überschritten wird. Die Höhe des Spannungsfalls wird in den Technischen Anschlussbedingungen (TAB) festgelegt, siehe Tab. 87.7. Bei räumlich ausgedehnten Strahlennetzen kann die Spannung am Leitungsende bei hoher elektrischer Last unzulässig absinken oder bei der Energieeinspeisung, z. B. aus Photovoltaikanlagen, und gleichzeitig geringem Verbrauch, unzulässig ansteigen. Mit der ansteigenden Verbreitung dezentraler Einspeisungen sind tageszeitabhängig gegenläufige Spannungsstellungen erforderlich. Diese werden durch regelbare Ortsnetztransformatoren realisiert, siehe Abb. 87.54. Hierbei wird die Oberspannungsseite des Ortsnetztransformators abhängig von der Spannungsdifferenz zwischen Unterspannungsausgang und einem Messpunkt am Ende der Niederspannungsleitung automatisch gestellt. Dafür wird ein motorbetriebener Stellschalter genutzt, der vorhandene Wicklungsanzapfungen an der Mittelspannungswicklung kontaktiert. In der Praxis wird bei hoher Netzbelastung die Ausgangsspannung des Transformators erhöht, damit die Spannung am Leitungsende ausreichend groß ist. Bei hohen Einspeiseleistungen am Leitungsende und der damit verbundenen Spannungsanhebung wird die Ausgangsspannung des Transformators verringert, siehe Abb. 87.55. Es ist der Einzel- und Mehrfachsensor-Betrieb von regelbaren Ortsnetztransformatoren zu unterscheiden. Die Erfassung der Spannungswerte auf der Unterspannungsseite kann entweder nur über einen einzelnen Sensor direkt am Ausgang der Sekundärwicklung oder zusätzlich mit Sensoren am Ende von Leitungszügen realisiert werden. Die Installation und Einbindung zusätzlicher, entfernter Sensoren für eine genauere Spannungsregelung verursacht zwar einen erheblichen Mehraufwand, ist jedoch bei Strahlennetzen in ländlichen Regionen oft die einzige Möglichkeit zur Vermeidung des Netzausbaus.

87.3.4.5 Dynamische Spannungsregelung

Während die zuvor beschriebenen Maßnahmen zur quasistationären (statischen) Spannungsregelung dienen, kann auch eine zeitlich schnellere Reaktion auf Spannungsänderungen erforderlich sein. Dies gilt auch für den Normalbetrieb, wenn schnelle Spannungsschwankungen unzulässige Effekte auslösen. Dazu gehören z. B. Flicker, die sich als wahrnehmbare Leuchtdichteschwankungen aufgrund von Spannungsschwankungen einstellen, da sich die Leuchtdichte L_V von Glühlampen proportional zum Quadrat der Spannung verhält:

$$L_V \sim U^2 = k \cdot U^2 \tag{87.61}$$

Andere Beleuchtungseinrichtungen wie Energiespar- oder LED-Lampen, verhalten sich grundsätzlich anders als Glühlampen, wodurch dieser Effekt zukünftig rückläufig sein wird. Jedoch ist auch das Drehmoment von Drehfeldmaschinen wie z. B. Asynchronmaschinen, abhängig von der Netzspannung, siehe Abb. 87.56:

$$\frac{M}{M_{\text{neu}}} = \left(\frac{U}{U_{\text{neu}}}\right)^2 \tag{87.62}$$

Spannungsschwankungen verursachen hierbei starke Drehmomentschwankungen, was sich negativ auf Fertigungsprozesse auswirkt.

Ein Beispiel für dynamische Störeffekte liefert der sich im Fehlerfall einstellende Spannungstrichter, siehe Abb. 87.57. In Abb. 87.57 links ist der ungestörte Spannungsverlauf eines vermaschten Netzes im Normalbetrieb gezeigt, Abb. 87.57 rechts zeigt den Spannungseinbruch bei Kurzschluss. Von oben gesehen verläuft dieser trichterförmig zum Kurzschlussort. Um die räumliche Ausdehnung des Spannungstrichters und damit die

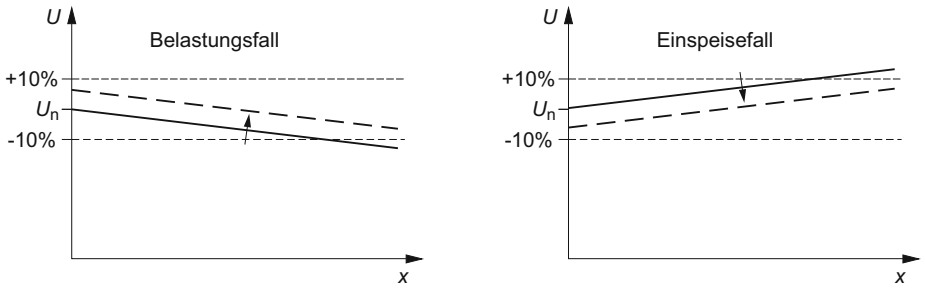


Abb. 87.55 Spannungsregelung mit regelbarem Ortsnetztransformator. *Links*: bei Belastung und *rechts*: bei Einspeisung, die *gestrichelten Linien* zeigen den Spannungsverlauf nach dem Stellvorgang an

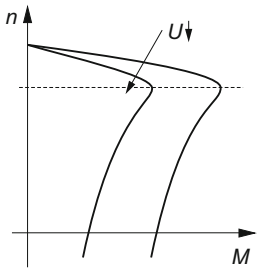


Abb. 87.56 Spannungsabhängige Drehzahl-Drehmoment-Kennlinie einer Asynchronmaschine

Anzahl möglicher Verbraucherausfälle zu vermindern, kann im Bereich der Flanken des Trichters Blindstrom in das Netz eingespeist werden, um einen Spannungsbeitrag gemäß (87.59) zu erhalten. Dies ist nur so lange erforderlich, bis die Schutztechnik den Fehler abgeschaltet hat, was einstellungsgemäß nach 150 Millisekunden der Fall ist. Die dynamische Einspeisung von Blindstrom im Fehlerfall kann nicht von der statischen Spannungsregelung erfüllt werden; sie wird nach definierten Einstellwerten von dezentralen Erzeugungseinheiten wie Windenergie- und Photovoltaikanlagen mit leistungselektronischer Netzkopplung erbracht, siehe Abschn. 87.3.4.3.

87.3.4.6 Systemdienstleistungen von Erzeugungseinheiten

Zur Sicherstellung der Netzstabilität werden beim Netzanschluss von Erzeugungseinheiten Anforderungen an deren Regelfähigkeiten und ihren Beitrag zu den Systemdienstleistungen, d. h. zur Frequenz- und Spannungshaltung, gestellt. Hierbei wird zeitlich zwischen quasistationären und dynamischen Anforderungen unterschieden. Im TransmissionCode 2007 des VDN (Verband der Netzbetreiber e.V. beim VDEW) wird quasistationäres Verhalten bei einem Frequenzgradienten von $\leq 0{,}5\,\%/\mathrm{min}$ bzw. einem Spannungsgradienten von $5\,\%/\mathrm{min}$ definiert. Als Erzeugungseinheiten gelten sowohl konventionelle Verbrennungskraftwerke als auch EEG-Erzeugungseinheiten wie Windenergie- und Photovoltaikanlagen, die nach dem Erneuerbare-Energien-Gesetz (EEG) vergütet werden. Alle Erzeugungseinheiten müssen folgende Hauptanforderungen erfüllen:

- Beteiligung an der Frequenzregelung nach Statiken (quasistationär)
- Wirkleistungsabgabe im dynamischen Kurzzeitbereich (dynamisch)
- Statische Spannungsregelung
- Durchfahren von Spannungseinbrüchen bei Netzfehlern und Einspeisung von Blindstrom zur Spannungsstützung (dynamisch).

87.3.4.6.1 Frequenzhaltung durch Wirkleistungsstellung

Zur Regelung der Frequenz muss wie im Abschn. 87.3.3 beschrieben die Wirkleistung verändert werden. Bei Überfrequenz muss im Hoch- und Höchstspannungsnetz die Leistung mit 1 % der Nennleistung pro Minute reduziert werden. Kraftwerke mit Leistungen ab 100 MW müssen sich an der Primärregelung beteiligen, EEG-Erzeugungseinheiten können davon befreit werden. Sinkt die Frequenz ab, wird zu wenig Wirkleistung in das Netz eingespeist. Diese kann bei Verbrennungskraftwerken durch die Turbinenregelung erhöht werden, wodurch der Brennstoffverbrauch ansteigt. Bei erneuerbaren Kraftwerken ohne Brennstoffzufuhr steht diese Möglichkeit nicht zur Verfügung, aber auch diese Kraftwerke müssen sich an den Systemdienstleistungen beteiligen. In Abb. 87.58 sind die Mindestanforderungen für die quasistationäre Wirkleistungsabgabe auf der Hoch- und Höchstspannungsebene abhängig von der

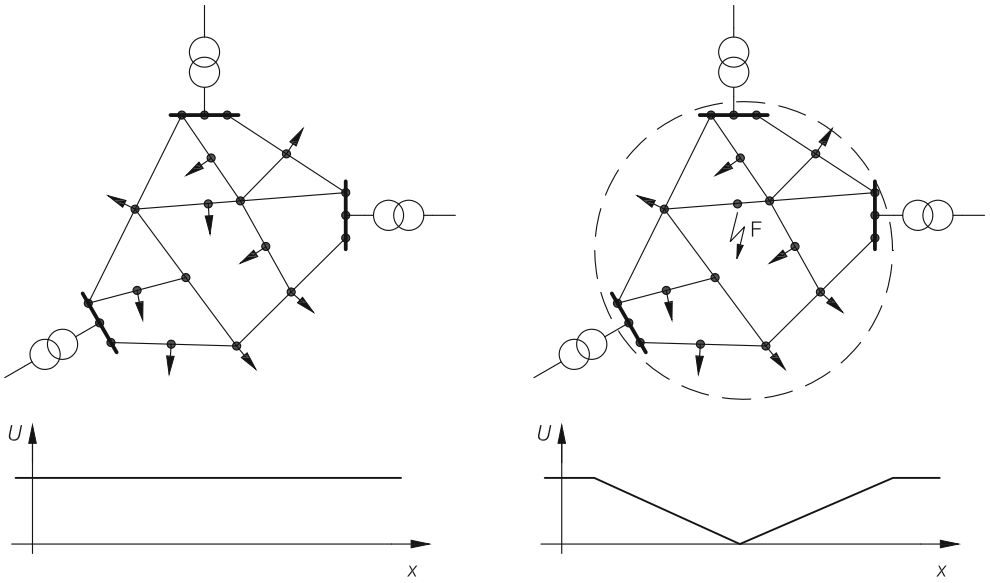


Abb. 87.57 Spannungstrichter im Fehlerfall

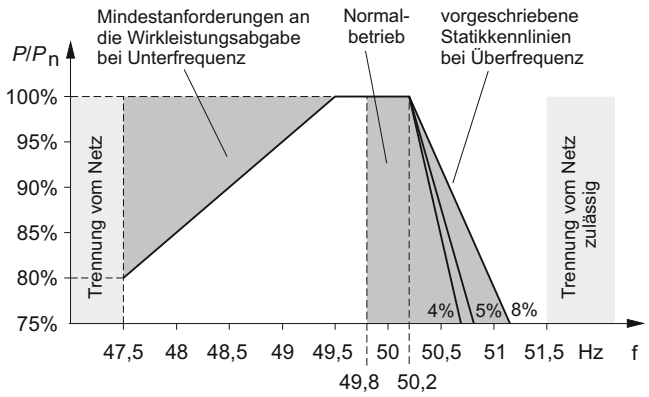


Abb. 87.58 Mindestanforderungen für die quasistationäre Wirkleistungsabgabe abhängig von der Netzfrequenz nach VDN TransmissionCode 2007

Netzfrequenz nach VDN TransmissionCode 2007 gezeigt. Bei abnehmender Frequenz darf die Wirkleistungsabgabe im Bereich ab 49,5 bis 47,5 Hz nur linear auf maximal 80 % abgesenkt werden. Bei einer Überfrequenz ab 50,2 bis 51,5 Hz muss die Leistung mit einer Statik von 4 bis 8 % und einem Leistungsgradienten von 0,5 P_n/Hz bis 0,25 P_n/Hz reduziert werden. EEG-Erzeugungseinheiten müssen bei einer Frequenz von mehr als 50,2 Hz die momentane Wirkleistung mit einem Gradienten von 40 % der momentan verfügbaren Leistung des Generators je Hertz absenken. Auf der Mittelspannungsebene gilt dies nach der BDEW-Mittelspannungsrichtlinie für alle Erzeugungseinheiten. Unterhalb von 47,5 Hz und oberhalb von 51,5 Hz ist die Trennung vom Netz zulässig.

Die Abb. 87.59 zeigt die Mindestanforderungen für die Wirkleistungsabgabe im zulässigen Spannungsbereich auf der Hoch- und Höchstspannungsebene abhängig von der Netzfrequenz nach VDN TransmissionCode 2007. Der Bereich der dauernd zu realisierenden Wirkleistungsabgabe liegt im Frequenzband von 49 bis 50,5 Hz und im Spannungsbereich von 85 bis 112 % der Nennspannung. Nach einer Netztrennung müssen die Kraftwerke im Inselbetrieb ihren Eigenbedarf decken und sich später bei einer Frequenz im Bereich von 48,5 bis 51,5 Hz wieder mit dem Netz synchronisieren können.

Sinkt die Netzfrequenz durch einen kurzzeitigen Leistungsmangel auf bis zu 49 Hz ab, müssen Erzeugungseinheiten das Netz stützen, indem sie weiter ihre Wirkleistung einspeisen, siehe Abb. 87.60. Hier ist eine Reaktion auf einen Frequenzgradienten von $> 0{,}5$ %/min gefordert, weshalb die Anforderung dem dynamischen Kurzzeitbereich zugeordnet wird. Der Schnittpunkt mit der Hüllkurve in Abb. 87.60 markiert, wie lange die vom ÜNB vorgegebene Leistung konstant gehalten werden muss. Im Beispiel sinkt die Frequenz auf 49,6 Hz ab, dann muss die vorgegebene Wirkleistung nach ca. 3 s wieder erreicht sein und ab dann bis ca. 51 s konstant gehalten werden. Ist der Leistungsmangel dann nicht behoben, findet an der Erzeugungsanlage eine Wirkleistungsabgabe gemäß Abb. 87.58 statt; parallel wird die Frequenz-Wirkleistungsregelung im Energieversorgungsnetz fortgesetzt.

87.3.4.6.2 Spannungshaltung durch Blindleistungsstellung

Im Abschn. 87.3.4 wurde bereits die Funktionsweise der Spannungsregelung mit einer Blindleistungseinspeisung und im Abschn. 87.3.4.5 wurde das Problem des Spannungstrichters im Fehlerfall beschrieben. Nun sollen die Anforderungen zur Bereitstellung von Blindleistung betrachtet werden. In Abb. 87.61 sind zwei von insgesamt drei Varianten für die Blindleistungsbereitstellung aus Erzeugungsanlagen auf der Hoch- und Höchstspannungsebene nach VDN TransmissionCode 2007 gezeigt. Die Übertragungsnetzbetreiber können aufgrund unterschiedlicher Netzanforderungen eine dieser Varianten vorgeben, deren Blindleistungswerte im gesamten Nennleistungsbereich bereitgestellt werden müssen. Dabei muss der Generator der Erzeugungsanlage weiterhin in jedem möglichen Arbeitspunkt gemäß seinem Leistungsdiagramm laut Herstellerangaben betrieben werden können. Die in Abb. 87.61a gezeigte Variante 1 weist den größten induktiven Blindleistungsstellbereich auf, die in Abb. 87.61b gezeigte Variante 3 den größten kapazitiven Stellbereich. Der Stellbereich der hier nicht gezeigten Variante 2 ist eine Schnittmenge der Varianten 1 und 3. Auf der Mittelspannungsebene muss nach der BDEW-Mittelspannungsrichtlinie

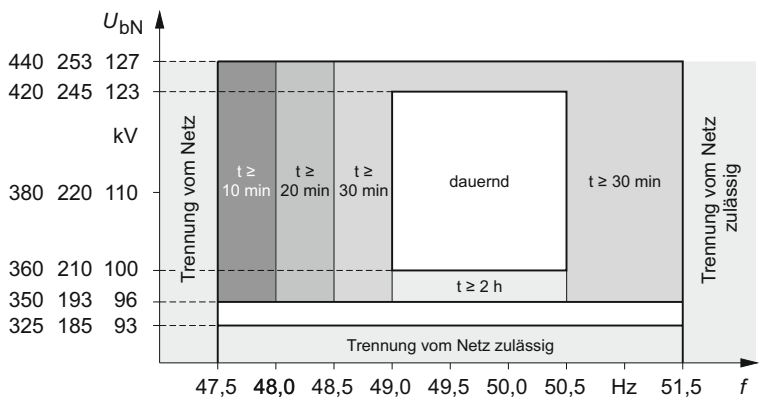


Abb. 87.59 Mindestanforderungen für die Wirkleistungsabgabe im zulässigen Spannungsbereich abhängig von der Netzfrequenz nach VDN TransmissionCode 2007

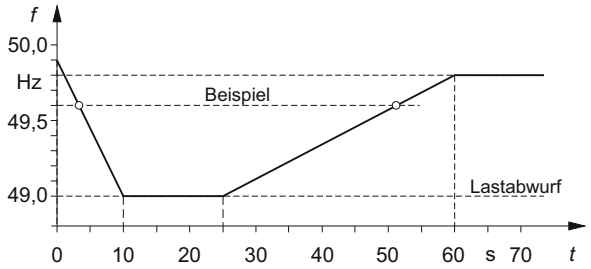


Abb. 87.60 Mindestanforderungen für die Wirkleistungsabgabe im dynamischen Kurzzeitbereich auf der Hoch- und Höchstspannungsebene nach VDN TransmissionCode 2007

der Leistungsfaktor im Bereich 0,95 induktiv und 0,95 kapazitiv einstellbar sein, alternativ kann eine leistungsabhängige Kennlinie für den Leistungsfaktor vorgegeben werden.

Treten Kurzschlüsse im Netz auf, bricht die Spannung abhängig vom Kurzschlussstrom I_k und der Kurzschlussimpedanz Z_k auf die Kurzschlussspannung U_k ein:

$$\underline{U}_k = \underline{I}_k \cdot \underline{Z}_k = \underline{I}_k \cdot (R_k + j\omega L_k) \qquad (87.63)$$

Bei Mehrfachspeisung ergibt sich der Kurzschlussstrom aus den Teilströmen der Einspeisungen. Die Größe der Kurzschlussimpedanz ist abhängig von der Entfernung zum Fehlerort und wird hauptsächlich durch die vorhandenen Transformatoren bestimmt. Somit bricht die Spannung bei einem Kurzschluss auf sehr geringe Werte ein, wenn die Entfernung zum Fehlerort klein ist oder sich kein Transformator zwischen Kurzschlussort und Netzanschlusspunkt von Erzeugungsanlagen befindet. Ist im Umkehrschluss ein Transformator zwischen Kurzschlussort und Netzanschlusspunkt vorhanden bzw. ist der Kurzschlussort weit entfernt, fällt der Spannungseinbruch geringer aus. Sowohl im TransmissionCode 2007 für das Hoch- und Höchstspannungsnetz als auch in der BDEW-Mittelspannungsrichtlinie werden dabei unterschiedliche Anforderungen an Erzeugungsanlagen vom Typ 1 mit direkt gekoppeltem Synchrongenerator und vom Typ 2 mit Netzkopplung über Frequenzumrichter gestellt. Anlagen vom Typ 1 müssen auch auf der Mittelspannungsebene kurzzeitige Spannungseinbrüche bis auf null durchfahren können, siehe Abb. 87.62. Die Zeitspanne von 150 ms ergibt sich aus der typischen Fehlerklärungszeit, in der ein Kurzschluss abgeschaltet wird und die Spannung wieder zur Verfügung steht.

An den Typ 2 werden im Bereich I identische Anforderungen wie an den Typ 1 gestellt, es ist keine Trennung vom Netz zulässig, siehe Abb. 87.63a. Spannungseinbrüche im Bereich II müssen ebenfalls durchfahren werden, in Absprache mit dem Netzbetreiber darf die Anlage sich kurzzeitig vom Netz trennen,

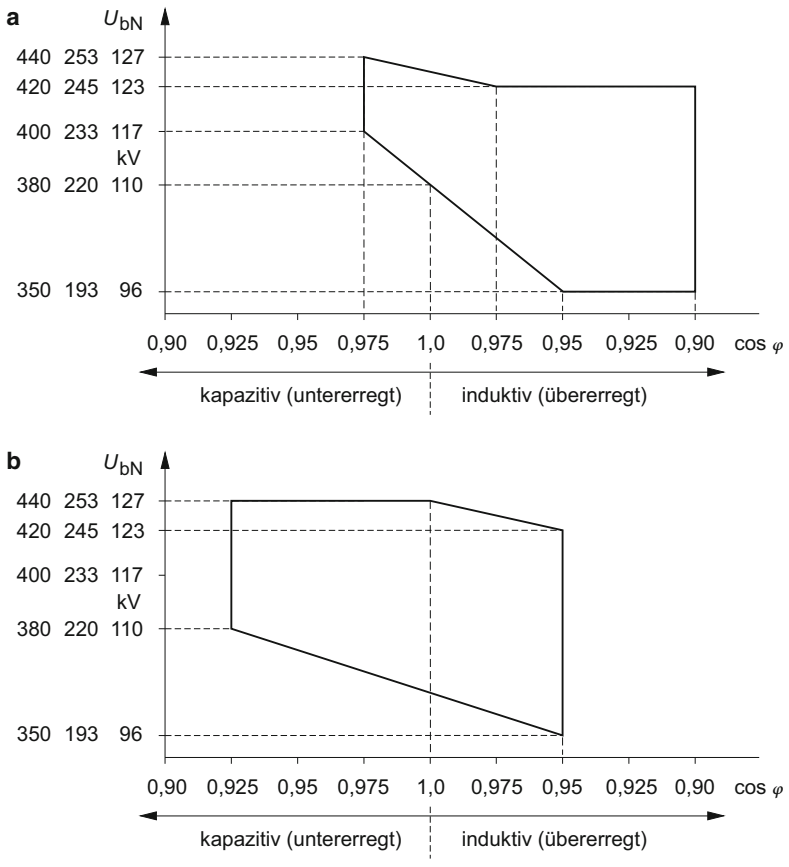


Abb. 87.61 Mindestanforderungen für die Blindleistungsbereitstellung auf der Hoch- und Höchstspannungsebene nach VDN TransmissionCode 2007. **a** Variante 1 mit größter induktiver und **b** Variante 3 mit größter kapazitiver Blindleistung

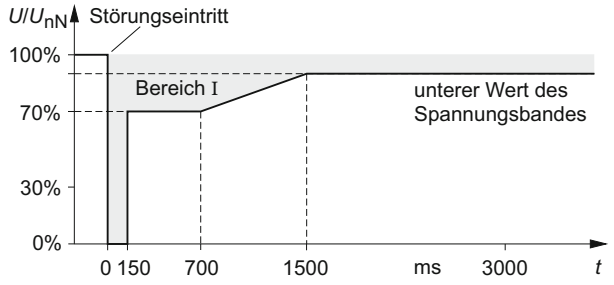


Abb. 87.62 Grenzlinie für einen Spannungseinbruch im Mittelspannungsnetz an einer Erzeugungsanlage vom Typ 1 mit direkt gekoppeltem Synchrongenerator nach BDEW-Mittelspannungsrichtlinie 2008

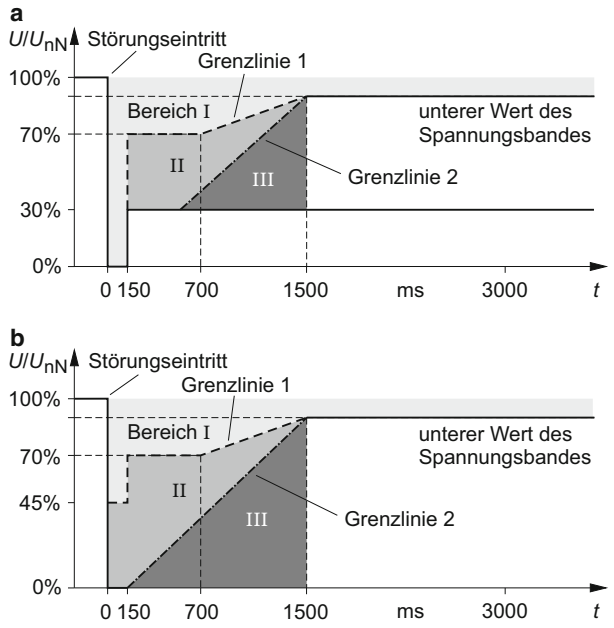


Abb. 87.63 Grenzlinien für einen Spannungseinbruch an einer Erzeugungsanlage vom Typ 2 mit Netzkopplung über Frequenzumrichter **a** im Mittelspannungsnetz nach BDEW-Mittelspannungsrichtlinie 2008, **b** im Hoch- und Höchstspannungsnetz nach TransmissionCode 2007

wenn sie sich nach 2 s wieder zuschaltet. Im Bereich III ist eine kurzzeitige Trennung der Anlage vom Netz zulässig.

Ähnliche Anforderungen gelten auch wie in Abb. 87.63b gezeigt für die Hoch- und Höchstspannung, allerdings müssen hier im Bereich I nur kurzzeitige Spannungseinbrüche bis auf 45 % der Nennspannung durchfahren werden. Der Grund dafür liegt in der zuvor beschriebenen höheren Kurzschlussimpedanz aufgrund eines Transformators zwischen physikalischem Anschlusspunkt einer Erzeugungsanlage und der Hoch- bzw. Höchstspannungsebene.

Erzeugungsanlagen mit regenerativen Energiequellen müssen darüber hinaus während des Spannungseinbruchs zur Spannungsstützung Blindleistung ins Netz speisen, um die räumliche Ausdehnung eines Spannungstrichters nach Abschn. 87.3.4.5 zu begrenzen. Dazu müssen sie innerhalb einer Anregelzeit von

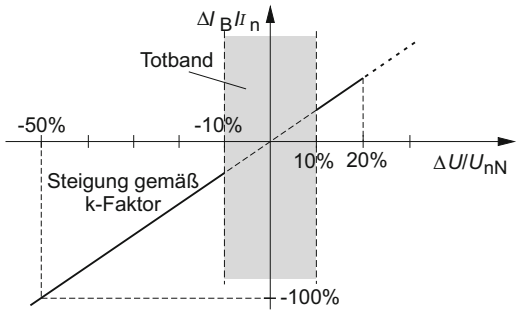


Abb. 87.64 Prinzip der Spannungsstützung bei Netzfehlern bei Erzeugungsanlagen mit regenerativen Energiequellen nach VDN TransmissionCode 2007

<20 ms Blindstrom einspeisen, die Steigung der dabei verwendeten Kennlinie wird abhängig vom Blindstrom I_B, dem Nennstrom I_n, der Spannungsdifferenz ΔU sowie der Nennspannung U_n mit einem k-Faktor beschrieben, siehe Abb. 87.64:

$$k = \frac{\Delta I_b / I_n}{\Delta U / U_n} \geq 2{,}0 \text{ p.u.} \tag{87.64}$$

Nach der Rückkehr in das Spannungstotband von $\pm 10\,\%$ muss die Spannungsstützung über weitere 500 ms aufrecht erhalten werden, danach muss die Blindleistungsabgabe nach wenigen Minuten dem vom Netzbetreiber vorgegebenen Sollwert entsprechen.

87.3.5 Herausforderungen bei einer Energieversorgung mit fluktuierender Erzeugung

Mit zunehmend dezentraler Energiewandlung aus fluktuierender Wind- und Solarenergie entstehen neue Anforderungen für innovative Lösungsansätze. Dabei soll stets eine hohe Versorgungssicherheit zu bezahlbaren Preisen gewährleistet bleiben. Mit diesem Ziel sind neuartige Strukturen für das gesamte Energieversorgungssystem zu entwickeln. Hierbei ist zu beachten, dass der Einsatz innovativer Einzeltechnologien oft nur sehr begrenzt wirkt, während systemtechnische Ansätze eine hohe Wirksamkeit aufweisen. Mit 181,2 GW Kraftwerksleistung im Jahr 2014 liegt der deutsche Kraftwerkspark weit oberhalb der Jahreshöchstlast von 81 GW, siehe Abschn. 87.2.2. Im Jahr 2014 sind bereits 35 GW Windenergie- und 38 GW Photovoltaikleistung installiert. Für den Endausbau allein der Windenergie On- und Offshore werden in Deutschland bis zu 160 GW erwartet. Beim Umbau der Energieversorgung zu einer CO_2-armen Erzeugung müssen Wind- und Sonnenkraft intensiv genutzt werden. Die dabei gelieferten wetterabhängigen und damit fluktuierenden Energiebeiträge schwanken stark. Aufgrund der hohen installierten Leistungen liegen die größten Probleme bei der Integration der Windenergie, da hierbei starke jahres- und tageszeitliche Schwankungen auftreten, siehe Abb. 87.65. Solche Schwankungen erschweren die Prognostizierbarkeit und Planbarkeit der Energiebeiträge und begründen eine Vielzahl

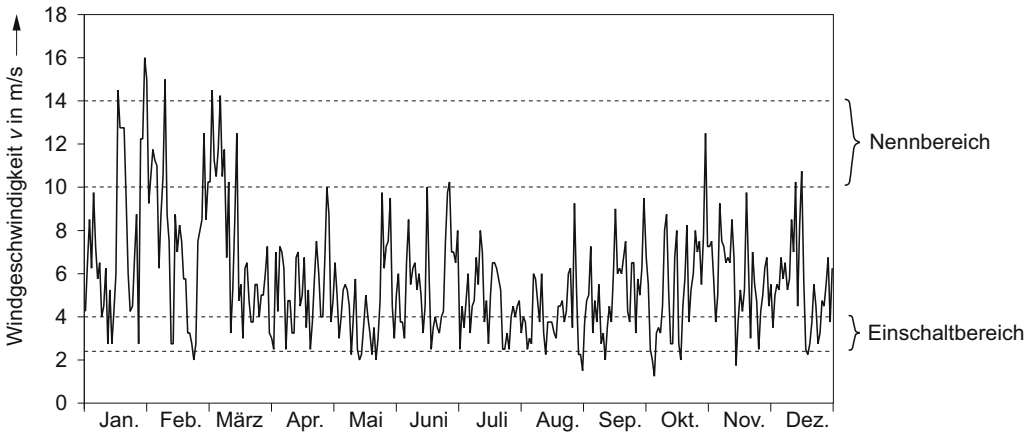


Abb. 87.65 Verlauf der Windgeschwindigkeit über ein Jahr

von Fragestellungen zu Versorgungssicherheit, Ersatzkraftwerken, Energiespeichern und Netzausbau.

Die Problemstellungen durch fluktuierende Energiebereitstellung lassen sich anhand der in Abb. 87.66 dargestellten Zusammenhänge von links beginnend systematisch klassifizieren:

1. **Erzeugung > Last:** Zur Sicherstellung der Frequenzstabilität muss die nicht benötigte Erzeugungsleistung entweder durch zusätzliche Verbraucher umgesetzt, in Energiespeichern gespeichert oder abgeschaltet werden.
2. **Prognoseabweichung:** Zur Absicherung der Differenz zwischen prognostizierter und eingespeister Erzeugungsleistung müssen andere Kraftwerke oder Energiespeicher herangezogen bzw. Verbraucher abgeschaltet werden.
3. **Flaute:** Es steht keine Erzeugungsleistung zur Verfügung, die benötigte Verbraucherleistung muss durch andere Kraftwerke oder Energiespeicher bereitgestellt werden.
4. **Gradient:** Zeitlich steile Anstiege oder Abfälle der Windleistung müssen durch die Regelung anderer Kraftwerke oder Energiespeicher kompensiert werden.

Zur technischen Beherrschung der vier Problemstellungen werden aktuell vier Technologie- und Maßnahmenansätze verfolgt, die nachfolgend detaillierter behandelt werden:

1. Netzertüchtigung und Netzausbau
2. Flexible Erzeugung
3. Flexibler Verbrauch durch steuerbare Lasten
4. Energiespeicherung

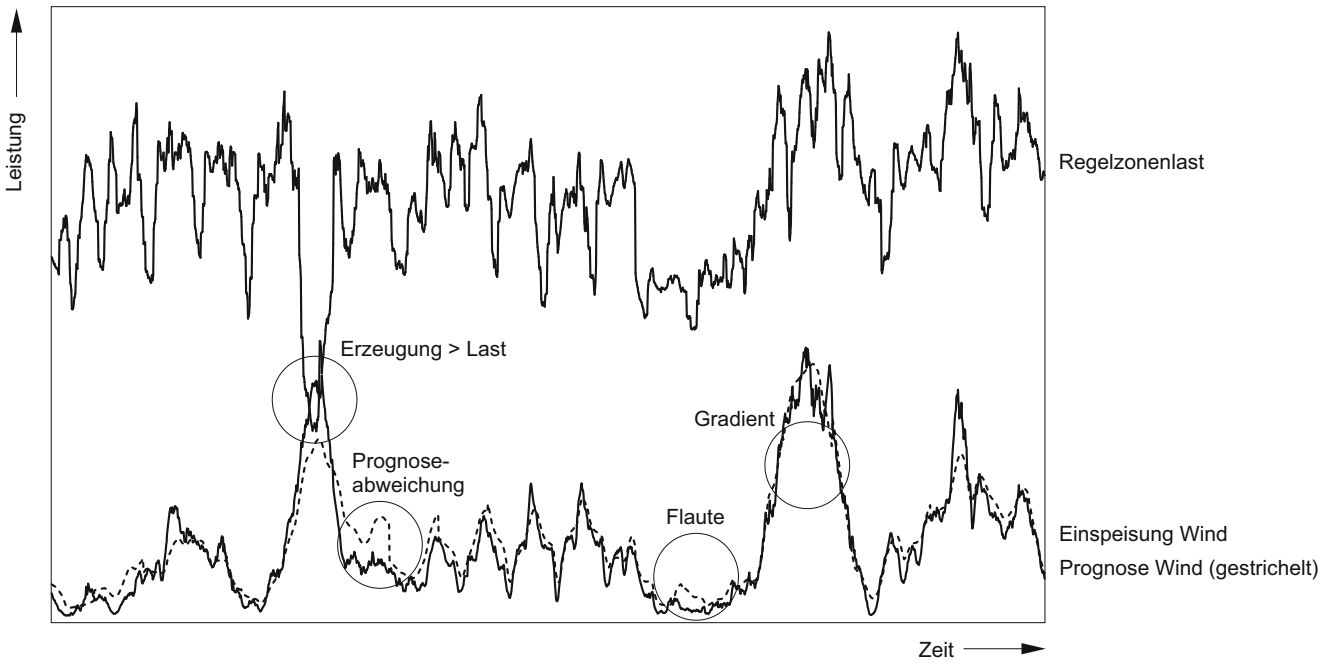
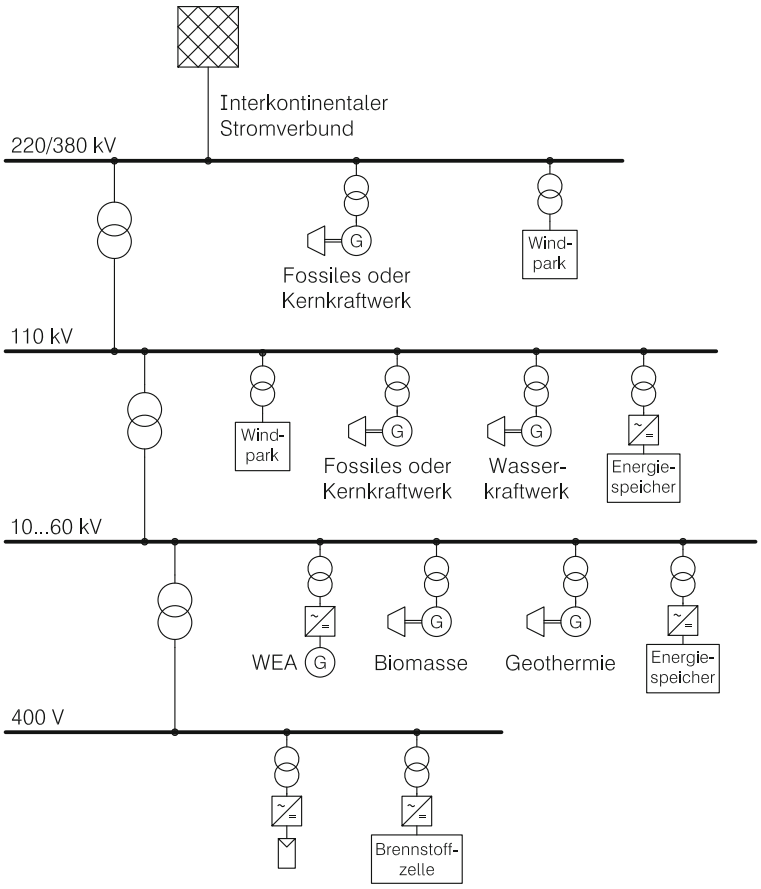


Abb. 87.66 Zeitlicher Verlauf von Regelzonenlast, Windprognose und Windeinspeisung



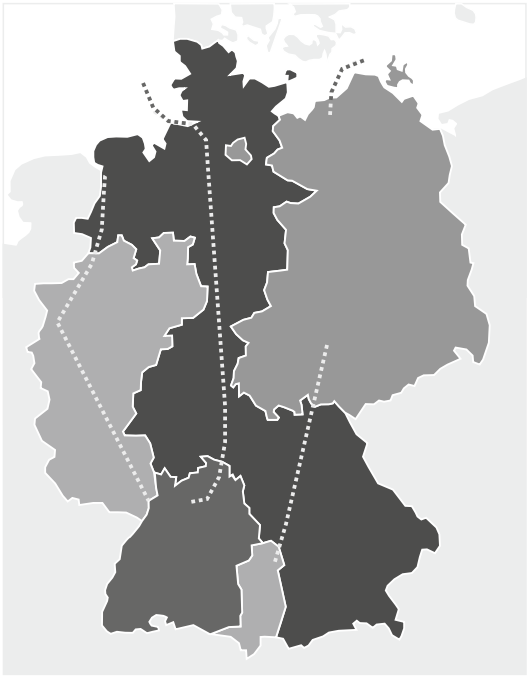


Abb. 87.67 *Links*: Struktur einer dezentralen Energieversorgung, *rechts*: Netzausbau nach Netzentwicklungsplan, Datenquelle: Bundesnetzagentur 2013

87.3.5.1 Netzausbau

Bei der Netzertüchtigung und dem Netzausbau geht es einerseits um die planmäßige Erneuerung und Schaffung von Transport- und Verteilungskapazitäten für elektrische Leistungen. Der Bedarf hierfür entsteht einerseits durch die dezentral verteilten regenerativen Energien, siehe Abb. 87.67 links. Dabei müssen neben dem klassischen Netzausbau auch neue Richtungen des Leistungsflusses, z. B. durch die Einspeisung von Photovoltaikanlagen auf der Niederspannungsebene, hin zu höheren Spannungsebenen technisch beherrscht werden.

In Deutschland besteht vor allem die Aufgabe, die Windkrafterzeugung im Norden Deutschlands mit den Verbraucherzentren im Süden zu verbinden. Darüber hinaus werden zukünftig auch Leistungsflüsse aus dem Süden in Richtung Norden erwartet, wenn die im Süden mit hoher Gesamtleistung installierte Photovoltaik mit maximaler Leistung einspeist. Daher enthält der Netzentwicklungsplan vier Gleichstromtrassen, wovon zukünftig drei über weite Entfernungen von Süd nach Nord geführt werden, siehe Abb. 87.67 rechts. Bei der Gleichstromübertragung fallen zwar höhere Installationskosten an, jedoch ist die wirtschaftliche Übertragungslänge wesentlich größer, da wegen der im Vergleich zur klassischen Drehstromübertragung fehlenden Blindleistung nur geringe Transportverluste auftreten.

Zur Erhöhung der elektrischen Transportkapazität kann die Spannungsamplitude erhöht werden. Damit ergibt sich mit der Netzimpedanz Z eine zur Spannungserhöhung quadratisch ansteigende Übertragungsleistung:

$$P = \frac{U^2}{Z} \tag{87.65}$$

In der Praxis stehen weitere Möglichkeiten für eine erhöhte Transportkapazität klassischer Drehstromsysteme zur Verfügung:

I **Temperatur-Monitoring:** Durch die Messung der Temperatur von Leiterseilen können diese bis 80 °C und damit stärker belastet werden als durch sichere Abschätzung nach DIN EN 50182 unter den dort angenommenen ungünstigen Kühlbedingungen mit einer Außentemperatur von 35 °C, einer Windgeschwindigkeit von 0,6 m/s und 100 % Sonneneinstrahlung. Dadurch kann die Leitungskapazität bei damit verbundenen höheren Verlusten um bis zu 50 % erhöht werden, wenn sich z. B. bei der Einspeisung höherer Ströme aus Windparks gleichzeitig mit der windbedingten Leistung eine gute Kühlung der Leiterseile gewährleistet ist. Gleichzeitig müssen jedoch die Leitungstransformatoren ausgetauscht und die Schutzeinrichtungen angepasst werden. Diese Methode wird selten eingesetzt, da die Ausrüstung

nur eines Leitersystems nicht ausreichend im Sinne der $(n-1)$-Ausfallsicherheit ist. Also müssen mit den damit verbundenen steigenden Kosten mindestens zwei parallele Leitungssysteme ausgerüstet werden oder die Leitung wird als sog. Einspeisenetz in alleiniger Verantwortlichkeit z. B. eines Windparkbetreibers genutzt.

II **Nutzung von Hochtemperatur-Leiterseilen:** Hierbei werden die Standard-Leiterseile mit Stahlseele und 80 °C Grenztemperatur gegen Ausführungen mit Kunststoffseele oder Keramikfaserseele mit Grenztemperaturen von 175 oder 210 °C ausgetauscht, siehe Abb. 87.74 im Abschn. 87.4.1.4 „Leiterseile“. Wegen der höheren Ströme und Temperaturen müssen die Leitungstransformatoren, die Seilaufhängungen und Isolatoren ausgetauscht sowie die Schutztechnik angepasst werden. Nach der Umrüstung kann bei höheren Verlusten bis zur doppelten Leistung übertragen werden. Wie beim Temperaturmonitoring ergibt sich eine Einschränkung bei der $(n-1)$-Ausfallsicherheit. Da die Kosten der Umrüstung im Bereich von Neubaukosten liegen, wird diese Technologie in Deutschland bisher nur in verschiedenen Testprojekten eingesetzt, in den USA, China und Indien existieren kommerziell genutzte Systeme.

III **Ersatz von Systemen durch Systeme mit höherer Spannung:** Es ist möglich, zwei 220-kV-Systeme auf einem Mast durch ein 400-kV-System zu ersetzen oder 400-kV-Systeme durch 500-kV-Systeme abzulösen. Dies muss wegen der $(n-1)$-Ausfallsicherheit mindestens jeweils auf zwei parallelen Trassen erfolgen. Der Bau von bisher in Deutschland nicht eingesetzten 500-kV-Systemen wird wegen der fehlenden Standardkomponenten auch zukünftig nur eine theoretische Option sein.

IV **Leitungsneubau:** Der Netzausbau ist das wirksamste Instrument zur Beherrschung dieser Problembereiche. Dies wurde in verschiedenen Studien bestätigt. Hierbei geht es ausschließlich um den Bau neuer Freileitungssysteme. Der Neubau von Leitungstrassen erhöht die Übertragungskapazität sehr wirksam, er war in den vergangenen Jahrzehnten wegen der langen Planungs- und Bauzeit von bis zu 12 Jahren jedoch nur ein langfristiges wirkendes Instrument. Mit dem Energieleitungsausbaugesetz wurden 2009 die gesetzlichen Grundlagen zum Bau von energiewirtschaftlich bedeutsamen Übertragungsnetzen geschaffen. Der Ausbaubedarf des Onshore-Stromnetzes wird von den vier Übertragungsnetzbetreibern unter Beteiligung der Bundesnetzagentur festgestellt, die zweite Fassung des Netzentwicklungsplans Onshore liegt seit 2013 vor. Daraus ergibt sich in den nächsten 10 Jahren ein Ausbaubedarf von 1700 km Hochspannungs-Drehstromleitungen und 2100 km Hochspannungs-Gleichstromleitungen, 4400 km des Übertragungsnetzes müssen verstärkt und optimiert werden. Der Netzentwicklungsplan Offshore wird vom Bundesamt für Seeschifffahrt und Hydrographie erstellt und liegt in seiner ersten Fassung seit 2013 vor. Durch das seit 2013 geltende Bundesbedarfsplangesetz („Zweites Gesetz über Maßnahmen zur Beschleunigung des Netzausbaus Elektrizitätsnetze“) wird sichergestellt, dass 36 energiewirtschaftlich bedeutsame Projekte in absehbaren Zeiträumen realisierbar werden.

Mit dem Netzausbau kann stets eine große Wirkung erzielt werden, da durch zusätzliche Transportkapazitäten Regionen mit alternierenden Energiebeiträgen für den notwendigen Leistungsaustausch besser untereinander gekoppelt werden können. Insbesondere bei weit auseinander liegenden Energieerzeugern mit Leistungen im Gigawatt-Bereich ist der Netzausbau alternativlos. Hingegen stellt sich bei weit verzweigten Netzen im ländlichen Raum mit gleichzeitig geringer Lastdichte bei einem starken Zubau von dezentralen Erzeugungseinheiten oft die Frage nach alternativen Verbrauchsszenarien zur Vermeidung des dann nicht wirtschaftlichen Netzausbaus.

Neben dem für das Gesamtsystem der elektrischen Energieversorgung wesentlichen Ausbau von Übertragungstrassen werden auch Lösungen für den Ersatz innerstädtischer Hochspannungskabel durch platzsparende Hochtemperatur-Supraleiter (HTS) getestet. Bisher existiert in Essen eine Teststrecke von 1 km mit einem 10-kV-HTS-Kabel als Ersatz für eine 110-kV-Kabelstrecke, wobei die Kosten geringer als bei konventionellen 110-kV-Kabeln, jedoch wesentlich höher als bei 10-kV-VPE-Kabeln sind. Zusätzlich zum dem Netzausbau kommt auch der Bestimmung der freien Aufnahmekapazitäten von Netzknoten große Bedeutung zu, z. B. für den Anschluss oder die Erweiterung von Windparks. Dazu ist wegen der leistungselektronischen Netzkopplung von Windenergieanlagen die Kenntnis der frequenzabhängigen Netzimpedanz erforderlich. Um diese zu bestimmen, muss das Netz elektrisch angeregt werden, beispielsweise durch die Pulsung einer definierten Last. Dies erfordert den Bau von dafür dimensionierten Messeinrichtungen insbesondere für die Mittelspannung bei 10 und 20 kV sowie die Hochspannung mit der 110-kV-Netzebene. In Abb. 87.68 links ist der getestete Prototyp einer Messeinrichtung für die Mittelspannung bis 20 kV gezeigt, in Abb. 87.68 rechts ist ein gemessener Frequenzverlauf der Netzimpedanz abgebildet. Die Netzimpedanz Z ergibt sich aus den Wirk- und Blindwiderständen:

$$\underline{Z} = R + j\omega L + \frac{1}{j\omega C} \tag{87.66}$$

Bedeutung der Netzimpedanz

Nicht sinusförmige Spannungen im Netz erzeugen an passiven Verbrauchern nicht sinusförmige Ströme und umgekehrt. Die Oberschwingungen von Strom $\underline{i}_{N,n}$ und Spannung $\underline{u}_{N,n}$ sind über die Impedanz des Netzes $\underline{Z}_N$ verknüpft:

$$\underline{i}_{N,n} = \frac{\underline{u}_{N,n}}{\underline{Z}_N} = \frac{\underline{u}_{N,n}}{R_N + j\omega L_N + \frac{1}{j\omega C_N}} \tag{87.67}$$

Im Gegensatz zur Mittel- und Hochspannungsebene ist der kapazitive Anteil der Netzimpedanz im Niederspannungsnetz im Vergleich zum ohmsch-induktiven Anteil vernachlässigbar klein. Für die Effektivwerte der Stromharmonischen erhält man somit:

$$I_{N,n} = \frac{U_{N,n}}{Z_N} = \frac{U_{N,n}}{\sqrt{R_N^2 + (n\omega L_N)^2}} \tag{87.68}$$

Bei vorhandenen Stromoberschwingungen können die Oberschwingungen der Netzspannung $U_{N,n}$ für jede Harmonische ermittelt werden:

$$U_{N,n} = I_{N,n} \cdot (R_N + j\omega L_N) \tag{87.69}$$



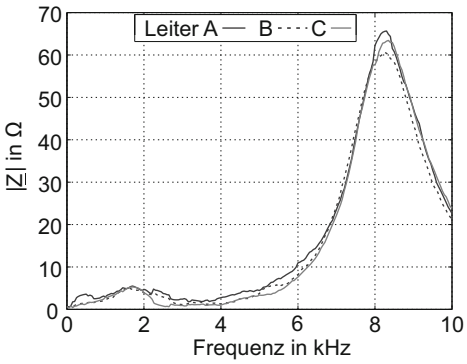


Abb. 87.68 Prototyp einer Messeinrichtung für die Mittelspannung bis 20 kV (*links*), gemessener frequenzabhängiger Impedanzverlauf (*rechts*)

In der VDE 0838 Teil 3, EN 61000-3-3 sind Bezugsimpedanzen angegeben, die bei 50 Hz nicht überschritten werden sollten:

Außenleiter-Null-Leiter

$$Z_{L,\mathrm{Null}} = (0{,}4 + j0{,}25)\,\Omega = 0{,}4\,\Omega + \omega \cdot 0{,}8\,\mathrm{mH} \qquad (87.70)$$

Außenleiter

$$Z_L = (0{,}24 + j0{,}15)\,\Omega = 0{,}24\,\Omega + \omega \cdot 0{,}478\,\mathrm{mH} \qquad (87.71)$$

Null-Leiter

$$Z_{\mathrm{Null}} = (0{,}16 + j0{,}1)\,\Omega = 0{,}16\,\Omega + \omega \cdot 0{,}318\,\mathrm{mH} \qquad (87.72)$$

Wegen $\underline{U}_N = \underline{I}_N \cdot \underline{Z}_N$ ist bei höherer Netzimpedanz und gleichem Oberschwingungsstrom eine höhere Oberschwingungsspannung zu erwarten. Deshalb wird oft die Netzimpedanz ermittelt, um Aussagen über den zu erwartenden Oberschwingungsgehalt der Spannung zu treffen. Bei Messungen im Versorgungsnetz muss die an den einzelnen Netzknoten unterschiedliche Impedanz des Netzes berücksichtigt werden. Dabei können auch die Frequenzen von Netzresonanzen festgestellt werden.

Einfluss von Netzresonanzen

Kapazitäten verursachen bei der Reihen- oder Parallelschaltung mit Induktivitäten Resonanzen im Netz. Das führt zu einer Impedanzverringerung bei der Reihenresonanz und Impedanzerhöhung bei der Parallelresonanz. Dadurch werden z. B. die Pegel einzelner Spannungsharmonischer überhöht, auch wenn die Ströme der zugehörigen Frequenz relativ gering sind. Liegt eine Reihenschaltung von Widerstand, Induktivität und Kapazität vor, wird die Impedanz nach (87.66) wie folgt ermittelt:

$$\underline{Z}_N = R_N + \underline{X}_L + \underline{X}_C = R_N + j\omega L_N + \frac{1}{j\omega C_N}$$

Bei der im Netz beim Verbraucheranschluss vorliegenden Parallelschaltung von Impedanzen wird die Admittanz Y als Kehrwert der Impedanz berechnet:

$$\underline{Y}_N = \frac{1}{R_N} + \frac{1}{\underline{X}_L} + \frac{1}{\underline{X}_C} = \frac{1}{R_N} - j\frac{1}{\omega L_N} + j\omega C_N \qquad (87.73)$$

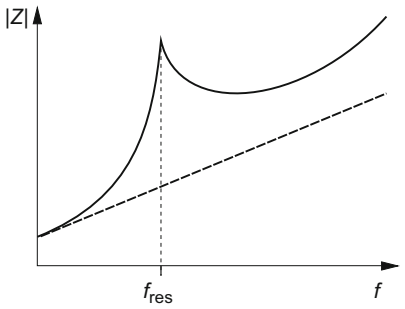


Abb. 87.69 Impedanzverläufe für eine ohmsch-induktive (*gestrichelt*) und ohmsch-induktiv-kapazitive Netzimpedanz

Resonanz tritt auf, wenn kapazitiver und induktiver Imaginärteil gleich groß und der resultierende Imaginärteil der Impedanz $\underline{Z}$ null ist:

$$j\omega L_N = j\omega C_N \qquad (87.74)$$

Mit $\omega = 2 \cdot \pi \cdot f$ erhält man für die Netzresonanzfrequenz $f_{\mathrm{res},N}$ mit der Netzinduktivität L_N und der Netzkapazität C_N für Reihen- und Parallelresonanz:

$$f_{\mathrm{res},N} = \frac{1}{2 \cdot \pi \sqrt{L_N \cdot C_N}} \qquad (87.75)$$

In Abb. 87.69 sind die Verläufe für eine ohmsch-induktive (gestrichelt) und ohmsch-induktiv-kapazitive Impedanz gezeigt. Enthält die Impedanz einen kapazitiven Anteil, kann sich wie in Abb. 87.69 gezeigt eine Netzresonanz nach (87.75) einstellen.

87.3.5.2 Flexible Erzeugung

Hierzu gehören sowohl die Flexibilisierung des konventionellen Kraftwerksparks als auch die virtuellen Kraftwerke. Konventionelle Kraftwerke werden durch eine schnellere Regelmöglichkeit flexibler einsetzbar, weil sie dann schneller den schwankenden Verbrauchskurven folgen können. Dies wird durch die

Auslegung von Kraftwerksblöcken mit im Vergleich geringerer Leistung möglich, z. B. 50 anstatt 500 MW. Bei kleineren Leistungen sind der Kessel sowie Rohrlängen und -durchmesser kleiner und weisen damit geringere thermische Zeitkonstanten auf. Neuauslegungen und Retrofit-Lösungen großer Kraftwerksblöcke mit Leistungen von 800 bis 1200 MW ermöglichen inzwischen auch eine bessere Regelfähigkeit.

Virtuelle Kraftwerke entstehen z. B. durch die Zusammenfassung mehrerer Kraftwerke mit unterschiedlichen Wandlungsprinzipien und/oder Energiespeicher. Das Ziel besteht darin, einen besser prognostizierbaren zeitlichen Leistungsverlauf realisieren zu können. Dafür bedarf es einer gemeinsamen Anlagensteuerung sowie der Einbindung der virtuellen Kraftwerke in die Netzführung des Verteil- und Übertragungsnetzes. Bei weiter abnehmendem konventionellem Kraftwerkspark müssen sich virtuelle Kraftwerke in steigendem Maße auch an der Netzregelung beteiligen.

87.3.5.3 Flexibler Verbrauch durch steuerbare Lasten

Die vier beschriebenen Problemstellungen nach Abb. 87.65 erfordern neben den Technologieansätzen einen Paradigmenwechsel für eine versorgungssichere elektrische Energieversorgung mit fluktuierenden Energien. Während bei einer konventionell strukturierten Energieversorgung die Energie bedarfsgerecht bereitgestellt wird, ist dies bei wetterabhängig fluktuierender Erzeugung nicht möglich. Hier wird das Energieangebot dann am besten genutzt, wenn der Verbrauch der Erzeugungscharakteristik folgt. Dies ist nicht für alle Verbraucher möglich, offensichtlich wird das bei der Beleuchtung oder abhängigen industriellen Produktionsprozessen. Daneben existiert eine Vielzahl unabhängiger Verbraucher in sog. unkritischen Prozessen, hierzu zählen Kühl- und Wärmeprozesse sowie viele Kreislaufprozesse mit großen Zeitkonstanten. Insbesondere der zukünftige Verbrauch fluktuierender Energie in zuschaltbaren industriellen Prozessen (Power to Gas, Power to Heat, Power to Process) besitzt ein hohes Ausbaupotenzial. Zur Umsetzung eines angebotsgesteuerten Verbrauchs müssen folgende Teilprobleme gelöst werden:

a) Bereitstellung von Informationen über das Energieangebot und den Preis
b) Zeitaufgelöste Erfassung des elektrischen Verbrauchs
c) Zentrale bzw. dezentrale Schalteinrichtungen für Verbraucher
d) Zusammenfassung (Pooling) von Verbrauchern
e) Zeitgenaue Zuordnung und Abrechnung von Energiemengen mit unterschiedlicher Wertigkeit

Diese Fragestellungen soll das zukünftige Smart Grid mit Hilfe von Smart Metern für die zeitaufgelöste Erfassung des elektrischen Verbrauchs lösen. Dafür sind hohe Investitionen in die Informations- und Kommunikationstechnik erforderlich. Darüber hinaus ergeben sich neue technische Herausforderungen für die Verarbeitung und Speicherung großer Datenmengen sowie neue Marktrollen mit hoher Verantwortung für den Datenschutz. Die breite gesellschaftliche Akzeptanz für die Steuerung elektrischer Verbraucher, insbesondere im Privatbereich, muss erst noch hergestellt werden.

87.3.5.4 Energiespeicherung

Es soll hier nur die Energiespeicherung im Energieversorgungsnetz betrachtet werden, d. h. die Speicherung von Leistungen im Megawatt-Bereich. Dafür stehen weiterhin nur wenige Technologien mit unterschiedlichen Reifegraden zur Verfügung, hierzu gehören die Pumpspeicherwerke, die Druckluftspeicherung und die Wasserstoffspeicherung.

Vor der detaillierten Technologiebetrachtung steht die Frage der notwendigen Speicherkapazität. Um den Einfluss der gesetzten Randbedingungen auf die Kapazität von Energiespeichern zu verdeutlichen, wird folgendes Beispiel genutzt. Bei einer angenommenen Vollversorgung mit regenerativen Energien in Deutschland besteht weiterhin die Aufgabe, die Jahreshöchstlast von 80 GW bereitzustellen. Damit dies möglich ist, müsste die Leistung unter der Annahme einer zweiwöchigen Windflaute über diesen Zeitraum gespeichert werden. Damit ergibt sich eine Speicherleistung von ca. 27 Terrawattstunden (1 Terrawattstunde entspricht 10 hoch 9 Kilowattstunden). Dies entspricht dem 450-fachen der heute in Deutschland verfügbaren Speicherenergie aus Pumpspeicherwerken von 60 GWh. In diesem Beispiel wurden mit dem kompletten Ausfall jeglicher Energieerzeugung bei Flauten bewusst ungünstige Randbedingungen gewählt. Die hiermit errechnete Erweiterung der Speicherkapazität wäre technisch und wirtschaftlich nicht realisierbar. Energiespeicher können offensichtlich nicht als alleiniger Lösungsansatz beim weiteren Ausbau der regenerativen Energie dienen. Eine zukünftige Energieversorgung darf zur Minimierung der Gefahr großflächiger Ausfälle nicht nur auf einer einzigen Wandlungstechnologie basieren und sollte schon allein aus Kostengründen auch nicht den zuvor errechneten hohen Speicherbedarf erzwingen.

Pumpspeicherwerke werden schon seit der ersten Installation im Jahre 1930 über viele Jahrzehnte zur Abdeckung von Lastspitzen eingesetzt und weisen im Vergleich aller Netzspeicher den höchsten technischen Reifegrad auf. Heutige Installationskosten liegen zwischen 800 und 2000 €/kW. In Deutschland sind Pumpspeicher mit einer Gesamtleistung von rund 7 GW installiert. Das neueste und größte Pumpspeicherwerk Goldisthal weist eine Leistung von 1060 MW und eine Speicherkapazität von 8480 MW h auf. Pumpspeicherwerke bestehen aus zwei Wasserbecken auf unterschiedlichen Höhenniveaus, zwischen denen sich ein Generator/Turbinensatz befindet. Bei einem Überangebot an Energie wird Wasser in das höher gelegene Becken gepumpt und dort als potenzielle Energie gespeichert. Diese kann bei Bedarf wieder in elektrische Energie gewandelt werden, indem das Wasser über die Turbine in das Unterbecken fließt. Die dabei auftretenden Verluste sind vergleichsweise gering, der Gesamtwirkungsgrad von Pumpspeicherwerken liegt bei großen Anlagen bei bis zu 80 %. Aufgrund der geringen Leistungsdichte von Wasser sind Pumpspeicherwerke jedoch nicht für die langfristige Speicherung großer Energiemengen geeignet, sie können z. B. nicht als saisonale Speicher zwischen Sommer und Winter genutzt werden.

Druckluftspeicher können in Verbindung mit Gasturbinen-Kraftwerken zum Ausgleich von Leistungsschwankungen eingesetzt werden. Dabei werden unterirdische Hohlräume, sog.

Kavernen, zur Speicherung von Druckluft genutzt. Die Installationskosten betragen 700 €/kW bis 800 €/kW ohne die Berücksichtigung der Kosten zur Erschließung von Kavernen. Die Energiedichte von Druckluft ist rund viermal höher als die von Wasser. In Deutschland existiert ein Druckluft-Kraftwerk in Huntorf bei Bremen, das eine Leistung von 290 MW über eine Zeit von zwei Stunden speichern kann. Bei Leistungsbedarf wird die gespeicherte Druckluft als Frischluft in eine mit Erdgas befeuerte Gasturbine geleitet und erhöht dabei deren Wirkungsgrad. Der Gesamtwirkungsgrad dieser Druckluftspeicherung beträgt 42 %, da die bei der Kompression entstehende Wärme verloren geht. Wird ein zusätzlicher Wärmespeicher eingesetzt, kann der Gesamtwirkungsgrad auf bis zu 70 % ansteigen. Aufgrund der vielfältigen Nutzungsmöglichkeiten von Kavernen zur Speicherung von Erdöl und Erdgas besteht eine starke Nutzungskonkurrenz, die den Zubau von Druckluftspeichern momentan behindert.

Wasserstoff-Speicher: Grundsätzlich kann Wasserstoff gasförmig in Kavernen oder Tanks bzw. flüssig in Tanks gespeichert werden. Er weist bei einem Druck von 70 MPa im Vergleich zu Wasser einen 1700-mal höhere und im Vergleich zu Druckluft eine 450-mal höhere Energiedichte auf. Deshalb ist die Speicherung von Wasserstoff eine Option auch für die saisonale Speicherung über mehrere Monate. Die Erzeugung von Wasserstoff kann über die klassische Niederdruck-Elektrolyse in großen Becken mit darin angeordneten Elektroden bei Wirkungsgraden von bis zu 70 % erfolgen oder auch mittels der Hochdruck-Elektrolyse in invers betriebenen Brennstoffzellensystemen bei Wirkungsgraden von maximal 40 %. Hierfür sind abhängig vom Reifegrad der Technologie und den Kosten der Gasverdichtung und -speicherung Installationskosten von 2000 €/kW bis 5000 €/kW erforderlich. Die Hochdruck-Elektrolyse weist ein besseres dynamisches Verhalten bei der Nutzung von stark fluktuierendem Windstrom für die Herstellung von Wasserstoff auf (Power to Gas), Niederdruckanlagen sind hierfür nicht einsetzbar. Zukünftige Brennstoffzellensysteme mit steuerbaren Membranen und verbessertem dynamischen Verhalten werden momentan erforscht.

Zur Speicherung kleinerer Mengen Wasserstoff kann das in seinem Betriebsdruck sehr variable Erdgasnetz genutzt werden, dem ohne Probleme bis zu einige Prozent Wasserstoff beigemischt werden können. Der Gesamtwirkungsgrad der Wasserstoffspeicherung sinkt nach der Elektrolyse noch durch die notwendige Verdichtung ab. Bei größeren Mengen ist zur Einspeisung in das Erdgasnetz eine weitere Reaktion erforderlich, die Methanisierung des Wasserstoffs, die den Wirkungsgrad nochmals um rund 30 % reduziert.

Neben den drei beschriebenen Speichertypen werden wegen ihres technologischen Fortschritts auch **Redox-Flow-Speicher** mit Leistungen bis zu einigen hundert Kilowatt immer interessanter für den Einsatz als Stundenspeicher in der elektrischen Energieversorgung. Sie bestehen aus zwei Tanks, zwischen denen eine Elektrolyt-Flüssigkeit hin und her gepumpt wird. Dabei läuft in einer Fließrichtung eine Oxidation mit Energieabgabe und in der anderen eine Reduktion mit Energieaufnahme ab. Der Wirkungsgrad kann bis zu 75 % betragen. Aufgrund ihrer kompakten Gestaltung und guten Skalierbarkeit sind Redox-Flow-Speicher auch grundsätzlich als Energiespeicher innerhalb von Stadtgebieten einsetzbar. Die Erforschung und Entwicklung chemisch unbedenklicher Elektrolytflüssigkeiten könnte solche Anwendungen zukünftig weiter beschleunigen.

Tab. 87.8 Kosten von Langzeitspeichern mit wöchentlichem Ladezyklus, DLS-Druckluftspeicher, PSW-Pumpspeicherwerk, Datenquelle VDE-Studie 2009

Speichertechnologie	Heutige Kosten in ct/kW h	Zukünftige Kosten in ct/kW h
Wasserstoff	23	9 (in > 10 Jahren)
DLS	38	22 (in > 10 Jahren)
PSW	3 bis 11	Nicht berücksichtigt

Tab. 87.9 Kosten von Langzeitspeichern mit stündlichem Ladezyklus, DLS-Druckluftspeicher, PSW-Pumpspeicherwerk, Datenquelle VDE.-Studie 2009

Speichertechnologie	Heutige Kosten in ct/kW h	Zukünftige Kosten in ct/kW h
Wasserstoff	24	11 (in > 10 Jahren)
DLS	6	3 (in > 10 Jahren)
PSW	2,5 bis 5,5	Nicht berücksichtigt

Die Wirtschaftlichkeit von Energiespeichern wird von den Investitions- und Betriebskosten, der Nutzungsdauer und -häufigkeit sowie primär von der Preisdifferenz aus Energiebezug und -lieferung bestimmt. In der Tab. 87.8 sind die Kosten von Langzeitspeichern mit wöchentlichem Ladezyklus gezeigt. Es wird deutlich, dass noch auf längere Sicht sowohl die Wasserstoffspeicherung als auch die Druckluftspeicherung weit höhere Kosten als die etablierte Pumpspeicherung aufweisen werden. Dies gilt auch bei stündlichen Ladezyklen, siehe Tab. 87.9. Da für Pumpspeicher geeignete geologische Randbedingungen vorhanden sein müssen, ist deren weiterer Ausbau zumindest in der konventionellen Ausführung stark begrenzt. Als Alternative dazu sind jedoch Pumpspeicher in künstlichen Strukturen wie z. B. in ehemaligen Tagebauen des Kohle- oder Kalksteinabbaus denkbar, um auch zukünftig größere Speicherkapazitäten zubauen zu können.

Für den Betrieb von Energiespeichern ergeben sich momentan viele Unsicherheiten. Dazu gehört die schwer abzuschätzende Entwicklung der Preisspanne für Energiemengen zur Ein- und Ausspeicherung. Wegen ihrer hohen Investitionskosten und den daraus folgenden langen Amortisationszeiten benötigen Speicher jedoch langfristig kalkulierbare Einnahmen, um sich verlässlich amortisieren zu können. Es ergibt sich zudem die Notwendigkeit, heute mit dem Speicherbetrieb verbundene Kosten wie z. B. Netznutzungsgebühren neu zu bewerten. Die Marktintegration von Energiespeichern ist momentan nicht gegeben, weshalb Speicher nicht wirtschaftlich betrieben werden können. Dazu gehört beispielsweise auch die zukünftige Definition der Verantwortlichkeiten beim Betrieb der Speicher.

Der Begriff der Energiewende beinhaltet neben der Stromversorgung auch die Wärmebereitstellung und die Treibstoffe für den Verkehr. In den letzteren zwei Sektoren wirken jedoch im Vergleich zur Stromerzeugung größere Zeitkonstanten, es besteht keine Notwendigkeit für zeitkritische Steuerungseingrif-

Tab. 87.10 Durchschnittliche Benutzungsstundenzahlen und gesicherte Leistung von Erzeugungstechnologien

Kraftwerksart	Stromgestehungskosten in ct/kW h_{el}[a]	CO_2-Intensität in g/kW h_{el}[a]	CO_2-Äquivalente in g/kW h_{el}[a]	Vollbenutzungsstunden pro Jahr[c]	Gesicherte Leistung in %[d]
Import-Steinkohle-Kraftwerk	4,0 … 5,0	897	949	6000 … 7000	93
Import-Steinkohle-Heizkraftwerk	2,5 … 3,5	508	622	6000 … 7000	93
Braunkohle-Kraftwerk	4,0 … 5,0	1142	1153	6000 … 7000	93
Braunkohle-Heizkraftwerk	2,5 … 3,5	703	729	6000 … 7000	93
Erdgas-GuD-Kraftwerk	4,0 … 5,0	398	428	6000 … 7000	93
Erdgas-GuD-Heizkraftwerk	3,5 … 4,5	116	148	6000 … 7000	93
Erdgas-Blockheizkraftwerk	7,0 … 8,0	5	49	6000 … 7000	93
Kernkraftwerk	4,5 … 5,5	31 … 61	32 … 65	6000 … 7000	93
Windpark onshore	6,5 … 8,0[b]	23	24	1800 … 2400	20
Windpark offshore	10,5 … 16,5[b]	22	23	3500 … 4500	40
Photovoltaik	14,0 … 20,5[b]	89	101	800 … 900	11
Solarthermisches Kraftwerk	19,0 … 23,0[b]	25	27	2000	30[e]
Wasserkraftwerk	5,0 … 10,0	39	40	4000 … 6000	50–80[f]
Biogas-Blockheizkraftwerk	6,0 … 8,0	-414	-409	6000 … 7000	93

[a] Quelle: Öko-Institut Darmstadt: Arbeitspapier Treibhausemissionen und Vermeidungskosten der nuklearen, fossilen und erneuerbaren Strombereitstellung, 2007
[b] Fraunhofer ISE: Studie Stromgestehungskosten Erneuerbare Energien, 2012
[c] Technisch mögliche Stundenzahl, die reale Nutzung kann davon marktabhängig stark abweichen
[d] Werte teilweise geschätzt
[e] Bei vorhandenem Energiespeicher
[f] Abhängig von der Wasserzuführung

fe wie bei der Netzregelung im elektrischen Versorgungsnetz. Dies ist der Grund dafür, dass viele technische Verbesserungsmaßnahmen zuerst im Bereich der Stromerzeugung eingesetzt werden. Jedoch wird stets die Kopplung zwischen den Systemen Strom, Wärme und Gas mit dem Ziel einer optimierten gesamten Energieversorgung betrachtet. Dies trifft auch auf die Energiespeicherung zu, hier wird die systemübergreifende Speicherung von Strom zu Wasserstoff (Power to Gas), Strom zu Wärme (Power to Heat) bzw. Strom zu prozessbezogener Speicherung (z. B. Power to Steel) genutzt. Durch die Steigerung der Energieeffizienz in allen Bereichen kann der Verbrauch gesenkt werden, wodurch weniger Kosten für die notwendigen technischen Anpassungsmaßnahmen in der Energieversorgung entstehen.

87.3.5.5 Versorgungssicherheit

Es ist erforderlich, die Frequenz innerhalb der vorgegebenen Grenzwerte zu regeln, wozu ein Gleichgewicht von erzeugter und verbrauchter Leistung eingestellt werden muss, siehe Abschn. 87.3.3. Dazu ist eine ausreichende Kraftwerksleistung notwendig, siehe Abb. 87.8 im Abschn. 87.2. Daneben muss sichergestellt werden, dass die erzeugte elektrische Energie zuverlässig die Verbraucher erreicht und Ausfälle vermieden werden. Dazu wird in der elektrischen Energieversorgung in Europa grundsätzlich das $(n-1)$-Kriterium für eine erhöhte funktionale Versorgungssicherheit angewendet. Die bedeutet, dass der Ausfall eines Betriebsmittels, z. B. einer Freileitung bei der Energieübertragung, noch nicht zum Ausfall der Versorgungsspannung führen darf. Für die Frequenzregelung sind in Deutschland die Übertragungsnetzbetreiber in den vier Regelzonen zuständig, siehe Abb. 87.45 rechts im Abschn. 87.3.2.1. Zur weiteren Erhöhung der Frequenzstabilität existieren Kuppelstellen zwischen den Regelzonen zum Leistungsaustausch. Dies gilt auch für die Kopplung der Ländernetze untereinander.

Für die Energieerzeugung ist sichergestellt, dass der Ausfall eines oder auch mehrerer Kraftwerke durch Regelvorgänge bei den verbleibenden Kraftwerken kompensiert werden kann. Bei der Sicherstellung der Versorgungsleistung spielen die Volllast-Benutzungsstundenzahlen der Kraftwerksarten eine entscheidende Rolle, siehe Tab. 87.10. Da nicht alle Kraftwerksarten in ihrer Leistung unabhängig steuerbar sind, wird zur Bewertung des jeweiligen Beitrags zur Sicherstellung der notwendigen Erzeugung die gesicherte Leistung verwendet. Diese beschreibt die zuverlässig verfügbare Leistung der Erzeugungsart. Naturgemäß ist dieser Leistungsanteil bei wetterabhängig fluktuierenden Energien geringer als bei kontrolliert durch Feuerung betriebenen Kraftwerksarten.

Dezentrale regenerative Erzeugungsanlagen übernehmen mit ihrem weiteren Zubau immer weitergehende Aufgaben zur Sicherstellung der Systemsicherheit. Dazu gehören das Durchfahren von Spannungseinbrüchen im Fehlerfall, die statische und dynamische Stützung der Netzfrequenz sowie die Beteiligung an der Spannungsregelung. Ein bisher nicht endgültig untersuchtes Problem beim Übergang zu einer überwiegend regenerativ versorgten Energieversorgung ist die Stabilität der Netzregelung ohne Großkraftwerke mit rotierenden Massen.

Trotz des in dieser Form noch nirgendwo vollzogenen Umbaus der deutschen Energieversorgung von einer auf verbrauchernahen Großkraftwerken basierenden Erzeugung zu einer klimafreundlichen regenerativen und damit wetterabhängigen fluktuierenden Erzeugung, d. h. der Energiewende, sind die Ausfallzeiten im europäischen Vergleich gering, siehe Abb. 87.70. Dies zeigt, dass diese Form einer umfassenden Transformation der Energieversorgung bisher ohne eine Verringerung der Ver-

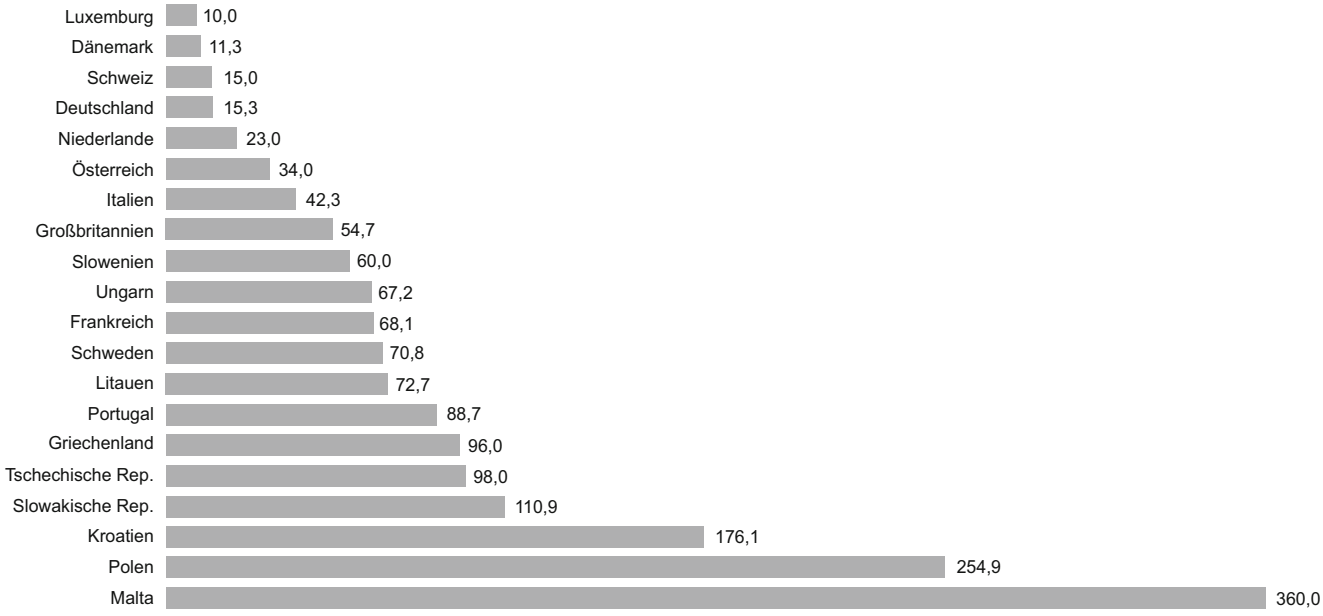


Abb. 87.70 Ländervergleich der durchschnittlichen Unterbrechung der Stromversorgung je Kunde in Minuten im Jahr 2013, nach CEER Benchmarking Report Update 2015

sorgungssicherheit umgesetzt werden konnte. Allerdings hat die Zahl der kurzfristigen Eingriffe der Übertragungsnetzbetreiber in den Kraftwerkseinsatz zur Vermeidung von Netzengpässen, das sog. Redispatch, in den letzten Jahren stark zugenommen. Durch solche Maßnahmen entstehen hohe Kosten, z. B. 132,6 Mio. € für Redispatch-Eingriffe an 232 Tagen des Jahres 2013 mit einem Gesamtvolumen von 4390 GWh.

87.4 Elektrische Betriebsmittel

87.4.1 Freileitungen

Zur Energieübertragung über große Entfernungen werden Freileitungen genutzt. Hierbei sollte zur verlustarmen Übertragung die Spannung in kV immer höher als die Entfernung in km sein. In Europa werden Spannungen ab der Mittelspannung und darüber hinaus in den genormten Spannungsebenen 110, 220 und 380 kV mit Freileitungen übertragen, außerhalb Europas werden bei größeren Entfernungen auch Spannungen von 500, 735, 750, 1000 und 1500 kV genutzt.

87.4.1.1 Aufbau

Freileitungen bestehen aus den Masten mit den zugehörigen Fundamenten und Erdungen, den an den Traversen aufgehängten Isolator-Garnituren mit Isolatoren aus Glas, Porzellan oder Kunststoff, einem oder mehreren Erdseilen an der Mastspitze, sowie den Leiterseilen. Die drei Leiterseile eines Drehstromsystems werden als Leitersystem bezeichnet. An der Mastspitze können ein oder zwei Erdseile angeordnet werden. Der grundsätzliche Aufbau einer Freileitung ist in Abb. 87.71 gezeigt.

87.4.1.2 Masten

Die Mastformen sind abhängig von der Spannungsebene und der Anzahl der geführten Systeme, siehe Abb. 87.72. Es wird

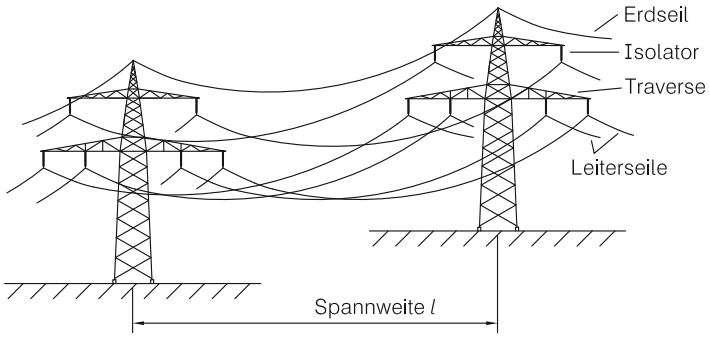


Abb. 87.71 Aufbau einer Freileitung

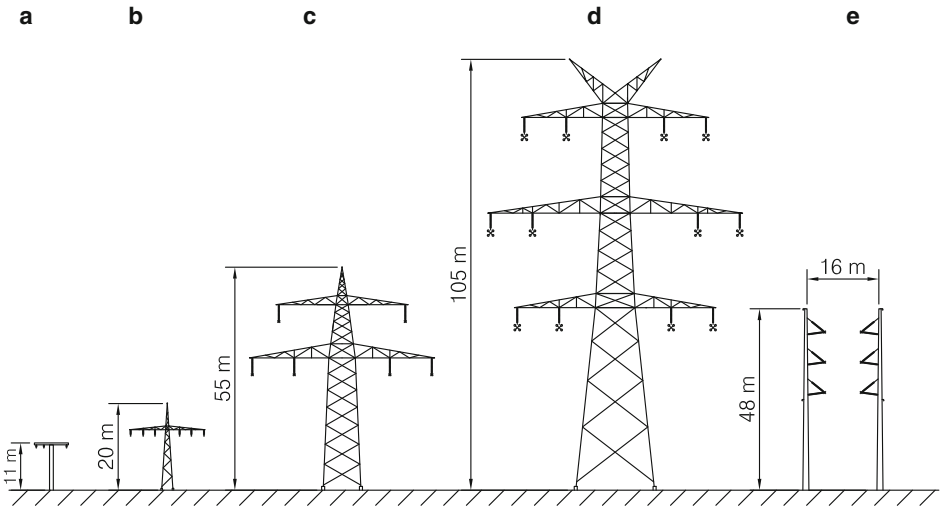


Abb. 87.72 Freileitungsmasten. **a** Betonmast für 20 kV Mittelspannung, **b** Einebenenmast für zwei Systeme 110 (Bildmaße) bis 380 kV, **c** Donaumast für zwei Systeme 110 bis 380 kV (Bildmaße), **d** Donaumast für vier Systeme und zwei Erdseile 110 bis 380 **kV** (Bildmaße), **e** Wintrack-Mastsystem als zukünftige 380-kV-Sondermasten für Windstromtrassen

Tab. 87.11 Typische Spannweiten von Freileitungsmasten

Nennspannung U_n in kV	Mittlere Spannweite l in m	Mastart
Bis 1	60 … 80	Stahlbeton
10 … 30	100 … 160	Stahlbeton
10 … 30	… 200	Gitter
110	250 … 350	Gitter
220	350 … 450	Gitter
380	400 … 500	Gitter

unterschieden zwischen Horizontalmasten, auch Einebenenmasten genannt, siehe Abb. 87.72a und Abb. 87.72b, Tonnenmasten mit einer Leitung pro Traversenseite (hier nicht gezeigt), Donaumasten mit mehreren Leitern pro Traversenhälfte, siehe Abb. 87.72c und Abb. 87.72d. Insbesondere für die Trassenführung zwischen großen Windparks und Umspannstationen sollen zur Erzielung einer höheren Akzeptanz Masten mit reduzierter Höhe und Trassenbreite eingesetzt werden, siehe Abb. 87.72e. Der geringere Mastabstand ergibt sich durch isolierende Traversen, mit der Nutzung von Stützisolatoren können die Seilkräfte besser beherrscht und damit der Mastabstand reduziert werden. Zur geraden Leitungsführung werden Tragmasten genutzt, bei denen die Leiterseile mit Tragklemmen unterhalb der Isolatoren fixiert werden. Wirken zusätzliche Horizontalkräfte auf die Leiterseile, wie bei der Seilführung vom Mast in ein Umspannwerk oder bei der Richtungsänderung von Seilführungen, werden Abspannmasten genutzt, bei denen die Leiterseile durch spezielle Keilabspannklemmen gehalten werden.

Mit Endmasten erfolgt die Zuführung der Leiterseile in Schaltanlagen, hierbei müssen hohe einseitige Kräfte aufgenommen werden. Der Abstand, d. h. die Spannweite l von Freileitungsmasten wird unter Beachtung des Durchhangs nach Abschn. 87.4.1.4.1 nach wirtschaftlichen Gesichtspunkten gewählt, siehe Tab. 87.11.

87.4.1.3 Isolatoren

Isolatoren können stehend in Stützen- oder hängend in Ketten- oder Langstabform ausgeführt werden, siehe Abb. 87.73. Bisher werden Stützenisolatoren aus Porzellan oder Glas nur in Spannungsbereichen bis 52 kV eingesetzt. Zukünftig werden sie jedoch auch für die Wintrack-Mastsysteme im Spannungsbereich bis 380 kV in Form von isolierenden Traversen vorgesehen, da aufgrund ihrer höheren Kraftaufnahme ein geringer Durchhang der Leiterseile und damit geringere Masthöhen realisierbar sind, siehe Abb. 87.72e. Ab dem Hochspannungsbereich werden meist Isolatorketten aus Glaskappen, oder Langstabisolatoren aus Keramik (Hartporzellan oder Steatit) eingesetzt. Darüber hinaus kommen auch Verbundisolatoren mit einem Kern aus glasfaserverstärktem Kunststoff (GFK) mit aufgespritzter Silikonschirmung zum Einsatz. Um die Sicherheit der Aufhängung zu erhöhen, werden Hängeisolatoren oft parallel ausgeführt, insbesondere bei der Trassenquerung von Straßen und Siedlungsgebieten. Bei höheren Spannungsebenen können entweder wie in Abb. 87.73 gezeigt Isolatorketten aus genormten Standard-Isolatoren oder auch durchgehende Langstabisolatoren eingesetzt werden.

87.4.1.4 Leiterseile

Leiterseile bestehen aus verseilten Drähten, siehe Abb. 87.74. Die einzelnen Lagen der Leiterseile werden vorrangig aus mechanischen Gründen mit entgegengesetztem Seilschlag ausgeführt. Als Leitermaterial wird im Nieder- und Mittelspannungsbereich Kupfer, Aldrey, das ist eine Legierung aus 99 % Aluminium und je 0,5 % Magnesium und Silizium, oder Aluminium verwendet. Im Hochspannungsbereich werden Einfach-Leiterseile oder Verbund-Leiterseile mit einer Stahlseele und Aluminium-Leitermaterial genutzt, siehe Abb. 87.74a, b. Aluminium weist bei gleichem Gewicht einen größeren Querschnitt und damit einen höheren Leitwert als Kupfer auf und ist deshalb wesentlich kostengünstiger, weshalb es bevorzugt

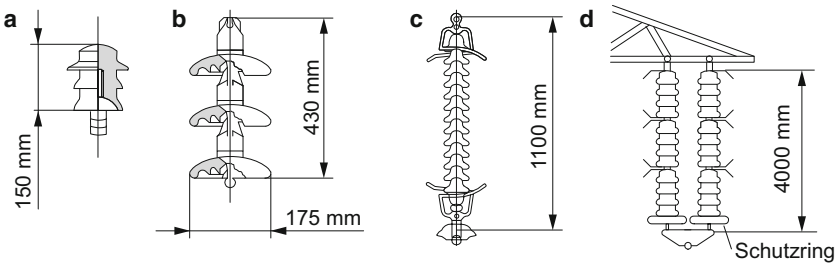


Abb. 87.73 Isolatoren. **a** Stützisolator, **b** Glaskappenisolator, **c** Langstabisolator, *rechts*: Doppel-Isolatorenkette einer 380-kV-Freileitung

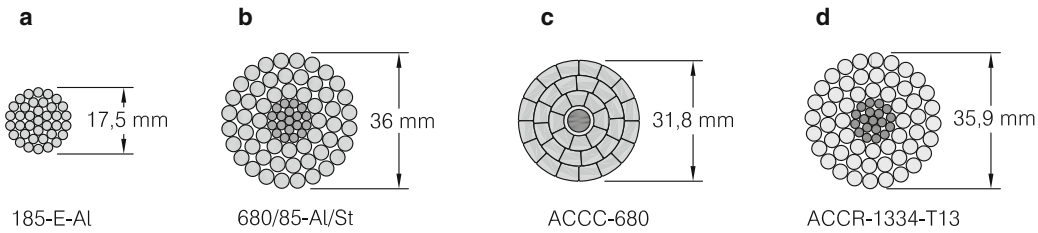


Abb. 87.74 **a** Einfach-Leiterseil, Dauerstrombelastbarkeit 520 A bei 80 °C, **b** Leiterseil mit Stahlseele, Dauerstrombelastbarkeit 1150 A bei 80 °C, **c** Hochtemperatur-Leiterseil mit sektorförmigen Leitern mit Kohlefaser-Verbundkern, Dauerstrombelastbarkeit 2069 A bei 175 °C, **d** Hochtemperatur-Leiterseil mit Keramikfaser-Verbundkern, Dauerstrombelastbarkeit 2327 A bei 210 °C

eingesetzt wird. Die Stahlseele bei Verbund-Leiterseilen erhöht bei größeren Querschnitten die mechanische Festigkeit. Verbund-Leiterseile werden nach DIN EN 50182 nach ihren Materialquerschnitten bezeichnet, z. B. 184-AL1/30-ST1A (früher 185/30-AL/St), womit der Querschnitt des Aluminium-Leitermaterials von 184 mm^2 und der Stahlseele von 30 mm^2 beschrieben wird. Standard-Leiterseile sind im Normalbetrieb bis zu einer Grenztemperatur von 80 °C belastbar, im Kurzschlussfall können nach DIN VDE 0103 kurzzeitig viel höhere Temperaturen von bis zu 200 °C auftreten. Sollen ohne eine Änderung des Leiterquerschnitts auf vorhandenen Trassen kurzzeitig höhere Leistungen übertragen werden, z. B. bei der Netzeinspeisung aus Windparks, können Hochtemperatur-Leiterseile eingesetzt werden. Hierfür stehen Leiterseile mit sektorförmigen Leitern und Kunststoff-Verbundseele für eine Grenztemperatur im Normalbetrieb von 175 °C zur Verfügung, siehe Abb. 87.74c. Leiterseile mit Keramikfaser-Verbundkern sind sogar bis 210 °C belastbar, siehe Abb. 87.74d. Hochtemperatur-Leiterseile werden im Vergleich zu Standardseilen teilweise abweichend bezeichnet, hierbei sind die Querschnittsflächen nicht immer wie in Abb. 87.74c aus der Bezeichnung abzuleiten. Beispielsweise weist das in Abb. 87.74d gezeigte Leiterseil ACCR-1334-T13 einen Querschnitt des Aluminium-Leitermaterials von 676 mm^2 und der Stahlseele von 86 mm^2 auf.

Bei stromabhängig quadratisch ansteigenden Verlusten kann mit Hochtemperatur-Leiterseilen bis zur doppelten Leistung übertragen werden. Zu beachten ist, dass eine Umrüstung von Freileitungen auf Hochtemperatur-Leiterseile weitere Umrüstungen von Seilaufhängungen, Isolatoren und Transformatoren erfordert und deshalb sehr kostenintensiv ist. In Deutschland existieren mehrere Teststrecken zum Betrieb von Hochtemperatur-Leiterseilen.

Bei Freileitungen werden die Induktivität und Kapazität durch die Anordnung der Leiter untereinander und zum Erdpotenzial bestimmt. Zur Vergleichmäßigung der Betriebswerte werden Freileitungen verdrillt, d. h. dass Ihre Lage zueinander an sog. Verdrillmasten so verändert wird, dass jede Leitung über der Gesamtlänge gesehen die gleiche Strecke in jeder Position war. Würden die Leiterseile über die gesamte Übertragungsstrecke an der gleichen Position bleiben, wären die Betriebswerte der Induktivität und Kapazität für jeden Leiter unterschiedlich und es würden Unsymmetrien in der Spannung auftreten. Ab Spannungen von 220 kV werden oft Bündelleiter verwendet, da hierbei an den Leitern geringere elektrische Randfeldstärken unterhalb von 17kV/cm auftreten und die verlustbehafteten Koronaeffekte, die bei Dunkelheit und feuchter Luft als Glimmen an den Freileitungen sichtbar sind, vermieden werden können. Bei 220-kV-Systemen werden meist Zweier-Bündel, bei 380 kV Dreier- oder Viererbündel eingesetzt, siehe Abb. 87.75. Zur Berechnung der Betriebswerte von Bündelleitern kann ein Ersatzradius r_{ers} errechnet werden, der abhängig von der Anzahl der Bündelleiter n, vom Leiterradius r, und dem Radius des Bündelleiters r_B ist und allgemein beschrieben wird mit:

$$r_{\mathrm{ers}} = \sqrt[n]{n \cdot r \cdot r_B^{n-1}} \qquad (87.76)$$

Während die Leiterseile zur Leistungsübertragung dienen, wird das Erdseil zum Schutz der Leiter gegen Blitzeinschlag an der Mastspitze befestigt und besitzt somit Erdpotenzial. Das Erdseil bildet einen Blitzschutzraum, der nach beiden Seiten als Kreisabschnitt mit dem Radius der doppelten Erdseilhöhe bis zum Erdboden abfällt. Wenn bei breiten Traversen und mehreren Systemen ein höherer Blitzschutz erforderlich ist, können sich auch zwei Erdseile an der Mastspitze befinden, siehe Abb. 87.72d. Erdseile werden meist erst oberhalb der Mittelspannung verwendet. Oft enthalten sie zusätzliche Licht-

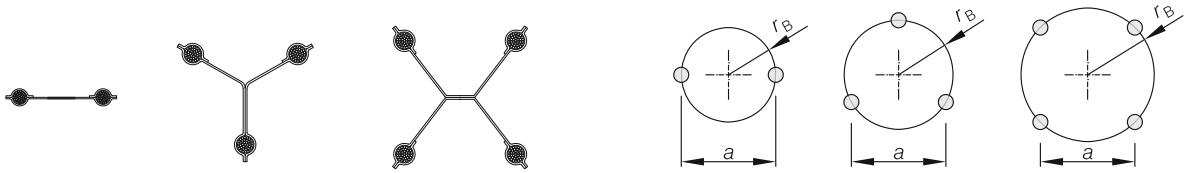


Abb. 87.75 Ausführung von Leiterbündeln mit Abstandhaltern als Zweier-, Dreier- und Viererbündel, Abstand a und Ersatzradien r_B

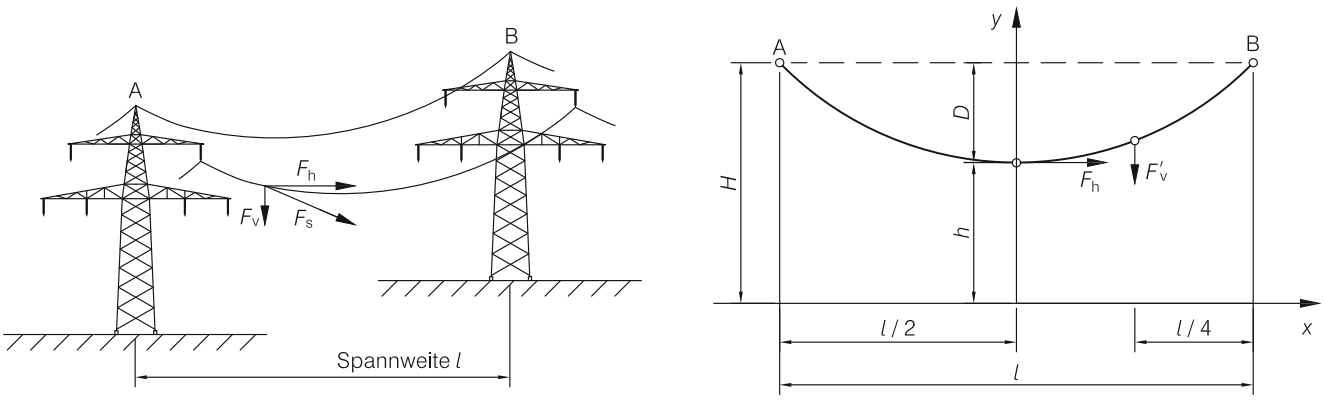


Abb. 87.76 Darstellung des Seildurchhangs

Tab. 87.12 Mindestabstände von Leiterseilen

Betriebsspannung U_b in kV	Mindestabstände d in m	
	senkrecht	waagerecht
Bis 123	6,00	3,00
Bis 245	6,75	3,75
Bis 420	7,80	4,80

wellenleiter zur Datenübertragung, die in einer Röhre innerhalb der Stahlseele geführt werden. Zur Dämpfung von meist windbedingten Schwingungen werden neben der Seilaufhängung Schwingungsdämpfer vorgesehen. Abhängig von der Spannungsebene ergeben sich die Mindestabstände der Leiterseile zur Erdoberfläche und gegenüber den Masten, siehe Tab. 87.12.

87.4.1.4.1 Seildurchhang

Auf das Leiterseil wirkt die in Abb. 87.76 links eingezeichnete Seilzugkraft F_S, die sich aus der Gewichts- bzw. Vertikalkraft F_V und der Zug- bzw. Horizontalkraft F_H errechnet:

$$F_S = \sqrt{F_H^2 + F_V^2} \tag{87.77}$$

Die mittlere Seilzugkraft errechnet sich aus der Horizontalkraft F_H, der Seillänge L und der Spannweite l:

$$\overline{F}_S = F_H \frac{L}{l} \tag{87.78}$$

Der Seildurchhang kann ermittelt werden, indem man eine sog. Kettenlinie betrachtet, bei der sich an jeder Stelle ein Momentengleichgewicht aus den Produkten von Horizontalkraft F_H und Durchhang D bzw. von Vertikalkraft F_V und Spannweite l einstellt. Für die Seilmitte würde dieses Gleichgewicht zwischen Horizontalmoment M_H und Vertikalmoment M_V lauten:

$$M_H = M_V = D \cdot F_H = F_V \frac{l}{2} \tag{87.79}$$

Bei der Spannweite $l/4$ stellt sich gemäß Abb. 87.76 rechts ein Vertikalmoment abhängig von der Vertikalkraft F'_V pro Längeneinheit dL des Seils ein:

$$M_H = M_V = D \cdot F_H = F'_V \cdot \frac{l}{2} \cdot \frac{l}{4} \tag{87.80}$$

Allgemein lässt sich für den Seildurchhang die Kettenlinie mit dem Cosinus hyperbolicus beschreiben, wobei der Parameter h wie in Abb. 87.76 rechts gezeigt der tiefste Punkt der Kettenlinie ist:

$$y = k(x) = h \cdot \cosh(b \cdot x) \tag{87.81}$$

Die Länge des Leiterseils L lässt sich als Bogenlänge l des Graphen von f über die in Abb. 87.76 gezeigten Punkte A und B errechnen:

$$l = \int_B^A \sqrt{1 + (f'(x)^2)} dx \tag{87.82}$$

87.4.1.4.2 Ersatzschaltbild von Freileitungen

Symmetrisch aufgebaute Freileitungssysteme werden mit dem Π- oder T-Ersatzschaltbild abgebildet, siehe Abb. 87.77. Als Elemente des Ersatzschaltbilds werden längenbezogene Größen, sog. Leitungsbeläge für die Kapazität C, die Induktivität

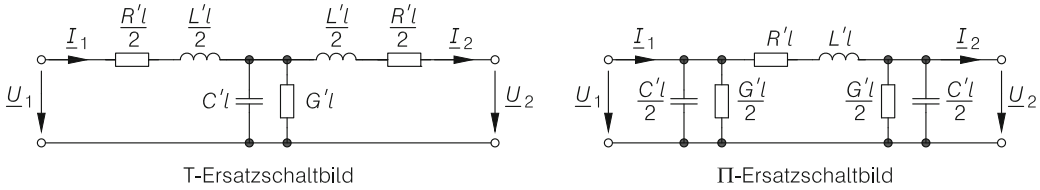


Abb. 87.77 Ersatzschaltbilder von Freileitungen

L, den Wirkwiderstand R und den Wirkleitwert G verwendet. Der Zahlenwert der Größen ergibt sich, wenn der jeweilige Leitungsbelag mit der Länge multipliziert wird, z. B. für die Kapazität C:

$$C = C' \cdot l \tag{87.83}$$

Während beim Π- Ersatzschaltbild das Ein- und Ausgangsverhalten allein durch die Leitungskapazitäten abgebildet wird und es damit gut für hohe Frequenzen geeignet ist, werden beim T-Ersatzschaltbild die Induktivitäten besser berücksichtigt. Im Bereich der Netzfrequenz sind beide Ersatzschaltbilder gleichwertig, sie sind grundsätzlich nur weit unter ihrer Eigenfrequenz gültig, im Bereich der Energieübertragung ist dies der Frequenzbereich bis ca. 1 kHz. Die Eigenfrequenz der Leitung Ω_L errechnet sich zu:

$$\Omega_L = 2\pi \cdot f_L = \frac{1}{\sqrt{L \cdot C/2}} \tag{87.84}$$

87.4.1.4.3 Wirkwiderstand von Leiterseilen

Der ohmsche Widerstand R_L (Resistanz-Realteil des komplexen Widerstands $\underline{Z}$) von Freileitungsseilen kann aufgrund der verlängerten Leiterführung durch den Seilschlag und Wirbelstromeffekten bei Verbundseilen nicht wie sonst bei Leitern üblich aus dem spezifischen Widerstand, der Länge und dem Querschnitt des Leiters errechnet werden, sondern ist den Tabellenwerten der Hersteller zu entnehmen, siehe Tab. 87.13, 87.14 und 87.15. Hierbei werden längenbezogene Werte R' in Ω/km angegeben, der Widerstandswert R kann daraus bei bekannter Länge l ermittelt werden:

$$R = R' \cdot l \tag{87.85}$$

Bei Gleich- und Wechselstromleitungen ist die Länge l gleich der doppelten Übertragungsstrecke und entspricht somit der Hin-und Rückleitung, im symmetrischen Drehstromnetz ist l gleich der einfachen Übertragungsstrecke anzusetzen. Da die Tabellenwerte für eine Umgebungstemperatur von 20 °C angegeben werden, muss der Widerstandswert bei abweichender Temperatur rechnerisch ermittelt werden, für den Temperaturbeiwert kann $\alpha \approx 0{,}004\,\mathrm{K}^{-1}$ angenommen werden:

$$R_\vartheta = R_{20}[1 + \alpha(\vartheta - 20\,^\circ\mathrm{C})] \tag{87.86}$$

Bei Bündelleitern errechnet sich der Betriebswirkwiderstand $R'_{b,B}$, indem durch die Anzahl der Teilleiter n dividiert wird:

$$R'_{b,B} = \frac{R'_L}{n} \tag{87.87}$$

Der Betriebsableitwert G'_b beschreibt die ohmsche Kopplung der Leiterseile und nimmt Werte im Bereich von 3 nS/km an.

Tab. 87.13 Dauerstrombelastbarkeit, Wirkwiderstände, Seilbruchkraft und Querschnitte von Kupferfreileitungsseilen (hart, blank) nach DIN EN 48201/1

Sollquerschnitt A in mm²	Nennquerschnitt A in mm²	Anzahl der Drähte	Draht-Ø D in mm	Seil-Ø D in mm	Seilgewicht in kg/km	Rechn. Seilbruchkraft F in kN	Rechn. Seilwiderstand R'_L bei 20 °C in Ω/km	Dauerstrombelastbarkeit I_d in A
10	10,02	7	1,35	4,10	90	4,02	1,8051	90
16	15,89	7	1,70	5,10	143	6,37	1,1384	125
25	24,25	7	2,10	6,30	218	9,72	0,7460	160
35	34,36	7	2,50	7,50	310	13,77	0,5264	200
50	49,48	7	3,00	9,00	446	19,84	0,3656	250
50	48,35	19	1,80	9,00	437	19,38	0,3759	250
70	65,81	19	2,10	10,50	596	26,38	0,2762	310
95	93,27	19	2,50	12,50	845	37,39	0,1949	380
120	116,99	19	2,80	14,00	1060	46,90	0,1554	440
150	147,11	37	2,25	15,80	1337	58,98	0,1238	510
185	181,62	37	2,50	17,50	1649	72,81	0,1003	585
240	242,54	61	2,25	20,30	2209	97,23	0,7527	700
300	299,43	61	2,50	22,50	2725	120,04	0,0610	800
400	400,14	61	2,89	26,00	3640	160,42	0,0456	960
500	499,83	61	3,23	29,10	4545	200,38	0,0365	1110
625	626,20	91	2,96	32,60		251,04	0,0292	

Tab. 87.14 Dauerstrombelastbarkeit, Wirkwiderstände, Seilbruchkraft und Querschnitte von Aluminium- Freileitungsseilen nach DIN 48201, EN 50182

Sollquerschnitt	Bezeichnung alt	Nennquerschnitt A in mm^2	Anzahl der Drähte	Draht-Ø D in mm	Leiter Seil-Ø D in mm	Seilgewicht m in kg/km	Rechn. Bruchkraft F in kN	Gleichstromwiderstand R'_L bei 20 °C	Dauerstrombelastbarkeit I_d in A
16-AL1	16	15,9	7	1,70	5,10	43,4	3,02	1,7986	110
24-AL1	25	24,2	7	2,10	6,3	66	4,17	1,181	145
34-AL1	35	34,4	7	2,50	7,5	94	5,78	0,833	180
49-AL1	50	49,5	7	3,00	9,0	135	7,94	0,579	225
48-AL1	50	48,3	19	1,80	9,0	133	8,45	0,595	225
66-AL1	70	65,8	19	2,10	10,5	181	11,32	0,437	270
93-AL1	95	93,3	19	2,50	12,5	256	15,68	0,308	340
117-AL1	120	117,0	19	2,80	14,0	322	18,78	0,246	390
147-AL1	150	147,1	37	2,25	15,8	406	25,30	0,196	455
182-AL1	185	181,6	37	2,50	17,5	500	30,54	0,159	520
243-AL1	240	242,5	61	2,25	20,3	670	39,51	0,119	625
299-AL1	300	299,4	61	2,50	22,5	827	47,70	0,097	710
400-AL1	400	400,1	61	2,89	26,0	1104	60,86	0,072	855
450-AL1	450	449,8	61	3,07	27,6	1246	68,70	0,065	910
500-AL1	500	499,8	61	3,23	29,1	1379	74,67	0,058	990
626-AL1	625	626,2	91	2,96	32,6	1732	95,25	0,046	1140
802-AL1	800	802,1	91	3,35	36,9	2218	118,39	0,036	1340
1000-AL1	1000	999,7	91	3,74	41,1	2767	145,76	0,029	1540

87.4.1.4.4 Induktivität von Leiterseilen

Die Induktivität von Leiterseilen wird als längenbezogener Wert L' in mH/km angegeben, die Induktivität L kann daraus bei bekannter Länge l ermittelt werden:

$$L = L' \cdot l \tag{87.88}$$

Die Leiterinduktivität einer Wechselstromleitung $L_{b,WS}$ errechnet sich mit dessen Radius r, dem Leiterabstand d, der Länge l und der Permeabilität μ_0:

$$L_{b,\mathrm{WS}} = \frac{\mu_0 l}{2\pi} \ln \frac{d}{r} \quad \mu_0 = 4\pi \cdot 10^{-4}\,\mathrm{H/km} \tag{87.89}$$

Für die längenbezogene Betriebsinduktivität $L'_{b,\mathrm{WS}}$ ergibt sich:

$$L'_{b,\mathrm{WS}} = \frac{\mu_0}{2\pi}\left(\ln \frac{d}{r} + 0{,}25\right) \tag{87.90}$$

Bei einer Drehstrom-Einfachleitung mit verdrillten Leitern wird die längenbezogene Betriebsinduktivität $L'_{b,\mathrm{DS}}$ über das geometrische Mittel der Leiterabstände d_m bestimmt:

$$L'_{b,\mathrm{DS}} = \frac{\mu_0}{2\pi}\left(\ln \frac{d_m}{r} + 0{,}25\right) \tag{87.91}$$

$$d_m = \sqrt[3]{(d_{12} \cdot d_{13} \cdot d_{23})} \tag{87.92}$$

Für die Berechnung der Betriebsinduktivität von Bündelleitern $L'_{b,B}$ ist deren mittlerer Abstand d_{mB} sowie der Ersatzradius r_{ers} einzusetzen, n ist die Anzahl der Bündelleiter:

$$L'_{b,B} = \frac{\mu_0}{2\pi}\left(\ln \frac{d_{mB}}{r_{\mathrm{ers}}} + \frac{1}{4n}\right) \tag{87.93}$$

Der Abstand d_{mB} beschreibt die Entfernung der Bündelleiter zu den Teilleitern eines anderen Bündels:

$$d_{mB} = \sqrt[4]{(d_{1B} \cdot d_{2B} \cdot d_{3B} \cdot d_4)} \tag{87.94}$$

Bei kleinem Abstand der Bündelleiter untereinander kann die Mittelung näherungsweise entfallen:

$$d_{1B} \approx d_{2B} \approx d_{3B} \approx d_{4B} \approx d \approx d_{mB} \tag{87.95}$$

Für den Ersatzradius r_{ers} kann mit dem Leiterradius r, und dem Radius des Bündelleiters r_B nach (87.75) angesetzt werden:

$$r_{\mathrm{ers}} = \sqrt[4]{r \cdot 4 \cdot r_B} \tag{87.96}$$

Bei Bündelleitern bewirkt der größere Ersatzradius r_B eine starke Verminderung der Induktivität, wodurch die Übertragungsfähigkeit verbessert wird. Die längenbezogene Betriebsreaktanz $X'_{b,B}$ wird aus der Betriebsinduktivität errechnet:

$$X'_{b,B} = \omega \cdot L_b \quad \omega = 2\pi f \tag{87.97}$$

Die Betriebsreaktanz von Drehstrom-Freileitungen liegt bei Einfachleitungen im Bereich von 0,35 bis 0,4 Ω/km, bei Doppelleitungen bei 0,3 Ω/km und bei Vierer-Bündelleitern bei 0,25 Ω/km.

87.4.1.4.5 Kapazität von Leiterseilen

Der Kapazitätsbelag von Leiterseilen wird als längenbezogener Wert C' in nF/km angegeben, die Kapazität C kann daraus bei bekannter Länge l nach (87.83) im Abschn. 87.4.1.4.2 ermittelt werden:

$$C = C' \cdot l$$

Damit kann die kapazitive Reaktanz X_C errechnet werden:

$$X_C = \frac{1}{\omega \cdot C} \quad \omega = 2\pi f \tag{87.98}$$

Tab. 87.15 Dauerstrombelastbarkeit, Wirkwiderstände, Seilbruchkraft und Querschnitte von Aluminium- / Stahl - ACSR Freileitungsseilen Typ AL1/ST1A nach DIN48204 und 48201/8, EN 50182:2001

Sollquerschnitt nach EN 50182	Bezeichnung alt DIN 48201	Nennquerschnitt Al/St in mm^2	Drähte Al Anzahl/Ø in St/mm	Drähte Stahl Anzahl/Ø in St/mm	Ø Seile/Leiter in mm	Seilgewicht gesamt in kg/km	Rechn. Seilbruchkraft F in kN	Gleichstromwiderstand R'_L bei 20 °C in Ω/km	Dauerstrombelastbarkeit I_d in A
15-AL/1/3-ST1A	16/2,5	15,3 / 2,54	6/1,80	1/1,80	1,80/5,40	61,6	5,80	1,8769	105
24-AL/1/4-ST1A	25/4	23,9/3,98	6/2,25	1 / 2,25	2,25/6,75	96,3	8,95	1,2012	140
34-AL1/6-ST1A	35/6	34,4/5,73	6/2,70	1/2,70	2,70/8,10	138,7	12,37	0,8342	170
44-AL1/32-ST1A	44/32	44/31,7	14/2,00	7/2,40	7,20/11,2	369,3	44,24	0,6574	–
48-AL1/8-ST1A	50/8	48,3/8	6/3,20	1/3,20	3,20/9,6	194,8	16,81	0,5939	210
51-AL1/30-ST1A	50/30	51,2/29,8	12/2,33	7/2,33	6,99/11,7	374,7	42,98	0,5644	–
70-AL1/11-ST1A	70/12	70/11,4	26/1,85	7/1,44	4,32/11,7	282,2	26,27	0,4132	290
94-AL1/15-ST1A	95/15	94,4/15,3	26/2,15	7/1,67	5,01/13,6	380,6	34,93	0,3060	350
97-AL1/56-ST1A	95/55	96,5/56,3	12/3,20	7/3,20	9,60/16,0	706,8	77,85	0,2992	–
106-AL1/76-ST1A	105/75	105,7/75,5	14/3,10	19/2,25	11,3/17,5	885,3	105,82	0,2742	–
122-AL1/20-ST1A	120/20	121,6/19,8	26/2,44	7/1,90	5,70/15,5	491,0	44,50	0,2376	410
122-AL1/71-ST1A	120/70	122,1/71,3	12/3,50	7/3,60	10,8/18,0	894,5	97,92	0,2364	–
128-AL1/30-ST1A	125/30	127,9/29,8	30/2,33	7/2,33	6,99/16,3	587,0	56,41	0,2260	425
149-AL1/24-ST1A	150/25	148,9/24,2	26/2,70	7/2,10	6,30/17,1	600,8	53,67	0,1940	470
172-AL1/40-ST1A	170/40	171,8/40,1	30/2,70	7/2,70	8,10/18,9	788,2	74,89	0,1683	520
184-AL1/30-ST1A	185/30	183,8/29,8	26/3,00	7/2,33	6,99/19,0	741,0	65,27	0,1571	535
209-AL1/34-ST1A	210/35	209,1/34,1	26/3,20	7/2,49	7,47/20,3	844,1	73,36	0,1381	590
212-AL1/49-ST1A	210/50	212,1/49,5	30/3,00	7/3,00	9,0/21,0	973,1	92,46	0,1363	610
231-AL1/30-ST1A	230/30	230,9/29,8	24/3,50	7/2,33	6,99/21,0	870,9	72,13	0,1250	630
243-AL1/39-ST1A	240/40	243,1/39,5	26/3,45	7/2,68	8,04/21,8	980,1	85,12	0,1188	645
264-AL1/34-ST1A	265/35	263,7/34,1	24/3,74	7/2,49	7,47/22,4	994,4	81,04	0,1095	680
304-AL1/49-ST1A	300/50	304,3/49,5	26/3,86	7/3,00	9,00/24,4	1.227,3	105,09	0,0949	740
305-AL1/39-ST1A	305/40	304,6/39,5	54/2,68	7/2,68	8,04/24,1	1151,2	96,80	0,0949	740
339-AL1/30-ST1A	340/30	339,3/29,8	48/3,00	7/2,33	6,99/25,0	1.171,2	91,71	0,0852	790
382-AL1/49-ST1A	380/50	381,7/49,5	54/3,00	7/3,00	9,00/27,0	1.442,5	121,30	0,0758	840
386-AL1/34-ST1A	385/35	386/34,1	48/3,20	7/2,49	7,47/26,7	1333,6	102,56	0,0749	850
434-AL1/56-ST1A	435/55	434,3/56,3	54/3,20	7/3,20	9,60/28,8	1.641,3	133,59	0,0666	900
449-AL1/39-ST1A	450/40	448,7/39,5	48/3,45	7/2,68	8,04/28,7	1.549,1	119,05	0,0644	920
490-AL1/64-ST1A	490/65	490,3/63,6	54/3,40	7/3,40	10,2/30,6	1852,9	150,81	0,0590	960
494-AL1/34-ST1A	495/35	494,4/34,1	45/3,74	7/2,49	7,47/29,9	1632,6	117,96	0,0584	985
511-AL1/45-ST1A	510/45	510,5/45,3	48/3,68	7/2,87	8,61/30,7	1765,3	133,31	0,0566	995
550-AL1/71-ST1A	550/70	549,7/71,3	54/3,60	7/3,60	10,8/32,4	2.077,2	166,32	0,0526	1020
562-AL1/49-ST1A	560/50	561,7/49,5	48/3,86	7/3,00	9,00/32,2	1.939,5	146,28	0,0515	1040
571-AL1/39-ST1A	570/40	571,2/39,5	45/4,02	7/2,68	8,04/32,2	1887,1	136,0	0,0506	1050
653-AL1/45-ST1A	650/45	653,5/45,3	45/4,30	7/2,87	8,61/34,4	2159,9	156,18	0,0442	1120
679-AL1/86-ST1A	680/85	678,6/86,0	54/4,00	19/2,40	12,0/36,0	2.549,7	206,56	0,0426	1150
1046-AL1/45-ST1A	1045/45	1045,6/45,3	72/4,30	7/2,87	8,61/43,0	3.248,2	21892	0,0277	1580

Für den Einzelleiter einer Wechselstromleitung wird die Betriebskapazität $C_{b,\mathrm{WS}}$ aus der Koppelkapazität C_K zwischen den Leitern und der Erdkapazitäten C_E zwischen Leitern und Erdpotenzial errechnet:

$$C_{b,\mathrm{WS}} = 2C_K + C_E \tag{87.99}$$

Die längenbezogene Betriebskapazität C'_b ergibt sich als Gesamtkapazität eines Leiters mit der Permittivitätszahl ε_0, der Leiterhöhe h, dem Leiterabstand d, dem Leiterradius r und dem Ersatzabstand D zu:

$$C'_{b,\mathrm{WS}} = \frac{2\pi\varepsilon_0}{\ln\frac{2hd}{rD}} \quad \varepsilon_0 = 8{,}854 \cdot 10^{-12}\,\frac{\mathrm{A\,s}}{\mathrm{V\,m}} \tag{87.100}$$

Der Ersatzabstand D errechnet sich mit der Leiterhöhe h und dem Leiterabstand d:

$$D = \sqrt{(2h)^2 + d^2} \tag{87.101}$$

Damit ergibt sich die Betriebskapazität $C_{b,\mathrm{WS}}$ zu:

$$C_{b,\mathrm{WS}} = \frac{2\pi\varepsilon_0 l}{\ln\frac{2hd}{r\sqrt{(2h)^2+d^2}}} \tag{87.102}$$

Bei der Drehstromleitung addieren sich die Teilkapazitäten der Leiter zur Betriebskapazität $C_{b,\mathrm{DS}}$:

$$C_{b,\mathrm{DS}} = \frac{C_1}{3} + \frac{C_2}{3} + \frac{C_3}{3} \tag{87.103}$$

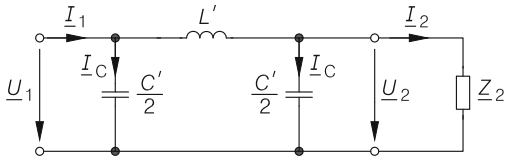


Abb. 87.78 Ersatzschaltbild der verlustlosen Freileitung

Zur Berechnung der Betriebskapazität der Drehstromleitung wird folgende Grundbeziehung für die Kapazität eines Zylinderkondensators genutzt:

$$C = \frac{2\pi\varepsilon_0 l}{ln\frac{r_2}{r_1}} \tag{87.104}$$

Gleichung (87.103) wird zur näherungsweisen Bestimmung der Betriebskapazität $C_{b,DS}$so modifiziert, dass für r_2 der Ersatzdurchmesser der Dreileiteranordnung d_m nach (87.91) eingesetzt wird, für r_1 kann der Leiterradius r eingesetzt werden:

$$C_{b,\mathrm{DS}} \approx \frac{2\pi\varepsilon_0 l}{\ln\frac{d_m}{r}} \tag{87.105}$$

Als Richtwerte für Betriebskapazitäten ergeben sich bei 20 kV Werte von 4 … 5 nF/km, bei 110 kV Werte von 8 … 9 nF/km, bei 220-kV-Zweierbündeln 12 nF/km und bei 380-kV-Viererbündeln 12 … 14 nF/km.

87.4.1.4.6 Betriebsverhalten von Freileitungen

Zur Beschreibung des Betriebsverhaltens von symmetrisch aufgebauten Freileitungen soll das Ersatzschaltbild der verlustlosen Freileitung aus Abb. 87.78 verwendet werden. Zu beachten ist, dass für Leitungen ab ca. 300 km Länge deren Wellenverhalten maßgeblich ist, d. h. dass die elektrischen Größen nicht nur zeit- sondern auch ortsabhängig sind. Die Leitung wird mit dem Lastwiderstand $\underline{Z}_2$ abgeschlossen, die Eingangsspannung $\underline{U}_1$ versorgt die Last und treibt den Strom $\underline{I}_1$. Am Ausgang stellen sich abhängig von $\underline{Z}_2$ die Spannung $\underline{U}_2$ und der Strom $\underline{I}_2$ ein.

Nachfolgend sollen vier Belastungsfälle beschrieben werden:

1. Leerlaufende Freileitung mit $\underline{Z}_2 \to \infty$
2. Zufällig abgeschlossene (nicht angepasste) Freileitung $\underline{Z}_2 \neq Z_W$
3. Gezielt abgeschlossene (angepasste) Freileitung $\underline{Z}_2 = Z_W$
4. Berücksichtigung des ohmschen Leitungswiderstands $R \neq 0$

Fall 1: Leerlaufende Freileitung

Zuerst soll der Leerlauf einer Freileitung, d. h. der Fall mit $\underline{Z}_2 \to \infty$ betrachtet werden, siehe Abb. 87.79. In diesem Fall tritt keine induktive Blindleistung auf ($Q_L = 0$), jedoch kapazitive Blindleistung beim Aufladen der Kapazitäten ($Q_C \neq 0$).

Die Leitung wird mit der Eingangsspannung $\underline{U}_1$ versorgt. Die Kapazitäten verursachen einen kapazitiven Ladestrom, der einen Spannungsfall über der Betriebsinduktivität verursacht. Die Ausgangsspannung ergibt sich als Summe der Eingangsspannung und der Spannungsfälle:

$$\underline{U}_2 = \underline{U}_1 + \underline{I}_C \cdot j\omega L = \underline{U}_1 + \underline{I}_C \cdot j\omega L' \cdot l \tag{87.106}$$

Da der Spannungsbeitrag durch das Produkt aus $\underline{I}_1$ und L' mit der Länge l zunimmt, steigt die Spannung zum Leitungsende hin immer weiter an, siehe Abb. 87.80. Dieser Effekt wurde nach seinem Entdecker als Ferranti-Effekt benannt. Der kapazitive Ladestrom nimmt zum Leitungsende hin ab.

Fall 2: Zufällig abgeschlossene (nicht angepasste) Freileitung

Nun wird die Leitung mit dem Lastwiderstand $\underline{Z}_2$ abgeschlossen, der Leistungsfaktor der Last ist $\cos\varphi = 1$. Die Eingangsspannung $\underline{U}_1$ versorgt die Last und treibt den Strom $\underline{I}_1$. Am Ausgang stellen sich abhängig von $\underline{Z}_2$ die Spannung $\underline{U}_2$ und der Strom $\underline{I}_2$ ein, siehe Abb. 87.81. In diesem Fall tritt sowohl induktive Blindleistung ($Q_L \neq 0$) als auch kapazitive Blindleistung auf; jedoch sind beide Anteile ungleich groß ($Q_L \neq Q_C$), es fließt z. B. wie in Abb. 87.82 gezeigt ein kapazitiver Strom.

In Abb. 87.82 sind die Zeigerbilder der elektrischen Größen für die Fälle kurze und lange Leitung gezeigt. Bei der kurzen Leitung liegt ein induktives Verhalten vor und es tritt ein Spannungsfall zum Leitungsende hin auf, bei der langen Leitung liegt ein kapazitives Verhalten vor und die Spannung steigt zum Leitungsende hin an. Auch wenn am Leitungsende wegen $\cos\varphi = 1$ der Last $\underline{Z}_2$ die Spannung $\underline{U}_2$ und der Strom $\underline{I}_2$ in Phase liegen, trifft dies nicht auf Spannung und Strom am Leitungsanfang zu. Der Grund dafür liegt darin, dass die kapazitive und induktive Blindleistung ungleich groß sind. Aufgrund der Welleneigenschaften langer Leitungen stellt sich abhängig von der Leitungslänge und unabhängig vom Phasenwinkel φ ein Verdrehwinkel θ in der Größe von 6°/100 km ein.

Fall 3: Gezielt abgeschlossene (angepasste) Freileitung

Im zuvor betrachteten Fall 2 hat sich gezeigt, dass ein zufälliger Abschluss der Freileitung durch die Last zu einem Ungleichgewicht von kapazitiver und induktiver Blindleistung im

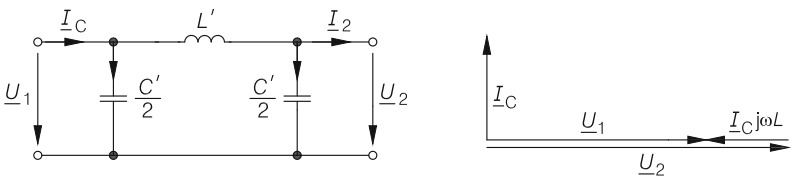


Abb. 87.79 Ersatzschaltbild und Zeigerbild der verlustlosen Freileitung im Leerlauf

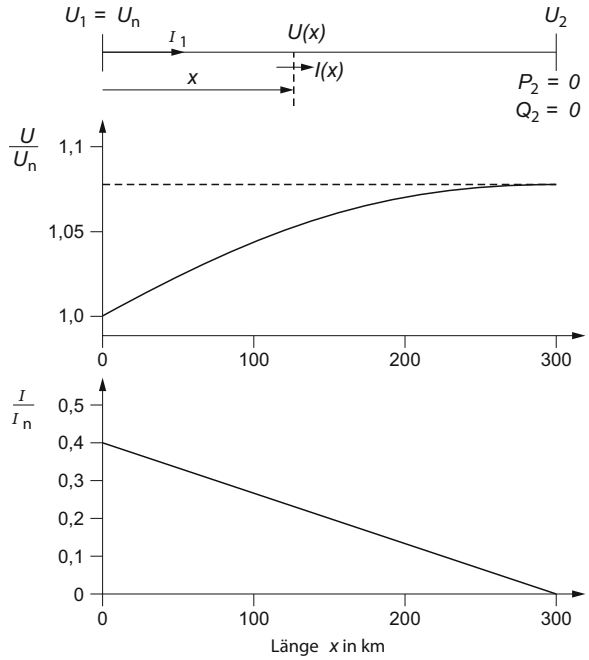


Abb. 87.80 Ortsabhängige elektrische Größen, Spannungs- und Stromverlauf einer langen Freileitung im Leerlauf

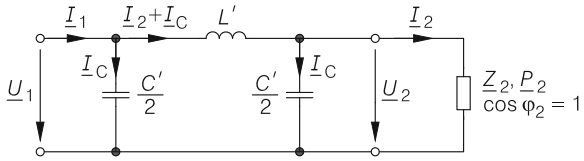


Abb. 87.81 Ersatzschaltbild der verlustlosen Freileitung mit nicht angepasster Belastung

Übertragungssystem und damit zu Spannungsabsenkungen oder -anhebungen führt. Daher soll nun der Abschlusswiderstand angepasst und so ausgewählt werden, dass die Blindleistungen auf der Übertragungsleitung im Gleichgewicht sind und keine Zufuhr der Blindleistung von außen erfolgen muss. Dies kann erreicht werden, wenn der Abschlusswiderstand $\underline{Z}_2$ so ausgewählt wird, dass er der Eingangsimpedanz $\underline{Z}_1$ der Leitung entspricht:

$$\underline{Z}_1 = \frac{\underline{U}_1}{\underline{I}_1} = \underline{Z}_2 = \frac{\underline{U}_2}{\underline{I}_2} \qquad (87.107)$$

Dazu wird der Abschlusswiderstand $\underline{Z}_2$ in der Größe des Wellenwiderstands Z_W gewählt; dieser wird aus den Leitungsbelägen L' für die Induktivität und C' für die Kapazität errechnet, siehe Abb. 87.83:

$$\underline{Z}_1 = \underline{Z}_2 = Z_W = \sqrt{\frac{L'}{C'}} \qquad (87.108)$$

Der Abschluss einer Leistung mit ihrem Wellenwiderstand wird als Leistungsanpassung bezeichnet, die Leitung wird dabei im sog. natürlichen Betrieb belastet. Sie nimmt hierbei eine Wirkleistung auf, die als natürliche Leistung P_{nat} bezeichnet wird:

$$P_{\mathrm{nat}} = \frac{U_{\mathrm{nN}}^2}{Z_W} \qquad (87.109)$$

Typische Werte für den Wellenwiderstand und die natürliche Leistung sind in der Tab. 87.16 angegeben. Es ist ersichtlich, dass die natürliche Leistung weit unterhalb der thermischen Grenzleistung liegt. Es wird später noch gezeigt, dass lange Leitungen zur Sicherstellung der Spannungsstabilität stets unterhalb ihrer natürlichen Leistung belastet werden.

Bei der Leistungsanpassung haben Strom und Spannung stets konstante Amplituden entlang der Leitung und liegen in Phase, es tritt jedoch ein mit der Leitungslänge zunehmender Verdrehwinkel zwischen Anfang und Ende der Leitung auf, siehe Abb. 87.84.

Fall 4: Berücksichtigung des ohmschen Leitungswiderstands

Ist der ohmsche Widerstand $R \neq 0$ und damit nicht vernachlässigt, treten zusätzliche Spannungsabfälle auf, in deren Folge der fließende Strom zum Leitungsende hin abnimmt.

Realisierung des Leitungsabschlusses mit dem Wellenwiderstand

Lange Leitungen ab ca. 300 km Länge werden mit ihrem Wellenwiderstand abgeschlossen, um auf ihnen ein Gleichgewicht von kapazitiver und induktiver Blindleistung einzustellen. Dies erfolgt wie in Abb. 87.85 gezeigt mittels einer einstellbaren Drosselspule an der Tertiärwicklung des Transformators am Ende der 380-kV-Leitung. Durch diese einstellbare Induktivität kann der Wellenwiderstand des Leitungszuges an die Last angepasst werden. Damit liegt auch bei Teillast immer ein natürlicher Betrieb vor. Die nachfolgende 110-kV-Leitung benötigt aufgrund ihrer kürzeren Länge keinen einstellbaren Abschluss am Leitungsende.

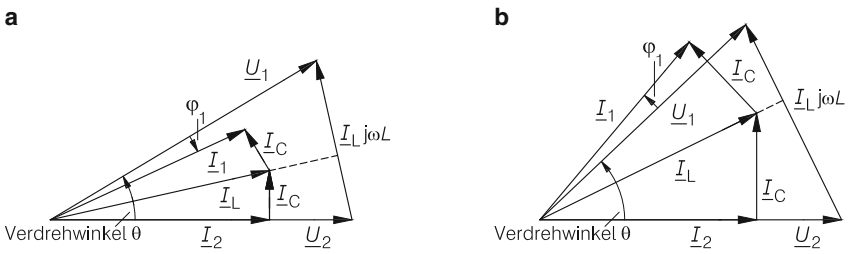


Abb. 87.82 Zeigerbilder der mit einem Widerstand Z abgeschlossenen verlustlosen Freileitung. **a** Kurze Leitung, **b** lange Leitung

Tab. 87.16 Wellenwiderstand und natürliche Leistungen von Freileitungen

Nennspannung U_n in kV	Seilquerschnitt A von Al/St in mm^2	Wellenwiderstand Z_W in Ω	Natürliche Leistung P_{nat} in MW	Thermische Grenzleistung P_{th} in MV A
10	50/8	330	0,3	3
20	120/20	335	2,7	14,2
110	240/40	380	32	123
220	2 × 240/40	276	175	492
380	4 × 240/40	240	602	1700

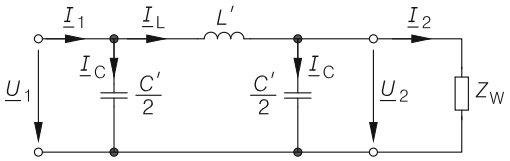


Abb. 87.83 Ersatzschaltbild der verlustlosen Freileitung mit angepasster Belastung

Wirkleistung als Funktion der Leitungslänge

Die übertragbare Wirkleistung wird durch die Leitungslänge bestimmt, siehe Abb. 87.86. Auf kurzen Strecken bis ungefähr 80 km ist wie im Bereich 1 des Bildes gezeigt allein die thermische Grenzleistung des Übertragungsmediums maßgebend. Im Bereich 2 zwischen 80 und 320 km entscheidet der Spannungsfall über die übertragbare Wirkleistung. Über 320 km verhalten sich Freileitungen bei der Leistungsübertragung stets kapazitiv, deshalb müssen im Bereich 3 Leitungen zur Sicherstellung der Spannungsstabilität immer unternatürlich betrieben werden, um Spannungsinstabilitäten zu vermeiden.

Betriebsarten von Freileitungen

Freileitungen können abhängig von der Art ihres Abschlusswiderstands natürlich, unternatürlich oder übernatürlich betrieben werden. Im natürlichen Betrieb der Leitung wird die Leitung mit ihrem Wellenwiderstand abgeschlossen: $\underline{Z}_2 = Z_W$, es wird die natürliche Leistung P_{nat} übertragen. Im unternatürlichen Betrieb der Leitung ist der Abschlusswiderstand größer als der Wellenwiderstand: $\underline{Z}_2 > Z_W$, damit weist die Leitung ein kapazitives Verhalten auf, das zu einem Spannungsanstieg am Leitungsende führt. In der Praxis wird dieser Effekt jedoch durch den ohmschen Leitungswiderstand abgeschwächt. Wegen des im Vergleich zum natürlichen Betrieb größeren Abschlusswiderstandes fließt ein geringerer Strom und es wird weniger Wirkleistung als im natürlichen Betrieb übertragen. Hingegen ist im übernatürlichen Betrieb der Leitung der Abschlusswiderstand kleiner als der Wellenwiderstand: $\underline{Z}_2 < Z_W$, die Leitung weist ein induktives Verhalten mit einem Spannungsfall zum Leitungsende hin auf. Dieser wird durch den ohmschen Leitungswiderstand noch verstärkt. Wegen des kleineren Abschlusswiderstandes fließt ein höherer Strom und es wird mehr Wirkleistung als im natürlichen Betrieb übertragen. Diese Betriebsart ist bei der möglichen Zuschaltung großer Konstantleistungslasten zu vermeiden, wie anschließend gezeigt wird.

Vermeidung von Spannungsinstabilitäten bei der Zuschaltung von Konstantlasten

Der durch das Zuschalten von Konstantleistungslasten auftretende Spannungsfall führt zu einer Stromerhöhung, die wiederum einen Spannungsfall verursacht; diese Abfolge führt zur

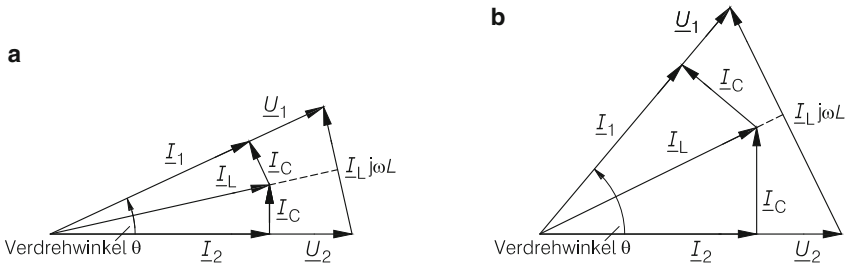


Abb. 87.84 Zeigerbilder der mit einem Widerstand Z abgeschlossenen verlustlosen Freileitung. **a** Kurze Leitung, **b** lange Leitung

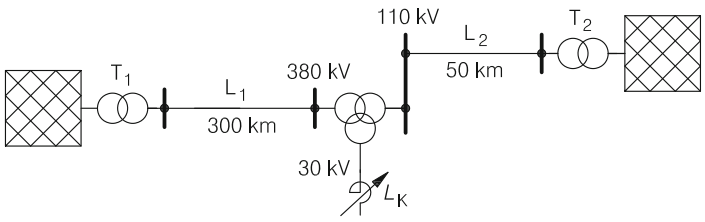


Abb. 87.85 Realisierung des Leitungsabschlusses mit dem Wellenwiderstand mit einstellbarer Drosselspule an der Tertiärwicklung des Transformators

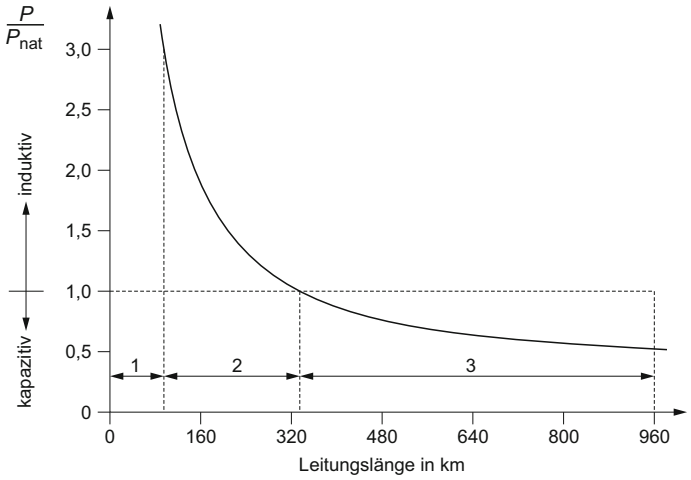


Abb. 87.86 Übertragene Wirkleistung als Funktion der Leitungslänge

Verletzung des Spannungsbandes und damit zur Abschaltung. Dies lässt sich vermeiden, indem lange Leitungen unternatürlich betrieben werden. In diesem Fall nimmt bei der Zuschaltung einer Konstantlast durch die Annäherung an den natürlichen Betrieb die im unternatürlichen Betrieb auftretende Blindleistung und damit der Blindstrom auf der Leitung ab, wodurch die Stromerhöhung durch die Konstantleistungslast kompensiert wird. Dies ist dadurch begründet, dass die Zuschaltung einer Last eine Verringerung des Abschlusswiderstands der Leitung verursacht.

Tab. 87.17 Eigenschaften von Kupfer und Aluminium

Parameter	Kupfer	Aluminium	Verhältnis Cu/Al
Dichte in kg/dm^3	8,9	2,7	3,2
Ausdehnungskoeffizient in 10^{-6}/K	17	24	0,7
Spezifische Wärmekapazität in J/g K	0,4	0,9	0,4
Schmelzpunkt in °C	1080	660	1,6
Wärmeleitfähigkeit in W/mK	380	220	1,7
Leitfähigkeit in m/Ω mm^2	58	36	1,6
Preis gerundet in €/kg	5	1,5	3,3

87.4.2 Kabel

Drehstromkabel stehen für Spannungen von 0,4 bis 380 kV zur Verfügung, sie werden bevorzugt im Bereich bis 110 kV eingesetzt. Gleichspannungskabel werden in Kombination mit Hochspannungs-Gleichstromanlagen für Spannungsbereiche bis zu 800 kV gefertigt, kommen jedoch insgesamt nur selten zum Einsatz und werden daher hier nicht vertieft. Von den im Abschn. 87.3.2.1 genannten Stromkreislängen im Drehstromnetz werden 90 bzw. 80 % als Kabel auf der Nieder- und Mittelspannungsebene ausgeführt, jedoch nur 9 bzw. 0,3 % auf der Hoch- und Höchstspannungsebene.

87.4.2.1 Leiterwerkstoffe und Leiterformen

Leiterwerkstoffe für Kabel sind Kupfer und Aluminium, siehe Tab. 87.17. Auch bei Kabeln gilt, dass Aluminium bei gleichem Gewicht einen größeren Querschnitt und damit einen höheren Leitwert als Kupfer aufweist und deshalb wesentlich kostengünstiger ist, weshalb es ab der Mittelspannungsebene bevorzugt eingesetzt wird. Auf der Niederspannungsebene werden wegen ihrer besseren Klemm- und Kontaktierungseigenschaften hauptsächlich Kupferleitungen eingesetzt, erst bei hohen Strömen werden Kabel mit mehrdrähtigen Rundleitern oder sektorförmigen Leitern aus Aluminium genutzt.

Leiterformen

Als Leiterformen kommen ein- und mehrdrähtige Rundleiter sowie sektorförmige Leiter zum Einsatz. Bei Querschnitten bis 35 mm^2 weisen Sektorleiter gegenüber Rundleitern kleinere Kabeldurchmesser auf.

Schirmleiter

Metallische Schirme aus Kupfer, Stahl oder Blei leiten den Ladestrom beim Einschalten, den Ableitstrom im Fehlerfall und dienen als Berührungsschutz.

87.4.2.2 Isolierstoffe

Zur Isolation von Kabeln stehen die Stoffe Polyvinylchlorid (PVC), Polyethylen (PE), Vernetztes Polyethylen (VPE), Ethylen-Propylen-Kautschuk (EPR) zur Verfügung. Ältere Mittel- und Hochspanungskabel weise eine öl- oder massegetränkte Papierisolierung auf, diese Art der Kabelisolierung wird jedoch aus Umweltschutzgründen heute nicht mehr eingesetzt. Wesentliche Eigenschaften der Isolierstoffe sind in der Tab. 87.18 gezeigt. Im Niederspannungsbereich wird die Schichtdicke der Isolierung vorrangig durch die mechanische Beanspruchung während der Verlegung bestimmt, mit steigender Spannung hängt die Isolierschichtdicke immer mehr von der elektrischen Feldstärke ab.

Tab. 87.18 Wesentliche Eigenschaften von Kabelisolierstoffen PVC – Polyvinylchlorid, PE – Polyethylen, VPE – Vernetztes Polyethylen, EPR – Ethylen-Propylen-Kautschuk, angepasst nach Constantinescu-Simon: Handbuch Elektrische Energietechnik 1997

Parameter	PVC	PE	VPE	EPR	Papier/Öl
Spezifischer Widerstand ρ in MΩ m	$> 10^8$	$> 10^{11}$	$> 10^{11}$	$> 10^8$	$> 10^8$
Verlustfaktor $\tan\delta$ in 10^{-3}	50	0,4	0,4	5	3
Permittivitätszahl ε_r	4	2,3	2,3	3	3,5
Versprödungstemperatur T_v in °C	−10	−60	−60	−50	Versprödet nicht
Zulässige Leitertemperatur T_L in °C	70	70	90	90	60 … 80
Zulässige Leitertemperatur T_{Lk} bei Kurzschluss in °C	140 … 160	150	250	250	140 … 180
Spezifischer Wärmewiderstand R_{th} in K/W	6,0	3,5	3,5	4–5	5,0
Dichte ρ in kg/dm^3	1,3	0,92	0,92	1,3	1,0
Zugfestigkeit β_Z in N/mm^2	16	15 … 20	15 … 20	6 … 10	gering
Reißdehnung ε in %	200 … 250	400 … 600	300 … 500	350 … 500	0
Flexibilität	Mittel	Mittel	Mittel	Hoch	Gering
Spannungsbereich	Niederspannung, Mittelspannung bis 20 kV	Alle Spannungsebenen	Mittel- Hoch- und Höchstspannung	Alle Spannungsebenen	Mittel- Hoch- und Höchstspannung
Anwendungsbereich: Isolierung von	Kabel	Nur Kabelmantel	Kabel	Isolatoren, Anschlüsse Garnituren	Kabel für AC und DC (HGÜ)

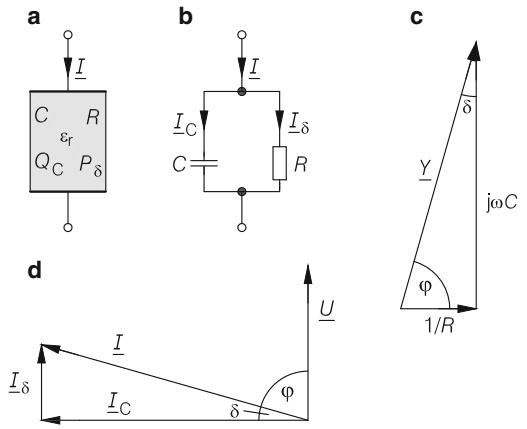


Abb. 87.87 Ermittlung des Verlustwinkels δ von Isolierstoffen. **a** Ersatzanordnung als Plattenkondensator, **b** Ersatzschaltbild des realen Kondensators, **c** Zeigerbild der Leitwerte, **d** Zeigerbild der Ströme

Verlustfaktor tan δ

Zur Bewertung der Isolationsqualität wird der Verlustfaktor $\tan\delta$ genutzt, der sich aus den ohmschen Verlusten des Isolierstoffs ergibt. Wird der Isolierstoff vereinfacht als Dielektrikum eines Plattenkondensators betrachtet, siehe Abb. 87.87a, lässt er sich durch das in Abb. 87.87b gezeigte Ersatzschaltbild des realen Kondensators abbilden. Zur Berücksichtigung der ohmschen Verluste wird ein Widerstand parallel zum Kondensator verschaltet. Damit ergibt sich das Zeigerbild in Abb. 87.87d, I_δ ist der Strom durch den Widerstand R, mit dem die im Dielektrikum umgesetzte Wirkleistung (dielektrische Verlustleistung) P_δ errechnet wird:

$$P_\delta = U \cdot I_\delta \tag{87.110}$$

Die kapazitive Blindleistung errechnet sich zu:

$$Q_C = U \cdot I_C = U^2 \cdot \omega C \tag{87.111}$$

Zur Berechnung des Verlustfaktors $\tan\delta$ können die Leitwerte gemäß Ersatzschaltbild aus Abb. 87.87b oder auch die Leistungen bzw. Ströme verwendet werden, siehe Abb. 87.87c und Abb. 87.87d:

$$\tan\delta = \frac{1/R}{\omega C} = \frac{P_\delta}{Q_C} = \frac{I_\delta}{I_C} \tag{87.112}$$

Der Winkel δ beschreibt die Abweichung von der 90°-Phasenverschiebung der idealen Kapazität, er ist abhängig vom eingesetzten Material sowie den Parametern Frequenz, Temperatur und Feldstärke:

$$\delta = 90° - |\varphi| \tag{87.113}$$

In der Praxis treten abhängig von der Isolationsqualität nur sehr kleine Verlustwinkel auf, z. B. $\delta = 0{,}05°$. Für die Hoch- und Höchstspannung werden nur Isolierstoffe mit kleinem Verlustfaktor $\tan\delta$, also kleinen dielektrischen Verlusten, eingesetzt. Aus (87.112) lässt sich direkt der Zusammenhang zwischen Verlustleistung und kapazitiver Blindleistung ableiten:

$$P_\delta = \tan\delta \cdot Q_C$$

Setzt man dies in (87.111) ein, erhält man für die Kabelverlustleistung P_δ die in der Praxis genutzte Beziehung:

$$P_\delta = \omega C \cdot \tan\delta \cdot U^2 \tag{87.114}$$

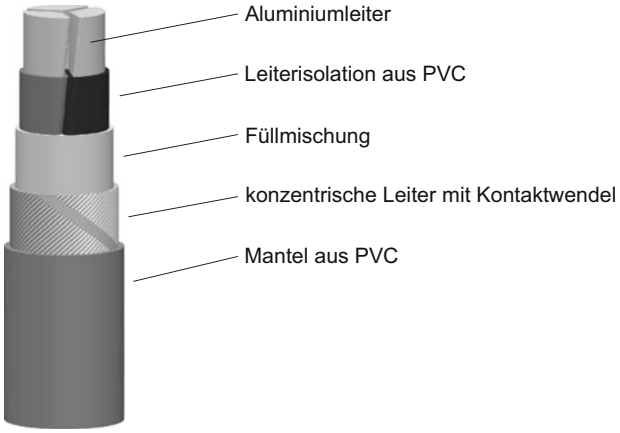


Abb. 87.88 Aufbau eines dreiadrigen Niederspannungskabel NAYCWY oder NA2XCWY mit konzentrischem Leiter und sektorförmigen Leitern, ohne konzentrischem Leiter NAYY, NA2XY

87.4.2.3 Aufbau von Kabeln

87.4.2.3.1 Kunststoffkabel

Im Niederspannungsbereich werden die Leiter rund oder bei größeren Querschnitten auch sektorförmig ausgeführt. Es werden oft Vierleiter-, aber auch Dreileiterkabel mit verseilten Leitern aus Kupfer oder Aluminium eingesetzt, siehe Abb. 87.88. Meist wird die Aderisolierung mit Polyvinylchlorid (PVC) realisiert, dass jedoch im Vergleich mit VPE-Vernetztem Polyethylen (VPE) nach Tab. 87.18 eine geringere Wärmebelastbarkeit aufweist und deshalb schrittweise durch VPE abgelöst wird. Die Leiter werden durch eine gemeinsame Umhüllung aus einer Füllmischung zusätzlich isoliert. Teilweise wird wie in Abb. 87.88 gezeigt als Schirmung eine leitfähige Schicht mit konzentrischem Leiter realisiert. Der Kabelmantel besteht aus PVC oder Polyethylen (PE), das eine höhere mechanische Festigkeit und Kältebeständigkeit aufweist.

Im Mittelspannungsbereich werden mehrdrähtige Leiter aus Kuper und Aluminium mit Isolierungen aus VPE und einem Mantel aus PE eingesetzt. VPE weist neben der höheren Wärmebelastbarkeit auch einen wesentlich geringeren Verlustfaktor $\tan\delta$ als PVC auf, siehe Tab. 87.18. Mittelspannungskabel werden bis 10 kV drei- und einadrig ausgeführt, darüber hinaus fast nur noch als Einleiterkabel. Für Nennspannungen ab 10 kV werden zur Vergleichmäßigung der Feldstärken und zur Vermeidung von Feldspitzen um den Leiter und die Isolierung zusätzlich eine innere und äußere Leitschicht aus halbleitendem Kunststoff vorgesehen, über der sich ein Quellvlies, ein Kupferschirm, eine Trennschicht und ein PE-Mantel befinden. Das Quellvlies quillt beim Eindringen von Wasser auf und verhindert die Ausbreitung von Feuchtigkeit in Längsrichtung, es wird deshalb auch als Längswasserdichtung bezeichnet. Das Eindringen von Wasser muss verhindert werden, erfolgt jedoch auch durch Diffusion über den Kabelmantel. Dringt Wasserdampf in die VPE-Kabelisolierung ein, lagern sich Wassermoleküle zwischen den vergleichsweise großen Kunststoffmolekülen an und bilden bei anliegender Feldstärke E durch die hohe Dielektrizitätszahl von Wasser Feldstärkespitzen im Dielektrikum, die wegen ihrer Form als „water trees“ bezeichnet werden, siehe Abb. 87.89. Dies führt zu Teilentladungen, die über längere Zeiträume elektrische Durchschläge verursachen können.

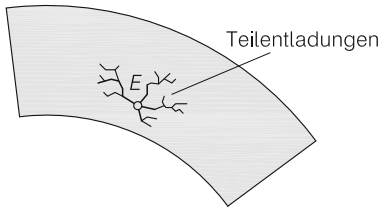


Abb. 87.89 Teilentladungen (water trees) beim Eindringen von Wassermolekülen in die Kunststoffisolation

Dieser Effekt wäre bei Hoch- und Höchstspannungskabeln stark ausgeprägt. Daher wird hier unter dem eigentlichen Kabelmantel aus PE ein zusätzlicher Aluminium-Schichtenmantel vorgesehen, der das Eindringen von Wasser in die Kabelisolierung verhindert und als Querwasserdichtung bezeichnet wird, siehe Abb. 87.90. Kabel für Hoch-und Höchstspannungen weisen ebenfalls die zuvor für Mittelspannungskabel beschriebenen zusätzlichen Isolationsmerkmale auf. Darüber hinaus haben sie über den Kupferschirm eine weitere Schicht aus leitfähigen Bändern, um das elektrische Feld zu vergleichmäßigen.

87.4.2.3.2 Massekabel

Bei dieser Ausführung wird die Isolation durch ölgetränktes Papier realisiert, die Tränkmasse hat den Begriff Massekabel geprägt. Massekabel werden im Niederspannungsnetz seit den siebziger Jahren nicht mehr und im Mittelspannungsnetz immer seltener verlegt. Ein großer Teil der verlegten Kabel ist jedoch aufgrund der Lebensdauer von 50 Jahren noch in Betrieb und wird schrittweise erst innerhalb der nächsten Jahre ausgetauscht. Einadrige Massekabel werden aufgrund ihrer guten Isolierstoffeigenschaften bei Gleichspannungen für HGÜs im Spannungsbereich bis 400 kV genutzt. Die Adern werden einzeln mit

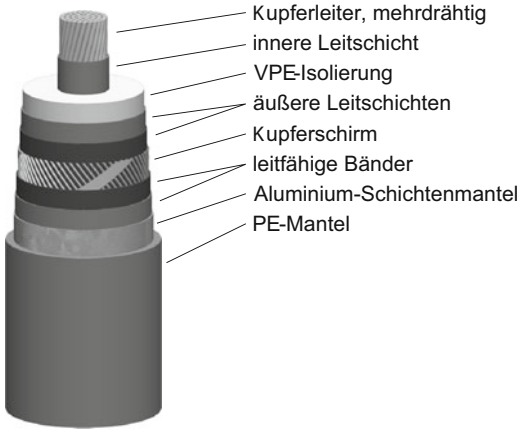


Abb. 87.90 Aufbau eines VPE-Höchstspannungskabels 2XS(FL)2Y 1x800 RM/50 220/380 kV



Abb. 87.91 Aufbau eines dreiadrigen Mittelspannungs-Gürtelkabels NKBA 3x95 RM 6/10 kV

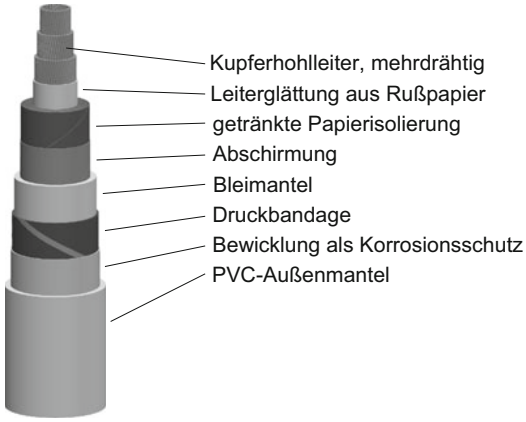


Abb. 87.92 Aufbau eines Ölkabels NÖKUDEY für Spannungen von 60 bis 400 kV

15 bis 30 mm breiten Papierbändern lagenweise geschichtet und überlappend umwickelt und anschließend mit zähflüssiger Masse oder Öl imprägniert. Diese flüssige Imprägnierung kann durch das Eindringen in Hohlräume zeitweilige auftretende Fehlerstellen beseitigen und somit Teilentladungen verhindern. Zusätzlich werden bei der geschichteten Isolierung Fehlerstellen der Einzelschicht durch benachbarte Schichten abgedeckt, dies wird als Schottung bezeichnet. Im Spannungsbereich bis 10 kV wird die Ausführungsart als Gürtelkabel eingesetzt, siehe Abb. 87.91. Hierbei werden die einzeln isolierten Leiter zusätzlich durch die Gürtelisolation aus öl- oder massegetränktem Papier umhüllt. Oberhalb der Niederspannung werden innere und äußere Leitschichten vorgesehen. Die innere Leitschicht wird durch leitfähiges Rußpapier realisiert, die äußere Leitschicht aus metallisiertem Papier, der sog. Höchstädter-Folie. Daraus abgeleitet werden diese Kabel im Spannungsbereich von 10 kV bis 30 kV als Höchstädter-Kabel bezeichnet. Zum mechanischen Schutz und gegen eindringende Feuchtigkeit werden ein Bleimantel, eine Druckschutzbandage, eine innere Schutzhülle, eine Stahlbandbewehrung und eine äußere Schutzhülle vorgesehen, siehe Abb. 87.91.

87.4.2.3.3 Öldruckkabel

Bei Ölkabeln werden die Leiter durch ölgetränktes Papier isoliert, das in Form von Papierbändern lagenweise geschichtet und überlappend gewickelt wird. Um Hohlräume auszufüllen und somit Teilentladungen zu vermeiden, wird durch Ölkanäle dünnflüssiges Öl in das Kabel gedrückt. In Abb. 87.92 ist ein einadriges Ölkabel gezeigt, dessen Bauform in weiten Spannungsbereichen von 60 bis 400 kV eingesetzt wird. Hier ist der Kupferleiter als Hohlleiter ausgeführt, der die Aufgabe des Ölkanals übernimmt. Als innere Leitschicht wird leitfähiges Rußpapier verwendet, auf der Isolation wird eine Schirmung aus Rußpapier oder metallisiertem Papier vorgesehen, die gemeinsam mit dem Bleimantel die Lade-und Ableitströme führt. Der Bleimantel dient gemeinsam mit der Druckbandage, der Bewicklung und dem PVC-Außenmantel dem mechanischen Schutz. Da bei mechanischen Beschädigungen Öl ausfließt, ist dieser Kabeltyp aus Umweltschutzgründen nicht mehr zulässig. Als Alternative stehen VPE-isolierte Kabel oder gasisolierte Kabel zur Verfügung.

87.4.2.3.4 Gasdruckkabel

Bei diesem Kabeltyp sind drei einadrige Massekabel gemeinsam in einem Stahlrohr angeordnet, das mit Stickstoff bei einem Druck von 15 bar befüllt ist. Beim Gasaußendruckkabel werden die Blei-Kabelmäntel von außen zusammengedrückt, wodurch vorhandene Hohlräume der massegetränkten Papierisolierung geschlossen werden und so die Isolierfähigkeit verbessert wird, siehe Abb. 87.93. Im Gegensatz dazu weisen die Adern des Gasinnendruckkabels keinen Kabelmantel auf, hier dringt der komprimierte Stickstoff in die Hohlräume der Papierisolierung ein und erhöht die Isolierfähigkeit abhängig vom Gasdruck. Dreiadrige Gasdruckkabel benötigen weniger Platz als einad-

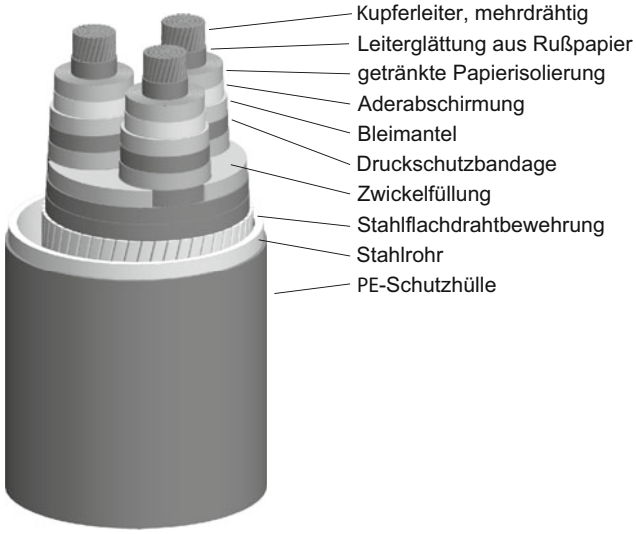


Abb. 87.93 Aufbau eines dreiadrigen Gasaußendruckkabels NPKDvFSt2Y 3x400 OM 64/110 kV

rige VPE-Hochspannungskabel, sie sind durch das umhüllende Stahlrohr mechanisch gut geschützt und robust gegenüber Beschädigungen bei Bauarbeiten. Sie werden deshalb bevorzugt im Innenstadtbereich eingesetzt.

87.4.2.4 Bezeichnung von Kabeln

Normkabel werden in ihrem Aufbau durch Buchstaben gekennzeichnet. Der erste Buchstabe N beschreibt Normkabel gemäß den VDE-Bestimmungen, danach folgt die Beschreibung der Kabelelemente von innen nach außen. Kupferleiter, getränkte Papierisolierungen und innere und äußere Leitschichten bei Kunststoffkabeln werden nicht aufgeführt. Nicht genormte, jedoch gebräuchliche Kurzzeichen werden in Klammern angegeben. In der Tab. 87.19 sind wichtige Kurzzeichen für Kabel aufgeführt.

Tab. 87.19 Bauartenkurzzeichen für Kabel nach VDE 0271

Kurzzeichen für Kabel mit Kunststoffisolierung	
A	Aluminiumleiter
I	Installationskabel (Gasinnendruckkabel)
Y	PVC-Isolierung (äußere und innere Schutzhülle)
2X	VPE-Isolierung
HX	Halogenfreie Polymer-Isolierung
C	Konzentrischer Leiter aus Kupfer
CW	Konzentrischer Leiter aus Kupfer, wellenförmig aufgebracht
CE	Konzentrischer Leiter aus Kupfer bei dreiadrigen Kabeln
S	Schirm aus Kupfer
SE	Schirm aus Kupfer bei mehradrigen Kabeln
(F)	Schirmbereich längswasserdicht
K	Bleimantel
Y	PVC-Mantel zwischen Kupferschirm bzw. konzentrischem Leiter und Bewehrung
F	Bewehrung aus verzinktem Stahlflachdraht
R	Bewehrung aus verzinktem Stahlrunddraht
G	Gegen-oder Haltewendel aus verzinktem Stahlband
2Y	PE-Mantel
(L)2Y	Querwasserdicht
-FE	Isolationserhalt im Brandfall
Kurzzeichen für Kabel mit Papierisolierung	
A	Aluminiumleiter
H	Höchstädter-Folie (Einzeladerabschirmung mit metallischem Papier)
HX	Halogenfreie Polymer-Isolierung
E	Schutzhülle mit eingebetteter Schicht aus Elastomerband oder Kunststofffolie
Y	Innere und äußere PVC-Schutzhülle
B	Bewehrung aus Stahlband
F	Bewehrung aus Stahlflachdraht
FO	Bewehrung aus Stahlflachdraht, offen
R	Bewehrung aus Stahlrunddraht
RO	Bewehrung aus Stahlrunddraht, offen
G	Gegen-oder Haltewendel aus Stahlband
A	Äußere Schutzhülle aus Faserstoffen (Jute)
Kurzzeichen für Öl- und Gasdruckkabel	
Y	Innere und äußere PVC-Schutzhülle
YV	Verstärkter PVC-Mantel
Ö	Ölkabel (Papier mit Öltränkung)
EK	Mehrfacher Bleimantel
KL	Aluminiummantel (gepresst, glatt)
D	Druckschutzbandage
V	Verseilte Adern
P	Gasaußendruckkabel (Papier mit Masseträänkung)
I	Gasinnendruckkabel (Papier mit Masseträänkung)
ST	Stahlrohr
Kurzzeichen für Kabel U_0/U 0,6/1 kV	
-J	Kabel mit grün-gelber Ader
-O	Kabel ohne grün-gelber Ader

87.4.2.5 Betriebsverhalten von Kabeln im Vergleich zu Freileitungen

Für Kabel können die in Abb. 87.77 gezeigten Ersatzschaltbilder von Freileitungen genutzt werden. Im Vergleich zu Freileitungen kann der ohmsche Anteil jedoch nicht vernachlässigt werden. Für Einleiterkabel kann man die Betriebskapazität anhand der Beziehung für die Berechnung der Kapazität eines Zylinderkondensators nach (87.104) abschätzen, indem für r_1 der Innenradius r_i und für r_2 der Außenradius r_a eingesetzt wird:

$$C = \frac{2\pi\varepsilon_0 l}{\ln \frac{r_2}{r_1}}$$

Nimmt man nun an, dass für den Vergleich von Freileitung und Kabel der Innenradius dem Leiter- bzw. Ersatzleiterradius entsprechen soll und der Außenradius bei der Freileitung dem Abstand zum Erdpotenzial und beim Kabel dem Abstand zum Kabelschirm entspricht, ergibt sich bei größerem Abstand eine kleinere Kapazität. Noch deutlicher wird dies, wenn man die Anordnung Freileitung-Erdboden bzw. Innenleiter-Kabelschirm sehr stark vereinfacht als Plattenkondensator mit der Fläche A und dem Abstand d betrachtet:

$$C = \varepsilon \frac{A}{d} \tag{87.115}$$

Wegen ihrer im Abschn. 87.4.1.4.5 ermittelten Kapazitäten im Bereich von nF/km kann bei Freileitungen eine wirtschaftlich Energieübertragung bis zu einer Länge vom 1000 km erfolgen. Betrachtet man ein dreiadriges Niederspannungs-Gürtelkabel, so wirken hier die Koppelkapazitäten C_K zwischen den Leiter sowie jeweils die Erdkapazitäten C_E zwischen Leiter und Kabelschirm. Daraus ergibt sich eine längenbezogene Betriebskapazität $C'_{b,K}$ von:

$$C'_{b,K} = 3C'_K + C'_E \tag{87.116}$$

Kabel weisen Kapazitätswerte im Bereich von 0,3 µF/km bis 0,4 µF/km auf. Durch den im Vergleich zu Freileitungen hohen Kapazitätsbelag wird schon im Leerlauf ein kapazitiver Ladestrom verursacht. Dies wird als Selbstauslastung eines Kabels bezeichnet. Mit steigender Belastung steigt die induktive Blindleistung an, im übernatürlichen Betrieb überwiegt dieser Anteil gegenüber der kapazitiven Blindleistung. Durch den Ladestrom werden auch bei Teillast Verluste verursacht, weshalb die wirtschaftliche Übertragungslänge im Hochspannungsbereich bei

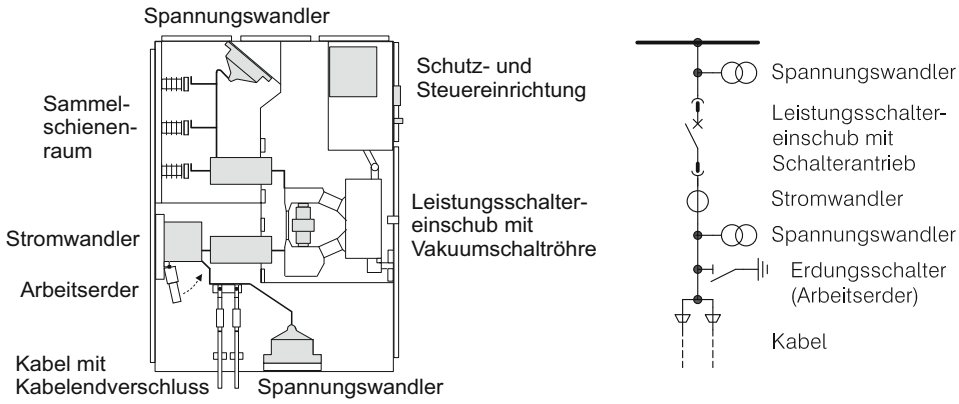


Abb. 87.94 Aufbau einer Mittelspannungs-Schaltanlage, *links*: mechanischer Aufbau, *rechts*: Schaltbild

ca. 30 km liegt. Bei der Netzanbindung von Offshore-Windparks werden auch größere Längen bis 120 km realisiert. Dazu müssen an den Leitungsenden Kompensationsdrosselspulen eingesetzt werden, in denen ohmsche Betriebsverluste auftreten. Solche Offshore-Windparkkabel werden abhängig von den Windverhältnissen nur in wenigen Wochen im Jahr mit Nennleistung betrieben.

87.4.3 Schaltanlagen

Schaltanlagen sind die Knotenpunkte des Energieversorgungsnetzes. Sie dienen der sicheren und flexiblen Schaltung und Verteilung von elektrischer Energie, indem sie über Freileitungen und Kabel angeschlossene elektrische Betriebsmittel und Verbraucher miteinander verbinden und voneinander trennen. Durch die flexible Verschaltung können im Fehlerfall, bei Überlast oder für Wartungsarbeiten Betriebsmittel spannungsfrei geschaltet und bei Bedarf die Netztopologie angepasst werden. Dazu werden die Betriebsgrößen über Messwandler gemessen und in Schutzgeräten ausgewertet. Mit einem Steuerungs- und Leitsystem werden die Anlagen überwacht und Fehlerzustände protokolliert.

87.4.3.1 Technologien und Einsatzbereiche

Als Technologien stehen luftisolierte Innenraum- und Freiluft-Schaltanlagen und gekapselte, mit Schwefelhexafluorid (SF_6) gasisolierte Schaltanlagen zur Verfügung. Das Isoliergas SF_6 wird wegen der durch die Paschen-Kurve beschriebene druckabhängige Isolierstofffestigkeit bei Drücken von 5 bis 10 bar eingesetzt und weist mit 80 kV/cm etwa das 2,4-fache Isoliervermögen gegenüber Luft mit 33 kV/cm auf. Nieder- und Mittelspannungsschaltanlagen werden immer in geschlossenen Gebäuden ausgeführt. Für die Niederspannung werden ausschließlich luftisolierte Schaltanlagen eingesetzt. Auf der Mittelspannungsebene findet man luftisolierte und gasisolierte Schaltanlagen. Für die Hoch- und Höchstspannung werden luftisolierte Freiluft-Schaltanlagen oder SF_6-gasisolierte Schaltanlagen eingesetzt, die nur 10 % des Platzbedarfs von Freiluftanlagen haben, jedoch wartungsintensiver sind. SF_6-Schaltanlagen werden zumeist als Innenraum-Ausführungen installiert, es sind jedoch auch Anlagen zur Installation innerhalb von Freiluftinstallationen verfügbar.

87.4.3.2 Aufbau von Schaltanlagen

Schaltanlagen werden in Form von Schaltfeldern aufgebaut. Nach ihrer Funktion werden diese beispielsweise als Eingangs-, Abgangs-, und Messfelder bezeichnet. Die Felder werden gegeneinander lichtbogensicher geschottet, d. h. ein Abbrand durch Lichtbogeneffekte bei einem Schalterversagen weitet sich im Regelfall nicht auf die benachbarten Felder aus. Als Leitermaterial für Schaltanlagen werden wegen der hohen Ströme unisolierte und meist rechteckförmige, aber auch runde Sammelschienen aus Kupfer oder Aluminium verwendet. Diese können als Einfach-, Doppel- oder Dreifachsammelschienen ausgeführt sein und werden auf isolierenden Stützern an der Rückwand der Schaltfelder montiert. Im unteren Bereich werden die Ein- und Abgänge von Kabeln angeschlossen, dort befinden sich auch die Messwandler für den Strom und die Spannung. Im oberen Frontbereich sind die Anzeigen der Strom- und Spannungswandler sowie die Steuerung für das Schaltfeld untergebracht. Das wesentliche Element, der Schalter, befindet sich in der Mitte des Schaltfeldes vor den Sammelschienen, siehe Abb. 87.94. Im Kurzschlussfall steigt der Druck in der Schaltanlage schlagartig an und könnte zu mechanischen Zerstörungen führen. Deshalb befinden sich bei luftisolierten Anlagen Druckentlastungsklappen auf der Oberseite der Schaltfelder, die bei Beanspruchung nach außen öffnen. Bei gasisolierten Schaltanlagen sind solche Klappen wegen des Betriebsdrucks nicht nutzbar, hier werden sog. Berstscheiben eingesetzt, die im Falle eines starken Druckanstiegs zerstört werden.

87.4.3.3 Schalter

Abhängig von der Spannungsebene und dem Einsatzzweck werden verschiedene Schalterarten in unterschiedlichen Technologien und Ausführungen benötigt. Schalter dienen neben ihrer eigentlichen Schaltfunktion auch als Schutzelement zur Abschaltung von Fehlerströmen bei Kurz- oder Erdschlüssen.

87.4.3.3.1 Anforderungen an Schalter

Ein Schalter muss für die elektrischen, thermischen und mechanischen Anforderungen seines Anwendungsspektrums ausgelegt werden.

Elektrische Anforderungen

Zu den elektrischen Anforderungen gehören bei Wechselstromschaltern die Stromlöschung, die stets im Nulldurchgang des Stromes erfolgt, die dauerhafte Isolierung der offenen Kontaktstelle sowie die Beherrschung von Überspannungen.

Zur sicheren Abschaltung muss insbesondere die thermische Belastung durch Lichtbogen beherrscht werden.

Thermische Belastung durch Lichtbogen

Wird der Stromfluss durch den Schalter unterbrochen, treten während des Schaltvorgangs Spannungsüberschläge auf, solange der Kontaktabstand die Spannung über den Schalterpolen noch nicht isolieren kann. Dabei kommt es durch eine Gasionisation zur Ausbildung eines Lichtbogens, dessen Plasma bei Temperaturen zwischen 6000 und 12.000 K brennt und erst ab 2000 K in Luft und 3000 K in SF_6 erlischt. Der Energieumsatz des Plasmas erfolgt zu 95 % in Wärme und zu 5 % durch Elektrodenabbrand und Wärmestrahlung. Stromschwache Lichtbogen verlöschen im Stromnulldurchgang, bei hohen Strömen muss eine Löschung erfolgen. Dies kann durch Kühlung des Lichtbogens und eine Wegverlängerung, z. B. durch Anströmung mit Gas oder das Einschieben von Schaltkammern mit mäanderförmigem Kanalverlauf erfolgen. Für das Schalten von Lichtbogen vorgesehene Schalter müssen sowohl in ihrer thermischen Kapazität als auch für den Elektrodenabbrand durch Lichtbogen dimensioniert werden und den Energieaustausch zwischen den Kontakten beherrschen können.

Mechanische Belastung

Beim Stromfluss treten hohe mechanische Kräfte auf, insbesondere im Kurzschlussfall und bei länger anstehenden Lichtbogen. Schalter müssen in ihrem Auslösemechanismus und ihrer Befestigung dafür ausgelegt sein. Um Schaltüberspannungen zu vermeiden, darf die Abschaltung nicht zu schnell erfolgen, in der Praxis werden Schaltzeiten von 30 ms realisiert. Schalter werden dafür mit Handantrieben oder Kraftantrieben über Magnete, Motoren, mechanische Federn oder Druckluft ausgerüstet.

87.4.3.3.2 Schalterarten

Sollen neben den Betriebsströmen auch Fehlerströme geschaltet werden können, kommen Leistungsschalter zum Einsatz. Aufgrund ihrer Dimensionierung für hohe elektrische, thermische und mechanische Beanspruchung weisen sie die im Vergleich höchsten Kosten auf. Günstigere Lastschalter können Betriebsströme schalten, sind jedoch nicht für das Schalten von Kurzschlussströmen geeignet. Deshalb werden sie zur Realisierung des Kurzschlussschutzes bis 1 kV mit Niederspannungs-



Abb. 87.95 Symbole der Schalterarten für Schaltanlagen

Hochleistungssicherungen (NH), über 1 bis 36 kV mit Hochspannungs-Hochleistungssicherungen (HH) und darüber hinaus mit übergeordneten Leistungsschaltern kombiniert. Trennschalter können hingegen keine Ströme schalten und dürfen deshalb nur im spannungslosen Zustand betätigt werden. Sie werden zur Schaffung sichtbarer Trennstellen und als Arbeitserder an abgeschalteten Betriebsmitteln eingesetzt. Zur Einsparung von Kosten werden Lasttrennschalter eingesetzt, bei denen im ausgeschalteten Zustand eine offene Trennstelle sichtbar ist. Die Symbole der in Schaltanlagen eingesetzten Schalterarten sind in Abb. 87.95 gezeigt.

Typischerweise werden vor und nach den Leistungsschaltern Trennschalter eingesetzt, um sichtbare Trennstellen für Wartungsarbeiten an Schaltanlagen zu schaffen. In Abb. 87.96 links ist ein typischer Kabelabzweig mit einem Leistungsschalter, den Trennschaltern zur Sammelschiene und zur Leitungsseite, dem Arbeitserder und den Strom- und Spannungswandlern gezeigt. Werden ausfahrbare Leistungsschalter genutzt, liegt nach dem Abschalten eine sichtbare Trennstelle vor und es sind keine zusätzlichen Trennschalter erforderlich, siehe Abb. 87.96 rechts.

87.4.3.3.3 Technologien und Einsatzbereiche von Schaltern

Niederspannung

Für die Niederspannung werden offene luftisolierte Leistungsschalter (Air Circuit Breaker – ACB), Kompaktleistungsschalter (Moulded Case Circuit Breakers – MCCB), Lasttrennschalter (Load Break Switches – LBS) mit und ohne vorgeschaltete Sicherung sowie Leitungsschutzschalter (Miniatur Circuit Breaker – MCB) eingesetzt. Die Leistungsschalter, auch als Motorschutzschalter bezeichnet, sind in ihrer thermischen oder magnetischen Auslösung einstellbar.

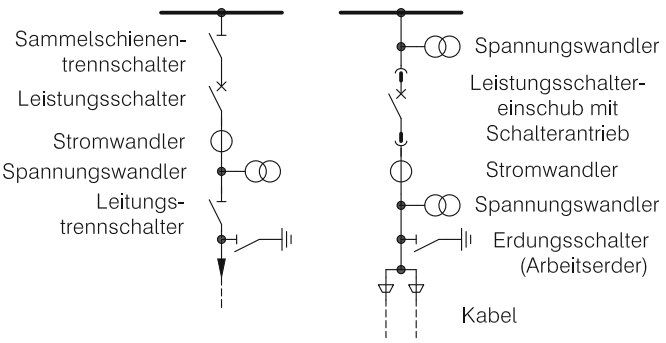


Abb. 87.96 Typischer Aufbau eines Abzweigs, *links*: mit Leistungsschalter, *rechts*: mit ausfahrbarem Leistungsschalter

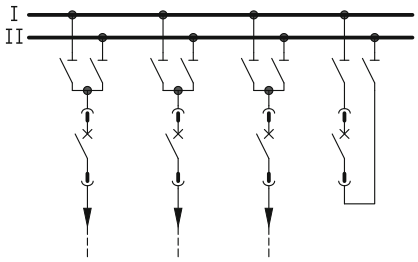


Abb. 87.97 Doppelsammelschienensystem

Mittelspannung

Hier werden bis 40 kV bevorzugt Vakuumschalter eingesetzt, die das kleinste Bauvolumen auf der Mittelspannung aufweisen. Nach der Paschenkurve liegt bei Vakuum das höchste Isoliervermögen luftisolierter Anordnungen vor. Außerdem weisen Vakuumschalter ein gutmütiges Kurzschlussverhalten im Vergleich zu den ebenfalls nutzbaren SF_6-Schaltern auf.

Hoch- und Höchstspannung

Auf der Hoch- und Höchstspannungsebene werden in Freiluftanlagen luftisolierte, Schalter und gekapselte Schalter verwendet. Bei Freiluftschaltern muss wegen der geringen Isolierfähigkeit von Luft die Isolierstrecke sehr groß dimensioniert werden, der Lichtbogen brennt frei und dehnt sich aufgrund der thermischen Effekte nach oben aus. Für Trennschalter wie Scheren- oder Drehtrennschalter ist dies unproblematisch, da sie leistungslos geschaltet werden. Bei hohen Spannungen kommt es vorrangig auf ein gutes Löschverhalten bei Lichtbögen an. Dafür werden gekapselte Schalter in der Ausführung als Druckluftschalter, SF_6-Schalter und seltener Ölschalter eingesetzt. Bei Druckluftschaltern wird der Lichtbogen von Luft mit einem Druck von 30 bar angeströmt, dabei gekühlt und im Weg verlängert und dadurch ausgeblasen. Dies gilt in ähnlicher Weise mit kleineren Volumina bei der Nutzung von SF_6. Bei den seit den siebziger Jahren durch SF_6-Schalter abgelösten Ölschaltern verdampft durch den Lichtbogen ein kleiner Teil des Öls und erzeugt damit eine Strömung, die den Löschvorgang einleitet.

87.4.3.4 Topologien von Schaltanlagen

Bei größeren Schaltanlagen werden die in Abb. 87.97 gezeigten Doppelsammelschienen verwendet. Hier sind Reinigung, Revision und Reparatur ohne Betriebsunterbrechungen realisierbar, da ein getrennter Betrieb der einzelnen Sammelschienen möglich ist. Im gezeigten Beispiel sind alle Leistungsschalter ausfahrbar und somit keine zusätzlichen Trennschalter erforderlich. Die vorhandenen Trennschalter dienen lediglich zur Vorwahl des Sammelschienensystems. Der Leistungsschalter-Zweig zwischen den Sammelschienen wird als Querkupplung bezeichnet. Bei noch größeren Anlagen werden auch Sammelschienen-Längstrennungen vorgesehen, wodurch sich die Flexibilität im Betrieb weiter erhöhen lässt.

Auch mit der in Abb. 87.98 gezeigten Umgehungssammelschiene US lässt sich die betriebliche Flexibilität weiter erhöhen. Indem die US-Sammelschiene über einen eigenen Abzweig

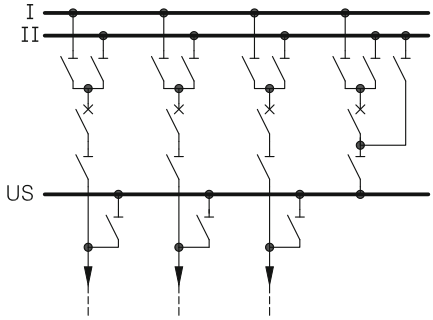


Abb. 87.98 Doppelsammelschienensystem mit Umgehungssammelschiene

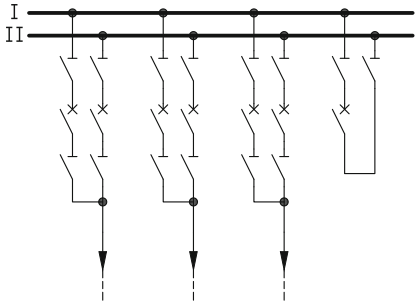


Abb. 87.99 Doppelsammelschienensystem mit 2-Leistungsschalter-Methode

versorgt wird, kann jeder Abzweig der Anlage ohne Unterbrechungen freigeschaltet werden. Wenn aufgrund zu hoher Kurzschlussströme galvanisch voneinander getrennte Teilnetze betrieben werden müssen, werden auch Dreifachsammelschienensysteme eingesetzt.

Bei der in Abb. 87.99 gezeigten Schaltanlage mit 2-Leistungsschalter-Methode sind für jeden Abzweig ein Leistungsschalter und zwei Trennschalter je Sammelschiene vorgesehen. Damit sind Freischaltungen der Sammelschienen und Abzweige sowie unterbrechungslose Umschaltungen zwischen den beiden Sammelschienen möglich.

Mit der in Abb. 87.100 gezeigten $1\frac{1}{2}$-Leistungsschalter-Methode wird im Vergleich zur 2-Leistungsschalter-Methode bei weniger Schaltern die gleiche Flexibilität erzielt. Auch hier sind Freischaltungen ohne Betriebsunterbrechungen möglich. Im Betrieb sind alle Schalter eingeschaltet, wodurch sich eine kostenoptimale Auslastung der Betriebsmittel ergibt. Die

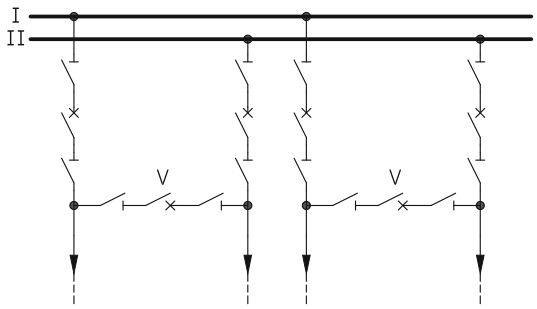


Abb. 87.100 Doppelsammelschienensystem mit $1\frac{1}{2}$-Leistungsschalter-Methode

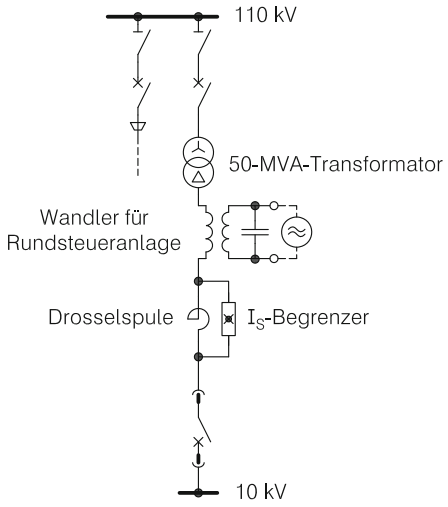


Abb. 87.101 Hochspannungsabzweig mit Transformator zur Mittelspannungsebene, Drosselspule mit parallel geschaltetem Stoßstrombegrenzer und Tonfrequenz-Rundsteueranlage (TRA)

Abzweige können über die Verbindungsschalter V zur jeweils anderen Sammelschiene durchgeschaltet werden.

87.4.3.5 Drosselspulen und Stoßstrombegrenzer

Zur Begrenzung des Kurzschlussstroms werden Drosselspulen eingesetzt, deren Induktivität erst im Fehlerfall wirksam wird. Um die ohmschen Verluste im Normalbetrieb zu vermeiden, werden ihnen Stoßstrombegrenzer parallel geschaltet, siehe Abb. 87.101. Dies ist ein pyrotechnisches Betriebsmittel, das aus einer Halterung mit einer parallel geschalteten Schmelzsicherung und einem Einsatz mit der Hauptstrombahn besteht. Die runde Hauptstrombahn ist in vier Segmente aufgeteilt, in der Mitte befindet sich eine Sprengkapsel. Bei Überschreitung eines Stoßstromwertes $\Delta i/\Delta t$ zündet, gesteuert über eine Auslöseelektronik, die Sprengkapsel. Dadurch werden die vier Segmente der Hauptstrombahn in der Mitte getrennt und nach beiden Seiten nach außen gedrückt. Der Strom kommutiert auf die parallele Sicherung, die kontrolliert abschmilzt, um einen Stromabriss und somit Überspannungen zu verhindern.

87.4.3.6 Tonfrequenz-Rundsteuerung in Schaltanlagen

In Abb. 87.101 ist auf der Mittelspannungsebene die Sendeanlage bzw. Einkopplungseinheit einer Tonfrequenz-Rundsteueranlage (TRA) gezeigt. Innerhalb der Mittel- und Hochspanungsnetze nutzen die Netzbetreiber solche Anlagen zur Übermittlung von Informationen, um z. B. Schaltgeräte anzusteuern, Beleuchtungseinrichtungen zu- und abzuschalten oder Tarifumschaltungen vorzunehmen. Das wird mit Hilfe von Steuerspannungen realisiert, die der Netzspannung aufmoduliert werden, was im Prinzip einer unidirektionalen Power Line Communication entspricht. Zur Nutzung der TRA werden eine Signaleinspeisung mit einer Sendeanlage und einer Ankopplungseinrichtung nach Abb. 87.101 sowie Empfängeranlagen an den Betriebsmitteln benötigt. Der Leistungsbedarf beim Betrieb von TRA ist hoch, bei Mittelspannung werden 80 bis 200 kVA und bei Hochspannung bis zu 2400 kVA für die Signaleinkopplung benötigt. Die Funktion der TRA kann nur gewährleistet werden, wenn die Steuerspannungen definierte Pegel aufweisen. Eine Abschwächung führt zum Totalausfall der TRA oder Fehlschaltungen. Tab. 87.20 zeigt die Nenn-Steuerfrequenzen der TRA und ihre oberen Sendepegel. Im Bereich dieser Frequenzen können nur geringfügige Oberschwingungspegel toleriert werden, die eine maximale Absenkung des Steuerspannungspegels von 10 bis 20 % nicht überschreiten. Als Alternative zur TRA kann auch die Funkrundsteuertechnik genutzt werden, die auf Langwellenfunkfrequenzen oder bei neueren Anlagen auch mit UKW-Funk betrieben wird.

Tab. 87.20 Nenn-Steuerfrequenzen und obere Sendepegel von TRA nach VDEW

Nenn-Steuerfrequenz in Hz	Oberer Sendepegel in %
110	1,7
168	1,7
183	3
194	2
206	2
217	3
228	2
270	3
283	3
317	3
383	3
425	3
485	4
600[a]	4
750	4
1050	3
1350	3
1600	2,5

[a] Einspeisung im Nullsystem

87.4.4 Netzformen und Schutz auf der Niederspannungsebene

Im Abschn. 87.3.1.3 wurden schon die elektrischen Größen und der allgemeine Aufbau von Drehstromnetzen beschrieben. Im Niederspannungsnetz werden die drei Phasenleiter L1, L2 und L3 sowie der Neutralleiter N zur Energieübertragung benötigt. Abhängig von der Netzform kann der Neutralleiter gleichzeitig die Funktion des Schutzleiters PE (protective earth) übernehmen und wird dann mit PEN bezeichnet. Auch die Art des Netzschutzes hängt von der Netzform ab.

87.4.4.1 Netzformen

Nach VDE werden die Netzformen TN, IT und TT unterschieden. Das TN-Netz wird weiter unterteilt in TN-S-, TN-C- und TN-C-S-Netze. In der Bezeichnung beschreibt der erste Buchstabe die Erdungsbedingung, T (Terra, Erde) steht für eine

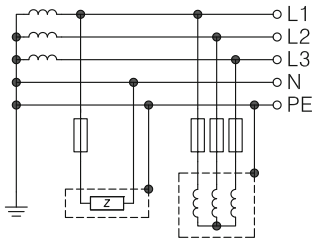


Abb. 87.102 Netzvariante Niederspannung TN-S mit separatem Neutral- und Schutzleiter

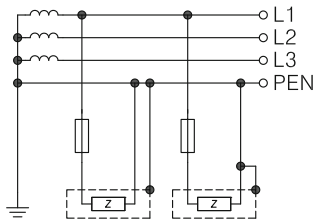


Abb. 87.103 Netzvariante Niederspannung TN-C mit kombiniertem Neutral- und Schutzleiter

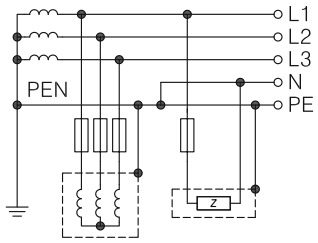


Abb. 87.104 Netzvariante Niederspannung TN mit kombiniertem und separatem Neutral- und Schutzleiter TN-C-S

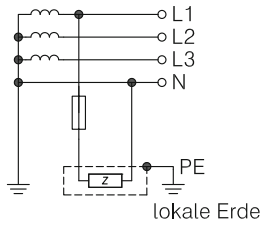


Abb. 87.105 Netzvariante Niederspannung TT mit lokalem Schutzleiter

niederohmige Erdung. Ist der erste Buchstabe ein I (isoliert), sind alle aktiven Teile von der Erdung isoliert. Mit dem zweiten Buchstaben wird die Art der Erdung von passiven leitfähigen Anlagenteilen beschrieben. Erfolgt die Erdung über den Neutral- oder Schutzleiter des Netzes, wird hierfür ein N verwendet. Wird unabhängig von einer bestehenden Erdung direkt lokal geerdet, wird dies mit einem T gekennzeichnet. In neuen Niederspannungs-Hausinstallationen ist die in Abb. 87.102 gezeigte Netzform TN-S mit separater Verwendung von Neutral- und Schutzleiter vorgeschrieben, da ansonsten der nach DIN VDE 0100-100 geforderte Fehlerstromschutz nicht realisiert werden kann, siehe Abschn. 87.4.4.2 „Personen- und Anlagenschutz“.

In einem TN-C-Netz wird ein PEN-Leiter verwendet, der gleichzeitig Neutralleiter N und Schutzleiter PE ist, siehe Abb. 87.103. Diese Netzform ist oft in älteren Hausinstallationen zu finden, für neue Installationen und nach wesentlichen Umbauten der elektrischen Hausanlage ist diese Netzform unzulässig, da kein Fehlerstromschutz realisiert werden kann. Darüber hinaus liegt bei unterbrochenem PEN-Leiter die Leiterspannung an leitfähigen Gehäusen von Verbrauchern an. In Niederspannungs-Verteilnetzen wird das TN-C-Netz weiter verwendet, allerdings muss der Leiterquerschnitt dabei 10 mm^2 bei Kupfer oder 16 mm^2 bei Aluminium betragen, damit ein Bruch des PEN-Leiters unwahrscheinlich wird.

Soll der Personenschutz in einem bestehenden TN-C Netz verbessert werden, kann der PEN-Leiter aufgetrennt werden und in Form von zwei separaten N- und PE-Leitern geführt werden. Dann liegt eine Kombination aus TN-C- und TN-S-Netz vor, die als TN-C-S bezeichnet wird, siehe Abb. 87.104. Die Erdung darf jetzt nur noch über den PE-Leiter erfolgen, da bei dessen Bruch im Gegensatz zur Verwendung des PEN-Leiters keine unzulässige Spannung an leitfähigen Gehäusen anliegen kann. Auch die Installation von Fehlerstrom-Schutzschaltern wird möglich. Es ist möglich, im TN-C-Netz ohne Veränderungen des Netzaufbaus speziell mit einem Fehlerstromschutz ausgerüstete Steckdosen zu installieren, bei denen die Auftrennung des PEN-Leiters und Fehlerstrom-Überwachung intern erfolgt. Hierbei bleibt der Personenschutz jedoch nur auf den einzelnen Steckdosenanschluss beschränkt.

Bei Leitungsführungen über große Entfernungen steigt der Widerstand mit der Leitungslänge an. Dabei kann das mitgeführte Sternpunktpotenzial des Neutralleiters über dem lokalen Erdpotenzial liegen und dann nicht mehr für die Schutzerdung geeignet sein. In diesem Fall ist ein TT-Netz mit einer lokalen Erdung erforderlich, die durch den zweiten Buchstaben T beschrieben wird, siehe Abb. 87.105. Der Anschluss aller passiven leitfähigen Teile muss jetzt direkt oder über den Schutzleiter PE an der lokalen Schutzerdung erfolgen, die meist durch den Fundamenterder gebildet wird. Die Größe des Erdungswiderstands wird durch Messung nachgewiesen, sie sollte wie auch bei TN-Netzen üblich unterhalb von 1 Ω liegen. Die Netzform TT ist bei weit abgelegenen Wohnsiedlungen und landwirtschaftlichen Betrieben zu finden.

Soll ein Netz mit erhöhter Verfügbarkeit betrieben werden und auch bei einem einpoligen Kurzschluss nicht abschalten, nutzt man ein IT-Netz (I-isoliert, T-Erde) mit einem isoliert geführten Schutzleiter, siehe Abb. 87.106. Die Verbraucher werden zwischen den Phasenleitern angeschlossen, der PE-Leiter wird für den Anschluss der Gehäuse und aller weiteren leitfähigen An-

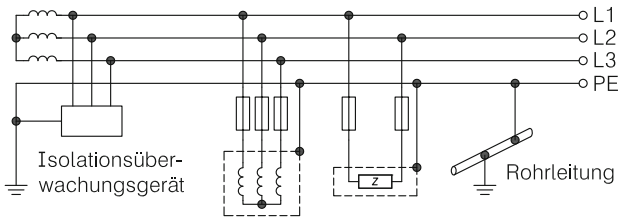


Abb. 87.106 Netzvariante Niederspannung IT mit isoliertem Schutzleiter

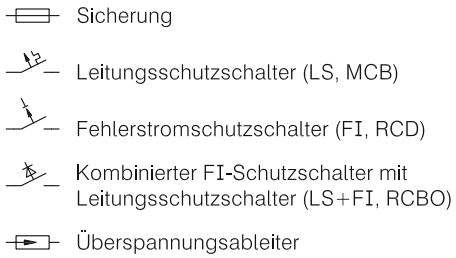


Abb. 87.107 Symbole von Schutzeinrichtungen

lagenbestandteile wie z. B. Rohrleitungen genutzt. Dies dient dem Potenzialausgleich zwischen den leitfähigen Teilen und ist zur Vermeidung unzulässiger Berührungsspannungen unbedingt erforderlich. Beim Auftreten eines Erdschlusses wird dieser über eine Isolationsüberwachung angezeigt, es erfolgt jedoch keine Abschaltung. Dazu würde es erst kommen, wenn ein zweiter Erdschluss eines anderen Phasenleiters zu einem Kurzschluss über den PE-Leiter führt. Wegen ihrer hohen Versorgungssicherheit wird diese Netzart im industriellen Bereich, in medizinischen Einrichtungen und bei Schiffsbordnetzen eingesetzt.

87.4.4.2 Personen- und Anlagenschutz

Es wird zwischen dem Personen- und Anlagenschutz unterschieden. Der Personenschutz umfasst den Schutz gegen unzulässige Berührungsspannungen sowie den Schutz gegen unzulässige Körperströme. Beim Anlagenschutz werden die Betriebsmittel gegen Zerstörung durch Kurzschlussströme, thermische Überlastung und Überspannungen geschützt. In Abb. 87.107 sind die Symbole der nachfolgend behandelten Schutzeinrichtungen gezeigt.

87.4.4.2.1 Personenschutz

Mit der zuvor im Abschn. 87.4.4.1 „Netzformen“ beschriebenen Schutzerdung mit Potenzialausgleich wird sichergestellt, dass keine unzulässigen Berührungsspannungen auftreten können. Zur Vermeidung unzulässiger Körperströme werden Fehlerstrom-(FI)-Schutzschalter eingesetzt, die auch als Residual-Current Circuit Breaker (RCCB) oder Residual-Current Device (RCD) bezeichnet werden. Die Phasen- und der Neutralleiter werden gemeinsam durch einen Stromwandler geführt, der Schutzleiter verläuft außerhalb. In Abb. 87.108 sind die Stromlaufpläne eines einpoligen und dreipoligen FI-Schutzschalters gezeigt. Im FI-Schalter wird die Stromsumme überwacht, die im Normalbetrieb null sein muss. Tritt ein Erdfehlerstrom auf, steigt die Stromsumme an und der FI-Schalter löst aus. Bei Auslöseströmen von 10 mA bis 30 mA liegt ein Personenschutz vor. Zur Vermeidung von Überhitzungen durch Fehlerströme und beim Anschluss von Geräten über lange Kabel mit kapazitiven Ladeströmen werden auch FI-Schalter mit höheren Ansprechwerten eingesetzt. Dieser muss von der Netzseite aus gesehen vor dem Überstromschutz, also vor den Leitungsschutzschaltern, installiert werden. Es werden auch kombinierte Leitungs- und FI-Schutzschalter eingesetzt, die als Residual

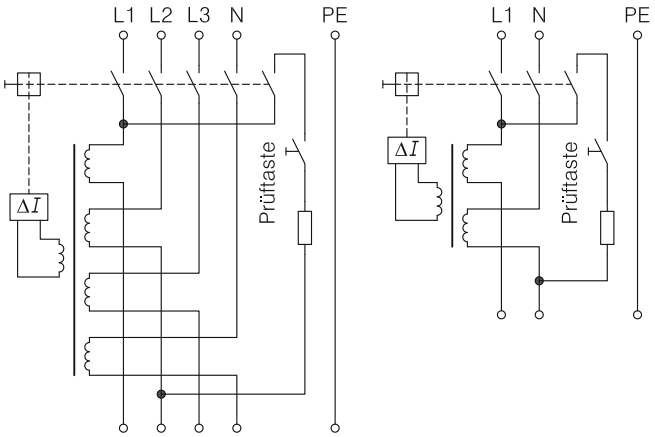


Abb. 87.108 Wirkprinzip und Stromlaufplan eines dreipoligen (*links*) und einpoligen (*rechts*) FI-Schutzschalters

current operated Circuit-Breaker with Overcurrent protection (RCBO) bezeichnet werden. Zu beachten ist, dass nach den Technischen Anschlussbedingungen (TAB) für Niederspannung nur Schutzgeräte eingesetzt werden dürfen, die ein Kurzschluss-Ausschaltvermögen von 6 kA aufweisen.

87.4.4.2.2 Anlagenschutz

Überstromschutz

Betriebsmittel können durch hohe Kurzschlussströme oder dauerhafte Überströme mit der dadurch verbundenen thermischen Überlastung geschädigt werden. Zur Sicherstellung einer ausreichenden Stromschutzfunktion muss das Grenzlastintegral der Schutzeinrichtung kleiner als das des zu schützenden Objekts sein:

$$\left(\int_0^{t_S} i^2(t)dt\right)_{\text{Schutz}} < \left(\int_0^{t_S} i^2(t)dt\right)_{\text{Objekt}} \qquad (87.117)$$

Dafür stehen Sicherungsautomaten (Leitungsschutzschalter) zur Verfügung, deren gebräuchlichste Auslösecharakteristiken in der Tab. 87.21 aufgeführt sind. Die in älteren Anlagen genutzten Schmelzsicherungen werden bei Neuinstallationen nicht mehr eingesetzt.

Sicherungsautomaten weisen jeweils eine im Zeit-Strom-Diagramm abfallende thermische und daran anschließende steile elektromagnetische Auslösekennlinie auf. In Abb. 87.109 sind die Strom-Zeit-Verläufe für die Charakteristiken B, C, D und K gezeigt.

Überspannungsschutz

Betriebsfrequente Unter- und Überspannungen werden durch Relais zur Spannungsüberwachung abgeschaltet. Zum Schutz elektrischer Betriebsmittel gegen hohe transiente Spannungen wird ein Überspannungsschutz nach DIN VDE 0185 vorge-

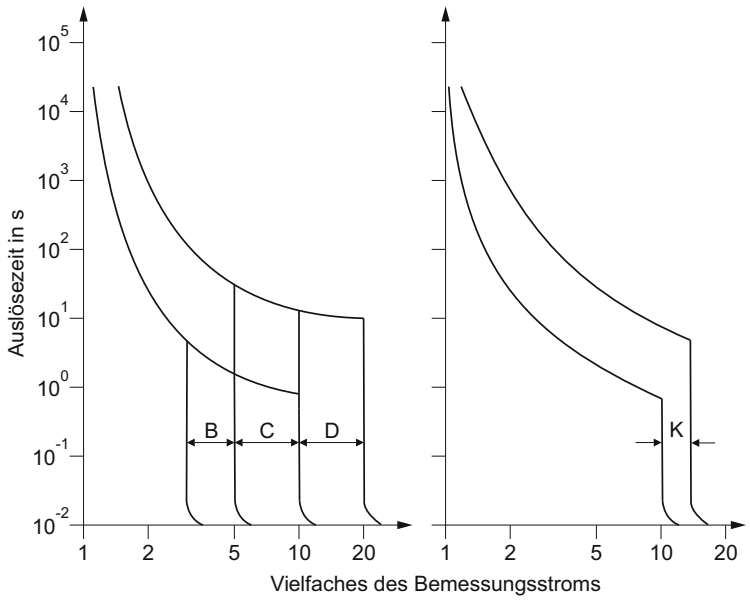


Abb. 87.109 Strom-Zeit-Kennlinien von Sicherungsautomaten

Tab. 87.21 Auslösecharakteristiken gebräuchlicher Sicherungsautomaten nach DIN VDE 0641/T11 und DIN VDE 0660-101

Charakteristik	Beschreibung	Thermische Auslösung bei	Elektromagnetische Auslösung bei
A	Nicht genormt: Halbleiterschutz, hohe Netzimpedanz	$1{,}13 \ldots 1{,}45 \cdot I_n$	$2 \ldots 3 \cdot I_n$
B	Standardschutz für Leitungen	$1{,}13 \ldots 1{,}45 \cdot I_n$	$3 \ldots 5 \cdot I_n$
C	Für höhere Einschaltströme	$1{,}13 \ldots 1{,}45 \cdot I_n$	$5 \ldots 10 \cdot I_n$
D	Für stark induktive oder kapazitive Lasten	$1{,}13 \ldots 1{,}45 \cdot I_n$	$10 \ldots 20 \cdot I_n$
E	„Exakt“, als Hauptleitungsschutzschalter	$1{,}05 \ldots 1{,}2 \cdot I_n$	$5 \ldots 6{,}25 \cdot I_n$
K	„Kraft“, für hohe Einschaltströme und sensible Überlastauslösung	$1{,}05 \ldots 1{,}3 \cdot I_n$	$8 \ldots 14 \cdot I_n$
Z	Halbleiterschutz, hohe Netzimpedanz	$1{,}05 \ldots 1{,}2 \cdot I_n$	$2 \ldots 3 \cdot I_n$

sehen, der den Blitz- und Geräteschutz umfasst. Für den Geräteschutz werden drei aufeinander aufbauende Schutzstufen mit Überspannungsableitern realisiert, siehe Abb. 87.107 unten. Der Grobschutz vom Typ 1 (alt Klasse B) wird in der Gebäudeeinspeisung vorgesehen. Er hat die Aufgabe, je nach genutzter Schutztechnologie die in das Gebäude einlaufende Restspannung bei Blitzeinschlag auf 1300 bis 6000 V zu begrenzen. Die Aufgabe des darauf folgenden Mittelschutzes vom Typ 2 (alt Klasse C) ist die Begrenzung der Überspannungen auf 600 bis 2000 V in den Etagenverteilungen. Der in speziellen Steckdosen, als Zwischenstecker oder in den Geräten realisierte Feinschutz vom Typ 3 (alt Klasse D) schützt die Geräte gegen Betriebsüberspannungen.

Danksagung Dem Laboringenieur Herrn Endrik Waldhaim der Professur für Elektrische Energiesysteme der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg danke ich für die freundliche Unterstützung bei der Erstellung der Abbildungen.

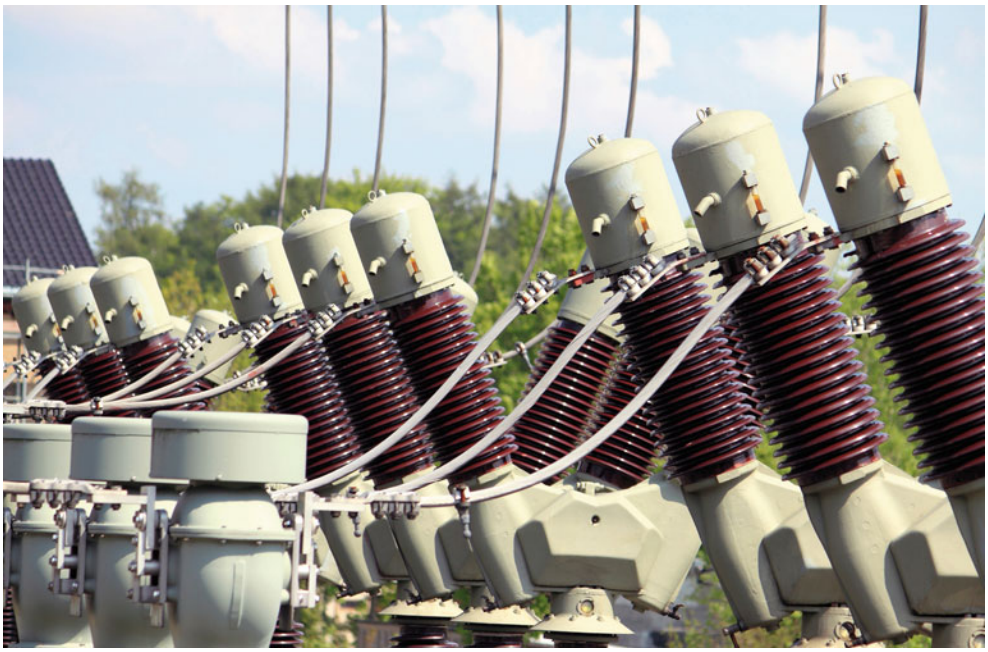
Literatur

1. Heuck, K.; Dettmann, K.-D.; Schulz, D.: Elektrische Energieversorgung, 9. Auflage, Wiesbaden: Springer Vieweg 2013
2. Constantinescu-Simon, L. (Hrsg.): Handbuch Elektrische Energietechnik, 2. Auflage, Wiesbaden: Vieweg 1997
3. Herold, G.: Elektrische Energieversorgung II, 2. Auflage, Wilburgstetten: J. Schlembach 2010
4. Crastan, V.: Elektrische Energieversorgung 1, 2. Auflage, Berlin, Heidelberg: Springer 2007
5. Kießling, F., Nefzger, P.,Kaintzyk, U.: Freileitungen, 5. Auflage, Berlin, Heidelberg: Springer 2001
6. Kämpfer, S., Kopatsch, G.: Schaltanlagen-Handbuch, 12. Auflage, Berlin: Cornelsen 2012
7. Konstantin, P.: Praxisbuch Energiewirtschaft, 2. Auflage, Berlin, Heidelberg: Springer 2009
8. Hosemann, G. (Hrsg.): Elektrische Energietechnik Band 3: Netze, Berlin, Heidelberg: Springer 2001
9. Schulz, D., Schulz, K.: Energiequellen und Kraftwerke, In: Informationen zur politischen Bildung Heft 319 Energie und Umwelt, Bonn: Bundeszentrale für politische Bildung 2013, Ausgabe 3/2013
10. Schulz, D.: Integration von Windkraftanlagen in Energieversorgungsnetze - Stand der Technik und Perspektiven für die dezentrale Stromerzeugung. Habilitationsschrift, Berlin, Offenbach: VDE-Verlag 2006

Elektrische Energieanwendung

88

Reinhard von Liebenstein



88.1 Kompensationsanlagen 1076
88.2 Beleuchtungsanlagen 1077
Literatur 1081

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_88

88.1 Kompensationsanlagen

In elektrischen Wechsel- und Drehstromnetzen entsteht neben der Wirkleistung auch Blindleistung, die im Netz zwischen Verbraucher und Erzeuger hin und her pendelt. Der durch Blindleistung verursachte Strom belastet die Zuleitungen, erzeugt Spannungsfälle auf den Zuleitungen und Stromwärmeverluste. Es ist daher anzustreben, möglichst wenig umhervagabundierende Blindleistung auf dem Netz zu haben. Einige Verbraucher benötigen aber Blindleistung (Motore, Transformatoren, auch Leitungen). Deshalb muss versucht werden, diese Blindleistung am Ort des Bedarfs bereitzustellen, ohne die Zuleitungen damit zu belasten. Da die meisten Verbraucher induktive Blindleistung benötigen, bietet sich die kapazitive Kompensation dieser Blindleistung an. Für die Leistungen gelten folgende Gleichungen (88.1) bis (88.3):

$$\textit{Wirkleistung} \qquad P = S \cdot \cos\varphi \qquad (88.1)$$

$$\textit{Blindleistung} \qquad Q = S \cdot \sin\varphi \qquad (88.2)$$

$$\textit{Blindleistung} \qquad Q = P \cdot \tan\varphi \qquad (88.3)$$

Aus den Formeln ist ersichtlich, dass der Winkel φ (Leistungsfaktor $\cos\varphi$) ein Maß für die Blindleistung ist. Der Winkel φ gibt eine Verschiebung zwischen Strom und Spannung wieder. Ist diese Verschiebung groß, wird im Netz viel Blindleistung benötigt. Es ist daher anzustreben, einen möglichst kleinen Winkel zwischen Strom und Spannung zu fahren. Da in den Netzen unterschiedliche Verbraucher verschiedene Lastwechsel aufweisen, verändert sich bei jedem Lastwechsel der Leistungsfaktor $\cos\varphi$ und somit auch der Winkel zwischen Strom und Spannung.

Es bieten sich drei Kompensationsarten an, und zwar die *Einzelkompensation*, die *Gruppenkompensation* und die *Zentralkompensation*.

Bei der Einzelkompensation (Abb. 88.1) wird der Kondensator direkt am Betriebsmittel angeschlossen und ist nur mit diesem Betriebsmittel gemeinsam am Netz.

Bei der Gruppenkompensation (Abb. 88.2) wird die Kompensationseinrichtung jeweils einer Verbrauchergruppe zugeordnet.

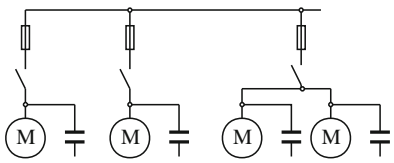


Abb. 88.1 Einzelkompensationsanlage

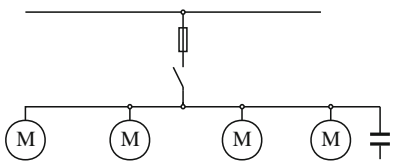


Abb. 88.2 Gruppenkompensationsanlage

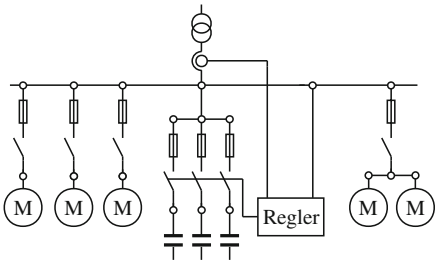


Abb. 88.3 Zentralkompensationsanlage

Die Zentralkompensation (Abb. 88.3) wird überwiegend als Blindleistungsregelanlage eingesetzt.

Sie ist normalerweise direkt einer Schaltanlage zugeordnet und kann verschiedene Kondensatorgruppen ein- oder ausschalten.

Welche dieser Kompensationsarten die für den jeweiligen Zweck geeignete ist, hängt von den Verbrauchern und der Belastungsart ab. Bei großen Verbrauchern mit konstantem Leistungsbedarf und großer Einschaltdauer ist eine Einzelkompensation wirtschaftlich. Bei vielen kleinen Verbrauchern mit unterschiedlichem Leistungsbedarf und wechselnder Einschaltdauer ist eine Zentralkompensation vorteilhafter. Die Einzelkompensation wird hauptsächlich bei größeren Motoren und Transformatoren eingesetzt. Hierdurch werden die Zuleitungen und Schaltelemente vom Blindstrom entlastet. Bei der Kompensation eines Motors ist darauf zu achten, dass bei Teilbelastung keine Überkompensation und beim Auslauf nach erfolgter Abschaltung keine Selbsterregung des Motors auftritt. Deshalb wird der Motor nur bis 90 % der Leerlaufblindleistung kompensiert. Im Vollastbetrieb ergibt diese Kompensation einen Leistungsfaktor von $\cos\varphi \approx 0{,}9$, im Leerlauf etwa 0,95 bis 0,98. Die erforderliche Kondensatorleistung ist

$$Q_C = 0{,}9 \cdot \sqrt{3} \cdot U \cdot I_0 \qquad (88.4)$$

I_0 Leerlaufstrom des Motors.

Eine weitere Möglichkeit, die Kondensatorleistung zu ermitteln, ist eine aus der Praxis abgeleitete Formel, wonach die Kondensatorleistung ungefähr das 0,35fache der Nennleistung des Motors betragen sollte ($Q_C \approx 0{,}35 \cdot P_N$). Anhand der Tab. 88.1

Tab. 88.1 Faktor zur Ermittlung der erforderlichen Kondensatorleistung je kW Wirkleistung

Vorhandener $\cos\varphi$	Gewünschter $\cos\varphi$				
	0,80	0,85	0,90	0,92	0,94
0,50	0,98	1,11	1,25	1,31	1,37
0,60	0,58	0,71	0,85	0,91	0,97
0,70	0,27	0,40	0,54	0,60	0,66
0,80	–	0,13	0,27	0,33	0,39
0,82	–	0,08	0,21	0,27	0,33
0,84	–	0,03	0,16	0,22	0,28
0,86	–	–	0,11	0,17	0,23
0,88	–	–	0,06	0,11	0,17
0,90	–	–	–	0,06	0,12
0,92	–	–	–	–	0,06

kann, ausgehend von dem vorhandenen Leistungsfaktor, die erforderliche kapazitive Blindleistung ermittelt werden, um den gewünschten Leistungsfaktor zu erhalten. Hierzu muss die vorhandene Wirkleistung mit dem ermittelten Faktor multipliziert werden.

Beispiel

Der vorhandene Leistungsfaktor von $\cos\varphi_1 = 0{,}6$ bei $P = 200\,\text{kW}$ soll auf einen Leistungsfaktor $\cos\varphi_2 = 0{,}9$ verbessert werden. Welche Kondensatorleistung muss dem Netz hinzugeschaltet werden?

$$Q_C = P \cdot \text{Faktor} = 200 \cdot 10^3\,\text{V A} \cdot 0{,}85 = 170 \cdot 10^3\,\text{V A}$$

Die Kondensatorleistung müsste dann 170 kvar betragen. ◀

Bei Festkompensation von Transformatoren kann der Kondensator sowohl auf der Oberspannungsseite wie auch auf der Unterspannungsseite angeschlossen werden. Nach den Richtlinien vom VDEW sind Kondensatorleistungen nach Tab. 88.2 auf der Unterspannungsseite des Transformators anzuschließen.

Werden Kondensatoren an der Unterspannungsseite des Transformators installiert, kann es bei Wegfall der induktiven Last zu Spannungserhöhungen kommen, wenn die Kondensatoren eingeschaltet bleiben. Nach (88.5) kann die Spannungserhöhung hinreichend genau berechnet werden.

Spannungserhöhung

$$\Delta u = u_k \cdot \frac{Q_C}{S_{NT}} \qquad \begin{array}{c|c|c|c} \Delta u & u_k & Q_C & S_N \\ \hline \% & \% & \text{V A} & \text{V A} \end{array} \qquad (88.5)$$

Beispiel

Ein Transformator mit $S_N = 315\,\text{kV A}$, $u_k = 4\,\%$ und $U_N = 10/0{,}4\,\text{kV}$ wird mit einer Kondensatorleistung von $Q_C = 180\,\text{kvar}$ beschaltet. Welche Spannungsüberhöhung tritt im Leerlauf auf?

$$\Delta u = u_k \cdot \frac{Q_C}{S_{NT}} = 4\,\% \cdot \frac{180 \cdot 10^3\,\text{V A}}{315 \cdot 10^3\,\text{V A}} = 2{,}28\,\% \quad ◀$$

Durch den Einsatz moderner Leistungselektronik sind in Drehstromnetzen sehr oft Oberwellen vorhanden. Durch Oberwellen kann es in kompensierten Netzen zu Resonanzerscheinungen kommen, die Überströme, erhöhte Spannungen und verzerrte Spannungsformen zur Folge haben.

In Netzen mit Oberwellen ist es deshalb erforderlich, Kompensationsanlagen mit verdrosselten Kondensatoren einzusetzen (Abb. 88.4). Bei verdrosselten Kondensatoren wird dem Kondensator eine Drosselspule vorgeschaltet. Die Resonanzfrequenzen des Kondensators und der Drosselspule sollten unterhalb der ersten Oberwelle des Netzes liegen.

Tab. 88.2 Erforderliche Kondensatorleistung in Abhängigkeit von Trafoleistung und Spannung

Bemessungsleistung des Transformators in kV A	Transformator mit Oberspannung Kondensatorleistung in kvar		
	5–10 kV	15–20 kV	25–30 kV
25	2	2,5	3
50	3,5	5	6
75	5	6	7
100	6	8	10
160	10	12,5	15
250	15	18	22
315	18	20	24
400	20	22,5	28
630	28	32,5	40

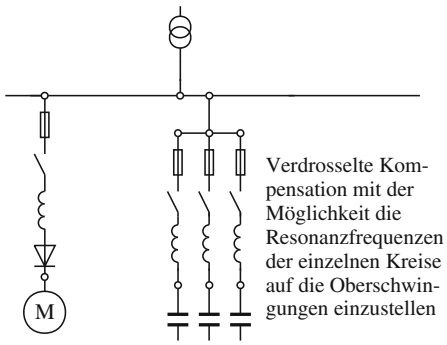


Abb. 88.4 Verdrosselte Zentralkompensationsanlage

Beispiel

Im Netz ist die 5. Oberwelle vorhanden, also 250 Hz. Die Resonanzfrequenz sollte also im Bereich von 200 bis 220 Hz liegen. Alle Frequenzen oberhalb der 5. Oberwelle erzeugen an der verdrosselten Einheit nur noch induktive Ströme. Im Netz ergeben sich keine Resonanzen mehr. ◀

Bei der Auslegung des Sicherungsorgans für Kondensatoranlagen sollte der Nennstrom der Sicherung den 1,6–1,8fachen Nennstrom der Kondensatoranlage haben.

88.2 Beleuchtungsanlagen

Beleuchtungsanlagen dienen dazu, Sicherheit und Arbeitsqualität zu erhöhen. Bei der Planung einer Beleuchtungsanlage ist darauf zu achten, dass die Beleuchtung dem jeweiligen Zweck unter optimalem Energieeinsatz entspricht. Wichtige Kriterien einer guten Beleuchtung sind Blendungsfreiheit, Leuchtdichteverteilung, Lichtfarbe, Gleichmäßigkeit und Farbwiedergabe. Kriterien zur Auswahl von Leuchten und Planung von Beleuchtungsanlagen werden von der „Fördergemeinschaft Gutes Licht“ entworfen und sind in der DIN 5035 sowie in der Publikation Nr. 3,5 der LiTG festgehalten.

Tab. 88.3 Mittelwerte für Leuchtdichten

Strahlendes Objekt	Leuchtdichte in cd/cm^2
Sonne	Ungefähr 150.000
Klarer Himmel	0,2 bis 1,2
Mond	ungefähr 0,25
Glühlampe klar	100 bis 2000
Leuchtstofflampe	0,3 bis 1,3
Xenon-Hochdrucklampe	15.000 bis 95.000
Gut beleuchtete Straße	$0{,}2 \cdot 10^{-3}$
Untere Grenze der Hellempfindlichkeit	$1 \cdot 10^{-8}$

88.2.1 Grundgrößen der Lichttechnik

Grundgrößen der Beleuchtungstechnik:

- *Lichtstrom* Φ in Lumen (lm),
- *Lichtstärke* I in Candela (cd),
- *Beleuchtungsstärke* E in Lux (lx).

Aus diesen Grundgrößen ergeben sich die Leuchtdichte und die Lichtausbeute nach (88.6) und (88.7).

$$\textit{Leuchtdichte} \qquad L = \frac{I}{A} \qquad \begin{array}{c|c|c} L & I & A \\ \hline \frac{cd}{m^2} & \text{cd} & \text{m}^2 \end{array} \qquad (88.6)$$

$$\textit{Lichtausbeute} \qquad \eta_{\text{Li}} = \frac{\Phi}{P} \qquad \begin{array}{c|c|c} \eta & \Phi & P \\ \hline \frac{\text{lm}}{W} & \text{lm} & \text{W} \end{array} \qquad (88.7)$$

Der *Lichtstrom* Φ ist das von der Lichtquelle nach allen Seiten abgestrahlte Licht. Die *Lichtstärke* I ist ein Maß für die Intensität der Lichtstrahlung. Die *Lichtausbeute* h_{Li} ist ein Maß für das Verhältnis von erzeugtem Lichtstrom zur eingesetzten elektrischen Leistung. Die *Leuchtdichte* L einer Lichtquelle oder einer beleuchteten Fläche ist für den im Auge entstehenden Lichtreiz verantwortlich und damit für den im Gehirn hervorgerufenen Helligkeitseindruck. Mittelwerte von Leuchtdichten sind in Tab. 88.3 zusammengefasst.

Die *Beleuchtungsstärke* E ist ein Maß für das auf eine Fläche treffende Licht. Die Beleuchtungsstärke beträgt 1 lx, wenn senkrecht einfallendes Licht mit einem Lichtstrom von 1 lm gleichmäßig auf eine Fläche von 1 m^2 trifft.

Da sich der Lichtstrom in der Praxis normalerweise nicht gleichmäßig über die Fläche verteilt, ist die nach der Flächenformel ermittelte Beleuchtungsstärke als *mittlere Beleuchtungsstärke* E_m aufzufassen. Die Beleuchtungsstärke in einem Punkt wird aus der Entfernung zur Lichtquelle und der Stärke der Lichtquelle berechnet (Abb. 88.5).

Beleuchtungsstärke

$$E = \frac{I}{r^2} \cdot \cos\varepsilon = \frac{I}{h^2} \cdot \cos^3\varepsilon \qquad (88.8)$$

Nach dem fotometrischen Entfernungsgesetz nimmt die Beleuchtungsstärke mit dem Quadrat der Entfernung ab. Dieses Gesetz gilt jedoch nicht für flächige Lichtquellen wie Leuchtenbänder oder Leuchtenflächen. Auch zusätzlich durch Begrenzungsflächen reflektiertes Licht wird durch die Punktbeleuchtungsformel nicht erfasst.

Tab. 88.4 Beleuchtungsstärken für verschiedene Raumarten

Art des Raumes	Beleuchtungsstärke in lx
Büroraum – normal	500
Büroraum – Großraumbüro	750 bis 1000
Zeichenraum	750
Verkehrswege in Gebäuden	100
Maschinenräume, Energieverteilräume	100
Räume zur Montage feinster Teile	1500

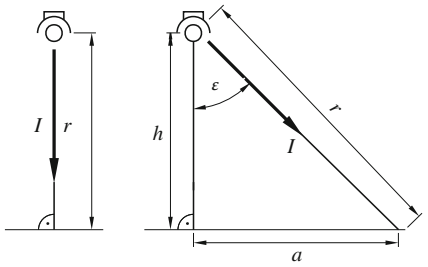


Abb. 88.5 Beleuchtungsstärke in einem Punkt bei senkrechtem und bei schrägem Lichteinfall

In DIN 5053 sind Beleuchtungsstärken für verschiedene Raumarten vorgegeben. Tab. 88.4 listet eine Auswahl der Raumarten auf.

Die Beleuchtung sollte so beschaffen sein, dass Gegenstände auf einer Arbeitsfläche gleichmäßig beleuchtet werden. Eine direkte oder indirekte Blendung sollte möglichst vermieden werden. Außerdem soll die Nennbeleuchtungsstärke in einem Raum (Raumzone) als Mittelwert in einer Höhe von 0,85 m über dem Fußboden konstant sein.

88.2.2 Lichtquellen

Die zur Lichterzeugung eingesetzten Lichtquellen werden als *Lampen* bezeichnet. Unter den Lampen gibt es eine große Auswahl, die von preiswerten Glühlampen mit eine Lichtausbeute von 20 lm/W bis hin zu Niederdruck-Natriumdampflampen mit einer Lichtausbeute von bis zu 183 lm/W reicht.

88.2.3 Glühlampen

Glühlampen stehen bis zu einer Leistungsaufnahme von 2000 W zur Verfügung. Das Glühlampenlicht wird zur allgemeinen Wohnraumbeleuchtung und für spezielle, stimmungsbetonte Lichteffekte eingesetzt. Die Lichterzeugung erfolgt durch Temperaturstrahlung eines auf etwa 2500 °C erhitzten Wolframdrahtes. Die Lebensdauer einer Glühlampe beträgt ungefähr 1000 h bei einer Lichtausbeute zwischen 9 bis 15 lm/W. Sie benötigt keine zusätzlichen Betriebsmittel. In neueren Glühlampen wird ein Halogen eingebracht. Die Lichterzeugung erfolgt wie bei der normalen Glühlampe; durch das Halogen wird eine

Kolbenschwärzung verhindert und die Lebensdauer der Lampe verdoppelt. Die Lichtausbeute liegt zwischen 15 und 25 lm/W. Die Halogen-Niedervoltlampen benötigen einen Transformator.

88.2.4 Leuchtstofflampen

Leuchtstofflampen haben Leistungsstufen zwischen 4 und 58 W. Die Lichterzeugung erfolgt durch elektrische Entladungen zwischen zwei Elektroden in einer Quecksilberdampfatmosphäre. Durch diese Entladungen wird unsichtbares UV-Licht erzeugt, das einen Leuchtstoff, der an der Innenwand der Leuchtstofflampe angebracht ist, zum Leuchten anregt. Die Lichtfarbe der Leuchtstofflampe kann je nach Leuchtstoffkombination verändert werden. Die Lichtausbeute dieser Lampen liegt zwischen 60 und 100 lm/W, die Lebensdauer bei ungefähr 7500 h. Eingesetzt werden Leuchtstofflampen für wirtschaftliche Beleuchtungszwecke in der Innen- und Außenbeleuchtung. Neben den Standard-Leuchtstofflampen werden die Kompakt-Leuchtstofflampen produziert. Kompakt-Leuchtstofflampen arbeiten wie die Standardlampen, sind jedoch erheblich kürzer oder flacher. Zur Erzeugung der Zündimpulse und zur Strombegrenzung werden Vorschaltdrossel und Starter benötigt (elektronisches Vorschaltgerät).

88.2.5 Entladungslampen

Unter Entladungslampen versteht man *Metalldampflampen*. Zu den Metalldampflampen gehören die Quecksilber-Hochdrucklampe (HQL), die Halogen-Metalldampflampe (HQI) und die Natrium-Hochdrucklampe (NAV).

Bei der *Quecksilber*-Hochdrucklampe wird die Lichterzeugung durch eine Entladung im Quecksilberdampf im Quarzbrenner bei Überdruck eingeleitet. Der Glaskolben ist mit einem rot strahlenden Leuchtstoff (Yttrium-Vanadat) versehen, der durch die UV-Strahlung zum Leuchten angeregt wird. Die Lichtausbeute dieser Lampe liegt bei 32 bis 60 lm/W, die Lebensdauer bei 9000 h. Zum Betreiben wird ein Vorschaltgerät benötigt. Der volle Lichtstrom wird nach ungefähr 4 bis 5 Minuten erreicht. Eine Wiederzündung ist erst nach Abkühlen der Lampe möglich.

Bei der *Halogen-Metalldampflampe* werden dem Quecksilber Halogenverbindungen bestimmter Metalle beigemischt. Die Lichterzeugung erfolgt durch Verdampfen der Metallhalogene in der Quecksilberatmosphäre. Die Lichtausbeute liegt zwischen 62 und 100 lm/W, die Lebensdauer zwischen 2000 und 6000 h. Zum Betrieb ist ein Vorschalt- und ein Zündgerät erforderlich. Der volle Lichtstrom steht nach 2 bis 4 Minuten zur Verfügung. Eine Wiederzündung kann nach 2 bis 15 Minuten erfolgen.

Die *Natrium-Hochdrucklampen* ergeben Licht, indem in einem Keramikbrenner Natriumdampfentladungen erzeugt werden. Der Keramikbrenner lässt in heißem Zustand Licht durch sein Kristallgitter, so dass eine starke Natrium-Linienstrahlung (monochromatisches Licht) entsteht. Die Lichtausbeute liegt zwischen 70 und 150 lm/W, die Lebensdauer bei 9000 h. Zum Betrieb ist ein Vorschaltgerät und bei einigen Typen ein Zündgerät erforderlich. Der volle Lichtstrom steht nach 5 bis 7 Minuten zur Verfügung, eine Wiederzündung erfolgt in weniger als einer Minute.

88.2.6 Leuchten

Leuchten haben die Aufgabe, den von einer Lichtquelle (Lampe) erzeugten Lichtstrom so zu beeinflussen, dass eine optimale Beleuchtung erreicht wird. Die Leuchte soll die Lampe und die übrigen elektromechanischen Innenteile gegen mechanische Beschädigung sowie vor Staub und Feuchtigkeit schützen. Bei der Auswahl der Leuchte, sorgfältiger Abstimmung von Beleuchtungsniveau, Lichtfarbe und Körperfarbe ist es empfehlenswert, sich anhand der Unterlagen der Hersteller zu vergewissern, dass die Leuchten nicht nur den lichttechnischen Anforderungen genügen, sondern auch den entsprechenden Gesetzen und Vorschriften über Leuchten und Beleuchtungsanlagen. Wichtige Gesetze und Vorschriften:

DIN VDE 0710, DIN VDE 0100; DIN VDE 0875 und DIN 40050.

Bei der Auswahl der Leuchte ist zu beachten, dass ein erheblicher Teil der Energie in Wärme umgewandelt wird. Diese Wärmeentwicklung ist bei der Auswahl und Montage der Leuchte zu berücksichtigen. In DIN VDE 0100 Teil 559 sind Kriterien niedergelegt, die der Errichter von Beleuchtungsanlagen beachten muss, damit die Anlage nicht Personen und Nutztiere durch gefährliche Körperströme und Sachen nicht durch zu hohe Temperaturen gefährdet. In der Norm wurden Klassen geschaffen, nach denen man die von der Leuchte ausgehende Gefährdung durch Wärmeentwickelung abschätzen kann. Folgende Einteilung wurde gewählt.

Kennzeichen 

Gilt nur für Leuchten mit Entladungslampen. Temperaturen von 130 °C im normalen Betrieb und 180 °C im Fehlerfall des Vorschaltgerätes werden nicht überschritten. Für direkte Montage auf Gebäudeteilen aus nichtbrennbaren, schwer- und normalentflammbaren Baustoffen nach DIN 4102 Teil 1.

Kennzeichen 

Gilt für Leuchten mit Glühlampen und Leuchten mit Entladungslampen. Leuchten mit begrenzter Oberflächentemperatur. Im Fehlerfall können keine Temperaturen auftreten, die zur Entzündung von brennbaren Stäuben und Fasern führen. Für Montage in Betriebsstätten geeignet, die durch brennbare Stäube und Fasern feuergefährdet sind.

Tab. 88.5 Reflexionsgrade

Farbe/Material	Reflexionsgrad	Farbe/Material	Reflexionsgrad
Weiß, lichtcreme	70–80 %	Hellelfenbein, RAL 1015	75 %
Hellgelb	55–65 %	Hellgrün, rosa	45–50 %
Pastellorange, RAL 2003	50 %	Himmelblau	40–45 %
Kieselgrau, RAL 7032	40 %	Braunbeige, RAL 1011	30 %
Ockergelb, olivgrün, hellbraun	25–35 %	Orange, zinnoberrot	20–35 %
Betongrau, RAL 7023	20 %	Dunkelgrün, -rot, -grau	10–15 %
Brillantblau, RAL 5067	10 %	Schwarz	4 %
Ahorn, Birke	50–60 %	Eiche hell	30–40 %
Nussbaum	15–20 %	Verputz, Gips	80 %
Fliesen, weiß	60–75 %	Ziegel, gelb	35–40 %
Beton	20–40 %	Ziegel, rot	10–25 %
Email, weiß	65–75 %	Aluminium, matt	55–60 %
Glas, klar	6–10 %		

Kennzeichen

Gilt für Leuchten und Entladungslampen. Temperaturen von 130 °C im normal Betrieb und 180 °C im Fehlerfall des Vorschaltgerätes werden an der Befestigungsfläche und an allen, der Leuchte benachbarten Flächen, nicht überschritten. Für Montage auf Einrichtungsgegenständen (Möbel, daher „M“) aus Werkstoffen, die in ihrem Brandverhalten nichtbrennbaren oder schwer- oder normalentflammbaren Baustoffen nach DIN 4102 Teil 1 entsprechen, selbst wenn sie beschichtet, furniert oder lackiert sind.

Kennzeichen 

Gilt für Leuchten mit Glühlampen und Leuchten mit Entladungslampen. Temperatur von 115 °C wird im Fehlerfall an der Befestigungsfläche sowie an allen der Leuchte benachbarten Stellen nicht überschritten. Für direkte Montage auf Einrichtungsgegenständen aus Werkstoffen, die in ihrem Brandverhalten nichtbrennbaren oder schwer- oder normalentflammbaren Baustoffen nach DIN 4102 Teil 1 entsprechen, auch wenn sie beschichtet, furniert oder lackiert sind.

88.2.7 Berechnung von Beleuchtungsanlagen

Bei der Berechnung von Beleuchtungsanlagen in geschlossenen Räumen spielt neben der Leuchtenauswahl und den Raumabmessungen auch die Raumgestaltung eine Rolle. Für die Auslegung der Raumbeleuchtung ist der *Raumwirkungsgrad* eine wichtige Größe. Er ergibt sich aus dem *Raumfaktor* und den *Reflexionsgraden* der Raumbegrenzung (Wände, Decken, Fußboden). Der Raumfaktor ist abhängig von den geometrischen Abmessungen des Raumes und deshalb sehr exakt zu berechnen. Für normale, rechteckige Räume gilt:

$$\text{Raumfaktor} \quad K_r = \frac{a \cdot b}{h \cdot (a + b)} \qquad \begin{array}{c|c|c} a & b & h \\ \hline \text{m} & \text{m} & \text{m} \end{array} \qquad (88.9)$$

Tab. 88.6 Beleuchtungswirkungsgrad η_{Bel} (%) in Abhängigkeit vom Raumfaktor und den Reflexionsgraden für eine Anbauspiegelraster-Leuchte

		Reflexionsgrad						
Decke		80	80	50	50	70	50	30
Wand		50	30	50	30	50	30	50
Boden (Nutzebene)		30	30	30	30	20	10	10
Raumindex	0,6	39	33	37	32	37	36	31
	0,8	48	42	46	41	46	44	39
	1,0	54	48	51	46	51	49	44
	1,25	61	55	57	53	57	54	50
	1,5	66	60	62	58	62	58	54
	2,0	71	66	66	62	66	61	58
	2,5	75	71	70	67	69	64	61
	3,0	79	75	73	70	72	67	64
	4,0	81	78	75	72	74	68	66
	5,0	84	81	77	75	76	70	67

Sind Räume rund oder n-eckig und nicht zu lang gestreckt, kann (88.10) angewendet werden:

$$K_r = \frac{A}{h \cdot 2 \cdot \sqrt{A}} \qquad \begin{array}{c|c} A & h \\ \hline \text{m}^2 & \text{m} \end{array} \qquad (88.10)$$

Soll in dem zu beleuchtenden Raum eine Leuchte mit indirekter Charakteristik eingesetzt werden, gilt (88.11):

$$K_r = \frac{3 \cdot a \cdot b}{2 \cdot h_1 \cdot (a + b)} \qquad (88.11)$$

K_r Raumfaktor; a Länge des Raumes; b Breite des Raumes; h Lichtpunkthöhe über Nutzebene; A Fläche des Raumes; h_1 Deckenhöhe über Nutzebene.

Die Bestimmung der Reflexionsgrade des Raumes ist in der Praxis nicht so ohne weiteres möglich, da die Raumgestaltung in der Planungsphase noch nicht eindeutig vorliegt. Es ist darum erforderlich, voraussichtliche Reflexionsgrade mit dem Bauherren abzustimmen. In Tab. 88.6 sind einige Reflexionsgrade aufgeführt.

Sind Raumfaktor und Reflexionsgrade bekannt, kann mit Hilfe der Herstellerlisten der *Beleuchtungswirkungsgrad* η_{Bel} be-

Tab. 88.7 Einteilung der Farbwiedergabeeigenschaften

Stufe der Farbwiedergabe	Beschreibung
1	Sehr gute Farbwiedergabe
1A	Erfüllt höchste Ansprüche
1B	Erfüllt hohe Ansprüche
2	Gute Farbwiedergabe
2A	Erfüllt die Ansprüche der üblichen Sehaufgaben in Büroräumen und Werkstätten
2B	Erfüllt viele Ansprüche
3	Weniger gute Farbwiedergabe
4	Die Erkennung von Farben ist stark eingeschränkt oder gar nicht möglich

Tab. 88.8 Lichtfarben und die Farbwiedergabeeigenschaft

Lichtfarbe	Ähnlichste Farbtemperatur	Stufe der Farbwiedergabe	Handelsbezeichnungen	
			Osram-Typ	Philips-Typ
Warmweiß (ww)	2700 K	1B	41	82
		1A	–	92
	3000 K	3	30	29
		1B	31	83
		1A	32	93
Neutralweiß (nw)	4000 K	2B	20	33
		2A	25	25
		1B	21	84
		1A	22	94
Tageslichtweiß (tw)	5000 K	1B	11	95

stimmt werden. Er ist das Produkt aus Leuchtenwirkungsgrad η_{Lb} und Raumwirkungsgrad η_R.

Beleuchtungswirkungsgrad $\eta_{\mathrm{Bel}} = \eta_R \cdot \eta_{\mathrm{Lb}}$ (88.12)

Die für einen Raum erforderliche Leuchtenzahl kann mit Hilfe von (88.13) ermittelt werden.

Leuchtenzahl

$$n = \frac{p \cdot E_m \cdot A}{\eta_{\mathrm{Bel}} \cdot z \cdot \Phi} \qquad \begin{array}{c|c|c|c|c|c|c} n & p & E & A & z & \eta_{\mathrm{Bel}} & \Phi \\ - & - & \frac{\mathrm{lm}}{\mathrm{m}^2} & \mathrm{m}^2 & - & - & \mathrm{lm} \end{array} \quad (88.13)$$

p Planungsfaktor > 1, berücksichtigt die Alterung; z Anzahl der Lampen pro Leuchte.

Die hier gemachten Angaben und Formeln geben nur einen kleinen Überblick in die Auslegung von Beleuchtungsanlagen. Für das einwandfreie Erkennen von Gegenständen ist die Lichtfarbe der Lichtquellen von großer Bedeutung. Außerdem wird dadurch ein harmonisches Farbklima erreicht, das sich auf Stimmung und Wohlbefinden des Menschen positiv auswirkt. Die Lichtfarbe der Lampen hängt stark von der Erzeugung des Lichtes ab. Damit unterschiedliche Lampen in ihrer Lichtfarbe verglichen werden können, wird sie durch die Farbtemperatur gekennzeichnet. Ferner wird eine Einstufung der Güte der Farbwiedergabe unternommen. In Tab. 88.8 sind die genormten Stufen der Farbwiedergabe aufgeführt.

Beispiel

Für einen Büroraum mit den Abmessungen Breite $a = 5\,\mathrm{m}$, Länge $b = 7\,\mathrm{m}$, Leuchtenhöhe über Nutzebene $h = 1{,}95\,\mathrm{m}$ und Raumhöhe $h_R = 2{,}8\,\mathrm{m}$ soll eine Beleuchtungsanlage ausgelegt werden. Für Büroräume beträgt die Nennbeleuchtungsstärke $E_m = 500\,\mathrm{lx}$.

Für den Raumfaktor ergibt sich bei direkter Lichtverteilung

$$K_r = \frac{a \cdot b}{h \cdot (a + b)} = \frac{5\,\mathrm{m} \cdot 7\,\mathrm{m}}{1{,}95\,\mathrm{m} \cdot (5\,\mathrm{m} + 7\,\mathrm{m})} = 1{,}49 \approx 1{,}5$$

Mit einer Leuchtstofflampe der Lichtfarbe weiß wird vom Hersteller ein Lichtstrom von 5400 lm angegeben. Die Reflexionsgrade der Decke (70 %), der Wände (50 %) und des Fußbodens (20 %) werden geschätzt. Für eine Spiegelrasterleuchte zum Deckenanbau ergibt sich entsprechend dem Raumfaktor und den Reflexionsgraden ein Beleuchtungswirkungsgrad von $\eta_{\mathrm{Bel}} = 0{,}62$. Die Anzahl der Leuchten berechnet sich jetzt zu

$$n = \frac{p \cdot E_m \cdot A}{\eta_{\mathrm{Bel}} \cdot z \cdot \Phi} = \frac{1{,}25 \cdot 500\,\frac{\mathrm{lm}}{\mathrm{m}^2} \cdot (5\,\mathrm{m} \cdot 7\,\mathrm{m})}{0{,}62 \cdot 1 \cdot 5400\,\mathrm{lm}} = 6{,}5 \text{ Leuchten}$$

Gewählt werden 7 Leuchten. ◄

Literatur

1. *Heuck/Dettmann/Schulz:* Elektrische Energieversorgung. Wiesbaden: Vieweg, 7. Auflage 2007
2. *div. Verfasser:* Schalten, Schützen, Verteilen in Niederspannungsnetzen. München: Publicis MCD; Siemens AG, 1997

Leistungselektronik XII



89 Halbleiter in der Leistungselektronik 1085
90 Leistungshalbleiter 1087
91 Einsatz von Leistungshalbleitern als Schalter 1109

Halbleiter in der Leistungs-elektronik

89

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_89

Die in diesem Buch im Teil V „Elektronik“ behandelten Elektronischen Bauelemente „Dioden“, „Transistoren“, „ Thyristoren“ und „Triacs“ finden Anwendung bei Strömen bis etwa 5 A und bei Spannungen bis etwa 100 V. Werden diese Werte wesentlich überschritten, befindet man sich im Anwendungsgebiet der Leistungselektronik. Es hat sich gezeigt, dass die genannten Bauelemente auch hier vorteilhaft eingesetzt werden können, wenn man sie den gegebenen Betriebsbedingungen anpasst. Dazu gehören u. a. die höheren elektrischen Feldstärken, die höheren Stromdichten sowie die höheren Verlustleistungen und die damit verbundenen Wärmeableitprobleme. Außerdem gibt es spezielle Schaltungen, die vornehmlich in der Energietechnik angewendet werden.

Elektronische Bauelemente werden in der Energietechnik bevorzugt eingesetzt als Schalter und als Signalverstärker.

Beispiel für den Einsatz als Schalter: Die Leistungsabgabe von z. B. Motoren kann durch Ansteuerung mit Rechteck-Spannungsimpulsen mit variablem Tastgrad $t_{\text{ein}}/T_{\text{Periode}}$ eingestellt werden (Gleichstromsteller). Dabei sind die mechanischen Rückwirkungen auf den Motor umso geringer, je höher die Frequenz $f = 1/T_{\text{Periode}}$ ist.

Beispiele für den Einsatz als Verstärker: Die Endstufen von Sendern mit Frequenzen im kHz- und MHz-Bereich können mit Leistungstransistoren aufgebaut werden; „Uhrensender“ DCF 77 mit einer Leistung von etwa 20 kW; terrestrische Sender für Digital-Fernsehen mit einer Leistung von z. B. 5 kW; Audio-Verstärker mit einer Leistung von 10 kW bei Musik-Großveranstaltungen

Die in der Leistungselektronik eingesetzten Bauelemente und die daraus aufgebauten Schaltungen sollten so betriebssicher aufgebaut werden, dass Ausfälle auf ein Minimum reduziert werden. Die Bauteile an sich sind dabei nicht der Kostenfaktor, vielmehr sind es die Ausfälle der angeschlossenen Geräte und die Folgeschäden, die enorme Kosten verursachen können.

Es sei darauf hingewiesen, dass die Hersteller von Leistungshalbleitern Datenblätter und Applikationsberichte vorrätig haben, die wertvolle Informationen z. B. zu den Themen Schaltungsbeispiele, Dimensionierung und Betriebssicherheit enthalten Besonders zur Optimierung der Leistungshalbleiter-Ansteuerung sollten diese Unterlagen zu Rate gezogen werden. Sie können oftmals direkt aus dem Internet geladen werden.

Im Folgenden werden zunächst bekannte Halbleiter-Bauelemente dargestellt, siehe auch Teil V „Elektronik“, die allerdings für die Anwendung in der Leistungselektronik optimiert sind. Es folgt deren Einsatz als Schalter, der häufigsten Anwendung in der Leistungselektronik, einschließlich der zum ordnungsgemäßen Betrieb erforderlichen Peripherie.

Leistungshalbleiter

90

Wilfried Plaßmann



90.1 Halbleiterdioden 1088
90.2 Bipolartransistoren 1091
90.3 Feldeffekttransistoren 1097
90.4 IGBT-Transistoren 1100
90.5 Thyristoren 1101

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_90

90.1 Halbleiterdioden

90.1.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte

Den prinzipiellen Aufbau einer *Halbleiterdiode* zeigt Abb. 90.1. Sie besteht in der einfachsten Ausführung aus einem PN-Übergang. Bei positivem, konstantem Potential der Anode (A) gegenüber der Kathode (K) ($U_{AK} > 0$) fließt der Durchlassstrom i_F, der von der angelegten Spannung und der in Reihe liegenden Impedanz $\underline{Z}$ (Last) bestimmt wird. Ist dagegen das Potenzial der Anode gegenüber der Kathode negativ ($U_{AK} < 0$), fließt der Sperrstrom i_R, der durch die Eigenschaften der Diode bestimmt wird und wesentlich von der Dotierung, der angelegten Spannung und der Diodentemperatur abhängt. Abb. 90.2 zeigt schematisch den Zusammenhang zwischen der Spannung u_F bzw. u_R an der Diode (Index F für Flussrichtung, Durchlassrichtung, Vorwärtsrichtung; Index R für Reverse, Sperrrichtung, Rückwärtsrichtung) und dem Strom i_F bzw. i_R in Durchlass- und Sperrrichtung. Die an der Diode auftretende Spannung in Durchlassrichtung liegt bei Siliziumdioden und den in der Leistungselektronik auftretenden Strömen im Bereich 1 ... 2,5 V. Bei einem Effektivwert des durch die Diode fließenden Stromes von z. B. 1000 A und einer Spannung in Durchlassrichtung von 2,5 V entsteht an der Diode eine Verlustleistung von 2,5 kW. Die dadurch entstehende Wärme muss, mit einigem Aufwand für die Kühlvorrichtung, abgeführt werden. Für den praktischen Einsatz der Diode sind die folgenden *Grenzwerte* wichtig:

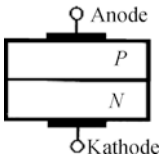


Abb. 90.1 Schematischer Aufbau einer Halbleiterdiode

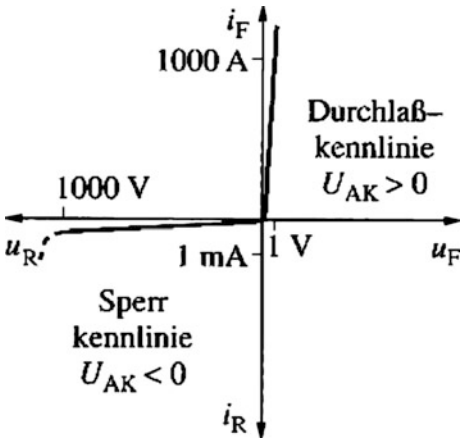


Abb. 90.2 Kennlinie einer Halbleiterdiode

- *Spitzensperrspannung* U_{DRM}: höchstzulässiger Augenblickswert der auftretenden Spannung im Sperrzustand in Vorwärtsrichtung;
- *Spitzensperrspannung* U_{RRM}: wie U_{DRM}, aber in Rückwärtsrichtung.

Da Dioden häufig bei sinusförmigen Spannungsverläufen mit einer Frequenz von 50 Hz eingesetzt werden, gibt man für diesen Betriebsfall den

- *Dauergrenzstrom* $I_{FAV(1)}$

an; der Index (1) steht für eine definierte Gehäusetemperatur, die dem Datenblatt zu entnehmen ist.

90.1.2 Dynamisches Verhalten, TSE

Das *dynamische Verhalten* gibt Auskunft über den zeitlichen Verlauf von Dioden-Spannung und -Strom bei Laständerung oder Änderung der den Lastkreis speisenden Spannung u_0. Besonders anschaulich lässt es sich einer Darstellung entnehmen, bei der die Spannung U_0 rechteckförmigen Verlauf mit Polaritätswechsel besitzt. Abb. 90.3 gibt den Verlauf von u_F, u_R, i_F und i_R in Abhängigkeit von U_0 wieder. Die eingetragenen Größen haben folgende Bedeutung:

U_0, u_0 die den Lastkreis speisende Spannung, hier rechteckförmiger Verlauf mit Polaritätswechsel,
u_F Spannung an der Diode in Vorwärtsrichtung,
u_R Spannung an der Diode in Rückwärtsrichtung,
i_F Strom durch die Diode in Vorwärtsrichtung,
I_R Strom durch die Diode in Rückwärtsrichtung,
u_{FM} Scheitelwert der Spannung im Sperrzustand in Vorwärtsrichtung,
u_{RM} wie u_{Fm}, aber in Rückwärtsrichtung,
i_{FM} Scheitelwert des Diodenstromes in Vorwärtsrichtung,
i_{RM} Scheitelwert des Diodenstromes in Rückwärtsrichtung,
t_{fr} Durchlassverzögerungszeit,
t_{rr} Sperrverzögerungszeit.

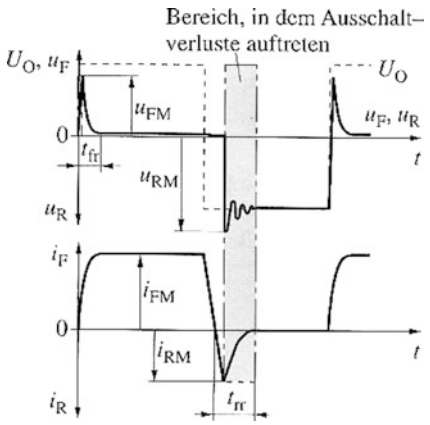


Abb. 90.3 Ein- und Ausschaltverhalten einer Halbleiterdiode

Nach Abb. 90.3 geht die Diode nach Anlegen einer Spannung in Durchlassrichtung innerhalb der Zeit t_{fr}, ca. 1 bis 20 µs, in den leitenden Zustand über. Dieses Verhalten wird durch die Größen u_{FM} und t_{fr} hinreichend genau beschrieben. Der Ausschaltvorgang beginnt mit dem Anlegen einer Spannung in Sperrrichtung. Der Stromfluss wird zunächst über den Nullpunkt hinaus in negativer Richtung fortgesetzt, da sich im Bereich des PN-Überganges noch ausreichend Ladungsträger befinden: ***T**räger**s**peicher**e**ffekt*, TSE. Erst wenn diese Ladungen abgebaut sind, verringert sich die Anzahl der am Stromtransport beteiligten Ladungsträger und führt zu einer Abnahme des Diodenstromes. Nach Ablauf der Zeit t_{rr}, ca. 10 … 60 µs, ist der Strom auf Werte abgesunken, die in der Regel vernachlässigbar sind.

Während des Betriebes können einzelne nicht periodische Stromimpulse auftreten, deren höchstzulässiger Augenblickswert durch den *Stoßstromgrenzwert* $I_{FSM(1)}$ festgelegt wird. Der Index (1) steht stellvertretend für vorgegebene Bedingungen für den zeitlichen Verlauf dieser Impulse, die dem Datenblatt zu entnehmen sind.

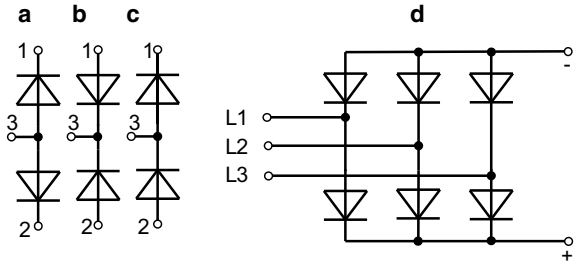


Abb. 90.4 Beispiele für Diodenmodule, **a** – **d** siehe Text

In den Datenblättern wird in der Regel ein Diagramm bzw. ein Wert für die maximal zulässige Verlustleistung des Halbleiters angegeben. Häufig wird ein Startwert (Diagramm) oder ein Wert (Tabelle) für eine Gehäusetemperatur von 25 °C des Halbleiters zugrunde gelegt, der aber mit den gängigen Kühlkörpern nicht erreichbar ist. Siehe dazu auch Abschn. 91.3 „Halbleiterkühlung“.

90.1.3 Verlustleistung

Beim Betrieb von Halbleiterdioden besteht eine wesentliche Aufgabe darin, die an der Diode entstehende Verlustwärme abzuführen. Dazu ist zunächst die an der Diode entstehende Verlustleistung P_V zu berechnen. Der weitaus größte Anteil entsteht im Durchlassbetrieb, der Anteil in Sperrrichtung kann i. Allg. vernachlässigt werden. Auf der sicheren Seite befindet man sich, wenn man den höchsten vorkommenden Diodenstrom $i_{F\,max}$ mit der zugehörigen Diodenspannung $u_{F\,max}$ multipliziert:

$$P_{V\,max} = u_{F\,max} \cdot i_{F\,max} \tag{90.1}$$

Wird der Halbleiter als Schalter eingesetzt mit dem Tastgrad t_0/T_0 und verlaufen u_F und i_F (nahezu) rechteckförmig mit diesem Tastgrad, kann P_V berechnet werden zu:

$$P_V \approx u_F \cdot i_F \cdot \frac{t_{0\,max}}{T_0} \tag{90.2}$$

Praktisch ist diese Formel für variables t_0 (z. B. Gleichstromsteller) mit der Nebenbedingung $0 < t_0 < T_0$, wobei für t_0 der größte mögliche Wert $t_{0\,max}$ einzusetzen ist.

Für periodische nicht rechteckförmige Signalverläufe von u_F und i_F muss P_V allgemein bestimmt werden nach der Formel

$$P_V = \frac{1}{T_0} \cdot \int_0^{T_0} u_F(t) \cdot i_F(t) dt \tag{90.3}$$

Wird der Halbleiter als schneller Schalter im oberen kHz- oder im MHz-Bereich eingesetzt, ist eine Überprüfung der zusätzlichen Verluste beim Übergang Sperren-Leiten und umgekehrt nach Abb. 90.2 und 90.3 sinnvoll.

90.1.4 Bauarten

Für spezielle Anwendungen sind Dioden entwickelt worden, die sehr kleine Sperrverzögerungszeiten t_{rr} (Schottkydioden) und kleine Ströme i_{RM} aufweisen. Sie besitzen dafür aber etwas höhere Durchlassspannungen und niedrigere Spitzensperrspannungen.

Werden *Hochleistungsdioden* bei niedrigen Frequenzen (z. B. 50 Hz) betrieben, wird zwischen die stark dotierten P- und N-Zonen ein schwächer dotiertes S-Gebiet eingefügt. Diese PSN-dotierten Dioden erreichen Durchlassspannungen von etwa 2,5 V bei 1000 A und Spitzensperrspannungen von etwa 4000 V.

Halbleiterdioden stehen für Ströme bis ca. 5000 A und Spannungen bis ca. 5000 V zur Verfügung. Für z. B. Stromrichter werden häufig mehrere Dioden benötigt. In einem Modul sind deshalb eine entsprechende Anzahl für einen bestimmten Verwendungszweck bereits zusammengefasst. Abb. 90.4 zeigt vier Beispiele. Modul b) findet Verwendung in der M2-Gleichrichterschaltung (Mittelpunktschaltung), c) als eine Hälfte einer B2-Gleichrichterschaltung oder in einer Spannungsvervielfacher-Schaltung, a) und b) zusammen in einer B2-Gleichrichterschaltung (Brückengleichrichter), d) in einer B6-Gleichrichterschaltung (3-Phasen-Gleichrichter). Die Dioden in a) und b) lassen sich parallel schalten und ähneln sich in ihren temperaturabhängigen Parametern, da sie auf einem Block thermisch gekoppelt untergebracht sind. Kennwerte zu a), b) und c): Spitzensperrspannung etwa 1400 V, maximaler Strom etwa 170 A (DD 171N-Series). Kennwerte zu d): Spitzensperrspannung etwa 1600 V, maximaler Strom etwa 200 A (A DD B6U 205 N16); [1, 2].

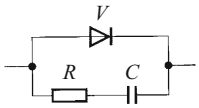


Abb. 90.5 TSE-Schutzbeschaltung

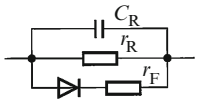


Abb. 90.6 Ersatzschaltbild einer Halbleiterdiode

90.1.5 Schutzschaltungen, TSE-Beschaltung

Sind die durch den Trägerspeichereffekt (TSE, s. Abschn. 90.1.2) vorhandenen Ladungen abgebaut, nimmt der Strom in Sperrrichtung innerhalb weniger µs von i_{Rm} auf nahezu Null ab (Abb. 90.3). Befinden sich im Lastkreis der Diode Verbraucher mit induktivem Verhalten, können durch diese großen Stromänderungen pro Zeiteinheit Spannungsimpulse auftreten, die den höchstzulässigen Augenblickswert der periodisch auftretenden Spannung, U_{RRM}, an der Diode überschreiten und sie unter Umständen zerstören. Eine wirksame *Schutzbeschaltung* wird in Abb. 90.5 gezeigt. Parallel zur Diode V liegt ein Kondensator C, der die vorhandenen Spannungsspitzen in der Amplitude verringert. Damit der Kondensatorstrom bei sehr großen Spannungsänderungen nicht zu groß wird, ist ein Widerstand R in Reihe geschaltet. Bei Anschlussspannungen im Bereich (230 ... 500) V empfehlen viele Hersteller, Widerstände im Bereich (30 ... 150) Ω und Kondensatoren im Bereich (0,1 ... 0,47) µF einzusetzen. Die TSE-Schutzbeschaltung sollte, unabhängig von der Last, in jedem Fall vorgesehen werden, da sie auch in der Lage ist, Spannungsspitzen über den Dioden durch andere Verursacher zu verringern (siehe auch Abschn. 90.1.7 „Zusammenschaltung von Dioden“ und Abschn. 91.5 „Schutzschaltungen“).

90.1.6 Sicherungen

Wird der Strom durch eine Diode zu groß, wird sie durch die zu hohe Stromdichte und die damit verbundene erhöhte Wärmeentwicklung zerstört. Zunächst verliert sie ihre Eigenschaft als Diode, und bei weiter steigender Temperatur schmelzen Bereiche innerhalb der Diode. Die Zerstörung kann bereits nach wenigen µs seit Beginn des Überlastfalles einsetzen. Kritisch dabei ist, dass der elektrische Widerstand eines Halbleiters mit steigender Temperatur sinkt und damit eine bestehende unvermeidbare Inhomogenität der Stromdichte weiter begünstigt wird. Besonders problematisch ist der Fall von Strom*änderungen*, weil dann die Stromdichte kurzzeitig (Bereich µs) besonders ungleichmäßig über der Fläche des PN-Überganges verteilt ist. Bereiche mit Übertemperatur senken den Widerstand und erhöhen damit den Stromfluss und die Temperatur überproportional (Durchbruch 2. Art). Ein Grenzwert für den Strom lässt sich nicht ohne Weiteres angeben, da die genannten Effekte nicht kalkulierbar sind und außerdem der zeitliche Verlauf des Stromes eine Rolle spielt. Ein Maß zur Abschätzung der entstehenden Verlustleistung und der damit verbundenen Wärmeentwicklung ist durch das *Grenzlastintegral ge*geben:

Grenzlastintegral

$$\int_{t_1}^{t_2} (i(t))^2 dt \text{ (Strom zeitabhängig)} \quad \text{bzw.} \quad I^2 \cdot t \text{ (Gleichstrom)} \tag{90.4}$$

Darin ist die Dauer des Stromes begrenzt (z. B. auf 2 ms oder 10 ms, weitere Einzelheiten siehe Datenblatt). Der errechnete Grenzwert hat die Einheit A^2s (z. B. 250.000 A^2s). Für sinusförmige 50-Hz-Netze wird häufig ein daraus bereits berechneter und auf eine Halbschwingung zugeschnittener *Grenzeffektivstrom* I_{FRMS} bei definierter Gehäusetemperatur angegeben. Die zwei genannten Grenzwerte sind für die Auswahl einer *Sicherung* entscheidend. Überschreitet innerhalb der vorgegebenen Zeit (z. B. 10 ms) das Grenzlastintegral den angegebenen Wert, muss die Sicherung den Stromkreis unterbrechen (siehe Abschn. 91.5.2 „Überstromschutz“).

90.1.7 Zusammenschaltung von Dioden

Durch geeignetes Zusammenschalten mehrerer Dioden können der Dauergrenzstrom und die Spitzensperrspannung einer Einzeldiode überschritten werden. Das in Abb. 90.6 gezeigte Ersatzschaltbild einer Diode gibt ihr statisches und dynamisches Verhalten für die meisten Anwendungsfälle ausreichend genau wieder.

Die Diode wird als ideal angenommen. Für die weiteren Elemente gilt:

- r_F Widerstand der Diode in Durchlassrichtung, ergibt sich aus dem Quotienten u_F/i_F im aktuellen Arbeitspunkt, stromabhängig, evtl. durch einen konstanten Wert angenähert (z. B. 2 mΩ);
- r_R Widerstand der Diode in Sperrrichtung, ergibt sich aus dem Quotienten u_R/i_R im aktuellen Arbeitspunkt, in erster Näherung als nahezu unabhängig von der angelegten Spannung angenommen;
- C_R Kapazität des PN-Überganges, hier speziell in Sperrrichtung, abhängig von u_R, evtl. durch einen konstanten Wert angenähert. Die Kapazität des PN-Überganges in Durchlassrichtung ist wegen $r_F \ll 1\,\Omega$ ohne Bedeutung.

1. Parallelschaltung von Leistungsdioden

Da die Kennlinien $i_F = f(u_F)$ der Einzeldioden nicht identisch sind (fertigungstechnisch bedingt, temperaturabhängig), ergibt sich eine unsymmetrische Stromaufteilung unter den Dioden in

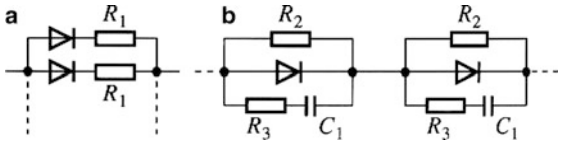


Abb. 90.7 Zusammenschaltung von Halbleiterdioden. **a** Parallelschaltung, **b** Reihenschaltung

Durchlassrichtung. Eine nahezu symmetrische Aufteilung des Gesamtstromes kann man dadurch erreichen, dass man

1. Dioden des gleichen Herstellers mit gleicher Typenbezeichnung verwendet;
2. Dioden mit angenähert gleicher Kennlinie aussucht;
3. die Dioden durch thermische Kopplung auf etwa gleicher Temperatur hält (Kennlinie $i_F = f(u_F)$ ist temperaturabhängig);
4. zusätzlich einen Widerstand R_1 (Abb. 90.7a) in Reihe zu jeder Diode anordnet.
   Wählt man R_1 so groß, dass etwa 0,4 V an ihm abfallen, weichen die Ströme der Dioden um weniger als 20 % voneinander ab. Nachteilig ist die an R_1 auftretende Verlustleistung, die die an der Last zur Verfügung stehende Leistung verringert und die unter Umständen eine gesonderte Abführung der entstehenden Wärme erfordert. Wirkungsweise von R_1: Die Spannung an R_1 der Diode mit dem größeren Strom wäre größer als die an R_1 mit dem kleineren Strom. Da die Summe $(u_F + u_{R1})$ beider Dioden gleich sein muss, muss die Spannung u_F der Diode mit dem größeren Strom sinken, was gemäß Kennlinie einen geringeren Strom zur Folge hat. Der andere Diodenkreis reagiert entsprechend u_F und damit stromerhöhend. Je konsequenter die Maßnahmen 1 bis 3 durchgeführt werden, desto kleiner kann der Wert vom Widerstand R_1 ausgelegt werden. Bei einem Halbleiterersatz sind die Punkte 1 und 2 von besonderer Bedeutung.

2. Reihenschaltung von Leistungsdioden

Bei der Reihenschaltung können Probleme während des Betriebes in Sperrrichtung auftreten. Im stationären Zustand verhalten sich die Teilspannungen an den Dioden wie die Sperrwiderstände r_R der Einzeldioden (s. Abb. 90.6 und 90.7b). Diese unterliegen aber großen Exemplarstreuungen und sind stark temperaturabhängig. Um eine nahezu gleichmäßige Spannungsaufteilung zu erhalten, gibt es folgende Maßnahmen:

1. Dioden des gleichen Herstellers mit gleicher Typenbezeichnung verwenden;
2. Dioden mit angenähert gleichem Widerstand r_R aussuchen;
3. die Dioden durch thermische Kopplung auf etwa gleicher Temperatur halten (r_R, Abb. 90.6, ist temperaturabhängig);
4. zusätzlich einen Widerstand R_2 (Abb. 90.7b) parallel zu jeder Diode anordnen.
   Sind diese Widerstände an beiden Dioden gleich groß, ergibt sich für den *statischen* Fall eine Halbierung der an der Reihenschaltung anliegenden Spannung. Bedingung ist, dass die Widerstände R_2 maximal 1/10 des Wertes vom kleinsten r_R haben, denn r_R unterliegt starken Exemplarstreuungen (herstellungsbedingt, temperaturbedingt). Da r_R meist größer ist als 100 kΩ, spielen die Verluste an R_2 kaum eine Rolle. Daneben ist noch eine weitere Maßnahme erforderlich. Ändert sich die Spannung über der Reihenschaltung der Dioden sehr schnell (Schaltvorgang), findet gemäß Ersatzschaltbild (Abb. 90.6) *vorübergehend* eine Spannungsaufteilung in umgekehrtem Verhältnis der Kapazitäten C_R zueinander statt. Diese Kapazitäten unterliegen Bauteilstreuungen und ändern sich stark während eines Schaltvorganges. Beim Übergang vom Leit- in den Sperrzustand entfernen sich die p- und n-Schichten voneinander und ändern gemäß $Q = C \cdot U$ die Kapazitäten der Dioden, aber nicht bei beiden völlig synchron. Beim Übergang vom Sperr- in den Leitzustand finden ähnliche Vorgänge mit teilweise umgekehrtem Vorzeichen statt. Auch wenn diese vorübergehende Spannungsaufteilung nur wenige µs andauert, kann sie den Halbleiter gefährden. Abhilfe schafft ein zu den Dioden jeweils parallel geschalteter Kondensator C_1 (Abb. 90.7b), dessen Wert mindestens 10mal so groß ist wie der größte an den Dioden auftretende Kapazitätswert. Der Widerstand R_3 begrenzt den Kondensatorstrom.

Für die genannten Bauteile folgen einige Angaben zur Dimensionierung: C_1 erfordert Werte im unteren nF-Bereich, R_3 Werte im Bereich (100 ... 1000) Ω.

Eine in der Regel vorhandene TSE-Beschaltung erfordert wesentlich größere Kapazitätswerte und wesentlich kleinere Widerstandswerte (s. Abschn. 90.1.5 und oben), so dass diese gleichzeitig auch eine wirksame Schutzschaltung für die Reihenschaltung von Leistungsdioden darstellt.

90.2 Bipolartransistoren

In der Leistungselektronik werden überwiegend *Silizium*-Transistoren eingesetzt, weil ihre maximale Kristalltemperatur im Bereich (150 ... 170) °C liegt und die Kühlung durch Wärmeabgabe in die Umgebung erfolgen kann. Die Kollektor-Emitter-Restspannung liegt bei (1 ... 2,5) V. Die Spannung an der Last wird um diesen Wert reduziert, was aber in der Regel von untergeordneter Bedeutung ist. Wie bei allen Leistungshalbleitern muss bei höheren Strömen für eine ausreichende Kühlung gesorgt werden. Germanium-Halbleiter dagegen finden evtl. dort noch Anwendung, wo die Versorgungsspannung niedrig ist (z. B. 12-V-Akku zum Betrieb eines Gleichstrom-Motors in einer Brückenschaltung mit 4 Transistoren zur Drehrichtungsumkehr), weil die Kollektor-Emitter-Restspannung unterhalb 1 V liegt. Allerdings ist die maximale Kristalltemperatur etwa (60 ... 70) °C, und damit wird die Kühlung bei Umgebungstemperaturen oberhalb 50 °C zum Problem. Germanium-Transistoren werden manchmal auch noch eingesetzt in Gegentakt-Endstufen von Niederfrequenz-Verstärkern, weil Temperaturen von 50 °C oder höher im häuslichen Bereich nicht erreicht werden, und wenn in Ausnahmefällen doch, dann kann eine Sicherheitsabschaltung der Endstufe erfolgen.

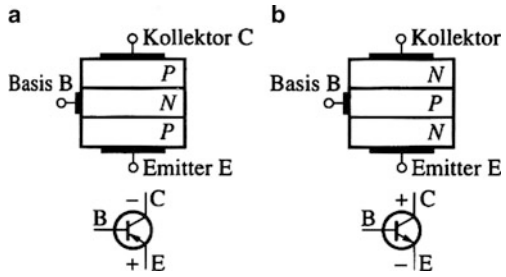


Abb. 90.8 Schematischer Aufbau eines Bipolartransistors. **a** PNP-Typ, **b** NPN-Typ

Von den zwei Typen PNP- und NPN-Transistor (Erläuterungen im nächsten Kapitel) werden die NPN-Typen häufiger eingesetzt, weil sie insgesamt etwas günstigere Daten aufweisen. In der Verstärkertechnik sind Leistungsendstufen häufig mit einem NPN- und einem PNP-Typ aufgebaut, die etwa die gleichen Eigenschaften haben und allgemein als „Pärchen“ bezeichnet werden. Durch diese Konfiguration vereinfacht sich der Aufwand für die Schaltung.

90.2.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte, SOAR

Abb. 90.8 zeigt den prinzipiellen Aufbau eines NPN- und eines PNP-Transistors. Ihre Wirkungsweise ist prinzipiell gleich. Der wesentliche Unterschied besteht darin, dass beim PNP-Transistor der Emitter das positivste, beim NPN-Transistor das negativste Potential besitzt. In der Leistungselektronik werden Transistoren vorwiegend im Schalterbetrieb oder in Leistungsverstärkern eingesetzt; im Schalterbetrieb in der Emitterschaltung (Abb. 90.9). Das Ausgangskennlinienfeld eines NPN-Transistors in Emitterschaltung zeigt Abb. 90.10. Wird in die Basis kein Strom eingespeist, d. h. $I_B = 0$, fließt ein nur geringer Kollektorstrom I_{CEV} der *Kollektorreststrom* (Basisanschluss offen, bei beschalteter Basis ist der Wert noch geringer; Strom ist temperaturabhängig). Er liegt bei Leistungstransistoren maximal bei etwa (0,5 ... 20) mA. Der Transistor befindet sich im AUS-Zustand (Sperrzustand). Damit wird der Betriebsfall „Schalter geöffnet“ für Anwendungen in der Leistungselektronik „nahezu ideal“ nachgebildet. Die am Transistor stehende Spannung zwischen Kollektor und Emitter, U_{CE}, ist fast gleich der Betriebsspannung U_G.Da der Kollektorreststrom im unteren mA-Bereich liegt, ist die am Transistor entstehende Verlustleistung im Sperrzustand klein (einige Watt oder weniger) und muss für die Halbleiter-Kühlung nicht berücksichtigt werden.

Durch Einspeisen eines ausreichend großen Basisstromes I_B in die Basis arbeitet der Transistor im EIN-Zustand (Schalter geschlossen, Leitzustand, Transistor durchgeschaltet). Der erforderliche Wert von I_B ergibt sich durch Division des Laststromes im Kollektorkreis (angenähert U_G/R_L, Abb. 90.9) durch die *Großsignal-Stromverstärkung* h_{21} (oder auch B, beides wird verwendet) des Transistors. Die verbleibende Spannung zwischen Kollektor und Emitter, die *Kollektor-Emitter-Restspannung* U_{CEsat}liegt im Bereich 1 bis 2,5 V. Sie ist in erster Näherung unabhängig vom durchfließenden Strom und kann auch durch Einspeisen eines höheren Basisstromes als erforderlich (Übersteuerung, erhöht die Ausschaltzeit, siehe Abschn. 90.2.2) nicht mehr wesentlich verringert werden. Damit lässt sich der Zustand *Schalter geschlossen* nicht ideal nachbilden.

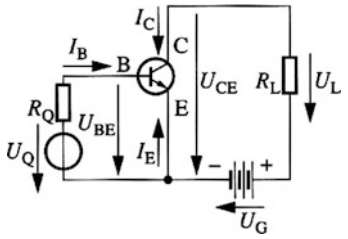


Abb. 90.9 NPN-Transistor als Schalter in Emitterschaltung

Die Restspannung U_{CEsat} verringert die an der Last zur Verfügung stehende Spannung U_L gegenüber der Betriebsspannung U_G um ca. 1 bis 2,5 V. Die am Transistor entstehende Verlustleistung kann z. B. Werte von 65 W (2N6032, $I_c = 50$ A, $U_{UCEsat} = 1{,}3$ V) erreichen.

Eine weitere Kenngröße ist die bereits angesprochene *Großsignal-Stromverstärkung* h_{21} (auch mit B bezeichnet). Sie hängt mehr oder weniger von der Größe des Kollektorstromes ab und unterliegt außerdem starken Exemplarstreuungen. Deshalb ist bei einer Schaltungsdimensionierung der kleinstmögliche Wert, $h_{21\,min}$ (B_{min}) einzusetzen (Datenblatt). Außerdem muss der sich daraus ergebende minimale Basisstrom $I_{B\,min}$ ($I_{B\,min} = I_c/h_{21\,min}$) mit einem „Sicherheits“-Faktor 1,2 ... 2 multipliziert werden, damit der Transistor wie gewünscht *sicher* durchschaltet. Andernfalls läuft man Gefahr, dass die Spannung U_{CE} nicht den MinimalwertU_{CEsat} erreicht und der Transistor durch die erhöhte Verlustleistung und die daraus resultierende Zunahme der Kristalltemperatur seine Funktionsfähigkeit verliert. Ein Nachteil dieser aus Sicherheitsgründen erforderlichen Übersteuerung ist, dass die Ausschaltzeit des Transistors zunimmt (s. Abschn. 90.2.2).

In Abb. 90.10 sind drei für den Betrieb eines Transistors als Schalter wichtige Grenzwerte eingetragen. Die höchstzulässi-

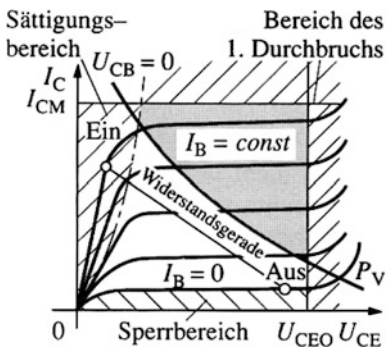


Abb. 90.10 Kennlinienfeld eines NPN Transistors mit SOAR-Bereich

ge *Vorwärts-Sperrspannung* U_{CE0} darf, auch nicht kurzzeitig, überschritten werden, da es sonst zu einem Lawinendurchbruch (zu hohe Feldstärke) bzw. zu einem Durchbruch 2. Art (Erklärung weiter unten) innerhalb des Transistors kommen kann. Der *höchstzulässige Mittelwert des Kollektorstromes*, I_{CM}, darf ebenfalls nicht überschritten werden. Die am Transistor erzeugte statische Verlustleistung muss unterhalb der *Verlustleistungshyperbel Pv* liegen. Dem Datenblatt des Transistorherstellers ist zu entnehmen, welche Bedingungen der eingetragenen Verlustleistungshyperbel zugrunde liegen; häufig wird von einer maximal zulässigen Sperrschichttemperatur für Siliziumtransistoren von (150 ... 170) °C, einer Umgebungstemperatur von (30 ... 50) °C und einem vorgegebenen Kühlkörper ausgegangen. Der Sperrbereich kann nicht verwendet werden, und der Sättigungsbereich ist zu vermeiden (Abschn. 90.2.2 „Dynamisches Verhalten“).

Damit ist der für den statischen Betriebsfall erlaubte Arbeitsbereich eines Transistors festgelegt; er wird *SOAR* (**s**afe-operating-**ar**ea) genannt und ist in Abb. 90.10 als nicht schraffierter und nicht grau unterlegter Bereich dargestellt. Für den dynamischen Betriebsfall gilt ein Sonderfall: Liegt die Widerstandsgerade mit ihrem mittleren Teil in dem grau unterlegten Bereich, die zwei Endpunkte aber im nicht gekennzeichneten Bereich, kann der Transistor für den Übergang zwischen den zwei Schaltzuständen diesen Bereich durchfahren, sofern das ausreichend schnell geschieht. Das heißt, die Schaltverluste und die Halbleiterkühlung müssen aneinander angepasst sein.

Die *Vorwärts-Sperrspannung* U_{CE0} gilt für den Fall, dass die Basis des Transistors offen (unbeschaltet) ist. Wählt man dagegen die Basis-Emitter-Spannung kleiner 0 V (NPN-Transistor), erhöht sich die Sperrspannung U_{CE0} auf den Wert: $U_{CEV} = (1{,}2 \ldots 2) \cdot U_{CE0}$. Allerdings gilt: $U_{BE\,min} \approx -5$ V (genauer Wert siehe Datenblatt).

90.2.2 Dynamisches Verhalten

Das *dynamische Verhalten* gibt Auskunft über den Verlauf von Kollektorstrom und Kollektor-Emitter-Spannung beim Übergang des Transistors vom Sperr- in den Leitzustand (Abb. 90.11) und umgekehrt (Abb. 90.12). Als Lastimpedanz wurde ein Wirkwiderstand eingesetzt.

1. Einschaltvorgang

Wird ein Basisstrom $i_B = i_{B1}$ in die Basis eingespeist, so vergeht bis zu einem merklichen Anstieg des Kollektorstromes (10 % vom Endwert) die *Verzögerungszeit* t_d; anschließend steigt der Kollektorstrom innerhalb der *Anstiegszeit* t_r auf 90 % des Endwertes. Gleichzeitig sinkt die Kollektor-Emitter-Spannung auf ca. 10 % des ursprünglichen Wertes ab. Zwei wichtige Folgerungen ergeben sich aus Abb. 90.11.

1. Der Transistor benötigt endliche Zeit (einige Mikrosekunden), um in den leitenden Zustand zu gelangen. Der Ansteuerung von Verbrauchern mit rechteckförmigen Spannungen und variablem Tastgrad zur Leistungseinstellung sind damit Grenzen bei (10 ... 30) kHz gesetzt

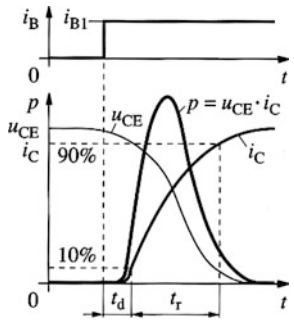


Abb. 90.11 Einschaltverhalten eines NPN-Bipolartransistors

2. Während des Überganges vom sperrenden in den leitenden Zustand entsteht am Transistor eine zeitabhängige Leistung p, deren *prinzipieller* Verlauf eingezeichnet ist. Sie ergibt sich zu $p = u_{CE} \cdot i_C$. Die unter dieser Kurve liegende Fläche ist ein Maß für die bei einem Einschaltvorgang entstehende Verlustleistung bzw. Verlustenergie.

2. Ausschaltvorgang

Der Übergang vom leitenden in den sperrenden Zustand ist beim Basisstrom dadurch gekennzeichnet, dass zunächst ein Strom i_{B2} für die Dauer t_s (*Speicherzeit*) aus der Basis herausfließt (Abb. 90.12). Während dieser Zeit ändert sich der Kollektorstrom kaum. Erst nach Ausräumen der Basisladungen (schraffierter Teil) beginnt der Basisstrom und damit auch der Kollektorstrom abzunehmen (*Abfallzeit* t_f), die Vorgänge ähneln jetzt denen beim Einschalten. Deshalb gelten die beim Einschaltvorgang dargestellten Folgerungen auch beim Ausschalten, allerdings mit dem wesentlichen Zusatz, dass beim Ausschalten eines Transistors eine *Speicherzeit* t_s auftritt, die um einiges größer ist als die *Verzögerungszeit* t_d des Einschaltvorganges (Faktor 2 bis 10). Aus diesem Grund werden spezielle Schaltungsmaßnahmen zur Verringerung der Speicherzeit t_s nach Abb. 90.13 angewendet.

Abb. 90.13a: Ist der Transistor im durchgeschalteten Zustand, verursacht der Basisstrom i_B an R_{B2} einen Spannungsabfall. R_{B2} und C müssen so dimensioniert werden, dass sich C während der Leitphase des Transistors durch den Basisstromabfall

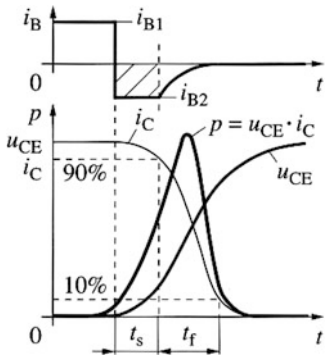


Abb. 90.12 Ausschaltverhalten eines NPN-Bipolartransistors

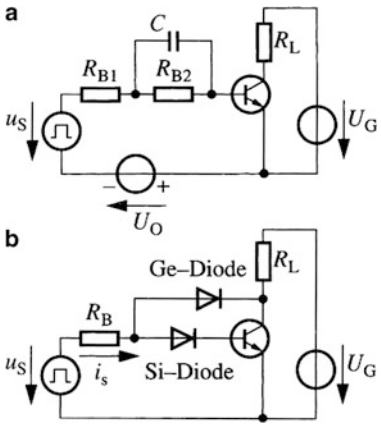


Abb. 90.13 Transistorschaltstufen zur Ausschaltzeitverkürzung, **a** mit Beschleunigungskondensator **b** mit Sättigungsschutzdioden

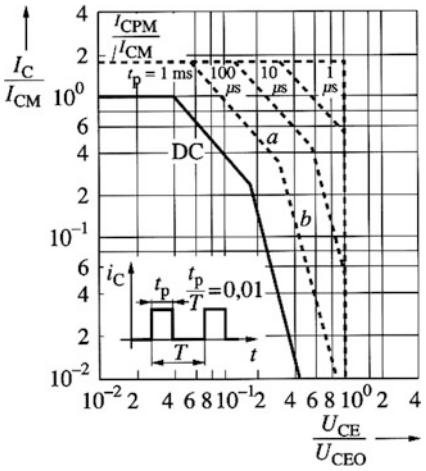


Abb. 90.14 SOAR-Diagramm eines Bipolartransistors

an R_{B2} auf etwa (1 ... 3) V auflädt. Das sollte angenähert der stationäre Endwert sein. Das Steuergerät für die Bereitstellung des Basisstromes muss diesen Spannungsabfall berücksichtigen. Der negative Pol der Kondensatorspannung liegt an der Basis des Transistors. Schaltet das Steuergerät den Transistor in den Sperrzustand durch $u_S = 0$, liegt durch die „ideale" Spannungsquelle mit der Spannung u_S und dem Innenwiderstand $R_{i(us)} \ll 1\,\Omega$ der positive Kondensatorpol am Emitter (U_0 sei 0) und der negative Anschluss an der Basis. Die Basisladungen des Transistors werden direkt durch die Kondensatorladungen kompensiert. Der Wert von R_{B1} ist als Begrenzungswiderstand deshalb ausreichend klein zu wählen oder kann evtl. ganz entfallen. Für C haben sich Werte im Bereich (50 ... 500) pF und für R_{B2} Werte im Bereich (5 ... 100) Ω bewährt. R_{B2} und u_S legen aber auch den Basisstrom für den Leitzustand des Transistors fest und sollten eine Übersteuerung vermeiden. Die Spannungsquelle mit der Spannung U_0 ((1 ... 2) V) ist bei Bedarf einzusetzen, sie beschleunigt den Entladevorgang zusätzlich. Hinweise: Die negative Spannung an der Basis des Transistors darf den Wert von etwa −5 V nicht unterschreiten (Datenblatt gibt den genauen Wert an). Die optimale Dimensionierung dieser Schaltung sollte durch Versuche ermittelt werden. Auch die Technischen Unterlagen der Transistorhersteller können eine wertvolle Hilfe sein.

Abb. 90.13b: Diese Schaltungsmaßnahme sorgt dafür, dass durch eine bewusste und begrenzte Übersteuerung sowohl ein beschleunigtes Einschalten (in der Regel Nebeneffekt) als auch vor allem eine gewisse Verringerung der Restspannung $U_{CE\,sat}$ und damit eine Verringerung der am Transistor entstehenden Verlustleistung erreicht wird. Das verringert den Aufwand für die Kühlung. Gilt: $0{,}7\,V + u_{BE} > 0{,}3\,V + U_{CE\,sat}$, fließt ein Teil des Stromes i_S über die Germaniumdiode und die Kollektor-Emitter-Strecke des Transistors. Damit wird der Basisstrom und damit die Übersteuerung begrenzt. Durch Reihenschaltung mehrerer Dioden (Schwellspannung Germaniumdioden etwa 0,3 V; Siliziumdioden etwa 0,7 V) kann eine Optimierung erreicht werden (evtl. experimentell ermitteln). Diese Maßnahme ist für den Schalterbetrieb bei niedrigen Frequenzen oder für den „statischen" Betrieb sinnvoll, weil der Verringerung der Restspannung eine Erhöhung der Speicherzeit t_S gegenübersteht.

Für Transistoren mit $U_{CE0} = 1000\,V$ und $I_{CM} = 150\,A$ liegt beispielsweise t_r bei etwa (2 ... 10) µs, t_s bei etwa (15 ... 30) µs. Spezielle schnelle Schalttransistoren mit $U_{CE0} = 400\,V$; $I_{CM} = 50\,A$ weisen z. B. folgende Werte auf: $t_r \approx 0{,}5\,\mu s$; $t \approx 4\,\mu s$.

Zur Kennzeichnung des SOAR-Bereiches eines Leistungstransistors im dynamischen Betrieb verwenden die Hersteller häufig das in Abb. 90.14 dargestellte Diagramm. I_{CPM} ist der höchstzulässige Mittelwert des Kollektorstromes bei Pulsbetrieb mit definiertem Tastgrad t_p/T. Er ist größer als der Wert I_{cm}. Bei Gleichstrombelastung wird der erlaubte Bereich durch die mit DC gekennzeichnete Linie begrenzt. Der Geradenteil a ergibt sich durch die Wärmeentwicklung infolge der auftretenden Verlustleistung, der Geradenteil b, um einen zweiten Durchbruch zu verhindern. Der Tastgrad t_p/T beträgt 0,01. Es handelt sich also um sehr kurze Einschaltdauern im Vergleich zur Periodendauer.

Der Durchbruch 2. Art entsteht dadurch, dass bei einem großen Stromanstieg pro Zeiteinheit die Stromdichte zunächst nicht homogen über der Kristallfläche verteilt ist. Flächen mit hoher Stromdichte erwärmen sich überproportional, was wiederum ihren Widerstand verringert und damit die Stromdichte weiter erhöht. Das kann zu einem Schmelzen lokaler Bereiche führen und damit den Halbleiter zerstören.

Enthält der Lastkreis auch Induktivitäten oder Kapazitäten, kann der Verlauf von i_C und u_{CE} erheblich von dem nach Abb. 90.11 bzw. 90.12 abweichen. Kritisch ist der in der Leistungselektronik häufig vorkommende Fall, dass die Lastimpedanz induktive Anteile, z. B. durch Motoren oder Transformatoren, enthält. Dann tritt im Ausschaltmoment des Transistors an der Induktivität eine Spannungsüberhöhung auf, die unter Umständen bewirkt, dass die Kollektor-Emitter-Spannung den Grenzwert U_{CE0} überschreitet und der Transistor zerstört wird. Deshalb sind dann geeignete Schutzbeschaltungen erforderlich (s. Abschn. 90.2.5).

90.2.3 Verlustleistung

Die Berechnung der Verlustleistung wird hier unter dem Problem der ausreichenden Halbleiterkühlung gesehen.

1. Transistor als Schalter

Die Gesamtverluste setzen sich aus 4 Anteilen zusammen: 1. Verluste beim Übergang vom AUS- in den EIN-Zustand; 2. Verluste im EIN-Zustand; 3. Verluste beim Übergang vom EIN- in den AUS-Zustand; 4. Verluste im AUS-Zustand.

Die Verluste unter Punkt 4 können vernachlässigt werden, weil der Kollektor-Reststrom im unteren mA-Bereich oder weniger liegt. Das ergibt bei 1 mA und einer Kollektor-Emitter-Spannung von 1000 V eine Verlustleistung von 1 W. Die Verluste in den 2 Umschaltzeiten, Punkte 1 und 3, sind im Prinzip als $p = p(t)$ in den Abb. 90.11 und 90.12 dargestellt. Formal ist die Gesamtleistung durch Integrieren von $p(t)$ über eine Periode und das Beziehen auf die Periode zu bestimmen. In der Praxis wird man auf Erfahrung oder Näherungslösungen zurückgreifen. Gut geeignet zur Erfassung sind die im Teil XIII „Nachrichtentechnik“, Kap. 97 „Filter“, Abschn. 97.3.2 „Tiefpassfilter“ genannten Simulationsprogramme. Die Praxis zeigt, dass sich erst bei höheren Frequenzen ab unterem kHz-Bereich evtl. zu berücksichtigende Anteile an der Gesamtleistung ergeben, weil das Maximum von $p(t)$ Leistungsanpassung zwischen Transistorwiderstand und Lastwiderstand (Lastimpedanz) bedeutet. Die Verluste nach Punkt 2, EIN-Zustand, sind in der Regel die größten. Sie lassen sich nach (90.2), mit entsprechend angepassten Bezeichnungen, berechnen.

Wird der Halbleiter als Schalter eingesetzt mit dem Tastgrad t_0/T_0 und verlaufen u_F und i_F (nahezu) rechteckförmig mit diesem Tastgrad, kann die Verlustleistung P_V berechnet werden zu:

$$P_V \approx u_{CE} \cdot i_C \cdot \frac{t_{0\,max}}{T_0} \tag{90.5}$$

Die Formel berücksichtigt, dass t_0 i. Allg. variabel ist (z. B. Gleichstromsteller) und dass dann dessen Maximalwert einzusetzen ist: $0 < t_0 < t_{0\,max} < T_0$.

2. Transistor als Leistungsverstärker

Die Spannungs- und Stromverläufe (u_{CE}, i_C) können beliebig und nichtperiodisch sein. Deshalb lässt sich die Verlustleistung dann nur nach der allgemeinen Formel berechnen:

$$P_V = \frac{1}{T}\left(\int (u_{CE}(t) \cdot i_C(t))dt + \int (u_{BE}(t) \cdot i_B(t))dt\right) \tag{90.6}$$

Da aber gilt:

$$I_C \gg I_B; \quad U_{CE} > U_{BE}, \tag{90.7}$$

vereinfacht sich (90.6) zu

$$P_V \approx \frac{1}{T}\left(\int (u_{CE}(t) \cdot i_C(t))dt\right). \tag{90.8}$$

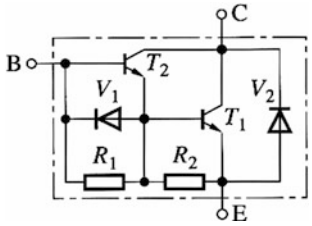


Abb. 90.15 Darlington-Leistungstransistor mit Zusatzelementen

Die Berechnung der Verlustleistung nach (90.8) ist in der Regel schwierig, da es sich häufig um stochastische Signale (z. B. Musik, Sprache) handelt. Dann führen Messungen oder sog. „Worst-Case-Abschätzungen“ zu brauchbaren Ergebnissen: $u_{CE\,max}$ aus Messung oder Datenblatt, $i_{C\,max}$ durch Versorgungsspannung und Lastimpedanz.

90.2.4 Bauarten

Bipolar-Leistungstransistoren werden auch in *Modulform* angeboten. Neben dem eigentlichen Transistor sind noch Zusatzelemente eingebaut, die als Schutzelemente wirken oder den Einsatz komfortabler gestalten. Beispiele dafür sind:

- 4 Transistoren in einem Modul, Einsatz z. B. in einer Brückenschaltung zur Drehrichtungsumkehr bei Gleichstrommotoren;
- Darlingtontransistor, erkennbar an einer Stromverstärkung von mehr als 1000 und einer Basis-Emitter-Schwellspannung von 1,4 V (2 Siliziumtransistoren), siehe dazu die folgenden Erläuterungen und Abb. 90.15;
- Diode (V_2) parallel zur Kollektor-Emitter-Strecke, siehe Abb. 90.15;
- eingebauter Optokoppler zur potentialgetrennten Ansteuerung;
- eingebauter Temperaturfühler. Er dient zur aktuellen Erfassung der Halbleitertemperatur.

Abb. 90.15 zeigt ein Beispiel für den internen Aufbau eines Transistormoduls. Die Diode V_2 soll verhindern, dass die Kollektor-Emitter-Spannung negativer als etwa −1 V werden kann (Schutzfunktion). Der Transistor T_1 ist der eigentliche Leistungstransistor. Bei Strömen oberhalb etwa 50 A liegt seine Stromverstärkung merklich unter 100, er benötigt also einen Basisstrom von 1 A oder mehr, um durchzuschalten. Der von außen über den Basisanschluss B in die Basis von Transistor T_2 eingespeiste Strom wird mit dessen Stromverstärkung (i. Allg. > 100) verstärkt und gelangt als Emitterstrom in die Basis von T_1. Theoretisch ist die resultierende Stromverstärkung dieser *Darlington-Schaltung* das Produkt der Stromverstärkungen von T_1 und T_2, in der Praxis ist sie geringer, liegt aber um den Faktor 10 … 100 über der von T_1. Damit beträgt der in den Basisanschluss B des Moduls einzuspeisende Strom einige mA. Der Nachteil dieser Schaltung liegt darin, dass beim Anlegen der Spannung 0 V an den Basisanschluss B (zum Sperren des

Transistors T_1) im Transistor T_2 der Kollektor-Emitter-Reststrom fließt, der direkt in die Basis von T_1 gelangt und diesen nicht vollständig zum Sperren bringt. Der Kollektor-Emitter-Reststrom von T_2 ist relativ hoch, weil er in etwa die gleiche – hohe – Temperatur hat wie T_1. Abhilfe schafft der Widerstand R_2, über den der Reststrom zum größten Teil (Anhaltswert etwa 9/10) umgeleitet wird. Allerdings wird auch der Steuerstrom für T_1 auf 1/10 reduziert, was aber durch den Stromverstärkungsfaktor $\gg 10$ von T_2 mehr als ausgeglichen wird. Über den Widerstand R_1 können die in der Basis von T_1 gespeicherten Ladungen beim Ausschalten von T_1 abfließen, weil T_2 bereits sperrt. V_1 verhindert, dass die Basis-Emitter-Spannung von T_2 kleiner $-0{,}7\,\mathrm{V}$ werden kann. Leitet V_1, wird R_1 „überbrückt“, und die Basisladungen von T_1 werden sehr schnell ausgeräumt. In das Modul ist also einiges Nützliche integriert worden.

Wichtig ist, dass besonders bei Leistungstransistoren das Datenblatt eingesehen wird, um Informationen über derartige Zusatzelemente zu erhalten. Hinweis: Einige Hersteller erwähnen nicht unbedingt oder aber sehr versteckt den Begriff Darlington-Transistor, dann erkennt man das an einer Stromverstärkung von mehr als 1000 und einer Basis-Emitter-Schwellspannung von 1,4 V (Siliziumtransistor) im Datenblatt!

Hier ein Beispiel für Leistungstransistoren der oberen Leistungsklasse: $U_{CE0} \approx 4000\,\mathrm{V}$; $I_{cm} \approx 1500\,\mathrm{A}$, beide Werte zusammen sind aber nicht von *einem* Transistor erreichbar. Durch einen Darlington-Transistor-Aufbau wird bei einem Kollektorstrom von 400 A noch eine Gesamtstromverstärkung von minimal 80 erreicht. Ein Ein-Aus-Schaltzyklus erfordert eine Zeit von ca. $(10 \ldots 20)\,\mu\mathrm{s}$, der Wärmewiderstand Kristall-Gehäuse liegt bei etwa $0{,}08\,\mathrm{K/W}$.

90.2.5 Schutzschaltungen

Die Beschädigung eines Transistors hat im wesentlichen drei Ursachen

1. Überschreiten der höchstzulässigen Vorwärts-Sperrspannung U_{CE0} bzw. U_{CEV}. Ursachen hierfür können Induktivitäten im Lastkreis und Überspannungen der Versorgungsspannung sein. Einen gewissen Schutz bieten Z-Dioden entsprechender Spannung parallel zu den Kollektor-Emitter-Anschlüssen des Transistors. Die Überspannungen bei einer „abrupten“ Unterbrechung des Stromflusses durch Induktivitäten können durch eine Diode, die sog. „Freilaufdiode“, parallel zum Verbraucher verhindert werden, Anode beim NPN-Transistor an den Kollektor, Kathode an die Betriebsspannung. Bei nur sehr kurzzeitig auftretenden Überspannungen hilft sowohl die bei den Dioden erwähnte TSE-Schutzschaltung als auch die Parallelschaltung eines spannungsabhängigen Widerstandes (Varistor). Z-Dioden und Varistoren können nur einen zeitlich begrenzten Schutz bis in den oberen ms-Bereich gewährleisten. Dann muss eine Überstromschutzeinrichtung (Schmelzsicherung oder dgl.) den Stromkreis unterbrechen.

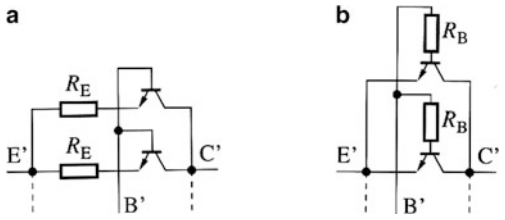


Abb. 90.16 Parallelschaltung von Bipolar-Leistungstransistoren, **a** mit Emitterwiderstand R_E zur Stromsymmetrierung, **b** mit Basiswiderstand R_B zur Stromsymmetrierung

2. Das Potenzial des Kollektors ist beim NPN-Transistor negativ gegenüber dem Emitter. Dieser Zustand entsteht häufig bei Verbrauchern mit induktivem Anteil und großen Stromänderungen pro Zeiteinheit. Er lässt sich dadurch vermeiden, dass eine Diode mit der Anode an den Emitter und mit der Kathode an den Kollektor angeschlossen wird (s. V_2 in Abb. 90.15). Bei einem PNP-Transistor sind die Polaritäten entsprechend zu ändern. Wird gemäß Punkt 1 eine Z-Diode vorgesehen, so erfüllt sie gleichzeitig die hier geforderte Schutzfunktion
3. Überschreiten des höchstzulässigen Mittelwertes des Kollektorstromes bei Gleichstrom (I_{CM}) oder bei Pulsbetrieb (I_{CPM}). Dieser Fall kann z. B. bei Kurzschluss im Verbraucherkreis auftreten. Da die maximale zeitliche Dauer für den Strom I_{CPM} vom Hersteller vorgegeben ist, dient das bei den Dioden behandelte Grenzlastintegral (90.5) zur Auswahl einer geeigneten Überstromschutzeinrichtung (Schmelzsicherung oder dgl.).

90.2.6 Parallelschaltung von bipolaren Transistoren

Die Zusammenschaltung von Transistoren beschränkt sich im Wesentlichen auf die Parallelschaltung zur Erhöhung des maximal möglichen Laststromes. Die Reihenschaltung ist prinzipiell möglich, erfordert aber einen erhöhten Aufwand in der Ansteuerelektronik, damit die Grenzwerte in keinem Fall überschritten werden.

Die einfachste Möglichkeit der Parallelschaltung, d. h. jeweils die Basis-, die Emitter- und die Kollektor-Anschlüsse mehrerer Transistoren zu verbinden, ist problematisch, da durch die nicht identischen Eingangskennlinien (Basisstrom als Funktion der Basis-Emitter-Spannung) und die unterschiedlichen Stromverstärkungen der Einzeltransistoren eine ungleichmäßige Stromaufteilung auftritt. Ungünstig ist auch, dass der Transistor mit dem höchsten Kollektorstrom die höchste Verlustleistung und damit Erwärmung erfährt. Dadurch verschiebt sich die Eingangskennlinie zu noch höheren Basis- und damit Kollektorströmen (positive Rückkopplung; $|\Delta T| = 1\,\mathrm{K} \Rightarrow |\Delta_{UBE}| \approx 2\,\mathrm{mV}$). In Abb. 90.16 sind zwei häufig benutzte Möglichkeiten gezeigt, eine mehr oder weniger gleichmäßige Stromaufteilung

zu erreichen. In Schaltung a) sollte der Spannungsabfall an R_E etwa (0,3 ... 0,8) V betragen. Um diesen Spannungsabfall wird die Spannung am Verbraucher verringert. Die Widerstände R_E müssen für eine Verlustleistung $P_V = (I_C)^2 \cdot R_E$ ausgelegt werden. Sie benötigen Platz, und die entstehende Wärme muss abgeführt werden. Diese Schaltung ist sehr wirkungsvoll, weil der Transistor mit dem größeren Strom ein größeres U_{RE} bewirkt und damit seine Spannung U_{BE} verringert und damit seinen Kollektorstrom wieder verkleinert (gängiges Verfahren zur Arbeitspunktstabilisierung von Transistoren). Schaltung b) bewirkt eine nur angenähert gleichmäßige Stromaufteilung (Erfahrungswert: Abweichung bis zu ±25 % vom Idealwert, Abweichung nimmt ab mit steigendem Wert von R_B). Die Kennlinien $I_B = f(U_{BE})$ und die Stromverstärkungen der Transistoren sind nicht gleich (Zufallsprozess bei der Fertigung). Sind die Widerstandswerte R_B gleich und werden von einigen Ω auf einige kΩ erhöht, geht die Spannungsansteuerung (über U_{BE}) gleitend über in eine Stromansteuerung (über I_B). Dadurch sinkt der Einfluss der Kennlinienabweichungen $I_B = f(U_{BE})$. Es bleibt der Einfluss der unterschiedlichen Stromverstärkungen. Deshalb werden die Widerstände R_B häufig als Potentiometer ausgelegt und die Stromaufteilungen durch Abgleich aneinander angenähert. Der Abgleich ist bei der Fertigung und auch bei einer eventuellen Ersatzteilbestückung durchzuführen. Allerdings darf pro Transistor nicht mehr der höchstzulässige Kollektorstrom I_{CM} zugrunde gelegt werden (Sicherheitsabstand). Dafür sind aber die Widerstände bzw. Potentiometer von den Abmessungen her kleiner und setzen außerdem eine weit geringere Verlustleistung in Wärme um als in Schaltung a). Außerdem ist die Spannung am Verbraucher um die genannten 0,3 ... 0,8 V gegenüber Schaltung a) größer.

90.2.7 Anwendungen

Bipolare Leistungstransistoren sind prinzipiell für Anwendungen als Linearverstärker und als Schalter geeignet. Allerdings stehen für Schaltanwendungen weitere Halbleiterbauelemente zur Verfügung (siehe Abschn. 90.3, 90.4 und 90.5), die für diese Aufgabe optimal geeignet sind und die Bipolartransistoren hier zunehmend verdrängen. Aber auch für die Anwendung in Leistungsverstärkern tritt das Halbleiterbauelement MOSFET zunehmend in den Vordergrund. Dieser Transistortyp erlaubt höhere Frequenzen und hat nicht das Problem des 2. Durchbruchs, was die Stromsymmetrierung bei der Transistor-Parallelschaltung in gewissem Rahmen automatisch erfüllt und höhere Ströme bei Impulsbetrieb zulässt (Abschn. 90.3.1). Weitere Unterschiede zum MOSFET sind in den folgenden Abschn. 90.3.1 und 90.3.2 dargestellt. Die Hauptanwendung für bipolare Leistungstransistoren als Schalter liegt damit im Bereich der 50-Hz-Stromversorgung bis 1000 V und einigen 100 A. Eine besondere Stellung nehmen die IGBT-Transistoren ein. Sie sind eine Kombination aus Feldeffekt- und Bipolar-Transistor mit dem Ziel, die Vorteile beider Transistortypen zu kombinieren. Siehe dazu Abschn. 90.4 „IGBT-Transistoren“.

90.3 Feldeffekttransistoren

90.3.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte

In der Leistungselektronik (z. B. Frequenzumrichter und Steller) werden vorwiegend *N-Kanal-Feldeffekttransistoren* mit isoliertem Gate (MOSFET, IGFET) vom Anreicherungstyp verwendet und als Schalter eingesetzt (Abb. 90.18). Gegenüber den *P-Kanal-Feldeffekttransistoren* haben sie in der Summe etwas bessere elektrische Eigenschaften. Der Drainstrom i_D wird durch Anlegen eines elektrischen Feldes zwischen Gate G und Source S beeinflusst. Für $u_{GS} = 0$ fließt nur der sehr kleine *Drainreststrom* I_{DSS}. Sein Wert steigt exponentiell mit steigender Transistortemperatur; bei Zimmertemperatur liegt er bei maximal einigen Milliampere. Mit steigender Spannung $u_{GS} > 0$ nimmt der Widerstand der Drain-Source-Strecke ab, er erreicht bei der *höchstzulässigen Gate-Source-Spannung* U_{GSM} den kleinsten Wert, der in der Größenordnung (10 ... 100) Ω liegt. Da dieser Wert für den Einsatz in der Leistungselektronik zu groß ist, hat man auf einem Substrat bis zu tausend Einzeltransistoren integriert und eingangs- und ausgangsseitig parallel geschaltet. Auf diese Weise werden *Drain-Source-Widerstände* $R_{DS(on)}$ von weniger als 0,1 Ω erreicht. Allerdings erhöht sich dadurch auch die Gesamtkapazität C_{GS} zwischen Gate und Source auf bis zu 10 nF. Bei der Fertigung wird häufig die in Abb. 90.18 eingezeichnete Diode zwischen Drain und Source integriert. Sie soll verhindern, dass der Drainanschluss negativeres Potential bekommt als der Sourceanschluss, weil dadurch der Transistor zerstört werden könnte.

Für den Betrieb als Schalter sind neben den drei bereits genannten Kenngrößen bzw. Grenzwerten noch die folgenden von Bedeutung: *Drain-SourceDurchbruchspannung* $U_{(BR)DSS}$; *höchstzulässiger Wert des Drainstromes* I_{DM}; *höchstzulässiger Augenblickswert des Drainstromes bei Pulsbetrieb mit definiertem Tastgrad* t_p/T, I_{DPM}. Die Darstellung der Grenzwerte geschieht von Seiten der Hersteller wiederum in einem SOAR-Diagramm, Abb. 90.17. Nähere Erläuterungen dazu siehe Abb. 90.14 mit dem erklärenden Text. Gegenüber den entsprechenden Werten bei Bipolartransistoren können bei Feldeffekttransistoren im Pulsbetrieb höhere Ströme und Verlustleistungen zugelassen werden. Das liegt an folgendem: Beim Übergang des Feldeffekt-Transistors vom Sperr- in den Leitzustand tritt auch bei ihm vorübergehend eine ungleichmäßige Stromdichte innerhalb des Halbleiters auf. Bereiche mit höherer Stromdichte erwärmen sich mehr als solche mit geringerer Stromdichte. Der Widerstand aber steigt mit steigender Temperatur, und das wirkt der ungleichmäßig verteilten Stromdichte entgegen. Dieser Effekt gilt sowohl für einen Einzeltransistor, für mehrere parallel geschaltete Transistoren auf einem Substrat (Verringerung des Widerstandes im EIN-Zustand) und für die Stromsymmetrierung bei mehreren extern parallelgeschalteten Transistoren zur Erhöhung des geschalteten Laststromes. Es entfällt damit die Gefahr des 2. Durchbruchs, denn im ungünstigen Fall wäre

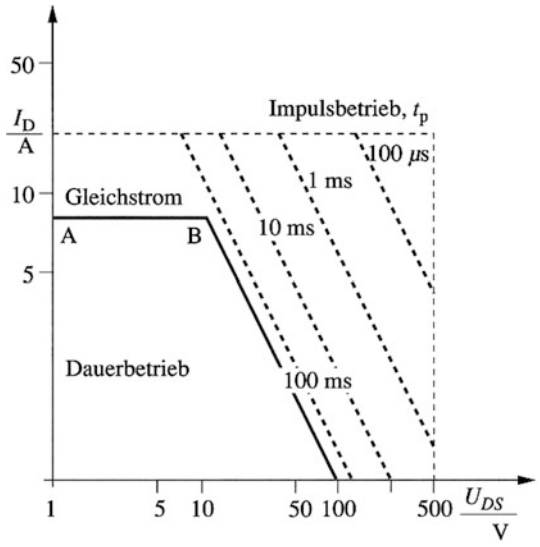


Abb. 90.17 SOAR-Diagramm eines Feldeffekttransistors, Periodendauer $T = 1$

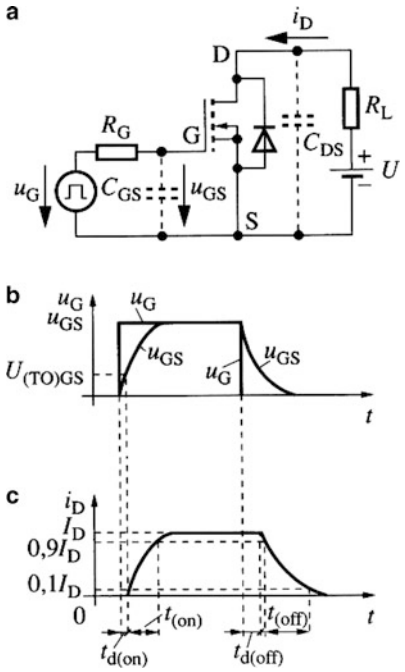


Abb. 90.18 Schaltverhalten eines Feldeffekttransistor. **a** Schaltungsaufbau, **b** Verlauf der Steuerspannung, **c** Verlauf des Drainstromes

die lokale Stromdichte so groß, dass Halbleiterbereiche schmelzen würden. Mehrere Hersteller bieten Transistoren der oberen Leistungsklasse mit folgenden Werten an: $U_{(BR)DSS} \approx 2000\,V$; $I_{DM} \approx 500\,A$; $R_{DS(on)} < 0{,}01\,\Omega$; $R_{thJG} < 0{,}05\,K/W$. Allerdings können die Maximal-Werte 2000 V und 500 A nicht von *einem* Transistor erfüllt werden.

Feldeffekttransistoren werden auch in Modulform angeboten, d. h., auf einem Substrat sind dann z. B. zwei oder vier Transistoren angeordnet. Damit lässt sich z. B. durch Einsatz eines solchen 4er-Moduls ein Gleichstrommotor durch Einbau in einer Brückenschaltung in seiner Drehrichtung einstellen. Da die Einzeltransistoren untereinander und damit vom Gehäuse elektrisch isoliert sein müssen, steigt der Wärmewiderstand dieser Module auf bis zu 0,5 K/W.

90.3.2 Dynamisches Verhalten

Im Prinzip werden Feldeffekttransistoren leistungslos gesteuert, und die Schaltzeiten sind niedriger als bei Thyristoren bzw. Bipolartransistoren. Für die Ansteuerung sind 2 Fälle zu unterscheiden: niedrige Frequenzen im Bereich bis 1 kHz und hohe Frequenzen im kHz-Bereich.

a) Frequenzen im Bereich < 1 kHz

Hat die Spannungsquelle mit der Spannung u_G in Abb. 90.18 den dargestellten rechteckförmigen Verlauf, so ergeben sich die eingezeichneten zugehörigen Spannungs- und Stromverläufe. Der Widerstand R_G beinhaltet den Innenwiderstand der Steuerspannungsquelle u_G. Die Spannung u_{GS} steigt gemäß einer e-Funktion an und fällt entsprechend ab mit der Zeitkonstanten $\tau = R_G \cdot C_{GS}$. Diese Zeitkonstante bestimmt wesentlich die Verluste in den Umschaltmomenten des Transistors, die aber wegen der relativ niedrigen Umschaltfrequenz gegenüber den Verlusten im leitenden Zustand häufig vernachlässigt werden können.

b) Frequenzen ≫ 1 kHz

Feldeffekttransistoren können bei diesen Frequenzen eingesetzt werden, weil sie gegenüber Bipolartransistoren die folgenden Unterschiede aufweisen:

1. Die Verzögerungszeiten t_d und die Anstiegszeiten t_r sind um den Faktor 2 bis 10 kleiner als bei Bipolartransistoren, und sie hängen jetzt wesentlich von der äußeren Beschaltung ab. In Abb. 90.18 bzw. Abb. 90.19 ist die Gate-Source-Kapazität C_{GS} eingezeichnet. Durch die Parallelschaltung vieler Einzeltransistoren auf einem Chip kann die resultierende Kapazität $C_{GS(ges)}$ Werte bis zu 10 nF annehmen, bei Sonderausführungen kann dieser Wert auch wesentlich höher sein. Diese Kapazität muss über den Innenwiderstand R_G der Spannungsquelle in kürzester Zeit umgeladen werden. Die die Umladung bestimmende Zeitkonstante τ ergibt sich mit $C_{GS(ges)} = 10\,nF$ und $R_G = 5\,\Omega$ zu:

$$\tau = R \cdot C_{GS(ges)} = 10\,nF \cdot 5\,\Omega = 50\,ns. \qquad (90.9)$$

Mit $u_G = 10\,V$ muss die Spannungsquelle einen Spitzenstrom von 2 A liefern. Ihr Innenwiderstand muss also sehr niedrig sein; bei Schaltfrequenzen im Bereich 10 kHz $\hat{=}$ 100 µs) weniger als 2 Ω (Gleichung II.9). Es werden zweckmäßigerweise nicht mehr die Zeiten t_d und t_r (Bilder II-11 und II-12) angegeben, sondern die in einem realen Aufbau erreichte *Einschaltverzögerungszeit* $t_{d(on)}$ und die *Einschaltzeit* $t_{(on)}$ bei einem gegebenen Wert für R_G.

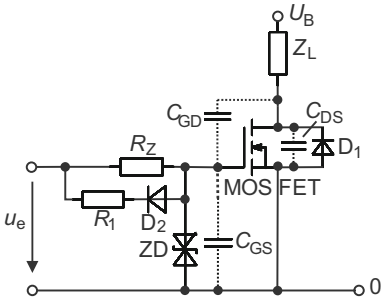


Abb. 90.19 MOSFET-Ansteuerung

2. Die Speicherzeiten t_s und die Ausschaltzeiten t_f sind ebenfalls um den Faktor 2 bis 10 kleiner als bei Bipolartransistoren. Deshalb bestimmt auch hier die externe Beschaltung ganz wesentlich diese Zeiten: Zu der beim Ausschalten durch den Eingangskreis nach Punkt 1 geschilderten Verzögerung kommt noch eine zweite hinzu, die dadurch entsteht, dass sich der Kondensator $C_{DS(ges)}$ über den Lastwiderstand R_L entladen muss. Auch hier gibt man i. Allg. die *Ausschaltverzögerungszeit* $t_{d(off)}$ und die *Ausschaltzeit* $t_{(off)}$ bei gegebenem R_L an.

Bei Frequenzen $\gg 1$ kHz bekommen die Gateansteuerung und der Widerstand R_Z in Abb. 90.19 eine besondere Bedeutung. Dieser Widerstand kann zusätzlich eingefügt werden, ist aber prinzipiell nicht erforderlich. Er vergrößert in jedem Fall die Schaltzeit, was aber manchmal sinnvoll sein kann, wenn beispielsweise die beim Schalten auftretende große Stromänderung di/dt an der Last oder im Netz Probleme verursacht. Weitere Beispiele: *Aus*schalten induktiver Lasten, *Ein*schalten kapazitiver Lasten, zu hohe hochfrequente Störsignalamplituden (Fourierzerlegung). In der Regel dauert auch das Einschalten ohmscher Lasten länger als das Ausschalten, verursacht durch den Miller-Effekt. Deshalb kann folgende Vorgehensweise sinnvoll sein: Der Innenwiderstand der Ansteuerquelle wird sehr niedrig gehalten ($< 1\,\Omega$, Standard-Ansteuer-ICs), und die Ein- und Ausschaltzeiten werden unabhängig voneinander an die Anforderungen angepasst. Abb. 90.19 zeigt ein Beispiel. Die Einschaltzeit wird durch R_Z vorgegeben, die Ausschaltzeit demgegenüber verkürzt durch die Parallelschaltung von R_Z und R_1 über D_2. Abb. 90.20 zeigt die sich bei Ansteuerung ergebenden Signalverläufe. Hinweis zu Abb. 90.20: Die einzelnen Diagramme haben unterschiedliche Amplituden und Einheiten. Sie wurden deshalb so gewählt, damit ihre *prinzipiellen* Verläufe übersichtlich und in zeitlicher Zuordnung zueinander dargestellt werden können. Das Einschalten wird verzögert durch den Miller-Effekt, wie der Verlauf von u_{GS} in Abb. 90.20 zeigt: u_{DS} fällt in sehr kurzer Zeit von U_B auf einige Volt, und damit muss die Kapazität C_{GD} zunächst umgeladen werden, was der Steuerspannung u_e entgegenwirkt (Bereich flacher Steigung von u_{GS}). Damit vergrößert sich die Einschaltdauer. Der gleiche Vorgang spielt sich prinzipiell beim Übergang des Transistors vom leitenden in den sperrenden Zustand ab mit dem Unterschied, dass die Umladung jetzt über Z_L erfolgen muss und nicht wie vorher über den leitenden Transistor ($|Z_L| > r_{DS}$); in Abb. 90.20 wiederum ein Bereich flacher Steigung von u_{GS}. Die Umladung nimmt also mehr Zeit in Anspruch und hängt jetzt auch wesentlich von R_Z ab. Um den Einfluss von R_Z zu verringern, liegt parallel zu ihm die Reihenschaltung aus D_2 und R_1. Sie ist wegen D_2 nur beim Ausschalten wirksam und schaltet R_1 (näherungsweise) parallel zu R_Z. In Abb. 90.20 sieht man am Diagramm von i_{RZ} (Einschalten, links im Diagramm, R_1 nicht wirksam) und den Diagrammen i_{RZ} und i_{R1} (Ausschalten, rechts im Diagramm) die Aufteilung der Ströme in diese 2 Anteile. Die Umladung nur über R_Z hätte mehr Zeit in Anspruch genommen und die Ausschaltzeit damit verlängert. Der Verlauf von u_{DS} und i_D ist interessant, denn damit können die Verluste beim Umschaltvorgang erfasst werden. Das Diagramm $p_{MOSFET} = p_{MOSFET}(t) = u_{DS}(t) \cdot i_{ZL}(t)$ zeigt den Momentanwert der im Transistor erzeugten Verlustleistung. Die Fläche unter der Kurve gibt die (Wärme-)Energie eines Umschaltvorganges an. Siehe dazu auch Abschn. 90.3.3 „Verlustleistung“.

Mit Feldeffekttransistoren stehen sehr schnelle Schalter zur Verfügung. So werden z. B. folgende Werte erreicht: $t_{(on)} \approx 50$ ns;

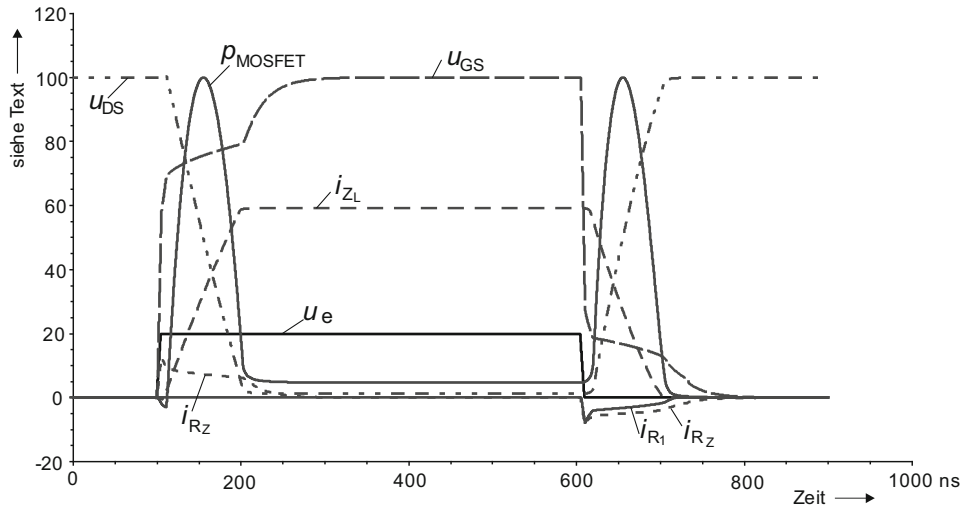


Abb. 90.20 Diagramme zu Abb. 90.19. Hinweise dazu im Text

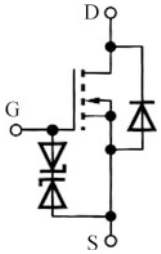


Abb. 90.21 Feldeffekttransistor mit interner Schutzbeschaltung

$t_{(\mathrm{off})} \approx 400\,\mathrm{ns}$ bei $R_G = 10\,\Omega$ (Abb. 90.18). Damit sind Pulsfrequenzen von mehr als 16 kHz möglich. Eine Hilfe bei der Schaltungsdimensionierung sind die im Teil XIII „Nachrichtentechnik“, Kap. 97 „Filter“, Abschn. 97.3.2 „Tiefpassfilter“ aufgeführten Simulationsprogramme.

90.3.3 Verlustleistung

Zur Berechnung der am Transistor entstehenden Verlustleistung können die beim Bipolartransistor verwendeten Gleichungen (90.6) bis (90.9) übernommen werden, wenn die Bezeichnungen entsprechend zugeordnet werden. Da die Gehäuse bei beiden Transistortypen häufig gleich sind, werden auch hier Wärmewiderstände von typisch (0,01 ... 0,2) K/W erreicht.

Eine grobe Abschätzung der durch Ein- und Ausschaltvorgänge entstehenden Verlustleistung $P_{\mathrm{dyn}}(f)$ im Transistor erhält man, wenn man in dem Produkt aus Drainstrom I_D (Schalter geschlossen) und Drain-Source-Spannung U_{DS} (Schalter geöffnet) während der Umschaltzeiten t_r bzw. t_f jeweils nur den halben Wert einsetzt und diesen Wert mit der Frequenz f multipliziert:

$$P_{\mathrm{dyn}}(f) \approx \frac{U_{\mathrm{DS}}}{2} \cdot \frac{I_D}{2}(t_r + t_f) \cdot f \qquad (90.10)$$

In dieser Gleichung sind nur die Verluste enthalten, die während des Umschaltens (Ein-Aus und Aus-Ein) auftreten; dazu müssen die Verluste addiert werden, die während des leitenden und während des gesperrten Zustandes entstehen. Für $u_{\mathrm{DS}} = 1000\,\mathrm{V}$, $i_D = 100\,\mathrm{A}$, $(t_r + t_f) = 10\,\mu\mathrm{s}$, $f = 50\,\mathrm{Hz}$ folgen daraus 12,5 W. Beim Einsatz der genannten Simulationsprogramme lässt sich der Verlauf von p_{MOSFET} integrieren und daraus ein genauerer Wert bestimmen als er sich aus der Näherungsrechnung nach (90.10) ergibt.

90.3.4 Schutzschaltungen

Ein Teil der für Feldeffekttransistoren erforderlichen Schutzschaltungen ist häufig schon in den Transistoraufbau integriert, siehe Abb. 90.21. Das ist zum einen die bereits erwähnte Diode nach Abb. 90.18, zum anderen eine Kombination aus zwei Z-Dioden, die verhindern soll, dass die Gate-Source-Spannung weder einen negativen Grenzwert unterschreitet noch den Wert U_{GSM} überschreitet. Eine zusätzliche externe Beschaltung muss sicherstellen, dass die Grenzwerte $U_{(\mathrm{BR})\mathrm{DSS}}$, I_{DM} und I_{DPM} nicht überschritten werden. Eine genaue Auskunft über die bereits integrierten Schutzschaltungen gibt das zum Transistor gehörende Datenblatt. Im Bedarfsfall lassen sich die bei den Bipolartransistoren dargestellten Schutzbeschaltungen auch bei Feldeffekttransistoren anwenden.

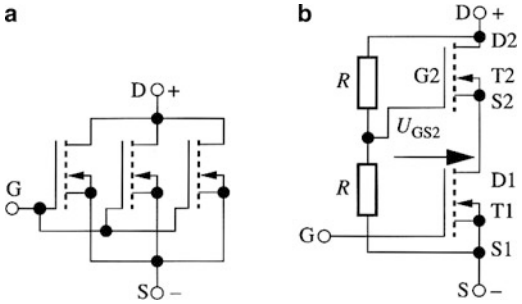


Abb. 90.22 Zusammenschaltung von Feldeffekttransistoren. **a** Parallelschaltung, **b** Reihenschaltung

90.3.5 Zusammenschaltung von Feldeffekttransistoren

1. Parallelschaltung

Im Gegensatz zu Bipolartransistoren können Feldeffekttransistoren durch Verbinden aller Gate- bzw. aller Source- bzw. aller Drain-Anschlüsse parallel geschaltet werden (Abb. 90.22a). Bei zunächst vorhandener unsymmetrischer Stromaufteilung der Einzeltransistoren erwärmt sich derjenige mit dem höchsten Drainstrom am meisten, vergrößert seinen Kanalwiderstand und verringert dadurch seinen Strom wieder, sodass sich die Teilströme der einzelnen Transistoren einander automatisch angleichen.

2. Reihenschaltung

Auch die Reihenschaltung ist nach Abb. 90.22b relativ einfach zu realisieren. Im gesperrten Zustand sorgen die Widerstände R für eine gleichmäßige Spannungsaufteilung unter den Transistoren. Wichtig ist die bei den Transistoren nicht in jedem Fall eingebaute Diode nach Abb. 90.19. Wenn nämlich T2 einen höheren Sperrwiderstand als T1 hat, schließen die Dioden von T2 Gate und Source kurz, sodass die über R festgelegte Spannungsteilung erzwungen wird. Beim Übergang von T1 in den leitenden Zustand sinkt das Drainpotenzial an D1 gegenüber dem von G2, und T2 wird automatisch leitend.

90.4 IGBT-Transistoren

Die Verbindung bestimmter Eigenschaften von bipolaren Transistoren und Feldeffekttransistoren führte zur Entwicklung einer neuen Transistorgeneration. Die geringe erforderliche Ansteuerleistung von Feldeffekttransistoren wurde mit der geringen Kollektor-Emitter-Restspannung eines Bipolartransistors kombiniert: Es entstand der IGBT.

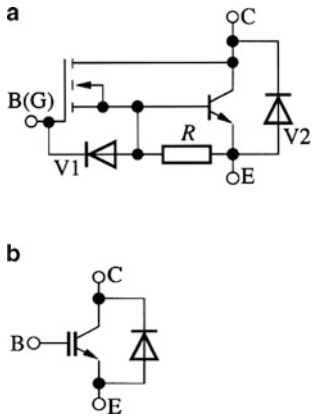


Abb. 90.23 Aufbau (**a**) und Schaltsymbol (**b**) eines IGBT

90.4.1 Aufbau von IGBT-Transistoren

1. Aufbau

Den Aufbau eines IGBT (*I*nsulated-*G*ate-*B*ipolar-*T*ransistor) zeigt Abb. 90.23a. Zusätzlich zu den zwei Transistoren sind noch Zusatzelemente eingebaut. V2 verhindert, dass der Kollektor C gegenüber dem Emitter E negativ wird. R sorgt für einen geringen nach außen wirkenden Kollektor-Reststrom des Bipolartransistors im Sperrzustand, und über V1 können die Basisladungen des Bipolartransistors beim Übergang vom leitenden in den sperrenden Zustand direkt ausgeräumt werden, wodurch sich die Speicherzeit verringern lässt. Es werden z. B. folgende Daten erreicht (Transistoren der mittleren Leistungsklasse): $U_{CE0} = 1200\,\text{V}$; $I_{CM} = 400\,\text{A}$; $t_r = 1\,\mu\text{s}$; $t_s = 3\,\mu\text{s}$; $t_f = 1\,\mu\text{s}$; $U_{CE\,sat} = 3\,\text{V}$; $R_{th\,JC} = 0{,}2\,\text{K/W}$. Die Bezeichnungen entsprechen denen eines Bipolartransistors, die Ansteuerung geschieht wie bei Feldeffekttransistoren.

IGBT werden auch in Modulform als Kombination von z. B. vier Einzeltransistoren zum Aufbau einer Brückenschaltung geliefert. Wegen der elektrischen Isolation der Einzeltransistoren untereinander liegt der Wärmewiderstand bei minimal etwa (0,05 ... 0,8) K/W.

Ein weiteres Beispiel für die Leistungsfähigkeit von IGBT-Halbleitern der oberen Leistungsklasse zeigen die auszugsweise aufgelisteten Daten eines Moduls mit 3 IGBTs auf einem Chip bei $\vartheta_J = 25\,°\text{C}$ (1MBI3600U4D-170, Fuji): max. 1700 V; max. 3600 A (Gleichstrom), $t_{on} \approx 1{,}8\,\mu\text{s}$, $t_{off} \approx 1{,}3\,\mu\text{s}$, $C_{GS} \approx 330\,\text{nF}$, $R_{th\,JG} \approx 0{,}0067\,\text{K/W}$; $U_{DS} \approx 2\,\text{V}$ (bei $\vartheta_J = 125\,°\text{C}$ und $I_D \approx 3500\,\text{A}$).

2. Schutzschaltungen

Die für einen IGBT-Transistor erforderlichen Schutzschaltungen setzen sich aus einer Kombination der Schutzschaltungen nach den Abb. 90.7 und 90.21 zusammen, wobei ein Teil bereits in den Transistoraufbau integriert sein kann.

3. Zusammenschaltung

IGBT können wie Feldeffekttransistoren in Reihe und parallel geschaltet werden, siehe auch Abb. 90.22. Bei der Parallelschaltung übernimmt der steuernde Feldeffekttransistor die Stromsymmetrierung, bei der Reihenschaltung sorgen die Z-Dioden zwischen Gate und Source mit den Widerständen R für eine symmetrische Spannungsaufteilung.

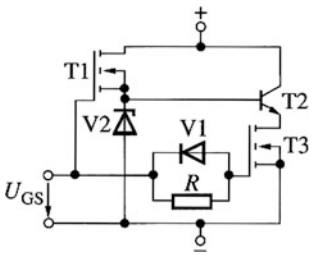


Abb. 90.24 BIMOS-Schaltstufe

90.4.2 BIMOS-Schaltstufe

Die Schaltzeiten des IGBT werden durch die Eigenschaften des Bipolartransistors bestimmt. Für besonders schnelle Schaltstufen wird eine Transistorkombination nach Abb. 90.24 eingesetzt. T1 ist für eine hohe Sperrspannung bemessen, hat aber einen relativ großen Drain-Source-Widerstand $R_{DS(on)}$, der in dieser Schaltung auch benötigt wird. T3 besitzt eine niedrige Sperrspannung, dafür aber einen niedrigen Bahnwiderstand. Im Einschaltmoment liefert T1 den Basisstrom für T2, der gleichzeitig durch $R_{DS(on)}$ von T1 begrenzt wird. T3 schaltet durch den Widerstand R verzögert gegenüber T1 ein (Zeitkonstante $\tau = R \cdot C_{GS}$), aber unverzögert über V1 aus. T2 und T3 dienen als Schalter. Im Ausschaltmoment sperrt T3 sofort, dadurch steigt das Emitterpotenzial von T2 stark an, und die Basis von T2 wird über die Z-Diode V2 in sehr kurzer Zeit entladen.

90.5 Thyristoren

90.5.1 Aufbau, statisches Verhalten, Kenngrößen, Grenzwerte

Ein Thyristor besteht aus vier Schichten abwechselnder Dotierung, PNPN. Abb. 90.25 zeigt diesen Aufbau und das Schaltzeichen. Über einen Steueranschluss G, das Gate, kann die Funktionsweise beeinflusst werden. Prinzipiell gilt, dass bei unbeschaltetem Gate der Thyristor, unabhängig von der Polarität der angelegten Spannung U_{AK}, stets sperrt: Für $U_{AK} > 0$ durch die Sperrschicht S2, Abb. 90.25, für $U_{AK} < 0$ durch die Sperrschicht S1 (S3 besitzt wegen der hohen Dotierung nur geringe Sperrfähigkeit). Der Thyristor lässt sich für $U_{AK} > 0$ dadurch in den leitenden Zustand versetzen, dass in die hochdotierte untere P-Schicht ausreichend Ladungsträger eingebracht werden. Ist diese Schicht räumlich genügend schmal, so diffundieren die Ladungsträger in die darüberliegende N-Schicht und heben

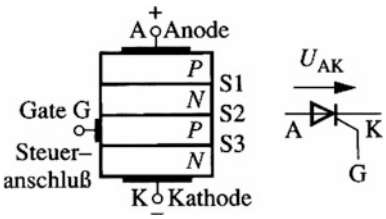


Abb. 90.25 Schematischer Aufbau und Schaltsymbol eines Thyristors (rückwärts sperrend)

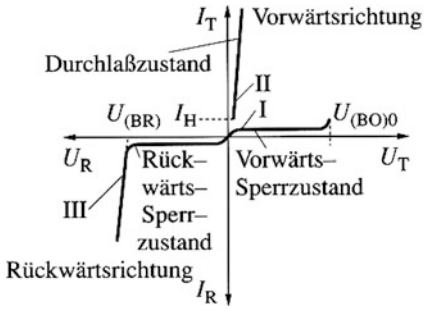


Abb. 90.26 Kennlinie eines rückwärts sperrenden Thyristors

die Sperrwirkung der Sperrschicht S2 auf. Der jetzt einsetzende Anoden-Kathoden-Strom sorgt für die Aufrechterhaltung des Stromflusses (*Zündvorgang*). Es genügt also, einmal kurzzeitig Ladungen in die Sperrschicht S2 einzubringen, um den Thyristor dauernd leitend zu machen. Eine Unterbrechung des Stromflusses ist nur dadurch zu erreichen, dass der Strom durch den Thyristor unter einen bestimmten Wert sinkt (z. B. durch den natürlichen Nulldurchgang des Netzwechselstromes). Der Übergang in den leitenden Zustand für $U_{AK} < 0$ ist nicht möglich.

Die Kennlinie in Abb. 90.26 hat drei voneinander unterscheidbare Bereiche, I, II und III. Der Bereich I wird *Vorwärts-Sperrzustand* (oder *Blockierzustand*) genannt und ist dadurch gekennzeichnet, dass die Anode positiv gegenüber der Kathode ist, der Gate-Anschluss unbeschaltet ist bzw. kein Strom in den Gateanschluss eingespeist wird und ein sehr geringer Strom, der Sperrstrom der Sperrschicht S2, fließt. Er steigt mit zunehmender Anoden-Kathoden-Spannung an. Beim Erreichen des kritischen Spannungswertes $_{U(BO)0}$, der *Nullkippspannung*, ist der Sperrstrom so groß geworden, dass der beschriebene Zündvorgang einsetzt. Dieser Vorgang, ausgelöst durch den Sperrstrom, heißt *Überkopfzünden*. Der Thyristor befindet sich jetzt im Durchlasszustand, seine Eigenschaften werden durch die Durchlasskennlinie, Bereich II, dargestellt. Die Anoden-Kathoden-Spannung, die Durchlassspannung, nimmt nur geringfügig mit steigendem Strom zu und beträgt weniger als 2 V bei Strömen von 1200 A.

Das Einbringen von Ladungen in die Sperrschicht S2 zum kontrollierten Zünden des Thyristors kann auf zwei Arten geschehen:

1. durch Einspeisen eines Stromes in den Gateanschluss; dies ist die gebräuchlichste Art der Zündung, da man Zeitpunkt und Dauer des Überganges in den leitenden Zustand durch den zeitlichen Verlauf des Gatestromes bestimmen kann;
2. durch Bestrahlen mit Licht; die Energie des Lichtes erzeugt zusätzliche Ladungsträger in der Sperrschicht S2, und bei ausreichender Anzahl wird der Thyristor ebenfalls gezündet. Anwendung findet dieses Verfahren im *Fotothyristor* (siehe Abschn. 90.5.8.5).

Die zwei folgenden Effekte können zu einem unkontrollierten und damit ungewollten Zünden eines Thyristors führen

1. durch den Sperrstrom selbst (Überkopfzünden; dabei hängt der Zündeinsatzpunkt unter anderem von der Temperatur ab; siehe auch Erläuterungen zu Abb. 90.26;
2. durch eine sehr große *Spannungssteilheit* (du/dt) der Anoden-Kathoden-Spannung. Da an der Sperrschicht S2 eine Spannung U steht und Ladungen Q vorhanden sind, kann man ihr gemäß $Q = C \cdot U$ eine Kapazität C zuordnen. Steigt nun die Anoden-Kathoden-Spannung und damit die Spannung an S2 in sehr kurzer Zeit (Bereich einige Mikrosekunden) merklich an, so muss sich der Kondensator auf den neuen Spannungswert aufladen, es fließt also ein zusätzlicher „Lade"-Strom in die Sperrschicht. Erreicht dieser Strom einen bestimmten Wert, wird der Thyristor gezündet.

Durch geeignete Schutzschaltungen bzw. Schaltungsdimensionierung wird erreicht, dass das Zünden eines Thyristors durch die zwei zuletzt genannten Effekte verhindert wird.

Ist der Thyristor gezündet und wird der fließende Strom ständig verringert, werden die durch den Stromfluss in die Sperrschicht S2 eingebrachten Ladungen irgendwann nicht mehr ausreichen, die Sperrfähigkeit aufzuheben: Der Thyristor erlischt, geht also in den Bereich I über (*Löschvorgang*). Derjenige Strom, der unterschritten werden muss, damit der Thyristor sicher löscht, ist der *Haltestrom* I_H, er ist in der Regel kleiner als 1/1000stel des maximal zulässigen Stromes durch den Thyristor.

Der Bereich III in Abb. 90.26 ist der *Rückwärts-Sperrzustand*, er ist gegeben durch die in Sperrrichtung betriebene Diode mit der Sperrschicht S1. Solange die Spannung am Thyristor die *Spitzensperrspannung in Rückwärtsrichtung*, U_{RRM}, nicht überschreitet, bleibt er gesperrt. Es gilt: $|U_{(BR)}| > |U_{RRM}|$. $U_{(BR)}$ ist die *Durchbruchspannung in Rückwärtsrichtung*.

Für den Betrieb sind außerdem folgende Kenn- bzw. Grenzwerte wichtig:

- U_{DRM}: *Spitzensperrspannung in Vorwärtsrichtung*, höchstzulässiger Augenblickswert der periodisch auftretenden Spannung im Sperrzustand in Vorwärtsrichtung;
- $I_{TAV(1)}$: *Dauergrenzstrom*, höchstzulässiger Mittelwert des Durchlassstromes (sinusförmige Halbschwingung bei 50 Hz, definierte Gehäusetemperatur);
- $I_{TSM(1)}$: *Stoßstromgrenzwert*, höchstzulässiger, nicht periodisch auftretender Augenblickswert eines einzelnen Stromimpulses bei definierter Gehäusetemperatur, Wiederholung erst nach definierter Mindestzeit zulässig;
- $I_{TRMS(1)}$: *Grenzeffektivstrom*, höchstzulässiger Effektivwert des Durchlassstromes bei definierter Gehäusetemperatur.

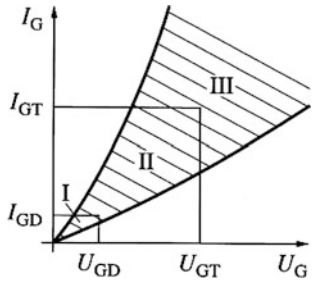


Abb. 90.27 Zünddiagramm eines Thyristors

Im Datenblatt des Herstellers stehen für (1) nähere Erläuterungen.

Daten für einen Thyristor (Serienfertigung, obere Leistungsklasse, TT1190N-Serie, Infineon, Auszug Datenblatt): $U_{DRM} \approx U_{RRM} \approx 1600\,\text{V}$; $I_{TAV(25\,^\circ\text{C})} \approx 1190\,\text{A}$; Durchlassspannung max. 2,05 V; Haltestrom min. 500 mA; Gate-Triggerspannung min. 2 V; Gate-Triggerstrom typ. 250 mA. Neueste Entwicklungen: 5 kA bei 8,5 kV und 6 kA bei 6,7 und 7,2 kV [7].

90.5.2 Dynamisches Verhalten

Die am häufigsten angewendete Art der Zündung ist das Einspeisen eines Stromes in den Gateanschluss. Wichtig ist dabei, dass eine bestimmte Anzahl Ladungen über das Gate in die Schicht S2 eingebracht wird. Nach einer *Zündverzögerungszeit* t_d von einigen Mikrosekunden breitet sich die leitende Strecke in einem Zeitraum von (10 ... 50) µs, der *Durchschaltzeit* t_r, über den gesamten Querschnitt aus. Während t_r tritt örtlich ein ungleichmäßiger Stromfluss und damit eine ungleichmäßige Stromdichte auf. Höhere Stromdichte hat eine höhere Temperatur zur Folge, was den Widerstand verringert und die Stromdichte weiter erhöht. Dieser positive Rückkopplungseffekt kann zum lokalen Schmelzen des Halbleitermaterials führen und den Halbleiter zerstören (2. Durchbruch). Man speist deshalb einen bis zum 10fachen Wert des minimal erforderlichen Zünd-Gatestromes ein, um diesen Übergangzustand so schnell wie möglich zu überwinden. Weiterhin wählt man für den Gateanschluss eine kammförmige Anordnung, um die Einspeisung parallel an mehreren Orten gleichzeitig vornehmen zu können. Abb. 90.27 zeigt ein Zünddiagramm, das Auskunft darüber gibt, in welchem Bereich sich die Gate-Kathoden-Spannung U_G und der Gatestrom I_G befinden müssen, um gar keine Zündung (Bereich I, wichtig für Störgrößen), eine mögliche (Bereich II, unerwünscht, da unsicher) oder eine sichere Zündung (Bereich III, $U_G \geq U_{GT}$; $I_G \geq I_{GT}$) zu erreichen. Die Verläufe von Gatestrom i_G, Spannung u_T am Thyristor und Strom i_T durch den Thyristor mit der dazugehörigen Verlustleistung $p = u_T \cdot i_T$ am Thyristor enthält Abb. 90.28. t_d und t_r sind die oben erwähnten Zündverzögerungs- bzw. Durchschaltzeiten. Wenn der Lastkreis, wie das häufig der Fall ist, Induktivitäten enthält, steigt der Strom nach der Zündung nur langsam an. Dann muss der Gatestrom solange fließen, bis der Laststrom den *Einraststrom* oder *dynamischen Haltestrom* erreicht hat. Dieser liegt bei

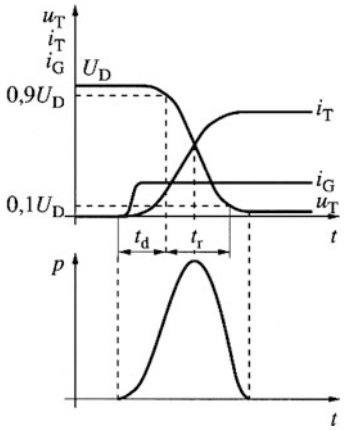


Abb. 90.28 Einschaltverhalten eines Thyristors

dem zwei- bis fünffachen Wert des Haltestromes eines Thyristors.

Es wurde oben bereits beschrieben, dass ein Zündvorgang auch durch eine zu hohe Spannungssteilheit (du/dt) hervorgerufen werden kann. Ein Grenzwert, die kritische Spannungssteilheit $(du/dt)_c$, wird in Datenblättern angegeben und liegt in der Größenordnung (200 ... 2000) V/µs. Dieser Wert soll von der Fertigung her möglichst groß sein (Störspannungsimpulse) und wird dadurch erreicht, dass man parallel zu der Strecke Gate–Kathode intern niederohmige Widerstände legt. Zwar wird dadurch die erforderliche Gate-Steuerleistung größer, der Thyristor benötigt aber weniger externe Schutzbeschaltung, damit die kritische Spannungssteilheit nicht überschritten wird. Das Ausschalten (Löschen) eines Thyristors kann nur durch Unterschreiten des Haltestromes geschehen. In Abb. 90.29 ist angenommen, dass die den Lastkreis speisende Spannung u_L eine periodische Wechselspannung ist. Nach dem Nulldurchgang von u_L ($t/T \geq 0{,}5$) fließt zunächst ein Strom in Rückwärtsrichtung, bis nach der *Ausräumzeit* t_{rr} die Ladungen aus der Sperrschicht abgeflossen sind. Für den Betrieb in Rückwärtsrichtung sperrt der Thyristor zwar, aber in der für das Sperren in Vorwärtsrichtung maßgebenden Sperrschicht S2 (Abb. 90.25) sind noch zu viele Ladungsträger vorhanden. Erst wenn diese nach Ablauf der *Freiwerdezeit* t_q ($t_q > t_{rr}$) ausgeräumt sind, kann ein Sperren in Vorwärtsrichtung erreicht werden. Die Zeit t_q liegt zwischen 400 µs (Netzthyristoren bei Betrieb mit 50 Hz) und 10 µs (spezielle Bauarten). Kleine Freiwerdezeiten sind z. B. für Wechselrichter mit Frequenzen oberhalb ca. 1 kHz erforderlich. Diese „*Frequenzthyristoren*“ sind in speziellen Fertigungsverfahren hergestellt worden, wobei u. U. eine etwas erhöhte Durchlassspannung in Kauf genommen werden muss.

Kleine Freiwerdezeiten haben auch die *asymmetrisch sperrenden* und die *rückwärts leitenden* Thyristoren. Ein asymmetrisch sperrender Thyristor besitzt in Rückwärtsrichtung nur eine vergleichsweise geringe Sperrfähigkeit (Größenordnung 20 V), dafür nehmen Freiwerdezeit und Durchlassspannung ab. Diese Leistungshalbleiter lassen sich allerdings nur dort einsetzen, wo auf die Sperrfähigkeit in Rückwärtsrichtung verzichtet werden

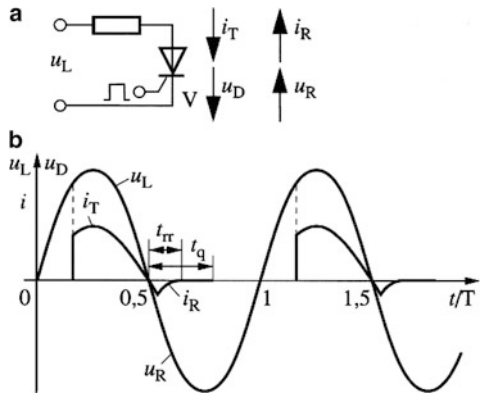


Abb. 90.29 Ausschaltverhalten eines Thyristors

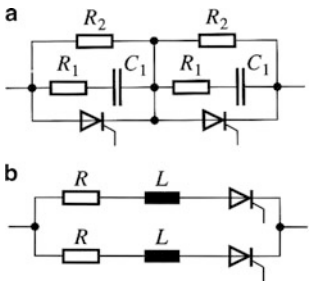


Abb. 90.30 Zusammenschaltung von Thyristoren mit Schutzbeschaltungen. **a** Reihenschaltung, **b** Parallelschaltung

kann. Bei rückwärts leitenden Thyristoren liegt zur Anoden-Kathoden-Strecke eine Diode in Antiparallelschaltung. Dadurch nehmen Freiwerdezeit und Durchlassspannung auf etwa das 0,5fache gegenüber symmetrisch sperrenden Thyristoren ab. Ein Betrieb in Rückwärtsrichtung ist auch hier nicht möglich.

90.5.3 Verlustleistung

Abb. 90.28 zeigt auch den prinzipiellen Verlauf der Verlustleistung $p(t)$ beim Zünden. Beim Löschen ergibt sich ein ähnlicher Verlauf, sodass sich die Verlustenergie am Thyristor pro Schaltzyklus durch Addition beider Flächen ergibt. Die Berechnung der am Thyristor entstehenden Gesamtverlustleistung geschieht gemäß (90.7) bis (90.9). Zur groben Abschätzung der Umschaltverluste allein kann auch wieder (90.10) verwendet werden

In den Datenblättern wird noch eine *kritische Stromsteilheit* $(di/dt)_c$ angegeben. Sie hat Bedeutung für die Einschaltverluste. Bei einer Induktivität im Lastkreis steigt der Strom nach dem Zünden verlangsamt an, sodass die stationäre niedrige Durchlassspannung am Thyristor bereits erreicht ist, wenn der Strom eine nennenswerte Größe erreicht. Damit sind die auftretenden Einschaltverluste klein. Wenn sich dagegen ein Kondensator im Lastkreis befindet (Kondensatorlöschung) und sich im Einschaltmoment über den Thyristor entlädt, dann steigt der Strom sofort stark an, noch bevor sich die Spannung am Thyristor wesentlich verringert. Daraus folgt eine höhere Einschaltverlustleistung. Bei hohen Wiederholfrequenzen kann der Thyristor thermisch zerstört werden. Die kritischen Werte liegen bei etwa (50 . . . 400) A/µs. Eventuell erforderliche Schutzmaßnahmen werden im Abschn. 90.5.4 vorgestellt.

90.5.4 Schutzschaltungen

Beim Einsatz von Thyristoren können einige oder auch alle der nachfolgend genannten acht kritischen Betriebsfälle eintreten, wenn sie nicht durch geeignete Schutzschaltungen verhindert werden:

1. Überschreiten der höchstzulässigen Spitzensperrspannung in Vorwärts- bzw. Rückwärtsrichtung (U_{DRM}, U_{RRM});
2. Überschreiten der höchstzulässigen Spitzensperrspannung in Vorwärts- bzw. Rückwärtsrichtung durch den Trägerspeichereffekt, insbesondere beim GTO-Thyristor während des Löschens bei induktiver Last;
3. ungleichmäßige Spannungsaufteilung, statisch und dynamisch, bei Reihenschaltung mehrerer Thyristoren;
4. ungleichmäßige Stromaufteilung, statisch und dynamisch, bei Parallelschaltung mehrerer Thyristoren;
5. Überschreiten der kritischen Spannungssteilheit (du/dt);
6. Überschreiten der kritischen Stromsteilheit (di/dt_c);
7. Überschreiten der höchstzulässigen Kristalltemperatur ϑ_J durch zu hohe Verlustleistung;
8. Überschreiten der höchstzulässigen Grenzströme (I_{TAV}, I_{TRMS}).

In Abb. 90.30 sind Schutzschaltungen für die Parallel- und Reihenschaltung von Thyristoren gezeigt. Diese Schutzschaltungen bieten auch teilweise einen mehr oder weniger wirkungsvollen Schutz gegen die genannten kritischen Betriebsfälle. Es folgen einige grundsätzliche Überlegungen zur Schutzbeschaltung, im konkreten Fall sind die Datenblätter, Applikationsschriften und Dimensionierungsvorschläge der Hersteller maßgebend.

Schaltung a) erfüllt die Forderungen der Punkte 2, 3 und 5, die von Punkt 1 teilweise, je nach Art der auftretenden Überspannung. Werden keine Thyristoren in Reihe geschaltet, entfällt R_2. R_1 und C_1 werden nach den Forderungen von Punkt 2 bzw. Punkt 5 dimensioniert und erfüllen dann gleichzeitig Punkt 3 mit (s. Ausführungen zu den Schutzmaßnahmen bei den Halbleiterdioden, Abschn. 90.1.5). Nach Schaltung b) sind zu jedem Thyristor ein Widerstand R und eine Induktivität L in Reihe geschaltet. Werden mehrere Thyristoren parallel geschaltet, so sorgt R für eine statische Stromsymmetrierung bei allerdings entsprechend höheren Verlusten. Eine zum Schutz des Thyristors in Reihe liegende Sicherung kann u. U. diese Aufgabe übernehmen. Die Induktivität L verbessert bei Parallelschaltung die Stromsymmetrierung bei Stromänderungen (Punkt 4). Sie kann auch das Erreichen der kritischen Stromsteilheit nach Punkt 6 verhindern; Einzelheiten dazu siehe im Abschn. 90.5.3, Verlustleistung. Werden keine Thyristoren parallel geschaltet, kann der Widerstand R entfallen. Um in der Schutzschaltung nach Abb. 90.30a eine ausreichende Schutzwirkung zu erhalten, wird der Widerstand R_1 im Bereich (50 . . . 200) Ω gewählt.

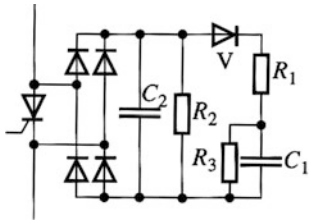


Abb. 90.31 Schutzbeschaltung eines Thyristors, Brückenschaltung

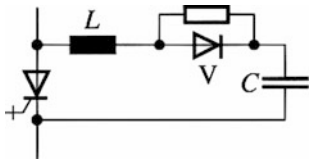


Abb. 90.32 Schutzbeschaltung eines GTO-Thyristors

Zündet der Thyristor, so entlädt sich der Kondensator (C_1 im Bereich (0,1 ... 0,47) µF) über den Thyristor und erhöht die Stromsteilheit (Punkt 6).

Abb. 90.31 zeigt eine häufig angewendete Schutzbeschaltung, die Brückenschaltung. Ihr Ziel ist es, die beim Zünden auftretenden Kondensator-Entladeströme zu vermeiden. C_2 dient zum Schutz vor einer zu großen Spannungssteilheit, er kann sich über R_2 entladen. R_1 und C_1 bieten Schutz vor Überspannungen durch den Trägerspeichereffekt bzw. anderen auftretenden Überspannungen. Die Diode V verhindert einerseits, dass ein momentan aufgeladener Kondensator C_1 die Schutzwirkung von C_2 beeinträchtigt, andererseits wird C_1 zu C_2 parallelgeschaltet, falls die Spannung an C_2 größer werden will als an C_1. Wird der Thyristor gezündet, kann sich wegen der Dioden-Brückenschaltung keiner der Kondensatoren über die Anoden-Kathoden-Strecke entladen und dadurch die Stromsteilheit erhöhen.

In der Schutzschaltung für GTO-Thyristoren nach Abb. 90.32 werden die Forderungen der Punkte 1 bis 7 relativ gut erfüllt: Zündet der Thyristor, kann sich der Kondensator C wegen der Induktivität L nur verzögert entladen. Bei auftretenden Überspannungen in Vorwärtsrichtung wird der Kondensator C über die Diode V direkt zum Thyristor parallelgeschaltet. Die Überschreitung der Grenzströme, Punkt 8, lässt sich durch den Einbau von Sicherungen weitgehend vermeiden. Eines muss aber stets bedacht werden: Diese Maßnahmen sind nur wirkungsvoll bei Störungen, die im Mikrosekundenbereich ablaufen und dann beendet sind. Siehe dazu auch Abschn. 91.5 Schutzschaltungen.

90.5.5 Zusammenschaltung von Thyristoren

Mehrere Thyristoren können sowohl parallel als auch in Reihe geschaltet werden, Abb. 90.30. Wichtig ist, dass die im vorigen Abschnitt erwähnten Schutzbeschaltungen vorgesehen werden, um ungleichmäßige Aufteilungen von Spannungen und

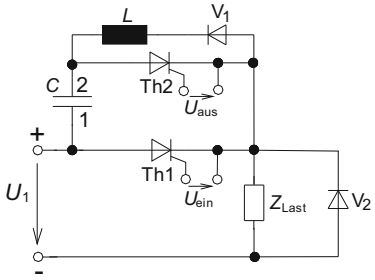


Abb. 90.33 Thyristor-Löschkreis, Beispiel 1

Strömen weitgehend zu verhindern. Es soll daran erinnert werden, dass zwischen dem Halbleiter und seinem Gehäuse eine elektrisch leitende Verbindung besteht (geringer Wärmewiderstand $R_{\text{th JG}}$), sodass z. B. bei einer Reihenschaltung nur durch zusätzliche elektrische Isoliermaßnahmen mehrere Thyristoren auf dem gleichen Kühlkörper montiert werden können.

90.5.6 Löschen von Thyristoren

Thyristoren können nur dadurch vom Leit- in den Sperr-Zustand gebracht werden, dass der Anoden-Kathoden-Strom einen Mindestwert, den Haltestrom, für (300 ... 800) µs unterschreitet, der bei Leistungsthyristoren im Bereich von etwa (0,5 ... 5) A liegt. Der genaue Wert ist dem Datenblatt des Halbleiters zu entnehmen. Wird der Thyristor mit Wechselspannung betrieben (z. B. 50-Hz-Versorgungsnetz), wechselt der Strom 50mal pro Sekunde die Polarität, ist also 100mal pro Sekunde für eine ausreichend lange Zeit ausreichend klein, um den Thyristor zu löschen. Diese Löschung geschieht automatisch, deshalb muss der Thyristor nach jedem Stromnulldurchgang wieder neu gezündet werden, in diesem Beispiel also 100mal pro Sekunde. Schaltet der Thyristor eine Gleichspannung, muss die Löschung durch externe Schaltungsmaßnahmen erzwungen werden. Es sind mehrere elektronische Schaltungen zur Löschung entwickelt worden, von denen 2 hier vorgestellt werden sollen. Gemeinsames Kennzeichen ist, dass der Strom durch den Thyristor zusammen mit einem zusätzlich eingespeisten Strom in entgegengesetzter Richtung den Haltestrom unterschreitet.

In Abb. 90.33 ist Th1 der Schaltthyristor, Th2 der Löschthyristor. Zuerst wird Th2 bei gesperrtem Th1 gezündet, und C lädt sich über Th2 und Z_{Last} auf (nahezu) U_1 auf; + an Punkt 1, − an Punkt 2. Der Aufladevorgang endet, wenn der Aufladestrom den Haltestrom von Th2 unterschreitet. Jetzt wird Th1 gezündet. C lädt sich über Th1 und V1 um, weil C und L einen Schwingkreis bilden: + an Punkt 2, − an Punkt 1. Ein weiteres Umladen verhindert die Diode V1. Zum Sperren von Th1 wird Th2 gezündet. C entlädt sich über Th2 und Th1. Dabei sind Kondensatorstrom und Laststrom entgegengesetzt gerichtet. Ist der Entladestrom von C gleichgroß oder größer als der Laststrom, wird der Haltestrom von Th1 unterschritten und Th1 löscht. Während der folgenden Sperrphase von Th1 kann sich C über Zünden von Th2 erneut umladen auf U_1: + an Punkt 1, − an Punkt 2. Für die Zündfolge muss das Steuergerät einen geeigneten Ablauf

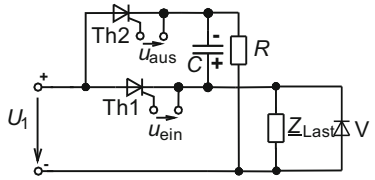


Abb. 90.34 Thyristor-Löschkreis, Beispiel 2

bereitstellen, und die Bauteile müssen entsprechend dimensioniert werden. Die Resonanzfrequenz des Schwingkreises ist so zu dimensionieren, dass sich C während der Leitphase von Th1 sicher umladen kann. C muss so groß sein, dass sein Entladestrom nach der Zündung von Th2 größer ist als der Laststrom.

Abb. 90.34 zeigt eine weitere Löschschaltung. Th1 leitet, Th2 sperrt zunächst. Der Kondensator C wird über R auf (nahezu) U_1 in der eingezeichneten Polarität aufgeladen. Soll Th1 gelöscht werden, wird Th2 gezündet. Der Entladestrom von C über Th1 und Th2 muss dazu größer sein als der Laststrom. Dann fließt ein „negativer" Strom durch Th1, der Haltestrom ist also sicher unterschritten. Sobald Th1 sperrt, löscht automatisch auch Th2, weil der Entladestrom von C Null wird. Die Bauelemente R und C müssen einige Bedingungen erfüllen. Der Strom durch R muss kleiner sein als der Haltestrom von Th2, andernfalls löscht Th2 nach Aufladung von C nicht. C muss ausreichend groß sein, um für ca. (300 ... 800) µs den „negativen" Strom durch Th1 aufrecht halten zu können. Die Zeitkonstante $R \cdot C$ muss es gestatten, dass sich der Kondensator in der Leitphase von Th1 auf (nahezu) U_1 aufladen kann.

90.5.7 Einsatz in kritischen Anwendungsfällen

Es kann sein, dass der Hersteller im Datenblatt darauf hinweist, dass er bei einem vorgesehenen Einsatz der Thyristoren in sicherheits- und ausfallkritischen Anwendungen wie in der Luftfahrt, der Medizintechnik und dgl. in den Herstellungsprozess eingebunden werden sollte, um die Risiken zu minimieren. Oder aber er rät vom Einsatz dieses Bauelementes in den genannten kritischen Anwendungen ab.

90.5.8 Spezielle Bauarten von Thyristoren

Neben den bereits erwähnten Bauarten „asymmetrisch sperrend" und „rückwärts leitend" haben sich weitere Sonderbauarten herausgebildet.

90.5.8.1 GTO-Thyristor

Prinzipiell lässt sich jeder Thyristor über einen negativen Gatestrom abschalten. Der GTO-(*G*ate-*T*urn-*O*ff-)Thyristor ist für diesen Anwendungsfall optimiert. Durch den negativen Gatestrom werden der Sperrschicht S2 (Abb. 90.25) so viele Ladungen entzogen, dass sich der Stromfluss nicht mehr aufrecht halten kann. Der PN-Übergang geht in den Vorwärts-Sperrzustand über. Der in das Gate einzuspeisende Strom zum

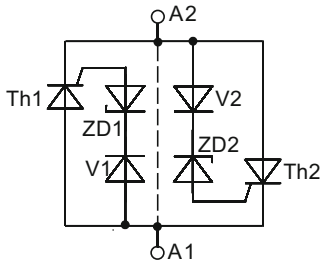


Abb. 90.35 Interner Aufbau eines Diacs

Löschen liegt in der Größenordnung 1/3 bis 1/5 des Laststromes. Durch die aufwändige Fertigung sind GTO-Thyristoren teuer, und sie werden deshalb vorwiegend bei Gleichspannung und großen Leistungen im Bereich $P > 500\,\text{kW}$ eingesetzt.

Abb. 90.32 und der zugehörige Text, Abschn. 90.5.4, stellen eine Schutzbeschaltung für einen GTO-Thyristor dar.

90.5.8.2 DIAC

Diacs (*Di*ode for *A*lternating *C*urrent) sind vom Prinzip her Thyristoren ohne Steueranschluss (Gate). Sie zünden beim Erreichen der Nullkippspannung in Vorwärtsrichtung (Überkopfzünden) und sperren in Rückwärtsrichtung. Wie ein Thyristor gehen sie beim Unterschreiten des Haltestromes in den sperrenden Zustand über. Damit die Bedingungen für das Überkopfzünden bei der Fertigung vorgegeben werden können (Zündspannung), wird in den Diac eine Z-Diode integriert. Außerdem ist sein Einsatz nur bei Wechselspannung sinnvoll. Er besteht deshalb nach Abb. 90.35 aus den 2 antiparallelgeschalteten gleichen Bauteilgruppen 1 (V1, ZD1, Th1) und 2 (V2, ZD2, Th2). Für $U_{A2} < U_{A1}$ sind V1, ZD1 und Th1 wirksam, für $U_{A2} > U_{A1}$ entsprechend V2, ZD2, Th2. Steigt das Potential vom Anschluss A_2 gegenüber Anschluss A1 an, sperrt der Diac solange, bis die Durchbruchspannung der Z-Diode ZD2 erreicht ist. V2 beginnt zu leiten, und bei weiterem Ansteigen des Potenzials wird ein zunehmender Strom in das Gate von Th2 eingespeist, er zündet. Baugruppe 1 ist über V1 gesperrt. Abb. 90.37a zeigt den prinzipiellen Kennlinienverlauf. In einer praktisch dimensionierten Schaltung ist es sinnvoll, den Gatestrom sprunghaft ansteigen zu lassen, um ein schnelles und sicheres Zünden zu erreichen. Weiterhin muss der Gatestrom begrenzt werden, und er muss unterhalb des Zündstromes von Th2 liegen, wenn Th2 den Haltestrom unterschreitet und löschen soll. Die Durchbruchspannung der Z-Diode, U_{RBr}, ist fertigungstechnisch vorwählbar und damit auch die Zündbedingung von Th2. Für $U_{A2} < U_{A1}$ arbeitet die Baugruppe 1 entsprechend. Der Bereich der Z-Dioden-Durchbruchspannung U_Z erstreckt sich bei industriell gefertigten Diacs in Sprüngen von 5 bis 10 V von etwa 15 bis 200 V. Der Bereich für Anwendungen ist eingeschränkt, weil der Diac nur bei einem über die Z-Dioden fest vorgegebenen Wert $|U_{A2} - U_{A1}| \approx U_Z + (0{,}7 + 2 \ldots 3)\,\text{V}$ zündet und diese Zündung nicht mit einem optimierten Zündimpuls erfolgt. Der jeweils nicht beteiligte Thyristor zündet nicht, weil er mit einem „negativen" Strom angesteuert wird, der zum Zünden aber nicht ausreicht, siehe GTO-Thyristor. Der maximal zuläs-

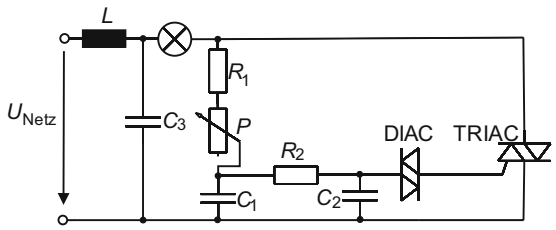


Abb. 90.36 Dimmerschaltung, siehe dazu auch Abb. 90.29

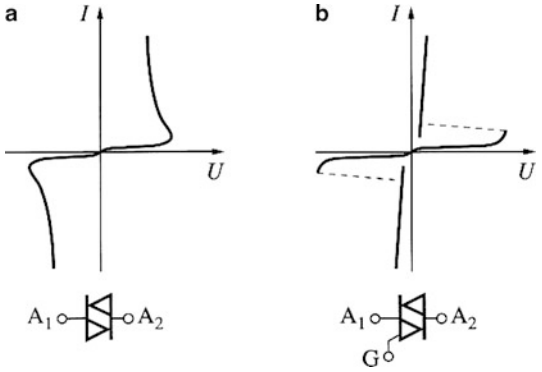


Abb. 90.37 Kennlinie und Schaltsymbol von **a** DIAC, **b** TRIAC

sige Laststrom liegt damit bei etwa 1 A. Insofern werden Diacs dem Kleinleistungsbereich zugeordnet.

Anwendungen:

- *Impulserzeugung allgemein*. Man erhält einen Impuls mit einem steilen Abfall und einem darauffolgenden Anstieg gemäß einer *e*-Funktion, wenn man einen Kondensator über einen Widerstand dauernd auflädt und parallel zum Kondensator einen Diac schaltet. Heute gibt es zur Impulserzeugung auf diese Anwendung optimierte Integrierte Schaltkreise.
- *Lebensdauererhöhung beim Leuchtmittel „Glühlampe“*. a) Wird ein Diac in Reihe zur „Glühlampe“ (Nennspannung 230 V) geschaltet, verringert sich der Effektivwert der Spannung um einige Prozent bei einer Durchbruchspannung der Z-Diode von 20 V, was zu einer wesentlichen Erhöhung der Lebensdauer führt. (Die verringerte Helligkeit lässt sich evtl. durch die nächst höhere Nenn-Wattzahl des Leuchtmittels ausgleichen.) b) In den normalen Ein-Aus-Schalter der Hausinstallation ist ein Integrierter Schaltkreis mit einer Dimmerschaltung integriert (siehe Abschn. 90.5.8.3 „Triacs“), der nach dem Ein-Schalten des Schalters den Dimmer selbsttätig innerhalb von etwa 1 s von dunkel bis maximal hell einstellt und damit den für Glühlampen als Kaltleiter sehr hohen Einschaltstrom (etwa 8- bis 10-facher Wert des Nennstromes im Heißzustand) begrenzt und damit lebensdauerverlängernd wirkt.
- *Zündimpulserzeugung in Energiesparlampen*. Die Zündimpulserzeugung in „Leuchtstoffröhren“ erfolgt überwiegend mit Glimmlampen (Haushalt, Kleingewerbe), in den im Haushalt eingesetzten „Energiesparlampen“ diverser Bauformen auch mit Diacs. Die „Energiesparlampen“ sind für die Anwendung im häuslichen Bereich ebenfalls auf dem Rückzug und werden zunehmend durch LED-Leuchtmittel ersetzt.

Abb. 90.36 zeigt das Schaltbild eines Dimmers. Das Prinzip besteht darin, dass ab jedem Nulldurchgang der sinusförmigen Netzspannung (50 Hz $\hat{=}$ 20 ms $\Rightarrow$ Nulldurchgang alle 10 ms) der Triac erst nach Ablauf einer über ein Potentiometer einstellbaren Zeit zwischen etwa 0,5 und 9 ms verzögert gezündet wird. Die Glühlampe wird also mit nahezu der gesamten Sinushalbwelle gespeist (0,5 ms Verzögerung, fast volle Helligkeit) bis herunter zur Speisung mit nur dem Anteil zwischen 9 und 10 ms (minimale Helligkeit). Der Leistungsverbrauch der Glühlampe darf nicht zu klein sein, weil es dann Probleme mit dem Haltestrom gibt. Dieses Verfahren wird auch mit *Phasenanschnittsteuerung* bezeichnet, siehe dazu auch Abb. 90.29. Dort ist der „verspätet“ einsetzende Stromfluss durch die Last und das automatische Löschen beim Stromnulldurchgang zu erkennen. Die Elemente C_3 und L in Abb. 90.36 dienen zur Verringerung der Oberschwingungsamplituden des nicht mehr sinusförmigen Stromverlaufes (Störsignalunterdrückung). Der Triac sei gesperrt, was bei jedem Stromnulldurchgang der Glühlampe automatisch geschieht. Der Kondensator C_1 lädt sich über R_1 und P nach einer *e*-Funktion auf. Die Spannung an C_2 folgt der an C_1 nur geringfügig verzögert, R_2 und C_2 sollen verhindern, dass kurze Störimpulse Diac und Triac zünden. Erreicht die Spannung an C_2 die Durchbruchspannung des Triacs (etwa (15 . . . 30) V), zündet dieser, seine Spannung sinkt innerhalb einiger µs auf wenige V, und am Gate des Triacs steht eine ausreichend hohe Spannung, um diesen zu zünden. C_1 und C_2 entladen sich über R_2, Diac und Triac und liefern den erforderlichen Zünd-Gatestrom. Über *P* kann die Aufladezeit und damit das Erreichen der Diac-Durchbruchspannung in Grenzen eingestellt werden. Schleifer von *P* in oberer Stellung heißt schnelle Aufladung – frühe Zündung – große Helligkeit; Schleifer von *P* in unterer Stellung entsprechend langsame Aufladung – späte Zündung – geringe Helligkeit.

Die Diacs haben, wie die Anwendungen zeigen, an Bedeutung verloren.

90.5.8.3 TRIAC

Ein *TRIAC* (*Tri*ode-*A*lternating-Current-Switch) besteht aus der Antiparallelschaltung zweier Thyristoren mit verbundenem Gate. Allerdings sind die zwei Thyristoren unterschiedlich aufgebaut, weil sich der Triggerimpuls zwangsweise stets auf eine der beiden Anoden bezieht. Der Zündvorgang ist damit nicht optimiert und verlängert die Zündzeit. Für kleine Leistungen bei einer Frequenz von 50 Hz und sinusförmigen Größen entstehen dadurch kaum Probleme. Der Triac ist damit für den Einsatz im 50-Hz-Wechselstromnetz gut geeignet, wobei die Löschung des leitenden Thyristors automatisch beim Stromnulldurchgang stattfindet. Damit lassen sich auch Verbraucher mit kombinierter ohmsch-induktiver Last ohne „Freilaufdiode“ oder ähnliche Schutzschaltungen betreiben (z. B. Universalmotoren in Bohrmaschinen, Heizung und Motoren in Waschmaschinen). Der Schaltungsaufbau ist sehr einfach, wie das Schaltbild einer *Dimmerschaltung* für das Leuchtmittel „Glühlampe“ zur Helligkeitseinstellung mittels Phasenanschnitt in Abb. 90.36, Abschn. 90.5.8.2, zeigt. Wichtig für eine sichere Betriebsweise ist ein niedriger Widerstand zwischen dem Gate-Anschluss und

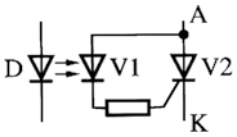


Abb. 90.38 Fotothyristor als Hilfsthyristor

der Kathode. Er ermöglicht die Zündung in beiden Schaltrichtungen. Abb. 90.37b zeigt die Kennlinie und das Schaltzeichen. TRIACS werden für Spannungen bis ca. 1500 V und Ströme bis ca. 150 A gefertigt und bevorzugt bei Netzfrequenz 50 Hz eingesetzt. Für höhere Ströme und höhere Schaltfrequenzen sind Thyristoren besser geeignet.

Anwendungen:

- *Schaltungen zur Leistungsreduzierung* (siehe auch Abschn. 90.5.8.2 DIAC). Beispiele: Beim Leuchtmittel „*Glühlampe*“ ist es wünschenswert, die Helligkeit einstellbar zu machen, bei Bohrmaschinen ist es praktisch, Drehzahl und Drehmoment vorwählen zu können. Die Heizleistung bei Waschmaschinen kann vorgewählt werden. Für diese Anwendungen wird entweder die Kombination aus Triac und Diac eingesetzt, oder es werden optimierte Integrierte Schaltkreise zur Ansteuerung der Triacs verwendet. Zwar ist die Glühlampe auf dem Rückzug, aber das Prinzip der Helligkeitseinstellung wird auch bei den Nachfolgern, den LED-Leuchtmitteln, erfolgreich eingesetzt, wenn auch mit abgewandelten elektronischen Verfahren.
- *Phasenanschnittsteuerung zur Leistungsreduzierung*. Eingesetzt im 4-Leiter-Drehstromnetz 230/400 V.

90.5.8.4 GATT

Beim *GATT* (*G*ate-*A*ssisted-*T*urn-Off-*T*hyristor, abschaltunterstützter Thyristor) wird die Freiwerdezeit dadurch verkürzt, dass nach Unterschreiten des Haltestromes eine negative Spannung an das Gate gelegt wird. Die in den Halbleiterschichten vorhandenen Ladungen können so in weniger als der halben Zeit gegenüber einem normalen Thyristor ausgeräumt werden.

90.5.8.5 Foto-Thyristor

Es wurde bereits darauf hingewiesen, dass auch durch Lichteinwirkung Ladungsträger erzeugt werden können. Thyristoren werden in der Regel so gefertigt, dass der Anodenanschluss, bei Hochleistungsthyristoren auch der Kathodenanschluss, mit Metallscheiben elektrisch verbunden ist, sodass ein guter Wärmeübergang zwischen dem Halbleiterkristall und einem Kühlkörper gewährleistet wird. Um das Potential des Thyristors von dem des Steuergerätes – auch aus Sicherheitsgründen – galvanisch zu trennen, führt man einen Lichtleiter mit dem einen Ende bis in die Nähe der Sperrschicht S2 (Abb. 90.25). Am anderen Ende befindet sich eine Leucht- oder Laserdiode, die mit einem kurzen Stromimpuls hoher Stromstärke (Dauer ca. (0,1 . . . 10) ms, Strom ca. (0,2 . . . 2) A) angesteuert wird. Da die Lichtleistung, die die Sperrschicht erreicht, im unteren Milliwatt-Bereich liegt, muss zur Zündung ein Hilfsthyristor nach Abb. 90.38 eingesetzt werden. Der Aufbau sieht dann so aus, dass der Hilfsthyristor V1 ein Fotothyristor kleiner Leistung ist, der den Leistungsthyristor V2 zündet. Manchmal ist V1 bereits in den Thyristoraufbau von V2 integriert. Über die Länge des Lichtleiters kann eine galvanische Trennung erreicht werden.

Einsatz von Leistungshalbleitern als Schalter

91

Wilfried Plaßmann



91.1 Schalter allgemein 1110
91.2 Halbleiter-Relais 1111
91.3 Halbleiterkühlung 1112
91.4 Ansteuerschaltungen 1115
91.5 Schutzschaltungen 1120

Teil XII

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_91

91.1 Schalter allgemein

Schalter dienen dazu, elektrische Kreise an einer vorgesehenen Stelle zu verbinden (Schalter EIN) oder aufzutrennen (Schalter AUS). In der Energietechnik werden in der Regel Verbraucher oder Energienetze ein- bzw. ausgeschaltet. Es gibt zwei Arten von Schaltern: 1. Elektronische Schalter (Halbleiterschalter) mit den Bauteilen Transistoren, Thyristoren, Triacs und deren Sonderformen und 2. Elektromechanische Schalter, die mit einem Elektromagneten aufgebaut sind, der mechanisch mittels Kraftwirkung Kontakte zusammenpresst oder trennt. Beide Schalterarten haben ihre Vor- und Nachteile, sodass letztlich der Einsatz den geeigneten Typ festlegt. Der *Steuerkreis* stellt die für das Betätigen des Schalters erforderliche Leistung zur Verfügung (Spannung, Strom), im *Lastkreis* wird die gewünschte elektrische Verbindung hergestellt oder unterbrochen.

Vorteile von elektronischen Schaltern, eingesetzt in der Leistungselektronik, sind im Wesentlichen:

- *Kurze Schaltzeiten* im mittleren µs-Bereich bzw. hohe Grenzfrequenzen im kHz-Bereich. Erst dadurch werden neue technische Lösungen möglich, z. B. Gleichstromsteller.
- *Keine mechanisch bewegten Teile* und damit keine Geräuschentwicklung, keine Erschütterungen, keine Materialermüdung.
- *Kein Kontaktprellen*.
- *Ansteuerbar über Rechnerprogramm*, d. h. viele Schaltvorgänge (EIN oder AUS) gleichzeitig oder gezielt zeitversetzt durchführbar; Notprogramme schnell und gezielt einleitbar, Fehlerauswirkungen (Kurzschluss) schnell und wirksam begrenzbar.
- *Kein mechanischer Kontakt-Verschleiß*, damit hohe Anzahl von Ein-Aus-Schaltzyklen. Siehe *kurze Schaltzeiten* oben.
- *Kein Lichtbogen beim Öffnen oder Schließen der Kontakte*. Wichtig für den Einsatz in explosionsgefährdeten Umgebungen. Trotzdem: Überspannungen beim Schalten induktiver Lasten vermeiden.
- *Eigensicherheit bei Schalter-(Halbleiter-)Defekt*. Darunter soll verstanden werden, welchen Zustand die Schalter bei Defekt (tritt in der Regel während eines Schaltvorganges ein) einnehmen. Beim Schalter-(Halbleiter-)Defekt gibt die Ausfallstatistik (evtl. Datenblatt oder Hersteller) eine Tendenz an, die aber zur Schadensbegrenzung nur bedingt hilfreich ist. Allerdings können durch zusätzliche oder schon vorhandene Sensoren, die auch zur intelligenten Ansteuerung von z. B. Thyristoren beim Zündvorgang eingesetzt werden, diese Defekte erkannt und geeignete Maßnahmen eingeleitet werden. Siehe Abschn. 91.4 „Ansteuerschaltungen“.
- *Eigensicherheit bei Spannungswiederkehr nach einem Stromausfall*. Hier kann durch eine geeignete Programmierung der Steuerung ein gewünschter Zustand gezielt eingenommen werden und auch ein komplexer geordneter Ablauf des Einschaltvorganges realisiert werden.
- *Vorteile der Rechnersteuerung nutzen*. Allgemein lässt sich sagen, dass die Nutzung der Rechnersteuerung Vorteile bringt insofern, als dass durch den Einbau geeigneter Sensoren trotz Aufwandes für die galvanische Trennung Fehler erkannt und geeignete Maßnahmen ergriffen werden können. Die Kosten hierfür stehen meist in keinem Verhältnis zu den in der Anlage auftretenden Schäden. Außerdem können die weiter unten dargestellten unterschiedlichen Schalterarten (mehrpolig, Umschalter, Stufenschalter) durch geeignete Ansteuerung mehrerer Halbleiterschalter beliebig nachgebildet werden.
- Zu beachten ist die Ausfallwahrscheinlichkeit von elektronischen Bauteilen. Diese steigt zwar mit steigender Anzahl von Kalt-Warm-Zyklen der Halbleiter, liegt aber erfahrungsgemäß über der von mechanischen Schaltern.

Nachteile von elektronischen Schaltern, eingesetzt in der Leistungselektronik, sind im Wesentlichen:

- *EIN-Widerstand*. Bei einem Leistungshalbleiter einige Ω; durch Parallelschaltung von z. B. bis zu 1000 Transistoren in einem Gehäuse: 100 µΩ bis einige mΩ. Die daraus resultierende Temperaturerhöhung des Halbleiters bei höheren Strömen muss durch Kühlung begrenzt werden.
- *AUS-Widerstand*. Durch die Parallelschaltung von bis zu 1000 Transistoren im Bereich kΩ bis einige MΩ. Die daraus resultierenden Sperrströme und die Temperaturerhöhungen der Halbleiter sind in der Leistungselektronik i. Allg. vernachlässigbar.
- *Galvanische Trennung*. Nur mit Zusatzschaltungen erreichbar. Hierbei sind 2 Fälle zu unterscheiden: Fall 1: Galvanische Trennung *Steuerkreis* ↔ *gesteuerter Kreis* (Lastkreis). Die Ansteuerung des Halbleiterschalters erfolgt häufig durch einen Rechner oder ein im Prinzip ähnliches Gerät. Dann muss der Lastkreis, sofern dort Spannungen oberhalb 24 V auftreten (genauer Wert siehe VDE-Vorschriften), vom Steuergerät galvanisch getrennt werden. Das ist möglich durch den Einsatz von Optokopplern oder Lichtleitern, erfordert aber einen erhöhten Aufwand (siehe VDE-Vorschriften).
  Fall 2: Im Kreis mit Spannungen oberhalb 24 V (genauer Wert siehe VDE-Vorschriften) muss ein Teil des Kreises von einem anderen Kreis elektrisch und damit galvanisch getrennt werden (z. B. Reparatur- oder Service-Arbeiten). Die galvanische Trennung kann durch Öffnen des *elektronischen* Schalters i. Allg. *nicht* erfolgen. Hier müssen die Kreise zwangsweise und vollständig getrennt werden (mechanischer Schalter, Sicherungen, Netzstecker ziehen, Anschlusskabel abklemmen – je nachdem, was geeignet ist; evtl. zusätzliche Zwangserdung des getrennten Kreises, siehe VDE-Vorschriften).
- *Empfindlichkeit gegenüber Störgrößen in Form von Spannungs- und Stromspitzen*. Prinzipiell sind Halbleiterschalter empfindlich gegenüber diesen Störgrößen, auch wenn sie nur kurzzeitig (ms- oder µs-Bereich) auftreten. Der negative Einfluss von Störgrößen kann dadurch verringert werden, dass für die Halbleiter ein geeigneter höherer Sicherheitsabstand zwischen den vom Hersteller angegebenen zulässigen Grenzwerten und den Betriebsgrößen eingeplant wird. Das ist oft nicht möglich, weil Grenzwert und Betriebswert nur geringfügig voneinander abweichen (es gibt keine geeigneten Halbleiter auf dem Markt, es soll aber ein Halbleiterelement wegen bestimmter Vorteile eingesetzt werden). Ev. kann man durch eine in Reihe zur Last geschaltete Induktivität Stromspitzen in ihrer Amplitude begrenzen. Ein parallel zur Last geschalteter Kondensator kann Spannungsspitzen begrenzen.

Beide Maßnahmen sind kombinierbar, sie bilden ein Tiefpasssystem. Siehe auch Abschn. 91.5 „Schutzschaltungen“.

- *Maximal schaltbare Leistung*. Die maximal schaltbaren Ströme liegen zurzeit bei etwa 3 bis 3,5 kA (Schalter EIN, Hochstrom-Typ, Serienfertigung), die Spannungen bei zurzeit maximal etwa 2 bis 3,5 kV (Schalter AUS, Hochspannungstyp, Serienfertigung). Damit liegen die maximal schaltbaren Leistungen bei etwa (50 ... 100) kW, da nicht beide genannten Maximalwerte von *einem* Halbleiter *gleichzeitig* erreicht werden. Es lassen sich Halbleiter parallel schalten zur Erhöhung des (Last-)Stromes und in Reihe schalten zur Erhöhung der (Sperr-)Spannung, siehe Zusammenschaltung bzw. Parallelschaltung bei den einzelnen Halbleitertypen in Kap. 90.

Hinweise und Bezeichnungen zu Schaltern:

1. Arbeitskontakt, Schließer, ohne Ansteuersignal ist der Schalter offen. NO (engl.) → *N*ormally *O*pen.
2. Ruhekontakt, Öffner, ohne Ansteuersignal ist der Schalter geschlossen. NC (engl.) → *N*ormally *C*losed.
3. Einpolig, zweipolig, usw. → ein, zwei oder mehr Kontakte schließen bzw. öffnen „gleichzeitig“, „galvanisch voneinander getrennt“ (Idealfall, Praxis siehe Datenblatt). SP, DP (engl. *S*ingle *P*ole, *D*ouble *P*ole usw.). Anwendung: Zweipolig, wenn Phase und Neutralleiter gleichzeitig ein- bzw. ausgeschaltet werden sollen. In der Energietechnik i. Allg. durch 2 Schalter nach Punkt 1 bzw. 2 nachgebildet.
4. Umschalter → Ein Kontakt schließt, ein weiterer öffnet „ gleichzeitig“. Sie haben i. d. R. *einen* gemeinsamen Anschluss (intern verbunden). Für den Umschaltvorgang gibt es zwei Schaltarten: 1. „ unterbrechend“, „nicht kurzschließend“, d. h., dass der öffnende Kontakt erst sicher geöffnet haben muss, bevor der andere Kontakt schließt; Beispiel: Umschalten zwischen zwei Spannungsquellen. 2. „nicht unterbrechend“, „kurzschließend“, d. h., dass der erste Kontakt erst öffnen darf, wenn der andere Kontakt schon sicher geschlossen hat; Beispiel: stromdurchflossene Spulen im Kreis. In der Energietechnik wird der Umschalter i. Allg. durch einen Schließer und einen Öffner mit geeigneter Ansteuerung nachgebildet.
5. Stufenschalter → *ein* Anschluss kann mit *einem von mehreren anderen* Anschlüssen verbunden werden. ST (engl.) *S*ingle *T*hrow (einstufig, steht im englischen auch für einen einpoligen Ein-Aus-Schalter), DT (engl.) *D*ouble *T*hrow (zweistufig) usw. Diese Schalterart wird in der Energietechnik durch die oben genannten Schalterarten nach Punkt 1 und 2 mit geeigneter Ansteuerung nachgebildet.
6. Manchmal werden für elektronische Schalter auch die Bezeichnungen „Relais“ oder „elektronische Relais“ oder „Halbleiter-Relais“ verwendet. Ob eine Lösung für die „Galvanische Trennung“ bereits in den Baustein integriert ist, zeigt das Datenblatt.
7. Werden elektrische Kreise mit Spulen *ausgeschaltet*, muss der Strom durch die Induktivitäten bis zum endgültigen Abbau des magnetischen Feldes weiterfließen können, um Überspannungen zu vermeiden. Folgende Bauteile sind dabei hilfreich: Widerstände, Dioden, Z-Dioden, VDR-Widerstände (spannungsabhängige Widerstände), Funkenstrecken, Glimmstrecken.
8. Werden elektrische Kreise mit Kondensatoren *eingeschaltet*, muss der Aufladestrom begrenzt werden. Das kann durch einen niederohmigen Reihenwiderstand erreicht werden, der die Funktionsfähigkeit der Schaltung nur unwesentlich beeinflusst. Der Widerstand kann nach erfolgter Aufladung evtl. auch überbrückt werden.

91.2 Halbleiter-Relais

Diese Baugruppe erfüllt die Funktionsweise eines Relais, indem eine Steuerspannung Leistungstransistoren, Thyristoren oder Triacs in den leitenden oder sperrenden Zustand schaltet. Zwischen der Steuerspannung (meist max. $|\pm 24|$ V; manchmal auch 230 V) und dem Lastkreis (meist ≥ 230 V) muss aus Sicherheitsgründen i. d. R. eine galvanische Trennung hergestellt werden. Für die Art der zu schaltenden Spannung ist entscheidend, welche elektronischen Schaltelemente im Relais verwendet worden sind:

1. *Ein Leistungs-Feldeffekttransistor*
   Bipolartransistoren sind wenig gebräuchlich. Das Relais hat eine ±-Kennzeichnung, und das Potential am „+“-Anschluss muss stets das positivere gegenüber dem anderen Anschluss sein. Eine häufig schon intern integrierte, zum Transistor parallel geschaltete, Diode schützt den Transistor vor Schäden durch Verpolung. Feldeffekttransistoren wirken im Leitzustand wie ein nahezu konstanter ohmscher Widerstand, der als „Kontaktwiderstand“ einfach kalkulierbar ist. Der Wert liegt i. Allg. bei (1 ... 5) Ω (Kleinleistungsbereich, ein Transistor) ; bei der internen Parallelschaltung mehrerer Transistoren (Leistungselektronik) auch im unteren mΩ-Bereich.
2. *Zwei Leistungs-Feldeffekttransistoren in Anti-Parallelschaltung*
   Die zu schaltende Spannung kann beliebige Polarität aufweisen. In Reihe zu jedem Transistor liegt eine Diode, um die zwei Polaritäten dem entsprechenden Transistor zuzuordnen. Diese Konfiguration kommt dem mechanischen Relais am Nächsten. Sie wird deshalb u. a. auch im Dual-In-Line-Gehäuse, 6-polig, mit Optokoppler zur galvanischen Trennung, für eine zu schaltende Spannung von maximal 230 V bei Strömen bis maximal 1 A gefertigt.
3. *Ein Thyristor*
   Der Thyristor zündet erst dann sicher, wenn die zu schaltende Spannung und die Lastimpedanz zusammen einen Strom ergeben, der den Einraststrom überschreitet. Beim Unterschreiten des Haltestromes löscht der Thyristor. Er muss, je nach Kurvenform der geschalteten Spannung, evtl. neu gezündet werden. Das Ausschalten muss über eine Löschschaltung für den Thyristor geschehen, was den Einsatz dieser Ausführung wesentlich einschränkt. Am Thyristor und damit zwischen den Schalterkontakten steht eine Spannung im Bereich von einigen Volt, die nur gering vom Strom abhängt.
4. *Zwei Thyristoren in Anti-Parallelschaltung*
   Es hängt vom Modell des Thyristors ab, ob die Sperrfähigkeit in Rückwärtsrichtung den Einsatz von 2 Dioden erspart.

Die zu schaltende Spannung kann beide Polaritäten besitzen, ansonsten gelten die bei *einem* Thyristor dargestellten Eigenschaften. Diese Konfiguration ist besonders geeignet zur Verwendung im 50-Hz-Stromversorgungsnetz bei niedrigen bis mittleren Leistungen, weil die Löschung automatisch beim Strom-Nulldurchgang erfolgt.

5. *Ein Triac*
   Er besitzt im Wesentlichen die Eigenschaften der Zwei-Thyristor-Konfiguration. Diese Variante wird häufig im 50-Hz-Stromversorgungsnetz bei kleinen bis mittleren Leistungen eingesetzt. Ein Optokoppler zur galvanischen Trennung ist i. d. R. schon integriert. Das Gehäuse kann auf einem Kühlkörper montiert werden. Die Ansteuerung ist einfach, siehe Abb. 90.36, Dimmerschaltung. Die Löschung erfolgt in jedem Nulldurchgang des Netzstromes mit anschließender erneuter Triggerung. Damit sind die Ausschalt- und Einschalt-Zeitpunkte vorgegeben als Vielfache von 10 ms.

Hinweis: Es kann sein, dass der Hersteller im Datenblatt darauf hinweist, dass er das betrachtete Halbleiter-Relais *nicht* für den Einsatz in der Luftfahrt, im medizinischen Bereich oder anderen kritischen Anwendungsfällen empfiehlt, in Bereichen also, in denen Zuverlässigkeit und Ausfallsicherheit eine besondere Rolle spielen.

Es folgen 3 Beispiele für Halbleiter-Relais aus der Serienfertigung (z. T. Näherungswerte, Datenblätter enthalten die verbindlichen Daten).

1. 400 V bei 1000 mA (PVX6012, Infineon, IR); 20 V bei 4 A (Wechselstrom) bzw. 6 A (Gleichstrom) (PVN012As, Infineon, IR). Die folgenden Werte gelten für beide Typen: Einschaltzeit 7 µs; Ausschaltzeit 1 µs; Widerstand im Einschaltzustand einige Ω bzw. 15 mΩ. Für die galvanische Trennung zwischen Steuer- und Lastkreis gilt: Isolationsspannung 4000 V, Widerstand 10^{12} Ω. Kapazität 1 pF. Diese Relais haben intern Feldeffekttransistoren eingebaut.
2. 48 bis 660 V Netzwechselspannung 50 Hz, 100 A, (SSRL660DC100, Omega), 1,6 V max. Spannungsabfall, Ein-Schaltzeit 20 ms, Aus-Schaltzeit 30 ms, Isolationsspannung 4000 V. Dieses Relais schaltet mit Thyristoren.
3. Für den Kleinleistungsbereich sind Bausteine entwickelt worden, die Ströme bis etwa 1 A bei Spannungen bis 230 V schalten können und einen integrierten Optokoppler zur Galvanischen Trennung haben. Sie sind in ein 6poliges Dual-In-Line-Gehäuse eingebaut. Es gibt sie mit einem FET für Gleichspannung und 2 FETs für Wechselspannung.

91.3 Halbleiterkühlung

91.3.1 Grundlagen, statisches Verhalten

Die Funktionsweise von Halbleiterbauelementen hängt entscheidend von PN-Übergängen ab, die in Sperrrichtung betrieben werden. Durch die während des Betriebes an ihnen auftretende Verlustleistung erwärmen sich diese Sperrschichten (engl. *J*unction). Überschreitet die *Sperrschichttemperatur* ϑ_J die *maximal zulässige Sperrschichttemperatur* $\vartheta_{J\max}$, verliert

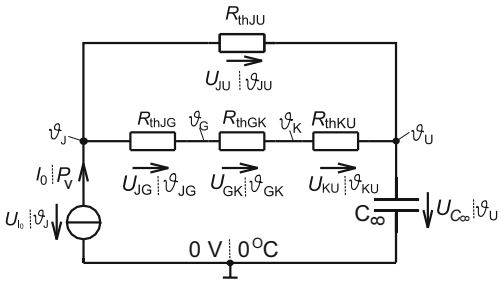


Abb. 91.1 Kombiniertes Ersatzschaltbild zur Halbleiterkühlung: Thermisch – Elektrisch; statisches Verhalten

das Halbleiterbauelement seine Funktionsfähigkeit. Durch geeignete Maßnahmen zur Wärmeabfuhr (Kühlung über Kühlkörper) kann erreicht werden, dass ϑ_J den Wert $\vartheta_{J\max}$ nicht erreicht. Die Grundlage der Berechnung von Kühlkörpern bilden die Gesetze des Wärmeströmungsfeldes. Vergleicht man die bestehenden Analogien zwischen dem Wärmeströmungsfeld und dem elektrischen Strömungsfeld, können die statischen und dynamischen Vorgänge eines Wärmeströmungsfeldes wie die eines elektrischen Strömungsfeldes behandelt werden. Die Analogien zeigt Tab. 91.1. Daraus ergibt sich das elektrisch-thermische Ersatzschaltbild nach Abb. 91.1. Hilfreich für die Analogie sind die Einheiten; ein Vergleich Wärmekapazität mit elektrischer Kapazität ergibt: $\mathrm{W\,s/K} \leftrightarrow \mathrm{A\,s/V}$. Damit lassen sich Temperatur und elektrische Spannung sowie Leistung und elektrischer Strom einander zuordnen. Die thermische Größe „Wärmestrom“ wird also behandelt wie die elektrische Größe „Leistung“ (die im Halbleiter entstehende Verlustleistung) für eine Stromquelle. Weiterhin sind Wärmewiderstand in K/W und elektrischer Widerstand in V/A einander zugeordnet. Ergebnis: Man berechnet ein elektrisches Netzwerk und ersetzt zum Schluss bei den Zahlenwerten die elektrischen Einheiten durch die zugeordneten thermischen Einheiten! Die in Abb. 91.1 eingezeichneten Wärmewiderstände haben folgende Bedeutung:

- *Wärmewiderstand* $R_{\mathrm{th\,JU}}$ zwischen *Halbleiterkristall* und *Umgebung*: Das ist der vorhandene Wärmewiderstand für den Fall, dass das Gehäuse des Leistungshalbleiters ohne irgendwelche zusätzlichen Maßnahmen zur Kühlung mechanisch montiert ist („frei in der Luft hängt“) und nur die daran vorbeistreichende Luft für Kühlung sorgt. Er wird in der Leistungselektronik nicht weiter berücksichtigt, da er häufig größer als 50 K/W ist und parallel zu der um Größenordnungen kleineren Widerstandssumme $R_{\mathrm{th\,JG}} + R_{\mathrm{th\,GK}} + R_{\mathrm{th\,KU}}$ liegt, siehe Abb. 91.1. Würde der Halbleiter nur auf diese Weise gekühlt, sorgte eine Verlustleistung von 3 W bei 50 K/W bereits für eine Kristalltemperatur von (3 W) · (50 K/W) = 150 °C *über* der Umgebungstemperatur!
- *Wärmewiderstand* $R_{\mathrm{th\,JG}}$ zwischen *Kristall* und *Gehäuse*: Er ist durch den internen Aufbau des Halbleiters fest vorgegeben und erreicht Werte kleiner 0,05 K/W.
  Er lässt sich dadurch klein halten, dass man dem Halbleiter ein relativ großes metallenes Teil-Gehäuse (Bodenplatte oder Kreisscheibe, einseitig oder oben und unten) mit geschliffener Oberfläche gibt und diejenige Kristallschicht, an der die

Tab. 91.1 Analogien zwischen thermischen und elektrischen Größen

Physikalische Größe → Analogie zur Elektrotechnik	FZ	EZ	EZ Sonderfall	Physikalische Größe	FZ	EZ
Temperatur → im Schaltbild eine Spannung an einer Stromquelle, einem Kondensator, einem Widerstand	ϑ	K	°C	Spannung	U	V
Temperatur-Differenz → im Schaltbild ein Spannungsabfall	$\Delta\vartheta$	K oder °C	°C	Spannungsabfall	ΔU	V
Wärmestrom → im Schaltbild eine Stromquelle mit dem Strom „(Verlust-)Leistung“	P	W	W	„Stromquelle“ mit dem Strom „Leistung, Verlustleistung“	P	W = V A
Wärmewiderstand → im Schaltbild ein Widerstand	R_{th}	K/W	°C/W	Widerstand	R	Ω = V/A
Wärmekapazität → im Schaltbild ein Kondensator	C_{th}	W s/K	W s/°C	Kapazität	C	F = A s/V
Eine *Wärmeinduktivität* gibt es nicht						

FZ: Formelzeichen, EZ: Einheitenzeichen, EZ Sonderfall: Bei den Berechnungen hier handelt es sich um Temperatur*differenzen*, und wegen $\Delta\vartheta = 1\,\text{K} \mathrel{\hat{=}} \Delta\vartheta = 1\,°\text{C}$ darf hier Kelvin durch Grad Celsius ersetzt werden.

höchste Verlustleistung entsteht, mit dem Gehäuseteil großflächig verbindet. Diese Kristall-Metall-Verbindung besitzt gute Wärmeleit-Eigenschaften und damit einen kleinen Wärmewiderstand $R_{\text{th JG}}$. Allerdings nimmt damit das Gehäuse das Potenzial des Halbleiterkristalles an.

- Wärmewiderstand $R_{\text{th GK}}$ zwischen *Gehäuse* und *Kühlkörper*: Da das Halbleitergehäuse das Potenzial des Kristalles hat, nimmt auch der Kühlkörper dieses Potential an, wenn der Halbleiter direkt auf dem Kühlkörper montiert wird. Wenn das nicht von Bedeutung ist, wird so vorgegangen: Bei der Montage kann sich zwischen Halbleitergehäuse und Kühlkörper z. B. durch Staub, Fremdkörper oder Unebenheiten eine dünne Luftschicht bilden, die einen vergleichsweise hohen Wärmewiderstand hat (Luft ist ein schlechter Wärmeleiter). Deshalb wird bei Leistungshalbleitern eine sog. Wärmeleitpaste zwischen Halbleitergehäuse und Kühlkörper aufgetragen, die selber einen sehr kleinen Wärmewiderstand hat. Sie soll die Luft verdrängen. Deshalb gilt: Die Flächen von Halbleiter und Kühlkörper sollen an der gemeinsamen Berührungsfläche eben sein, die Flächen müssen gereinigt werden, die Wärmeleitpaste ist dünn und flächendeckend aufzutragen, die Befestigungsschrauben sind fest anzuziehen. Für diesen Wärmewiderstand kann man nur einen Erfahrungswert angeben, der auch von der verwendeten Wärmeleitpaste und der Schichtdicke abhängt. Er kann, bei sorgfältiger Montage, im Bereich (0,01 … 0,1) K/W angenommen werden. Soll der Kühlkörper nicht das Potenzial des Halbleitergehäuses annehmen, werden dünne Isolierscheiben aus Glimmer oder Kunststoff oder aber spezielle Wärmeleitpasten, die gleichzeitig elektrisch isolieren (Kombi-Wärmeleit-Isolier-Pasten), zwischen Halbleiter-Gehäuse und Kühlkörper angebracht. Auch hier sollte Wärmeleitpaste verwendet werden, wenn nicht die Isolierpaste zum Einsatz kommt. Die Werte für den Wärmewiderstand der Isoliermaterialien kann man dem Datenblatt entnehmen, wobei stets ein Zuschlag für den Auftrag der Wärmeleitpaste erfolgen muss. Für die Isolierpaste gibt das Datenblatt ausführliche Montagehinweise.
- *Wärmewiderstand* $R_{\text{th KU}}$ zwischen *Kühlkörper* und *Umgebung*: Er ist durch den verwendeten Kühlkörper gegeben und hängt im Wesentlichen von folgenden Faktoren ab:
  1. der Größe des Kühlkörpers: Der Wärmewiderstand nimmt mit wachsenden Abmessungen ab, oberhalb einer gewissen Größe nur noch geringfügig.
  2. der Lage des Kühlkörpers im Raum: Es soll soviel Luft wie möglich ohne Wirbelbildung am Kühlkörper entlangströmen können.
  3. der Form des Kühlkörpers: Hier haben sich z. B. Profile mit rippenförmigem Aufbau (Abb. 91.2) bewährt, durch den die Oberfläche vergrößert wird und die Wärmeabgabe durch Konvektion und Strahlung begünstigt. Die Industrie bietet vielfältige Profile und Formen an, weil evtl. nur ein vorgegebenes Platzangebot besteht. Auskunft geben die Datenblätter der Hersteller.
  4. der Oberflächenfarbe des Kühlkörpers: Kühlkörper mit geschwärzter Oberfläche haben einen etwas geringeren Wärmewiderstand als solche mit metallisch blanker Oberfläche.
  5. der Art der Kühlung: Selbstkühlung, Fremdkühlung oder Flüssigkeitskühlung, siehe weiter unten.

Die Forderung nach ausreichender Kühlung von Halbleiterbauelementen führt in der Regel auf die Berechnung des Wärmewiderstandes $R_{\text{th KU}}$ des Kühlkörpers. Dazu geht man von folgenden gegebenen Größen oder Erfahrungswerten aus (Abb. 91.1):

1. der *Umgebungstemperatur* ϑ_U, von der man annimmt, dass sie sich durch den Kühlvorgang nicht ändert und deren Maximalwert man entweder kennt oder geeignet annimmt. So wird häufig für die *maximal auftretende* oder *zu erwartende Umgebungstemperatur* $\vartheta_{U\,\text{max}}$ ein Wert im Bereich (30 … 70) °C eingesetzt. Elektrotechnisch entspricht das dem „unendlich großen“ Kondensator C_∞ in Abb. 91.1, der seine Spannung $U_{C\infty}$ trotz dauernder Aufladung mit I_0 nicht ändert.
2. der im Halbleiter erzeugten Verlustleistung P. Sie wird entweder z. B. nach den im Kap. 90 angegebenen Gleichungen (90.5) bis (90.8) berechnet oder abgeschätzt.
3. dem Wärmewiderstand $R_{\text{th JG}}$ laut Datenblatt.
4. der maximal zulässigen Sperrschichttemperatur $\vartheta_{J\,\text{max}}$ laut Datenblatt.
5. einem der Praxis entnommenen Erfahrungswert für $R_{\text{th GK}}$ von z. B. 0,05 K/W (ohne elektrische Isolierung) oder 0,1 K/W (mit elektrischer Isolierung). Nützliche Hinweise zur Festlegung des Wertes bieten die Datenblätter der Hersteller der beteiligten Komponenten.

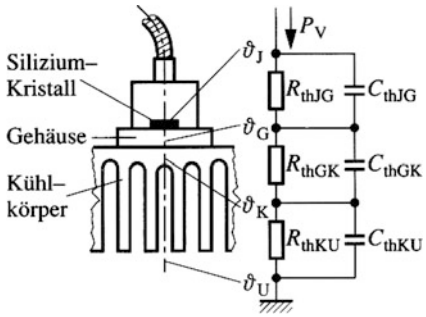


Abb. 91.2 Leistungshalbleiter (Schraubzelle) mit Kühlkörper und thermischem Ersatzschaltbild

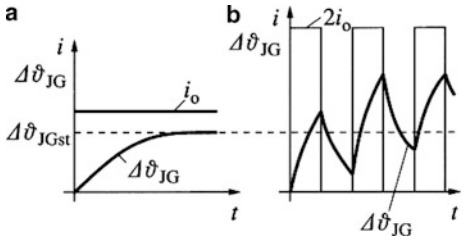


Abb. 91.3 Verlauf der Temperatur zwischen Sperrschicht und Gehäuse eines Leistungshalbleiters bei Gleichstrom (**a**) und bei Impulsstrom (**b**)

Ein Spannungsumlauf im Netzwerk nach Abb. 91.1 liefert (elektrische Bezeichnungen links von den gestrichelten Linien):

$$I_0 \cdot (R_{\text{th JG}} + R_{\text{th GK}} + R_{\text{th KU}}) + U_{C\infty} \approx U_{I0} \qquad (91.1a)$$

Dabei wurde folgende Näherung verwendet: $(R_{\text{th JG}} + R_{\text{th GK}} + R_{\text{th KU}}) \ll R_{\text{th JU}}$, und daraus folgt $[(R_{\text{th JG}} + R_{\text{th GK}} + R_{\text{th KU}}) \parallel (R_{\text{th JU}})] \approx (R_{\text{th JG}} + R_{\text{th GK}} + R_{\text{th KU}})$

Für den entsprechenden „thermischen Umlauf" folgt (thermische Bezeichnungen rechts von den gestrichelten Linien):

$$P_v \cdot (R_{\text{th JG}} + R_{\text{th GK}} + R_{\text{th KU}}) + \vartheta_U \approx \vartheta_J \qquad (91.1b)$$

(Näherung wie unter (91.1a)).

Daraus lässt sich $R_{\text{th KU}}$ berechnen und ein geeigneter Kühlkörper (Form, Profil, Mindestabmessungen) auswählen:

$$R_{\text{th KU}} = \frac{1}{P_v} \cdot (\vartheta_J - \vartheta_U) - (R_{\text{th JG}} + R_{\text{th GK}}) \qquad (91.2)$$

Zweckmäßigerweise setzt man in dieser Formel die „ungünstigsten" Werte ein und erhält den größten möglichen Wärmewiderstand des Kühlkörpers, der aus Sicherheitsgründen um 10 bis 20 % unterschritten werden sollte:

$$R_{\text{th KU(max)}} = \frac{1}{P_{v(\text{max})}} \cdot (\vartheta_J - \vartheta_{U(\text{max})}) - (R_{\text{th JG}} + R_{\text{th GK}}) \qquad (91.3)$$

Andererseits kann ein vorhandener Kühlkörper auf seine Eignung überprüft werden, wenn man die vorhandenen Werte in (91.1b) einsetzt und damit die sich ergebende Kristalltemperatur berechnet. Auch hier ist ein Sicherheitszuschlag bei der Umgebungstemperatur sinnvoll.

In Abb. 91.2 sind die in Abb. 91.1 eingezeichneten Wärmewiderstände und Temperaturen am Schema eines auf einem Kühlkörper montierten Halbleiters gezeigt. Die zusätzlich skizzierten Kapazitäten werden im nächsten Kapitel beschrieben.

91.3.2 Dynamisches Verhalten

Bisher wurde angenommen, dass der Leistungshalbleiter mit Gleichspannung betrieben wird und thermisch der eingeschwungene Zustand vorliegt. Die in Abb. 91.2 *eingezeichneten* Wärmekapazitäten C_{th} sind dann von Interesse, wenn die Halbleiter mit periodischen Wechsel-Spannungen und -Strömen betrieben werden. Der zeitliche Verlauf der Temperaturen ϑ_J, ϑ_G und ϑ_K hängt dann von den Zeitkonstanten $\tau_{\text{JG}} = R_{\text{th JG}} \cdot C_{\text{th JG}}$, $\tau_{\text{GK}} = R_{\text{th GK}} \cdot C_{\text{th GK}}$ und $\tau_{\text{KU}} = R_{\text{th KU}} \cdot C_{\text{th KU}}$ ab. Das thermische Verhalten im Halbleiter selbst ist durch die Zeitkonstante τ_{JG} gegeben. Abb. 91.3a zeigt den Verlauf der Temperaturänderung $\Delta\vartheta_{\text{JG}}$ an der Sperrschicht bei Belastung durch einen Gleichstrom ($i(t) = i_0$). Setzt zum Zeitpunkt $t = 0$ der Stromfluss ein, steigt die Temperatur an der Sperrschicht verzögert an, da die Wärmekapazität zwischen Sperrschicht und Gehäuse zunächst einen Teil des Wärmestromes durch Speicherung abzieht (Stichworte elektrotechnisch: Kondensatoraufladung, Energiespeicher). Dieser Verlauf lässt sich *auch* erklären durch die Annahme eines zeitveränderlichen Wärmewiderstandes, den man als *transienten inneren Wärmewiderstand* $Z_{\text{th JGt}}$ bezeichnet. Sein Wert ist zu Beginn des Stromflusses gering; er geht innerhalb kurzer Zeit (etwa 1 ... 2 s) in einen größeren stationären Endwert über, siehe Kurvenverlauf DC in Abb. 91.4. In Abb. 91.3b wird der Leistungshalbleiter mit einem Impulsstrom der Größe $2 \cdot i_0$ und dem Tastgrad 1/2, also dem gleichen arithmetischen Mittelwert i_0 wie bei Gleichstrom nach Abb. 91.3a, betrieben. Der Kurvenverlauf von $\Delta\vartheta_{\text{JG}}$ zeigt, dass sein Maximalwert den Wert aus Abb. 91.3a pro Periode zeitweise wesentlich überschreitet. Die Größe des Scheitelwertes hängt ab vom Tastgrad und der Stromflussdauer. Die Kurvenschar im Abb. 91.4 stellt den transienten inneren Wärmewiderstand $Z_{\text{th JGt}}$ in Abhängigkeit vom Stromflusswinkel Θ dar: $\Theta = 0 \rightarrow$ kein Stromfluss; $\Theta = 180 \rightarrow$ Tastgrad 1/2;

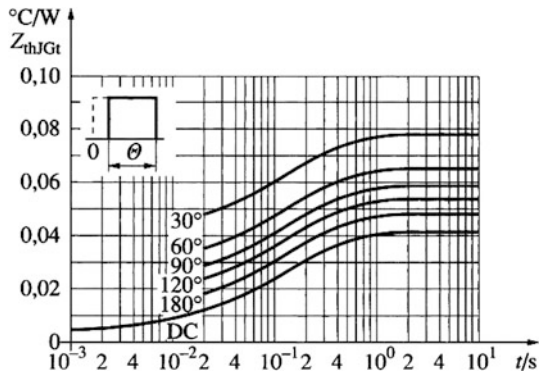


Abb. 91.4 Transienter innerer Wärmewiderstand $Z_{\text{th JGt}}$ als Funktion der Zeit und des Stromflusswinkels

Tab. 91.2 Typische Werte für Wärmewiderstände

Gehäuse-bauform	Wärmewiderstand in K/W			
	$R_{\text{th JG}}$	$R_{\text{th GU}}$		
		Selbstkühlung Luft	Fremdkühlung Luft (Gebläse)	Flüssigkeitskühlung
Schraubzellen	2,5 … 0,8	5,0 … 1,0	2,0 … 0,4	0,08
Scheibenzellen	0,1 … 0,03	0,5 … 0,2	0,4 … 0,08	0,02

$\Theta = 360 \rightarrow$ Gleichstrom. Das Diagramm ist dann praktisch, wenn der Stromflusswinkel einen konstanten Wert hat (weniger häufig) oder in gegebenen Grenzen variabel ist. Dann wählt man den größten Stromflusswinkel zur Festlegung des Wärmewiderstandes. Aus Abb. 91.4 lässt sich auch ablesen, welchen Einfluss die Periodendauer der Stromimpulse auf $Z_{\text{th JGt}}$ hat: bis 20 ms gilt der Ersatzwert für 20 ms (entspricht 50 Hz oder mehr, Anwendung 50-Hz-Energieversorgungsnetz), ab etwa 1,5 s der stationäre Endwert.

In gleicher Weise wird ein *zeitabhängiger transienter äußerer Wärmewiderstand* $Z_{\text{th GUt}}$ definiert, der sich durch den verwendeten Kühlkörper ergibt. Sein Verlauf entspricht dem aus Abb. 91.4, der stationäre Endwert wird aber erst nach (10 … 1000) s erreicht. Häufig ist ein Diagramm nach Abb. 91.4 auch im Datenblatt für den Kühlkörper abgebildet, sodass für den aktuellen Anwendungsfall der Wärmewiderstand bestimmt und in (91.1b) eingesetzt werden kann.

Tab. 91.2 zeigt typische Werte für Wärmewiderstände. Dass $R_{\text{th GU}}$ angegeben ist, liegt daran, dass Hersteller von Leistungshalbleitern vorgefertigte Kühlmodule liefern können, die sie für einen minimalen Wärmewiderstand optimiert haben und für die sie dann den resultierenden Wärmewiderstand $R_{\text{th GU}}$ angeben ($R_{\text{th GU}} = R_{\text{th GK}} + R_{\text{th KU}}$).

- *Selbstkühlung*: Diese Art der Kühlung liegt vor, wenn der (die) Leistungshalbleiter auf einem Kühlkörper montiert ist (sind) und die Wärme darüber an die Umgebung abgeführt wird.
- *Fremdkühlung*: Hier sorgt ein zusätzliches Element am Kühlkörper, meist in Form eines Gebläses, für einen erhöhten Luftstrom und damit für eine vergrößerte Wärmeabfuhr, was sich rein formal in der Verringerung des Widerstandes $R_{\text{th KU}}$ niederschlägt. Wie effektiv ein Gebläse ist, hängt wesentlich ab von seiner Drehzahl, seiner Flügelform, seinem Rotor-Durchmesser und seiner Ausrichtung zum Kühlkörper. Zur Einschätzung der Wirksamkeit ist Erfahrung notwendig, oder es sind Messungen am fertigen Objekt erforderlich. Die Hersteller geben häufig die pro Stunde geförderte Luftmenge des Gebläses an, ein Wert, der sich nur schwerlich in eine Abnahme des Kühlkörper-Wärmewiderstandes umrechnen lässt. Hier haben sich temperaturabhängige Widerstände, montiert direkt am Halbleitergehäuse oder nahe neben dem Halbleiter auf dem Kühlkörper, bewährt.
- *Flüssigkeitskühlung*: Die entstehende Wärme wird von einer vorbeiströmenden Flüssigkeit (z. B. Wasser) aufgenommen und anschließend an die Umgebung abgegeben. Hierbei wird ausgenutzt, dass die vorbeiströmende Flüssigkeit eine hohe Wärmekapazität und einen kleinen Wärmewiderstand Halbleitergehäuse-Flüssigkeit bzw. Kühlkörper-Flüssigkeit hat. Die erwärmte Flüssigkeit kann eine hohe Fließgeschwindigkeit haben, und sie wird zu entfernten Orten geleitet, wo sie ausreichend schnell abgekühlt werden kann. Auch die *Siedekühlung* wird angewendet, bei der ein wesentlicher Anteil der Wärmeabfuhr durch das Verdampfen einer Flüssigkeit entsteht. Der Dampf kondensiert an einer weiter vom Halbleiter entfernten kühleren und für das Kondensieren gut geeigneten Stelle, und die Flüssigkeit wird dem Kreislauf erneut zugeführt (*Wärmerohrkühler*, „*heat pipes*“). Es gibt einige geeignete Flüssigkeiten mit Siedepunkten zwischen etwa 40 und 100 °C. Die Hersteller geben einen Wärmewiderstand $R_{\text{th KU}}$ oder $R_{\text{th GU}}$ an, der das gesamte Kühlsystem umfasst und eventuelle Vormontagen berücksichtigt.

91.4 Ansteuerschaltungen

91.4.1 Ansteuerschaltungen allgemein

Unter Ansteuerschaltungen werden hier elektronische Schaltungen verstanden, die am Ausgang ein Signal zur Verfügung stellen, mit denen Leistungshalbleiter direkt am Gate bzw. an der Basis angesteuert werden können.

Für die Ansteuerschaltungen ist die folgende Unterteilung nützlich:

1. Die eigentliche Ansteuerschaltung für den Einsatz bei Frequenzen < 1 kHz.
2. Die eigentliche Ansteuerschaltung für den Einsatz bei Frequenzen im KHz- und unteren MHz-Bereich.
3. Die elektrische und mechanische Integration der Ansteuerschaltung in den Leistungsteil.

Für alle 3 Punkte gilt generell: Die Ansteuerung sollte mit speziell entwickelten optimierten Integrierten Schaltungen erfolgen. Diese berücksichtigen, dass Spannungen und Ströme vorgegebene Verläufe in Abhängigkeit von der Zeit besitzen und vorhandene Grenzwerte nicht über- bzw. unterschreiten. Und sie bieten auch, je nach Modell, nützliche Zusatzfunktionen.

Zu Punkt 1: Die Ansteuerung kann i. Allg. mit Integrierten Schaltungen ohne zusätzliche weitere Optimierungsmaßnahmen erfolgen.

Zu Punkt 2: Die Ansteuerung muss evtl. noch optimiert werden. Es sei an die Ansteuerschaltung in Abb. 90.19 erinnert, bei der eine Möglichkeit zur Ansteuersignal-Optimierung gezeigt ist. Allgemein sollten die von den Halbleiterherstellern herausgegebenen Applikationsberichte unbedingt zu Rate gezogen werden.

Zu Punkt 3: Die Ansteuerschaltungen benötigen eine Energieversorgung, in der Regel eine galvanische Trennung zur Befehlseingabe und eine elektrische Verbindung zwischen Ansteuer- und Lastkreis.

Die gesamte Ansteuerung setzt sich i. d. R. aus vier Baugruppen zusammen:

1. *Externes Steuergerät oder Befehlseingabe*, z. B. Rechner, Ablaufsteuerung, Bedienelemente, Notauslöser. Die ausgegebenen Signale liegen häufig als Digitalsignal vor mit den

zwei Logikpegeln 5 V und 0 V. Die Frequenz liegt bei mehr als 10 MHz und ist damit für die Anwendungen in der Leistungselektronik ausreichend.

2. *Galvanische Trennung*. Ein Potentialunterschied von bis zu 400 V ermöglicht die Verwendung von Optokopplern. Sie werden als Integrierter Baustein in einem 6-poligen Dual-In-Line-Gehäuse gefertigt und enthalten intern eine Leuchtdiode (LED) und einen Fototransistor oder eine Fotodiode. Es gibt sie auch mit VDE-Zulassung. Für Spannungen > 400 V bieten sich Lichtleiter an mit der Impulseingabe über eine Laserdiode und der Impulsabnahme über eine Fotodiode. Der eigentliche Lichtleiter kann der geforderten Potentialtrennung durch seine Länge angepasst werden. Optokoppler und Lichtleiter benötigen auf der Empfängerseite meist noch eine Impulsaufbereitung.
3. *Stromversorgung für den Ansteuerbaustein*. Die Stromversorgung muss i. d. R. eine Spannung im Bereich (15 ... 30) V und einen Strom von (1 ... 10) A zur Verfügung stellen. Das kann über einen Netztransformator, ausgeführt als Trenntransformator, aus dem 230-V-Wechselstromnetz erfolgen. Je höher die Potentialdifferenz ist, desto aufwändiger werden die Transformatoren. Man kann die Versorgung aber auch aus demjenigen Netz entnehmen, das die Last speist. Die Spannung muss dann nur auf den erforderlichen Wert (15 ... 30) V herabgesetzt werden. Die galvanische Trennung ist an dieser Stelle nicht erforderlich. Woher die Stromversorgung auch bezogen wird, es muss sichergestellt sein, dass sie für den Ansteuerbaustein gepuffert ist, damit dieser bei einem wo auch immer auftretenden Fehler (Speisekreis, Lastkreis, ...) noch die Möglichkeit hat, dringend erforderliche Schaltfunktionen durchzuführen.
4. *Eigentliche Ansteuerschaltung und ihr elektrischer Anschluss an den Leistungshalbleiter*. Hier sollen die Probleme beim Anschluss der Ansteuerschaltung an den Leistungshalbleiter dargestellt werden. Folgende Näherungswerte werden den Berechnungen zu Grunde gelegt: 10 cm Cu-Verbindungsleitung haben einen Widerstand von etwa 2,2 mΩ (1 mm Durchmesser) bzw. etwa 0,09 mΩ (5 mm Durchmesser) und eine Induktivität von etwa 0,1 µH (Wert soll hier durchmesserunabhängig sein). Der sog. Masseanschluss des Ansteuerbausteines muss *direkt* an den Anschluss Emitter bzw. Source bzw. Kathode des Leistungshalbleiters angeschlossen werden. Schließt man ihn z. B. 10 cm vom Anschluss Emitter bzw. Source bzw. Kathode des Leistungshalbleiters an die Zuleitung zum Halbleiter an, addiert sich der Spannungsabfall auf der Zuleitung zum Ansteuersignal des Leistungshalbleiters. Allerdings liegt der Wert wegen des großen Leiterquerschnittes im unteren mV-Bereich. Kritischer sind die 10 cm Leitungslänge beim Schaltvorgang. Werden z. B. 100 A innerhalb 500 ns geschaltet, ergibt sich eine Spannung von $u = L \cdot \frac{di}{dt} \approx 0{,}1\mu H \frac{100\,\mathrm{A}}{500\,\mathrm{ns}} \approx 20\,\mathrm{V}$. Diese Spannung überlagert sich der Ansteuerspannung des Leistungshalbleiters während des Umschaltvorganges und beeinflusst ihn wesentlich. Dazu kommen noch die Auswirkungen des Miller-Effektes, siehe Abschn. 90.3 „Feldeffekttransistoren“.

Die elektrischen Verbindungsleitungen zwischen dem Ausgang des Ansteuerbausteines und den Anschlüssen Gate und Source (Basis und Emitter bzw. Gate und Kathode) erfordern besondere Aufmerksamkeit. Der Einsatz eines Integrierten Bausteines ist von Vorteil, weil die angesprochenen Leitungen sehr kurz gehalten werden können, auf ganzer Länge verdrillbar sind und der kompakte und symmetrische Aufbau die elektrische Einstreuung von Störgrößen minimiert. Ihre Amplituden steigen mit steigenden Amplituden von geschalteter Spannung und geschaltetem Strom, siehe dazu auch Teil X „Messtechnik“, Kap. 82 „Messdatenaufbereitung“, Abschn. 82.1 „Verringerung der Störeinflüsse von außen“. Somit gelangt die optimale Signalform der Ansteuerschaltung nahezu unverändert auf den Steuereingang des Leistungshalbleiters.

Die Ansteuersignale sind keine Digitalsignale, sie haben vielmehr Analogcharakter. Optokoppler bzw. Lichtleiter sind für die Übertragung von Analogsignalen nicht optimal ausgelegt bzw. nicht geeignet, ihre Übertragungsgeschwindigkeit für Digitalsignale ist aber allemal ausreichend. Insofern sollte die galvanische Trennung so verwirklicht werden, dass entweder der Bezugspunkt der Ansteuerschaltung auf dem Potential vom Emitter bzw. von der Kathode bzw. vom Source-Anschluss des Leistungshalbleiters liegt und die Potentialtrennung vor den Eingang der Ansteuerschaltung platziert wird, oder bei niedrigen Potentialunterschieden von maximal 400 V ein Netztransformator als Trenntransformator die Netzspannung von 230 V auf (15 ... 30) V für die Energieversorgung der Ansteuerschaltung heruntertransformiert.

91.4.2 Ansteuersignale für Bipolartransistoren

Für einen Bipolartransistor im gesperrten Zustand sind der Basisstrom, die Basis-Emitter-Spannung und damit die bereitzustellende Leistung Null. Soll er dagegen leiten, ist ein Strom I_B in die Basis einzuspeisen, dessen Größe sich zu

$$I_B = m \cdot \frac{I_{C\,\mathrm{max}}}{B_{\mathrm{min}}} = I_{B(\mathrm{on})} \tag{91.4}$$

ergibt (Erläuterungen siehe Text).

I_{cmax} ist der maximal mögliche Kollektorstrom, der bei minimaler Lastimpedanz und bei maximaler Spannung (Scheitelwert der den Lastkreis speisenden Spannung) auftritt. Für die *Großsignal-Stromverstärkung* $B(h_{21})$ ist der kleinste mögliche Wert, B_{min}, einzusetzen, da B sehr großen Exemplarstreuungen unterliegt und auch vom Kollektorstrom und der Temperatur abhängt. Zusätzlich führt man noch einen *Sicherheitsfaktor m* ein, der im Bereich 1,2 bis 2 liegt. Der so bestimmte Basisstrom sollte nicht unterschritten werden, da andernfalls die Kollektor-Emitter-Restspannung $U_{\mathrm{CE\,sat}}$ u. U. nicht ihren Minimalwert erreicht und als Folge davon die Temperatur im Transistor auf unzulässige Werte ansteigt. Diese Art der Ansteuerung hat allerdings den Nachteil, dass in der Regel ein zu großer Basisstrom fließt, was wiederum die Ausschaltzeit verlängert und dadurch eventuell die Betriebsweise beeinträchtigt.

Die genannten Probleme lassen sich durch den Einsatz spezieller integrierter Schaltkreise lösen, die für bestimmte Anwendungsfälle optimiert sind. Ihre *prinzipielle* Wirkungsweise soll

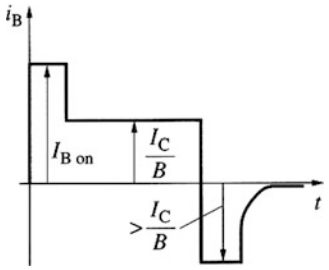


Abb. 91.5 Verlauf eines optimierten Basisstromes für einen Leistungstransistor

anhand der Abb. 91.5 und 91.6 erläutert werden. Der Basisstrom i_B von Leistungstransistoren hat den in Abb. 91.5 dargestellten Verlauf. Zum Übergang vom sperrenden in den leitenden Zustand wird zunächst ein Strom $i_B \geq i_{B(on)}$ nach (91.4) eingespeist, um eine kurze Einschaltzeit zu erhalten. Anschließend wird der Basisstrom auf den minimal erforderlichen aktuellen Wert $(I_c)/(B)$ verringert, damit keine überflüssigen Ladungen in der Basis gespeichert werden. Das setzt eine Überwachung der Kollektor-Emitter-Restspannung $U_{CE\,sat}$ voraus, und das geschieht in der Funktionseinheit F1 in Abb. 91.6. Mit Beginn des Ausschaltvorganges wird ein Strom aus der Basis herausgeführt, der betragsmäßig größer ist als der im Leitzustand eingeprägte Strom $(I_c)/(B)$, um die gespeicherten Ladungen so schnell wie möglich auszuräumen. Steigt die Spannung zwischen Basis und Emitter in negativer Richtung merklich an, muss der Stromfluss unterbrochen werden, Funktionseinheit F2. Die über dem Widerstand R_M anstehende Spannung ist dem durch den Transistor fließenden Strom proportional. Sie wird in der Funktionseinheit F3 ausgewertet und führt beim Erreichen eines kritischen Wertes zum dauernden Sperren des Transistors. Über die Funktionseinheit F4 können Grenzwerte für die Einschalt- bzw. Ausschalt-Dauer vorgegeben werden. F5 erfasst die Temperatur des Transistorkühlkörpers. Die Steuereinheit F6 wertet die eingehenden Informationen aus und stellt den Basisstrom i_B entsprechend ein. Über den Eingang E1 werden die Steuersignale zum Ein- bzw. Ausschalten des Transistors einer Anpassschaltung zur Pegelanpassung, F7, zugeführt. F8 schließlich erzeugt die Versorgungsspannung für die Steuerschaltung.

Es gibt Ansteuerschaltungen, die den Widerstand R_M nicht erfordern. Die Überwachung des Stromgrenzwertes erfolgt dann in der Funktionseinheit F1: Steigender Laststrom bedeutet eine zwar geringe aber auswertbare Zunahme der Kollektor-Emitter-Restspannung $U_{CE\,sat}$, sodass durch deren Überwachung ein vorgegebener Stromgrenzwert erkannt wird und der Transistor gesperrt werden kann.

Eine galvanische Trennung von Steuer- und Lastkreis geschieht hier sinnvollerweise dadurch, dass man die zur Ansteuerung an E1 anzulegenden Digitalsignale bereits über Optokoppler oder Lichtleiter zuführt und die eigentliche Steuerschaltung elektrisch mit dem Leistungstransistor verbindet.

91.4.3 Ansteuersignale für Feldeffekttransistoren und IGBT

Im Prinzip benötigen Feldeffekttransistoren keine Steuerleistung am Gate. Durch die Parallelschaltung von bis zu 1000 Einzeltransistoren zu einem „Gesamt“-Leistungstransistor ent-

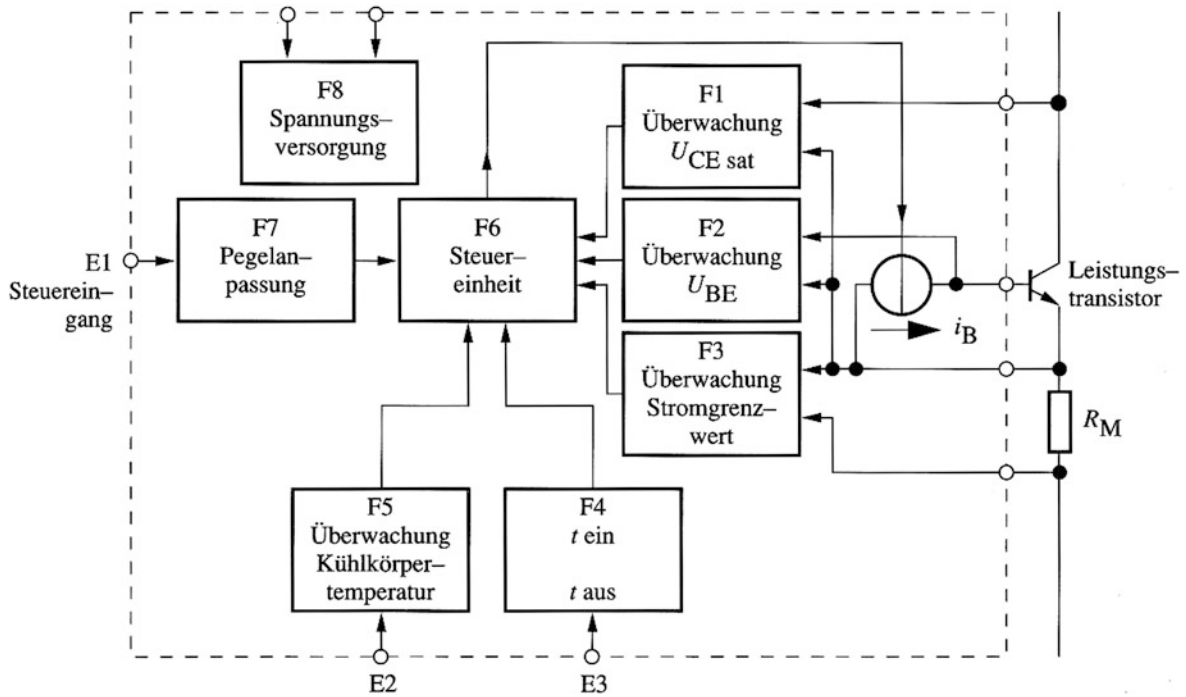


Abb. 91.6 Blockschaltplan einer Steuerschaltung zur Erzeugung des optimalen Basisstromes nach Abb. 91.5

steht aber zwischen Gate und Source eine Kapazität von bis zu 10 nF (häufiger Wert, auch > 50 nF möglich; Datenblatt!), die innerhalb einer Zeit von weniger als 1 µs von etwa 0 V auf (12 ... 15) V und umgekehrt umgeladen werden muss. Eine Überschlagsrechnung liefert: Ein Feldeffekttransistor hat 10 nF Gate-Source-Kapazität und soll in 100 ns von 12 auf 0 V umgeladen werden. Dann fließt ein Strom von $i = C \cdot \frac{du_C}{dt} = 10\,\text{nF} \cdot \frac{12\,\text{V}}{100\,\text{ns}} = 1{,}2\,\text{A}$. Diesen Strom muss die Ansteuerschaltung bei jedem Umschaltvorgang für etwa 100 ns liefern, und damit muss ihr Innenwiderstand $\ll 1\,\Omega$ sein. Auch hier sollte man für die Ansteuerung speziell entwickelte Schaltungen einsetzen. Mit derartigen Bausteinen können beispielsweise folgende Werte erreicht werden: Bei einer Lastkapazität (Gate-Source-Kapazität) von 10 nF steht ein Spitzenstrom für den Umladevorgang von 6 A zur Verfügung, die Anstiegs- und Abfallzeiten betragen maximal 80 ns, die Verzögerungszeit zwischen Eingangs- und Ausgangssignal liegt bei 40 ns. Legt man an den Eingang dieser Ansteuer-Bausteine 5 V, wird der Leistungstransistor in den Ein-Zustand gesteuert, bei 0 V schaltet er aus. Die Ansteuerung kann also durch digitale Logikschaltungen erfolgen. Die Bausteine selbst benötigen eine Versorgungsspannung im Bereich (15 ... 30) V. Sie sind sinnvollerweise (siehe Punkt 4.1 oben) mit dem Leistungstransistor elektrisch verbunden. Durch den Einsatz von Optokopplern kann eine galvanische Trennung zum Digital-Eingang des Ansteuerkreises erreicht werden. Ist die Betriebsfrequenz der Transistoren niedrig, z. B. 50 Hz, genügt es in der Regel, das Gate mit einem Rechteckimpuls anzusteuern, wenn man die Umladung der Gate-Source-Kapazität durch einen Treiberbaustein mit den oben genannten Eigenschaften (Innenwiderstand $\ll 1\,\Omega$, Umladestrom bis 2 A möglich) betreibt. Wird der Transistor dagegen bei Frequenzen > 1 kHz betrieben, muss das Gate-Ansteuersignal evtl. noch optimiert werden, siehe Abb. 90.19. Es kann sinnvoll sein, das Gatesignal zum Sperren, zumindest für den *Übergang* vom Leiten zum Sperren, nicht nur zu 0 V, sondern negativ zu wählen, soweit das von der Transistorphysik her zulässig ist (Anhaltswert: $U_{GS} > -5\,\text{V}$, genauer Wert siehe Datenblatt). Der Grund dafür ist, dass beim Ausschalten die Drain-Source-Spannung von nahezu 0 V auf die Betriebsspannung des geschalteten Elementes ansteigt und sich dieser positive Impuls über die Drain-Gate Kapazität, verstärkt mit dem Miller-Effekt, auf das Gate überträgt und damit dem Ausschaltvorgang entgegenwirkt, Auch aus diesem Grund ist der Innenwiderstand des Ansteuer-Bausteines $\ll 1\,\Omega$ zu realisieren.

91.4.4 Ansteuersignale für Thyristoren

Zur Zündung von Thyristoren reichen im Allgemeinen kurze Stromimpulse aus, die in das Gate eingespeist werden. Für ihren zeitlichen Verlauf und ihre Amplituden gelten folgende Anforderungen:

1. Nach Abb. 90.27 müssen der Gatestrom bzw. die Gatespannung so groß gewählt werden, dass man sich im Bereich III befindet, d. h. im Bereich sicherer Zündung.
2. Für die Mindestdauer des Zündimpulses gibt der Hersteller entsprechende Werte an; sie liegen im Bereich (10 ... 500) µs. Bei induktiver Last muss diese Mindestdauer u.U. soweit verlängert werden, bis der dynamische Haltestrom (Einraststrom) sicher erreicht ist.
3. Der Anstieg des Gatestromes pro Zeiteinheit soll so groß wie möglich sein, um ein schnelles ganzflächiges Zünden der Sperrschicht zu erreichen. So schreiben die meisten Hersteller Werte $di_G/dt > 1\,\text{A/µs}$ vor.
4. Eine Beanspruchung der Gate-Kathoden-Strecke in Sperrrichtung von weniger als −5 V ist nicht zulässig.
5. Während des Betriebes in Rückwärtsrichtung dürfen Steuerimpulse am Gate, gleich welcher Polarität, nicht auftreten.

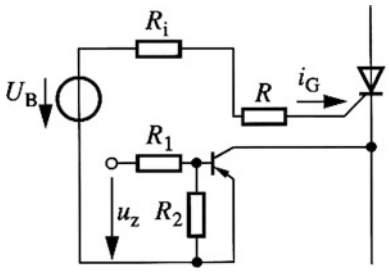


Abb. 91.7 Leistungsthyristor mit Transistor-Zündstufe

Eine häufig verwendete Zündstufe zeigt Abb. 91.7. Durch Anlegen einer Spannung u_z schaltet der Transistor durch, und in das Gate fließt ein Strom i_G von näherungsweise

$$i_G \approx \frac{U_B}{R_i + R} \tag{91.5}$$

R_i Innenwiderstand der speisenden Quelle.

Durch Reihenschaltung von R wird ein geeigneter Wert für i_G eingestellt. Der Spannungsteiler aus R_1 und R_2 dient der Erhöhung der Störsicherheit. Je geringer der Wert von R_2 ist, desto weniger können Störimpulse zu unerwünschter Zündung des Thyristors führen. Der Spannungsteiler R_1, R_2 ist so zu bemessen, dass bei gegebenem u_z ein zum Durchschalten des Transistors ausreichender Basisstrom fließt. Eine galvanische Trennung zwischen Ansteuerkreis und Thyristor lässt sich durch einen Transformator erreichen (Abb. 91.8). Er arbeitet nahezu im sekundärseitigen Kurzschluss. Mit dem Widerstand R_1 wird der Gatestrom eingestellt. R_2 und V2 ermöglichen den Stromfluss nach Abschalten des Transistors (Transformator hat induktive Komponente; V2 als „Freilaufdiode“). V1 sorgt dafür, dass die Gatespannung nicht negativ wird. Der Transformator arbeitet als Impulsübertrager. Bei rechteckförmiger Basis-Emitter-Spannung u_z am Transistor ergibt sich auf der Sekundärseite des Transformators ein Gatestrom gemäß Abb. 91.9.

Die Steilheit des Stromanstieges wird durch die Zeitkonstante τ_e erfasst:

$$\tau_e = \frac{L_\sigma}{R_1 + R_i + R_W} \approx \frac{T_e}{2} \tag{91.6}$$

L_σ Streuinduktivität des Transformators, R_w Transformator-Wicklungswiderstand, R_i Innenwiderstand der Spannungsquelle

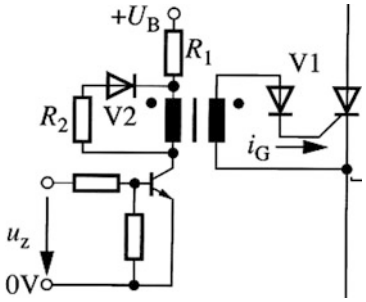


Abb. 91.8 Leistungsthyristor mit Transistor-Zündstufe und Impulsübertrager zur galvanischen Trennung

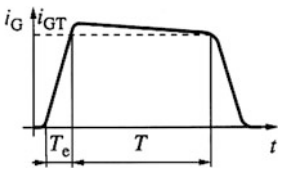


Abb. 91.9 Zeitlicher Verlauf des Gatestromes bei Ansteuerung durch eine Zündstufe nach Abb. 91.8

U_B, sekundärseitige Größen auf die Primärseite umgerechnet, T_e Anstiegszeit.

Da die untere Grenzfrequenz eines Transformators nicht Null ist, sinkt der Strom auf der Sekundärseite mit einer Zeitkonstanten τ_a gemäß

$$\tau_a = \frac{L_H}{R_{\text{ges}}} \tag{91.7}$$

Darin ist R_{ges} die Parallelschaltung von $(R_1 + R_i + R_w)$ mit dem Widerstand der Gate-Kathoden-Strecke im leitenden Zustand. L_H ist die Hauptinduktivität des Transformators. Der Transformator hat das Übersetzungsverhältnis 1 : 1.

Um den Mindeststrom i_{GT} für die gesamte Dauer T des Zündimpulses bereitzustellen, ist zu Beginn des Zündvorganges ein höherer Wert einzuspeisen.

Übersteigt der Potentialunterschied zwischen Primär-und Sekundärwicklung etwa 600 V, steigt durch den erforderlichen größeren Abstand zwischen beiden Wicklungen die Streuinduktivität L_σ. Damit sinkt die Steilheit des Stromanstieges auf Werte $< 1\,\text{A}/\mu\text{s}$. Man wählt deshalb ein Konzept, wie es in Abb. 91.10

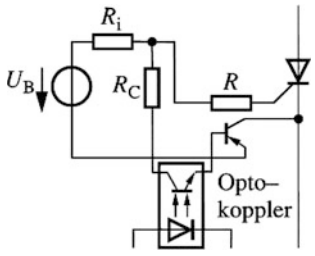


Abb. 91.10 Leistungsthyristor mit Transistor-Zündstufe und Optokoppler zur galvanischen Trennung

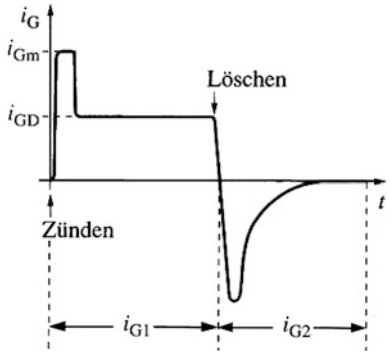


Abb. 91.11 Verlauf des Gatestromes beim GTO-Thyristor, Zünd- und Löschvorgang

dargestellt ist. Die Spannungsquelle mit der Spannung U_B muss potenzialfrei bereitgestellt werden. Das kann beispielsweise durch einen Transformator geschehen, dessen Sekundärspannung gleichgerichtet und stabilisiert wird. Neben ausreichender Spannungsfestigkeit werden keine weiteren Forderungen an ihn gestellt. Die galvanische Trennung vom Zündsignal wird durch den Optokoppler erreicht. Diese Schaltung hat damit den Vorteil, dass Gateströme einstellbarer Amplitude (über R bei gegebenem R_i) und beliebiger Dauer zur Verfügung stehen.

91.4.5 Ansteuersignale für GTO-Thyristoren

Gegenüber einem Steuerteil für Thyristoren muss dasjenige für GTO-Thyristoren einige Besonderheiten aufweisen:

1. Der Zündstrom ist prinzipiell höher.
2. Zum Löschen ist ein negativer Stromimpuls in das Gate einzuspeisen. Sein Amplitudenwert liegt bei 1/3 bis 1/5 der Amplitude des fließenden Anodenstromes.
3. Um ein vorzeitiges Löschen bei geringen Lastströmen zu verhindern, ist es häufig notwendig, einen Dauerstrom in das Gate einzuspeisen.

Aus Punkt 3 folgt, dass die Verwendung eines Impulsübertragers nicht ohne weiteres möglich ist. Der typische Gatestromverlauf besteht aus i_{G1} und i_{G2} und hat einen Verlauf wie in Abb. 91.11 dargestellt. Er wird im Steuerteil durch zwei Spannungsquellen erzeugt, Abb. 91.12.

Zum Zünden wird zunächst ein Gatestrom i_{Gm} mit ausreichender Amplitude und Dauer erzeugt, anschließend wird er auf einen Dauerstrom i_{GD} verringert. Zum Löschen speist man einen negativen Gatestrom ein. Dabei geht die Gate-Kathoden-Strecke bereits in den Sperrzustand über, bevor der Thyristor vollständig sperrt. Zu diesem Zeitpunkt zieht die Induktivität L durch ihre Induktionsspannung weiterhin Strom aus dem Gate, wobei die Gate-Kathoden-Strecke für einige Mikrosekunden durchbricht. Durch diese Stromflussverlängerung wird der GTO-Thyristor sicher gesperrt.

Abb. 91.12 zeigt auch, wie die Potentialtrennung zwischen dem Steuerteil und der Ansteuerung des GTO-Thyristors erreicht

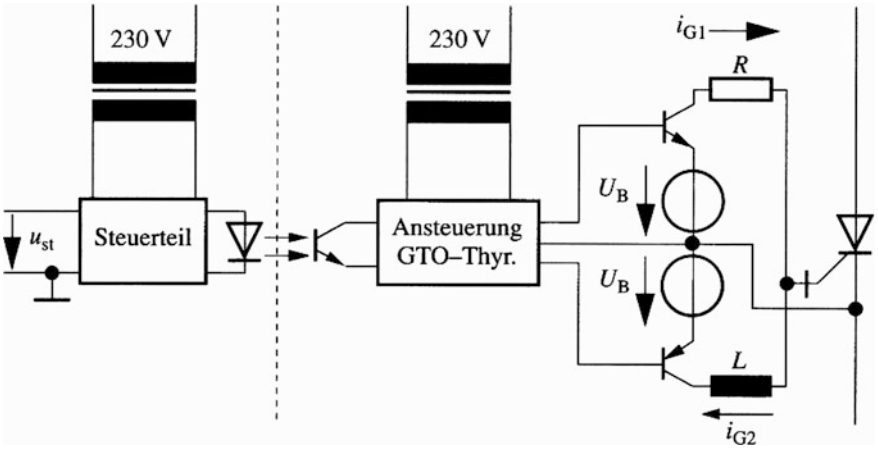


Abb. 91.12 Blockschaltplan einer Ansteuerschaltung zur Erzeugung des Stromverlaufes nach Abb. 91.11 mit galvanischer Trennung zwischen Steuer- und Leistungsteil

wird. Die eigentlichen Steuersignale werden über Optokoppler übertragen, zwei Transformatoren sorgen für eine potenzialfreie Bereitstellung der Versorgungsspannung beider Funktionseinheiten.

91.5 Schutzschaltungen

Der Zweck von Überspannungs- und Überstrom-Schutz ist, Spannungen und Ströme oberhalb eines Grenzwertes auf einen zulässigen Wert zu reduzieren oder zu Null zu erzwingen. Der Grund ist, dass die zu hohen Werte Geräte, Verbraucher, elektronische Schaltungen und dgl. beschädigen (könnten). Die – theoretische – Idee, durch entsprechende Widerstände und Reihen- bzw. Parallelschaltung zu den gefährdeten Teilen die zu hohen Werte zu reduzieren oder umzuleiten, scheitert an zwei Dingen. 1. Die Überspannungen und Überströme treten oft innerhalb von ns bis µs auf, und die zu schützenden Elemente können innerhalb dieser Zeit bereits zerstört werden, und 2. die „Vernichtung" der zu hohen Spannungen bzw. Ströme erfordert Absorber mit evtl. sehr großer Energieaufnahme. Daraus folgt das Grundprinzip der Schutzschaltungen: Erkennen des kritischen Zustandes innerhalb von ns, (nahezu) „unverzögerter" Energieabbau mit kleinen preiswerten zuverlässigen Elementen ohne deren Zerstörung solange, bis ein für solche Fälle optimiertes Organ mit konstruktionsbedingter höherer Reaktionszeit (Verzögerungszeit) eine Totalabschaltung vornimmt.

91.5.1 Überspannungsschutz

Die bisher betrachteten Überspannungen wurden durch den Einsatz der Leistungshalbleiter selbst, d. h. „im Innern", hervorgerufen (z. B. Trägerspeichereffekt, Reihenschaltung mehrerer Halbleiter). Daneben können auch von außen erzeugte Überspannungen über das speisende Wechsel- oder Gleichstrom-Netz auf die Halbleiter einwirken, die z. B. durch externe Schaltvorgänge, atmosphärische Störungen, elektrische Überschläge oder induktive Kopplungen entstehen. Um eine Beschädigung der Halbleiter durch diese äußeren Überspannungen zu verhindern, wird neben der Schutzschaltung gegen innere Überspannungen eine zweite gegen äußere Überspannungen eingesetzt. Im allgemeinen schützen die hier vorgestellten Maßnahmen den gesamten Schaltungsaufbau „hinter" der Schutzschaltung. Ein wirksamer Überspannungsschutz ist nach dem Prinzip einer zeitlichen Staffelung aufgebaut. Zunächst sprechen Schutzelemente an, die sehr schnell reagieren (Bereich einige ns), aber ein geringes Energieabsorptionsvermögen haben [3]. Sie werden vor ihrer Zerstörung durch Überhitzung durch Elemente mit verzögerter Ansprechzeit (Bereich unterer µs-Bereich) und mit erhöhter Energieabsorption abgelöst [6]. Bevor *diese* Elemente zerstört werden, trennt ein weiteres letztes Element mit der höchsten Ansprechzeit (ms- bis s-Bereich) den gefährdeten Kreis endgültig von der Überspannung ([9] und Text zu Abb. 91.15). Dieses Element schaltet sicher ab: Schmelzsicherung mit angepasster Auslösecharakteristik, im Kleinleistungsbereich evtl. auch Sicherungsautomaten (ermöglichen mehrfaches Abschalten). Ein Problem beim Überspannungsschutz ist, dass nur sehr wenig über Dauer und Form der Störgrößen bekannt ist. Die Erfahrung zeigt, dass Störungen häufig entweder sehr kurzzeitig auftreten (unterer µs-Bereich bis unterer ms-Bereich, z. B. durch Schaltvorgänge) oder längerfristig (ab einigen Millisekunden, z. B. Beschädigung der Last). Insofern hat sich die genannte Staffelung in der Praxis bewährt: Häufig auftretende, sehr kurzzeitige, Überspannungen werden durch Energievernichtung begrenzt, und die angeschlossene Schaltung (Last) bleibt weiter funktionsfähig. Tritt die Überspannung für längere Zeiträume auf (ab mittlerer ms-Bereich), liegt ein kritischer Störfall vor, der die Funktionsfähigkeit des angeschlossenen Systems nicht mehr sicherstellt und die Abschaltung eine zwingend notwendige Maßnahme ist. Für den Fall des endgültigen Abtrennens wird letztlich das Auslösen der vorgeschalteten Sicherung vorausgesetzt. Dazu ein wichtiger Hinweis: Beim Ersetzen der Sicherung ist unbedingt darauf zu achten, nur solche mit den vom Entwickler der Schutzeinrichtung vorgegebenen Werten (Auslösestrom, Auslösecharakteris-

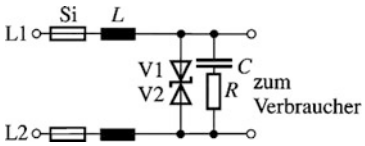


Abb. 91.13 Schutzbeschaltung gegen äußere Überspannungen, Teil 1

tik) einzusetzen, weil diese Werte für den Überspannungsschutz optimal ausgesucht worden sind! Abb. 91.13 zeigt ein Beispiel zum Überspannungsschutz. Die Z-Dioden V1 und V2 besitzen eine Durchbruchspannung, die nur geringfügig über dem Scheitelwert der Spannung zwischen L1 und L2 liegt. Der bei Überspannung auftretende zunehmende Stromfluss durch diese Dioden und die Last wird durch die Induktivität L zunächst soweit verlangsamt, dass entweder die Überspannung abgeklungen ist, bevor die Dioden zerstört werden, oder bei länger andauernder Überspannung die Sicherung Si rechtzeitig den Stromkreis unterbricht. Weiterhin verringert der Ladestrom des Kondensators C durch den Spannungsabfall an L die Amplitude einer Überspannung. R begrenzt den Kondensatorstrom. Es wird vorausgesetzt, dass der Spannungsabfall an L und deren Phasenverschiebung die nachfolgende Apparatur nicht wesentlich beeinflussen. Die Reaktionszeit (Verzögerungszeit) der Z-Dioden liegt im Bereich einiger µs, die sich durch den Einsatz von Lawinendioden verkürzen lässt. Das sind Z-Dioden mit sehr geringen Reaktionszeiten beim Übergang vom gesperrten in den leitenden Zustand. Die heute verwendeten CMOS-Bausteine der Nachrichten- und Datentechnik können bereits durch Überspannungsimpulse von der Dauer einiger ns zerstört werden. Einen Schutz bieten hier die speziell entwickelten Suppressordioden, auch mit *T*ransient *V*oltage *S*uppressor-Dioden (TVS-Dioden) oder *T*ransient *A*bsorption *Z*-Dioden (TAZ-Dioden) bezeichnet. Sie besitzen eine Verzögerungszeit von einigen ns, haben ein vergrößertes Energieabsorptionsvermögen und bei höheren Strömen eine steilere Kennlinie $i = f(u)$ als Z-Dioden. Die Durchbruchspannung beträgt maximal einige hundert Volt. Sie werden in uni- und in bipolarer Ausführung gefertigt. Es gibt spezielle Suppressordioden, die auch für die Energietechnik einsetzbar sind: Durchbruchspannung 6 bis 800 V; 600 W Impulsleistung bei einem 10/1000 µs-Impuls (Erklärung siehe weiter unten bei den Varistoren); Verzögerungszeit 1 ps. Weitere Einzelheiten siehe [3]. Die Suppressordioden haben die Z-Dioden als Überspannungsschutz abgelöst, das Schaltzeichen ist für beide gleich.

Eine evtl. mögliche Methode zur Überspannungsbegrenzung beim Einsatz von Thyristoren als Schalter besteht darin, diese, sofern sie sperren, beim Überschreiten eines kritischen Spannungswertes zwischen Anode und Kathode durch Z-Dioden oder Varistoren „notzuzünden“. Die Überspannung tritt damit an der Last auf, aber diese Variante darf folglich nur dann angewendet werden, wenn sie dort keinen Schaden verursacht. Die vorgeschaltete Sicherung sollte so schnell wie möglich auslösen. Da diese Zündung sehr schnell (Bereich einige Mikrosekunden oder weniger) erfolgen muss, verwendet man auch hier entsprechend schnelle Triggerschaltungen.

Abb. 91.14 zeigt weitere Möglichkeiten zur Überspannungsbegrenzung. Die hier gezeigten Maßnahmen lassen sich einzeln oder in Kombination einsetzen. Die Sicherung ist wiederum das Element, dass den Stromkreis unterbricht, bevor bei anstehender länger andauernder Überspannung die eingesetzten Überspannungselemente durch Überhitzung Schaden nehmen und die Überspannung nicht weiter abbauen können.

- Der Kondensator C_1 und die Spule L zusammen dienen dazu, Überspannungen und Oberschwingungen, die im Verbraucher erzeugt worden sind, zu verringern, bevor sie in das Netz gelangen.
- Die Spule L verlangsamt den Strom*anstieg* auf der Lastseite. Sofern der Spannungsabfall, die Phasenverschiebung und die Wärmeentwicklung nicht von Bedeutung sind, kann die Effektivität und die Funktionssicherheit nachfolgender Schutzorgane erhöht werden.
- Der Gasableiter GS besteht aus einem gasgefüllten Glasgefäß mit 2 Elektroden. Beim Überschreiten einer über das Gasgemisch vorwählbaren Spannung zünden die Gasableiter. Wird der Strom extern auf Werte $\ll 1$ A begrenzt, findet eine Glimmentladung statt (bekannt von der Glimmlampe) und die Spannung wird auf (50 ... 150) V begrenzt. Dieser Effekt ist hier wenig hilfreich. Schließt man dagegen den Gasableiter direkt an die zu überwachende Leitung an, entsteht bei einem Strom ab etwa 1 A eine Bogenentladung. Die Spannung, die sich einstellt, kann über das Gasgemisch im Bereich (80 ... 4500) V voreingestellt werden [4], [5]. Durch die u. U. hohe Wärmeentwicklung muss die Sicherung den Stromkreis rechtzeitig unterbrechen. Die Ansprechzeit beträgt (1 ... 2) µs, sie steigt mit steigender Spannungsanstiegszeit du/dt und sinkender Umgebungstemperatur. Die Gasableiter dienen als sog. Grobschutz. Nach einer Zündung müssen sie, abhängig vom durchgeflossenen Strom, evtl. ausgetauscht werden. Der für den Hochspannungssektor vorgesehene Gasableiter SA-xxx-STD hat z. B. folgende Daten: Zünd-Gleichspannung max. 7200 V; Ableitstrom 5 kA (durchschnittlich 5 Zyklen) bzw. 10 kA (1 Zyklus); Ansprechzeit 50 ns; Kapazität 1 pF; Isolationswiderstand $10^9\ \Omega$. Die Hochstromausführung 2063-xx-A weist folgende Werte auf: Zünd-Gleichspannung (230 ... 800) V; Ableitstrom 100 kA (1 Zyklus); 60 kA (5 Zyklen). Einzelheiten siehe Datenblätter [4], [5]).
- Der Kondensator C_2 und die Spule L verringern die Amplituden sehr kurzer Störspannungsimpulse (Tiefpass).
- Der Widerstand R_V ist ein Varistor (spannungsabhängiger Widerstand, VDR-Widerstand, *V*oltage *D*ependent *R*esistor). Bis zu einer durch die Mischung und den Aufbau voreinstellbaren Spannung ist der Widerstand sehr groß (100 kΩ oder größer), darüber hinaus hat er einen sehr kleinen Widerstand (einige Ω). Die Kennlinie verläuft gemäß $i \sim u^n$ mit $n = 2$ bis $n > 30$. Bei entsprechendem Aufbau sind sie in der Lage, auch energiereichere Überspannungen zu begrenzen. Allerdings können nur Überspannungen bis zu einigen µs abgefangen werden. Dazu einige Daten eines Varistors mit hoher Energieaufnahme [6], BB-Serie: Energieabsorption 10.000 J; Spannungsbereich 1100 bis 2800 V (AC), 1400 bis 3500 V (DC); Spitzenstrom 70 kA bei einem 8/20 µs-Impuls; 10.000 J bei einer Dauer von 2 ms. Hinweis: Der 8/20 µs-Impuls ist ein in IEC 61000-4-5 und IEC 61643-321 festgelegter „Norm-Test“-Impuls, mit dem

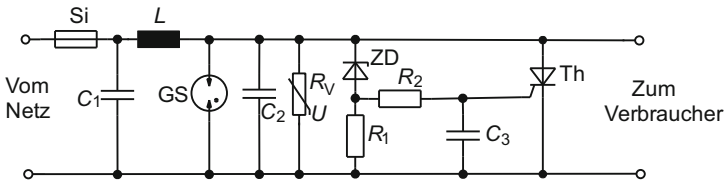


Abb. 91.14 Schutzbeschaltung gegen äußere Überspannungen, Teil 2

auch die Störspannungsfestigkeit von elektrischen Geräten geprüft werden kann. In einer ersten groben Näherung steigt seine Amplitude in 8 µs (10 µs) nahezu linear auf 100 %, und in einem abgerundeten Übergang fällt sie ähnlich einer *e*-Funktion so ab, dass sie nach 20 µs (1000 µs) seit Anstiegsbeginn auf 50 % abgefallen ist. Die Werte in Klammern gelten für den oben angeführten 10/1000 µs-Impuls. Je nach externem Aufbau und der Möglichkeit der Wärmeabgabe an die Umgebung kann er für einige µs sehr große Energien aufnehmen. Seine Ansprechzeit liegt im unteren ns-Bereich, er ist also sehr schnell. Damit kann er Überspannungen fast unverzögert solange abbauen, bis andere trägere Überspannungsschutzorgane mit der Möglichkeit größerer Energieübernahme oder Abschaltung ansprechen.

- Die Schaltung aus R_1, R_2, C_3, ZD und Th bildet eine sog. Crowbar-Schaltung (engl., Brecheisen). Sobald die zu überwachende Spannung die Spannung der Z-Diode überschreitet, beginnt sie zu leiten und speist Strom in das Gate des Thyristors, der daraufhin zündet. Damit wird die Spannung zu nahezu 0 V erzwungen („kurzgeschlossen"), eine sehr wirkungsvolle Art des Überspannungsschutzes. Der Thyristor kann auf einem Kühlkörper montiert werden und die Leistung solange aufnehmen, bis die Sicherung den Stromkreis endgültig unterbricht. Die Ansprechzeit liegt im Bereich (10 ... 100) µs, ist also relativ groß. Das liegt daran, dass der Strom durch die Z-Diode ebenfalls in dieser zeitlichen Größenordnung ansteigt. Damit wird der Thyristor nicht optimal schnell gezündet. Durch den Einsatz von Lawinendioden kann diese Zeit verkürzt werden. Es gibt deshalb speziell für diese Anwendung entwickelte Integrierte Schaltungen, sog. Watchdogs (engl., Wachhunde). Sie haben eine eigene Stromversorgung und sehr schnelle Komparatoren. Die Thyristoren werden, wenn bei kritischen Anwendungen gewünscht, dauernd mit einem Strom knapp unterhalb des minimalen Zündstromes betrieben. Im Schadensfall wird innerhalb von wenigen ns ein Strom von mehreren 100 mA bis zu einigen A für (10 ... 300) ms in das Gate eingespeist. Damit zündet der Thyristor sehr schnell und sicher. Dieses Schaltungsprinzip ist u. a. für Schaltnetzteile entwickelt worden, die aus einer höheren Spannung eine niedrigere erzeugen (Abwärtswandler), und bei denen ein Schadensfall mit großer Wahrscheinlichkeit so aussieht, dass die höhere Spannung auf die Niederspannungsseite „durchgeschaltet" wird, verursacht durch einen Kurzschluss im Leistungstransistor, der in Längsrichtung zwischen Ein- und Ausgang angeordnet ist.

91.5.2 Überstromschutz

Beim Überstromschutz sind zwei Überlastfälle zu unterscheiden:

- Fall1: ein Halbleiterstrom, der kurz- oder langzeitig den Nenn- bzw. Dauergrenzstrom (I_N bzw. I_{AVM}) um ein Vielfaches übersteigt, und
- Fall2: ein Halbleiterstrom, der den Nenn- bzw. Dauergrenzstrom (I_N bzw. I_{AVM}) nur geringfügig, aber für längere Zeit, überschreitet.

In der Hausinstallation sind die Leitungsschutzschalter (Sicherungsautomaten) auf diese zwei Überlastfälle ausgelegt: Ein elektromagnetischer Auslöser erfasst Fall 1, ein thermischer Auslöser Fall 2.

Zu hohe Ströme in Leistungshalbleitern entstehen u. a. durch Überlast oder Beschädigung des Verbrauchers sowie durch Fehler in den Zuleitungen (Kurzschluss). Weiterhin sind sie möglich durch Ansprechen von Schutzeinrichtungen („Notzündung" bei energiereichen Überspannungen oder Ansprechen einer oder mehrerer der in den Abb. 91.13 und 91.14 aufgeführten Schutzeinrichtungen). Da die Wärmekapazität von Leistungshalbleitern gering ist, müssen superflinke Sicherungen zur Überstromabschaltung bei Leistungshalbleitern verwendet werden. Maßgebend für die Auswahl einer geeigneten Sicherung ist das die Stromwärme erfassende Integral

$$\int_0^{t_A} (i_K(t))^2 dt \tag{91.8}$$

Erläuterungen siehe Text.

Darin ist i_K der die Sicherung durchfließende Strom und t_A die Zeit bis zum Auslösen der Sicherung. Der Wert, der sich nach (91.8) ergibt, muss stets kleiner sein als das Grenzlastintegral des Halbleiters (s. (90.4)). In Abb. 91.15 sind zwei Kennlinien eingetragen. Die mit H1 bezeichnete gibt an, wie lange ein bestimmtes Vielfaches des Nenn- bzw. Dauergrenzstromes (I_N bzw. I_{AVM}) den Halbleiter durchfließen darf, ohne ihn zu beschädigen, die andere, mit Si gekennzeichnet, nach welcher Zeit die Sicherung bei gegebenem Strom schmilzt (Schmelzsicherung). Der Halbleiterschutz ist um so wirkungsvoller, je größer die Differenz beider Kennlinien für steigende Vielfache des Stromes I_N bzw. I_{AVM} ist. Damit der Halbleiter im Kurzschlussfall

auf der Verbraucherseite nicht durch den großen Strom und die damit verbundene hohe Stromdichte zerstört wird, muss die in Abb. 91.14 eingezeichnete Induktivität L verwendet werden: Der Stromanstieg wird verlangsamt, sodass die kritische Stromdichte innerhalb des Halbleiters noch nicht erreicht ist, wenn die Sicherung aufgrund der Wärmeentwicklung nach (91.8) den Stromkreis unterbricht.

Auch PTC-Widerstände (Positiv Temperature Coeffizient, Kaltleiter) sind einsetzbar. Bis zu einer über die Materialmischung bestimmbaren Bezugstemperatur T_0 sinkt der Widerstand mit steigender Temperatur (bis dahin also NTC-Verhalten, Heißleiterverhalten), ab T_0 steigt der Widerstand exponentiell so an, dass bei $T_0 + 2\,°C$ der Widerstand auf das 10^3- bis 10^6-fache angestiegen ist. Die Widerstandswerte werden auch im unteren Ω-Bereich gefertigt und können dann mit Strömen oberhalb 1 A beaufschlagt werden. Bei höheren Strömen kann der Einsatz eines Stromteilers zu Erkennung des zu hohen Stromes eingesetzt werden. Das nutzbare Zeitfenster liegt im Millisekunden- bis unterem Sekunden-Bereich.

Bei den heute zunehmend verwendeten integrierten Bausteinen zur Ansteuerung von Leistungshalbleitern sind häufig zwei Zusatzschaltungen zur Erhöhung der Betriebssicherheit eingebaut, s. auch Abschn. 91.4.2. Die eine erfasst laufend die Temperatur des Kühlkörpers, auf dem der Halbleiter montiert ist, die zweite wertet den Strom durch den Halbleiter aus. Überschreitet einer der Werte einen vorzuwählenden Grenzwert, wird der Halbleiter in den nichtleitenden Zustand gesteuert. Die Kühlkörpertemperatur wird mit temperaturabhängigen Widerständen erfasst, der Strom durch Messen der Spannung an einem mit dem Leistungshalbleiter in Reihe liegenden Widerstand. Bei Transistoren ist auch die Erfassung eines unzulässigen Stromanstieges durch Messen der Kollektor-Emitter-Spannung möglich.

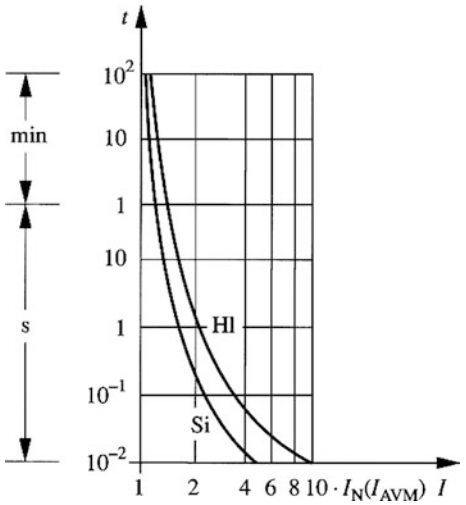


Abb. 91.15 Überstromschutz. *H1* Überstromkennlinie eines Leistungshalbleiters, *Si* Schmelzkennlinie der vorgeschalteten Sicherung

Die thermische Überwachungsschaltung bietet einen Schutz vor Zerstörung bei nicht zu großer Überlast ($I < (1{,}2 \ldots 2) \cdot I_N$), die über einen längeren Zeitraum (Sekunden bis Minuten) auftritt. Die Stromüberwachung dagegen kann bereits eine beliebige Überlast, die für mindestens $(10 \ldots 100)\,\mu s$ auftritt, erkennen. Dadurch können Transistoren relativ schnell abgeschaltet werden, bei Thyristoren dagegen ist das erst im nächsten Stromnulldurchgang möglich, sofern sie an Wechselspannung betrieben werden. Es bleibt zu überlegen, ob nicht eine Löschschaltung, wie unter Abschn. 90.5.6 beschrieben, vorgesehen werden sollte.

Nachrichtentechnik XIII



92 Grundlagen der Nachrichtenübertragung 1127
93 Zweitore, Vierpole 1147
94 Leitungen 1161
95 Antennen 1193
96 Modulation 1203
97 Filter 1229
98 Empfängerschaltungstechnik 1245
99 Ton- und Bildübertragung 1249
100 Mehrfachübertragung – Multiplexverfahren 1257
101 Richtfunktechnik 1261
102 Nachrichtenübertragung über Satellit 1265
103 Nachrichtenübertragung über Lichtwellenleiter (LWL) 1269
104 Funkmesstechnik 1277
105 Elektroakustische Wandler 1283
106 Vermittlungstechnik 1293
107 Kommunikations- und Datennetze 1299
108 Optimierte Nachrichten- und Datenübertragung 1309

92 Grundlagen der Nachrichtenübertragung

Wilfried Plaßmann



92.1 Prinzip der elektrischen Nachrichtenübertragung 1128
92.2 Aufgaben der Nachrichtentechnik 1129
92.3 Grundbegriffe 1129
92.4 Nachricht, Information und Signal 1130
92.5 Kenngrößen der Übertragungsstrecke 1143

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_92

92.1 Prinzip der elektrischen Nachrichtenübertragung

Die Nachrichtentechnik hat die Aufgabe, *Nachrichten* auszutauschen. Nachrichten sind z. B. Fragen, Beobachtungen und Befehle; der Begriff wird im Abschn. 92.3 weiter aufgeschlüsselt. Unter dem Begriff *Nachrichtenübertragung* wird hier die *elektrische* Nachrichtenübertragung verstanden, denn zur Übermittlung von Nachrichten werden Spannungen und Ströme sowie elektrische und magnetische Felder eingesetzt. Die Übertragung beginnt nach Abb. 92.1 mit der Nachrichtenquelle, die die Nachrichten aussendet, und endet mit der Nachrichtensenke, die sie empfängt. Zwischen beiden liegt das elektrische *Nachrichtenübertragungssystem*, dessen Funktionseinheiten dargestellt sind und anhand der Rundfunk-Sprachübertragung erläutert werden. Das Sprachsignal in Form von Luftdruckänderungen wird mit einem Mikrofon in proportionale Spannungsänderungen umgewandelt und verstärkt. Da viele unterschiedliche Nachrichten gleichzeitig übertragen werden (große Zahl von Rundfunksendern mit unterschiedlichen Programmen), muss der Frequenzbereich des Sprachsignals (15 Hz bis 15 kHz) in einen anderen Frequenzbereich umgesetzt werden. Das geschieht im Modulator. Anschließend stellt der Sender die für die Übertragung erforderliche Leistung (kW bis MW) bereit und sorgt über eine Antenne für die Aussendung in Form eines elektromagnetischen Feldes. Dieses wird von der Antenne des Empfängers aufgenommen. Da viele Nachrichten gleichzeitig, aber mit unterschiedlichen Frequenzbereichen, an der Empfängerantenne vorhanden sind, filtert ein selektiver Empfänger einen bestimmten Frequenzbereich und damit eine bestimmte Nachricht (den gewünschten Rundfunksender) aus. Der Demodulator setzt den empfangenen Frequenzbereich wieder in den Originalfrequenzbereich (15 Hz bis 15 kHz) um. Das Signal wird verstärkt und einem Lautsprecher zugeführt, der das elektrische Signal in ein Schalldrucksignal umwandelt, damit es vom Ohr (Empfänger) aufgenommen werden kann.

Bei der Nachrichtenübertragung ist es wünschenswert, die folgenden Bedingungen zu erfüllen:

1. Die Übertragung soll über sehr große Entfernungen unbeeinflusst von Hindernissen (Häusern, Bergen) möglich sein. Häufig wird zusätzlich ein lückenloser Empfang innerhalb eines vorgegebenen geografischen Gebietes gefordert (Rundfunk- und Fernsehempfang).
2. Es sollen viele unterschiedliche Nachrichten gleichzeitig übertragen werden, die der Empfänger aber wieder eindeutig voneinander trennen kann.
3. Der Empfänger kann entscheiden, ob er die Nachricht empfangen will oder nicht.
4. Die Übertragung von Daten und deren Weiterverarbeitung soll einfach möglich sein.

Die elektrische Nachrichtentechnik erfüllt diese Forderungen in nahezu idealer Weise: Die überbrückbaren Entfernungen reichen um den Erdball, bis zum Mars und darüber hinaus (Punkt 1). Weltweit senden einige Tausend Sender ihre Programme, ohne dass es zu nennenswerten gegenseitigen Beeinflussungen kommt (Punkt 2). Durch Ausschalten der Energieversorgung kann der Empfang von Nachrichten beendet werden. Im Vergleich dazu kann beispielsweise bei akustischer Nachrichtenübertragung durch Lautsprecher der Empfänger nicht so ohne Weiteres den Empfang beenden (Punkt 3). Die übertragenen Daten liegen bereits in einer für die Weiterverarbeitung optimalen Form vor (Punkt 4).

Die Übertragung geschieht entweder *drahtgebunden* (z. B. Fernsprechverkehr im Nahbereich, lokale Datennetze) oder *drahtlos* durch elektromagnetische Felder (z. B. Rundfunk- und Fernsehübertragung, Fernsprechverkehr im Fernbereich, Satellitenkommunikation). Für die drahtlose Übertragung werden Frequenzen im Bereich von etwa 10 kHz bis zurzeit 250 GHz benutzt. Tab. 92.1 gibt einen Überblick über die Nutzung dieses Frequenzbereiches durch die Rundfunk- und Fernsehsender, die bis etwa 1 GHz den weitaus größten Teil davon belegen. Daneben sind auch Frequenzzuteilungen für den Amateur- und den Poli-

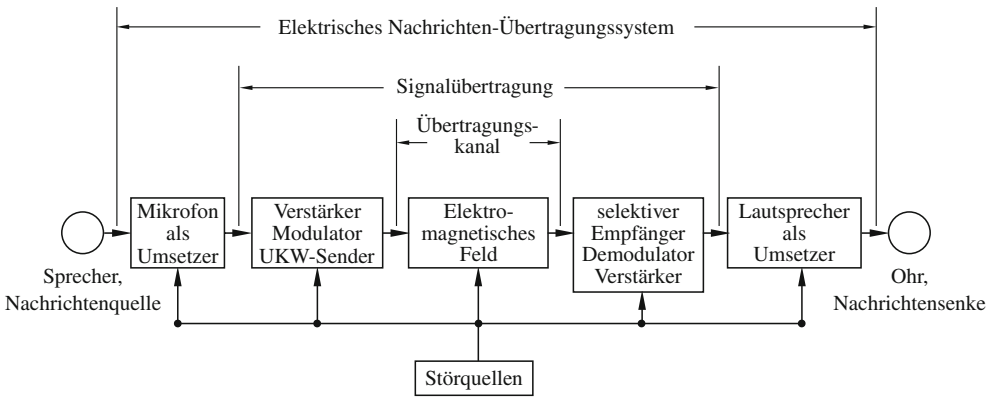


Abb. 92.1 Prinzip der Nachrichtenübertragung am Beispiel der Rundfunk-Tonübertragung

Tab. 92.1 Frequenz- und Wellenlängenbereiche und deren Nutzung durch einige wichtige Dienste

Bezeichnung	Frequenz		Wellenlänge	
	Definition	Verwendet	Definition	Verwendet
Myriameterwellen (VLF)	3 … 30 kHz		100 … 10 km	
Radiowellen (RF)	ab 30 kHz		≤ 10 km	
Kilometerwellen (LF, LW)	30 … 300 kHz		10 … 1 km	
Rundfunk auf „Langwelle“		150 … 285 kHz		2 … 1,05 km
Hektometerwellen (MF, MW)	300 … 3000 kHz		1000 … 100 m	
Rundfunk auf „Mittelwelle“		525 … 1605 kHz		571 … 187 m
Amateurfunk		1875 kHz		160 m
Dekameterwellen (HF, KW)	3 … 30 MHz		100 … 10 m	
u. a. Rundfunk auf „Kurzwelle“		6 … 19 MHz		49 … 16 m
Amateurfunk		3,5 … 30 MHz[a]		80 … 10 m[a]
Meterwellen (VHF)	30 … 300 MHz		10 … 1 m	
Fernsehen Band I		41 … 68 MHz		7,3 … 4,4 m
Polizei		86,5 MHz		3,5 m
UKW-Rundfunk		87 … 108 MHz		3,5 … 2,9 m
Amateurfunk		144 … 146 MHz		2,08 … 2,05 m
Fernsehen Band III		174 … 230 MHz		1,7 … 1,3 m
Hochfrequenzwellen	300 … 300 GHz		1 m … 1 mm	
Dezimeterwellen (UHF)	300 … 3 GHz		100 … 10 cm	
Fernsehen Band IV/V		470 … 960 MHz		64 … 31 cm
Amateurfunk		430 … 440 MHz		70 … 68 cm
Amateurfunk und Ortungsfunkdienst		1250 … 1300 MHz; 2300 … 2350 MHz		24 … 23 cm; 13 … 12,7 cm
Zentimeterwellen (SHF) oder Mikrowellen	3 … 30 GHz		10 … 1 cm	
Satellitenfunk		10,95 … 12,75 GHz		2,7 … 2,4 cm
Strahlung	ab 3000 GHz		≤ 100 mm	

[a] Für die Funkamateure stehen in diesem Bereich mehrere schmale Frequenzbereiche zur Verfügung.
Erläuterungen der Abkürzungen siehe Text.

zeifunk eingetragen. Insgesamt gibt es noch eine Vielzahl von weiteren Anwendern, denen Frequenzbereiche zugeteilt worden sind. Als Beispiel seien hier aufgeführt: Feste und bewegliche Seefunkdienste, Navigationsfunkdienste, Raumforschung und Weltraumfunkdienste, Normalfrequenz- und Zeitzeichenfunkdienste, Wetterfunkdienste. Weiterhin gibt es Frequenzen, die speziellen Verwendungen vorbehalten sind, z. B. Grubenalarmfunk (68,010–68,030 MHz), Funkstellen in Rettungsfahrzeugen (243,000 MHz), Weltraum-Notruf-Frequenz (20,007 MHz), Seefunk-Not- und Anruffrequenz (2,182 MHz).

Die in Tab. 92.1 verwendeten Abkürzungen haben folgende Bedeutung (nach DIN 40015):

VLF *V*ery *L*ow *F*requencies
RF *R*adio *F*requencies
LF *L*ow *F*requencies
MF *M*edium *F*requencies
HF *H*igh *F*requencies
VHF *V*ery *H*igh *F*requencies
UHF *U*ltra *H*igh *F*requencies
SHF *S*uper *H*igh *F*requencies
EHF *E*xtremely *H*igh *F*requencies
LW Langwelle
MW Mittelwelle
KW Kurzwelle
UKW Ultrakurzwelle
UV Ultraviolettstrahlung

92.2 Aufgaben der Nachrichtentechnik

Ein wesentliches Problem bei der Nachrichtenübertragung liegt darin, dass die Nachrichten durch Störungen verfälscht werden. Deshalb entwickelt die Nachrichtentechnik Verfahren, mit denen diese Verfälschungen so klein wie möglich gehalten werden können. Für die nicht vermeidbaren Verfälschungen wird gefordert, dass sie erkennbar sind und sich korrigieren lassen. Aber auch bei einer ungestörten Nachrichtenübertragung treten Verfälschungen auf. Sie werden durch den begrenzten Frequenzbereich des Übertragungskanals verursacht. Hier muss die Nachrichtentechnik Lösungen finden, mit denen eine möglichst große Anzahl von Nachrichten mit einem vertretbaren Maß an Verfälschungen übertragen werden kann.

92.3 Grundbegriffe

Die Nachrichtentechnik bekam ihre ersten Impulse durch den weltweiten Einsatz des Telefons, aber erst 1948 begründete *Shannon* mit seiner Informationstheorie die moderne Nachrichtentechnik. Es werden folgende Begriffe verwendet (auszugsweise DIN 44300):

Nachricht: Sie besteht aus der Information und dem Signal. Sie wird unverändert weitergegeben.
Information: Sinngehalt der Nachricht (*was* mitgeteilt werden soll).
Signal: Die physikalische Realisierung der Nachricht (*wie* es mitgeteilt wird).
Daten: Sie werden, im Gegensatz zu Nachrichten, verändert und weiterverarbeitet.
Analoge Signale: Kontinuierlich über der Zeit dargestellte Signale.
Digitale Signale: Zu diskreten Zeitpunkten dargestellte Signale; die Amplitudenwerte sind codiert.

Beispiel 1

Jemand hat sich einen Hund zugelegt und möchte das einem anderen mitteilen. Die *Information* „Ich habe mir einen Hund zugelegt" kann durch unterschiedliche *Signale* übermittelt werden: Das Wort „Hund" als geschriebener Text, ein Bild des Hundes, das Wort „Hund" im Morsealphabet oder beliebige andere Zeichen. Wichtig ist, dass die im Signal vorhandenen Zeichen vorher zwischen Sender und Empfänger in ihrer Bedeutung vereinbart worden sind, sonst „versteht" der Empfänger die Nachricht nicht. ◀

92.4 Nachricht, Information und Signal

Um zu zeigen, welche Signalverläufe zur Nachrichtenübertragung geeignet sind, muss zunächst der Begriff „Informationsgehalt" erläutert werden.

92.4.1 Informationsgehalt

Bereits Morse hatte in seinem Alphabet den am häufigsten vorkommenden Buchstaben „e" mit einem Punkt, also mit der kürzesten Codierung, dargestellt. Um zu zeigen, dass das tatsächlich sinnvoll ist, wird zunächst ein Maß dafür gesucht, wie „wichtig" eine Nachricht (ein Zeichen) ist. Da einerseits die Nachricht aus mehreren Zeichen (Buchstaben, Ziffern) zusammengesetzt ist, andererseits die Wichtigkeit der Nachricht von persönlichen Dingen beeinflusst wird und damit nicht wertneutral ist, kann sich die Bewertung nur darauf beziehen, *wie häufig* die *einzelnen Zeichen* auftreten. Die zu definierende Größe wird *Informationsgehalt I* genannt und muss, um nicht mit den Gesetzen der Wahrscheinlichkeitsrechnung im Widerspruch zu stehen, die drei folgenden Bedingungen erfüllen: 1. Der Informationsgehalt ist um so kleiner, je größer die Wahrscheinlichkeit p ist, dass dieses Zeichen auftritt. 2. Der Informationsgehalt eines Zeichens mit der Wahrscheinlichkeit $p = 1$ (das Zeichen tritt sicher auf) ist Null. 3. Der Informationsgehalt voneinander unabhängiger Zeichen soll sich addieren.

Diese drei Forderungen werden durch die folgende Formel erfüllt:

Informationsgehalt

$$I(x_i) = \mathrm{ld}\,\frac{1}{p(x_i)} \quad \text{in bit} \tag{92.1}$$

x_i die einzelnen Zeichen; ld Logarithmus zur Basis 2; $p(x_i)$ Wahrscheinlichkeit für das Auftreten des Zeichens x_i.

Die Wahl der Basis 2 ist nicht zwingend, aber naheliegend, da die binäre Kodierung mit ihren zwei Zuständen sehr häufig angewendet wird.

Aus (92.1) ergibt sich eine wichtige Aussage: *Es handelt sich nur dann um eine echte Information, wenn der Empfänger sie nicht schon vorher kennt.*

Beispiel 2

Die Wahrscheinlichkeit, 6 „richtige" Zahlen im Lotto „6 aus 49" ausgewählt zu haben, ist ungefähr $1/14.000.000$. Mit (92.1) ist der Informationsgehalt der Nachricht, tatsächlich 6 richtige Zahlen zu haben, etwa $\mathrm{ld}(14.000.000) \approx 23{,}7$ bit. In dem hohen Zahlenwert spiegelt sich wider, wie wichtig diese Nachricht für den Empfänger ist. Demgegenüber ist die Wahrscheinlichkeit, *keine* 6 richtigen Zahlen getippt zu haben, etwa $(1 - 1/14.000.000) \approx 0{,}99999993$. Daraus folgt der Informationsgehalt zu ungefähr $\mathrm{ld}(1/0{,}99999993)$ bit $\approx \mathrm{ld}(1{,}00000007)$ bit $\approx 0{,}0000001$ bit. Der sehr kleine Zahlenwert zeigt, dass diese Nachricht unwichtig ist, weil man sie „so gut wie sicher erwartet hat". Wendet man die Ergebnisse auf das Morsealphabet an, so war es sinnvoll, dem am häufigsten vorkommenden Buchstaben „e" mit der Auftrittswahrscheinlichkeit $p(e) \approx 14\,\%$ die kürzeste Codierung, einen Punkt, zuzuordnen. Ein einzelner Punkt kann bei der Übertragung am ehesten „verloren gehen". Da aber $I(e)$ nur etwa 2,84 bit ist, trägt ein fehlender Buchstabe „e" nur geringfügig zur Nachrichtenverfälschung bei. Anschaulich liegt das daran, dass man diesen Buchstaben am ehesten in verfälschte Texte einfügt, weil er „sowieso erwartet" wird und man damit erfahrungsgemäß die Verfälschung schnell und einfach korrigiert (siehe auch Abschn. 108.1.1, „Quellencodierung"). ◀

92.4.2 Signale in der Nachrichtentechnik

Abb. 92.2 stellt eine Übersicht über die in der Elektrotechnik verwendeten Signalverläufe unter dem Gesichtspunkt dar, ob sie zur Informationsübertragung geeignet sind. Da ein Signal nur dann Information enthält, wenn sein Verlauf vorher nicht bekannt ist, sind auch nur die stochastischen Signale für die Nachrichtentechnik geeignet.

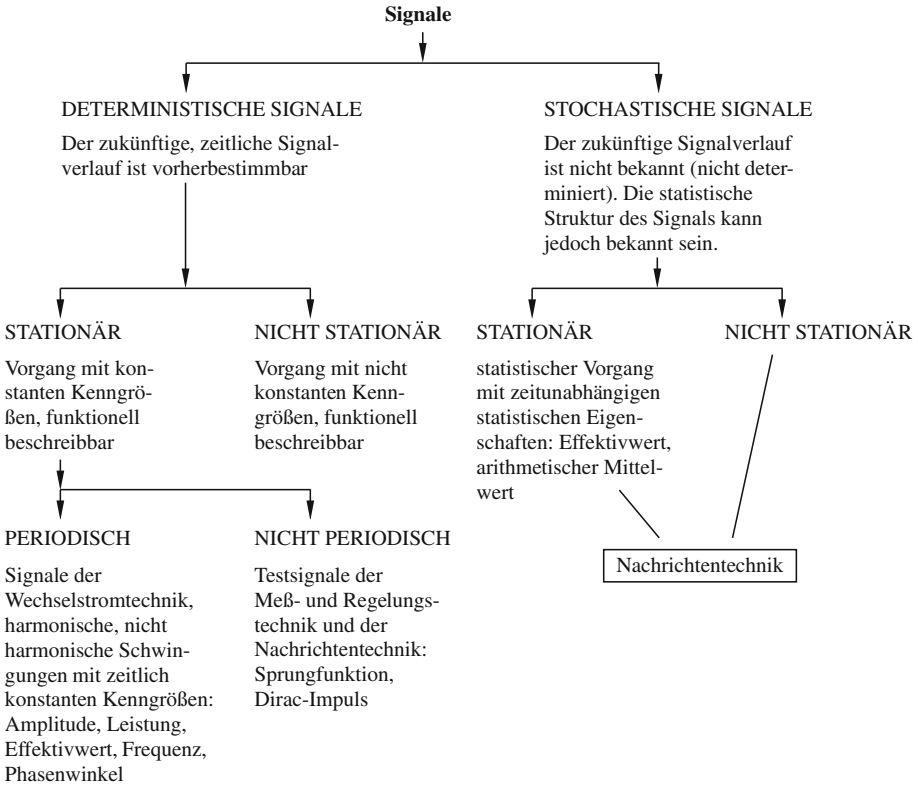


Abb. 92.2 Signale in der Nachrichtentechnik

92.4.3 Entropie

Da sich eine Nachricht in der Regel aus mehreren Zeichen (Buchstaben, Ziffern) zusammensetzt und der Informationsgehalt der einzelnen Zeichen nach (92.1) bestimmt wird, ergibt sich der Informationsgehalt der vollständigen Nachricht durch die Bildung der Summe aus dem Informationsgehalt der einzelnen Zeichen, multipliziert mit der Wahrscheinlichkeit ihres Auftretens. Dieser Wert wird *Entropie* genannt und mit H bezeichnet:

Entropie oder mittlerer Informationsgehalt

$$H = \sum_i p_i \cdot \mathrm{ld}\,\frac{1}{p_i} \quad \text{in bit} \tag{92.2}$$

p_i Wahrscheinlichkeit für das Auftreten des Zeichens x_i.

Der Wert ist stets positiv. Er wird maximal, wenn alle Zeichen x_i mit der gleichen Wahrscheinlichkeit p_i auftreten: $p_i = p_0$. Man nennt diesen Wert *Entscheidungsgehalt* H_0.

Entscheidungsgehalt

$$H_0 = H_{\max} = \sum_1^n p_0 \,\mathrm{ld}\,\frac{1}{p_0} = \frac{1}{n} n \,\mathrm{ld}\,\frac{1}{p_0} = \mathrm{ld}\,\frac{1}{p_0} \tag{92.3}$$

alle Zeichen gleich wahrscheinlich mit der Wahrscheinlichkeit p_0.

92.4.4 Redundanz

Sind die Wahrscheinlichkeiten p_i für das Auftreten der Zeichen x_i nicht gleich, ergibt sich eine Entropie, die kleiner ist als der Maximalwert H_0. Die Differenz zwischen H_0 und H wird *Redundanz* R („Weitschweifigkeit“) genannt, die relative Redundanz r ergibt sich durch Division durch H_0:

$$\textit{Redundanz} \qquad R = H_0 - H \quad \text{in bit} \tag{92.4a}$$

$$\textit{Relative Redundanz} \qquad r = \frac{H_0 - H}{H_0} \quad \text{in bit} \tag{92.4b}$$

Die Redundanz ist der Anteil, der keine Information enthält und damit überflüssig ist. Sie verlängert die Zeitdauer zur Übertragung, gestattet andererseits aber eine Fehlererkennung und -korrektur (Abschn. 108.1). Das wird am folgenden Beispiel 3 gezeigt.

Beispiel 3

Wenn in der deutschen Sprache die 26 Buchstaben mit gleicher Wahrscheinlichkeit ($p_i = 1/26$) aufträten, folgte daraus $H = H_0 = \mathrm{ld}\,26 = 4{,}7$ bit. Nach *Küpfmüller* ergibt sich aber $H \approx 1{,}3$ bit. Das liegt daran, dass 1. die Buchstaben nicht gleichwahrscheinlich auftreten (z. B. e häufiger als k), 2. bestimmte Buchstabenfolgen häufiger

sind als andere (z. B. „ei“ häufiger als „kp“ oder „ein“ häufiger als „dol“) und 3. bestimmte Buchstabenfolgen sehr unwahrscheinlich sind (z. B. „aai“ oder „sfk“). Damit ergibt sich die relative Redundanz $r \approx (4{,}7 - 1{,}3)/4{,}7 \approx 72\,\%$. Das heißt, dass nur 28 % der Sprache Information enthalten und 72 % überflüssig sind. Der positive Aspekt der Redundanz liegt darin, dass auch dann eine nahezu fehlerfreie Sprachübermittlung möglich ist, wenn die Übertragung gestört ist: In einer lärmerfüllten Fabrikhalle kann man sich relativ fehlerfrei verständigen, obwohl man wegen der Geräusche nur einen Bruchteil der gesprochenen Worte vollständig versteht. ◄

92.4.5 Informationsfluss

Der *Informationsfluss F* gibt den pro Zeiteinheit übermittelten mittleren Informationsgehalt an:

$$\textit{Informationsfluss} \quad F = \frac{H}{T_m} \quad \text{in} \ \frac{\text{bit}}{\text{s}} \tag{92.5}$$

T_m mittlere Zeit zur Übertragung eines Nachrichtenelementes in s, siehe (92.6).

Da ein Nachrichtenelement meist aus mehreren Symbolen (Zeichen) besteht, berechnet sich T_m zu:

$$T_m = \sum_i p_i T_i \quad \text{in s} \tag{92.6}$$

T_i Dauer zur Übertragung eines Symbols mit der Wahrscheinlichkeit p_i in s.

92.4.6 Kanalkapazität, Dynamik

Die *Kanalkapazität C* ist der *maximale Informationsfluss*, der über einen *gegebenen Nachrichtenkanal ohne Fehler* übertragen werden kann:

Kanalkapazität

$$C = F_{\max} = \left(\frac{H}{T_m}\right)_{\max} = 2 \cdot B \cdot D \quad \text{in bit/s} \tag{92.7a}$$

T_m mittlere Zeit zur Übertragung eines Nachrichtenelementes nach (92.6) in s, B Bandbreite des Signals in Hz, D Dynamik (siehe Abschn. 92.4.7 und (92.7b)).

$$\textit{Dynamik} \quad D = \text{ld}\, N \quad \text{in bit} \tag{92.7b}$$

N Anzahl der diskreten Amplitudenstufen, siehe Abschn. 92.4.7.

Weiterhin gilt:

$$C = 2 \cdot B \cdot \text{ld} \sqrt{\frac{P_S}{P_N} + 1} = B \cdot \text{ld}\left(\frac{P_S}{P_N} + 1\right) \quad \text{in bit/s} \tag{92.7c}$$

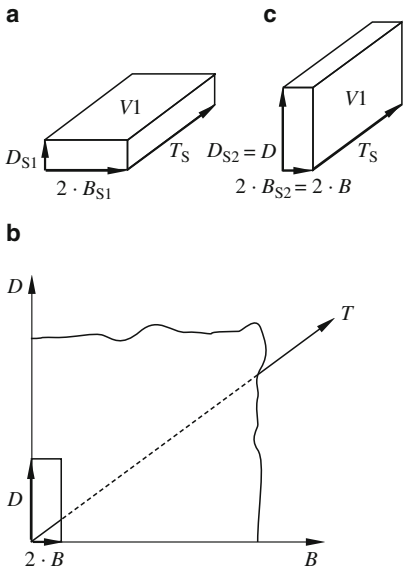


Abb. 92.3 Nachrichtenquader: **a** gegebenes Signal mit D_{S1}, B_{S1}, T_S, **b** Kanal mit gegebenem B, D, **c** an den Kanal (**b**) angepasstes Signal (**a**)

P_S Signalleistung in W, P_N Rausch- bzw. Störleistung in W, B Bandbreite des Signals in Hz, D Dynamik (siehe Abschn. 92.4.7 und (92.7b)).

Der Zusammenhang zwischen der Kanalkapazität C und den Größen B, P_S und P_N wurde von *Shannon* abgeleitet, die Erklärung für die Dynamik D folgt im anschließenden Abschn. 92.4.7. Die Kanalkapazität eines gegebenen Übertragungskanals hängt von dessen technischer Realisierung ab und ist eine Konstante, die gleichzeitig ein Maß für seine Leistungsfähigkeit ist.

92.4.7 Nachrichtenquader

Ein Übertragungskanal kann nur dann ein zeit- und amplitudendiskretes Signal (Begriffe siehe Abschn. 92.4.8) *fehlerfrei* übertragen, wenn es eine maximale Frequenz f_g und eine maximale Anzahl N unterschiedlicher Amplitudenwerte besitzt, die binär mit $D_S = \text{ld}\, N$ bit codiert werden. Liegt ein kontinuierliches Signal vor, lässt es sich auf ein zeit- und amplitudendiskretes Signal zurückführen. D_S ist die *Dynamik* des Signals. Es wird nach dem Abtasttheorem von *Shannon* theoretisch mit $2 \cdot f_g$, praktisch mit $\geq 2{,}2 \cdot f_g$ abgetastet. Für seine Übertragung wird die Zeit T_S benötigt. Dann kann man die Gesamtheit des Signals darstellen als das „Volumen $V1$“ eines Quaders mit den „Seitenlängen“ Bandbreite $2 \cdot B_{S1}$, Dynamik D_{S1} und Zeit T_S, Abb. 92.3.

Der *Kanal*quader ist in seinem „Querschnitt“ $2{\cdot}B{\cdot}D$ vorgegeben. Der „Querschnitt“ des *Signal*quaders muss so gewählt werden, dass er durch die Öffnung des Kanalquaders „hindurchpasst“, ohne im Idealfall Lücken zu lassen. In Abb. 92.3 wurde die zu große Bandbreite B_{S1} des gegebenen Signals verringert und die

Dynamik D_{S1} entsprechend vergrößert und damit der Kanalkapazität angepasst: $2{\cdot}B_{S1}{\cdot}D_{S1} = 2{\cdot}B{\cdot}D$. Gelingt dies bei gleicher Zeit T_S im Signal- und im Kanalquader, wie in Abb. 92.3 dargestellt, spricht man von *Echtzeitübertragung*. Für die Modulation und die Codierung von Signalen wird die Anpassung zwischen Signal und Kanal wie folgt vorgenommen (Signalgrößen mit Index „S“, Kanalgrößen ohne Index):

$C = C_S$ *Echtzeitübertragung*. Dies wird für die folgende Zusammenstellung vorausgesetzt. ($C > C_S$ ist zwar möglich, wird aber vermieden, da dann die teuren Kanalkapazitäten nicht voll ausgenutzt werden.)

$B = B_S; D = D_S$ *Einseitenband-Amplitudenmodulation*. Die Bandbreite (Frequenzbereich des Signals) wird nur in einen anderen Frequenzbereich verschoben.

$D_S > D$ *Frequenzmodulation*, *Pulscodemodulation*. Der für den Kanal zu große Dynamikbereich des Signals wird verkleinert und in die größere Bandbreite des Kanals umgesetzt.

$D_S < D$ *Pulscodemodulation*. Die zur Verfügung stehende größere Dynamik des Kanals wird dadurch ausgenutzt, dass mehrere Abtastwerte des Signals in einem Codewort zusammengefasst werden. Damit ist für die Kanalbandbreite ein kleinerer Wert als für die Signalbandbreite erforderlich: $B_S > B$.

Beispiel 4

Ein Kanal hat die Bandbreite 4,5 kHz und kann 256 Amplitudenwerte übertragen. Dann ist seine Dynamik $D = \mathrm{ld}(256)\,\mathrm{bit} = 8\,\mathrm{bit}$ und seine Kanalkapazität $C = 2 \cdot B \cdot D = 2 \cdot 4{,}5 \cdot 10^3 \cdot (\mathrm{ld}\,256)\,\mathrm{bit/s} = 72 \cdot 10^3\,\mathrm{bit/s}$. Soll über diesen Kanal ein Signal mit 1024 Amplitudenstufen übertragen werden, ohne dass die zur Übertragung notwendige Zeit geändert wird (Echtzeitübertragung), lässt sich die Signalbandbreite B_S bestimmen: $C = 2 \cdot B \cdot D = 2 \cdot 4{,}5 \cdot 10^3 \cdot (\mathrm{ld}\,256)\,\mathrm{bit/s} = 2 \cdot B_S \cdot (\mathrm{ld}\,1024)$. Daraus folgt: $B_S = f_S = 3{,}6\,\mathrm{kHz}$. (Es wurde, wie in der Praxis häufig anzutreffen, $B_S = f_S$, d. h. Signalbandbreite gleich höchster Signalfrequenz, gesetzt.) ◀

92.4.8 Signale im Zeitbereich: Analog, digital, kontinuierlich, diskret

In Abb. 92.4 sind die in der Nachrichtentechnik auftretenden Signalverläufe $s(t)$ im Zeitbereich dargestellt ($s(t)$: Spannungs-, Strom- oder Feldverlauf). Es wird vorausgesetzt, dass das zu übertragende Signal (z. B. Sprachsignal) in *analoger* Form nach Abb. 92.4a vorliegt, d. h. zu *jedem* Zeitpunkt existiert ein eindeutig zugeordneter Amplitudenwert, und *alle* Amplitudenwerte innerhalb eines Wertebereiches sind möglich. *Shannon* hat gezeigt, dass es unter bestimmten Voraussetzungen aus-

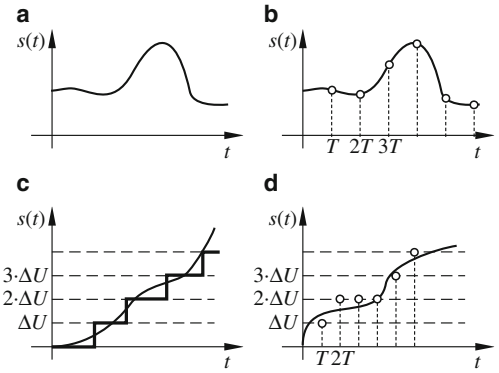


Abb. 92.4 Signaldarstellungen im Zeitbereich. **a** zeit- und amplitudenkontinuierlich, analog, **b** zeitdiskret, amplitudenkontinuierlich, analog, **c** zeitkontinuierlich, amplitudendiskret, digital, **d** zeit- und amplitudendiskret, digital

reicht, den Signalverlauf nur noch zu diskreten Zeitpunkten zu erfassen (Abb. 92.4b). Das analoge Signal wird *abgetastet*. Ist die Zeitdauer t_a der Abtastung viel kleiner als der Abstand T der Abtastzeitpunkte, kann man die Signale mehrerer Nachrichten gleichzeitig über einen Kanal übertragen, indem man die einzelnen Signale in dem Zeitbereich T nacheinander anordnet (*Zeitmultiplexverfahren*, siehe Kap. 100, „Mehrfachübertragung“).

Der Amplitudenwert von analogen Signalen oder Daten kann auf dem Weg vom Sender zum Empfänger durch Störungen verfälscht werden. In kritischen Fällen wandelt man deshalb den analogen Amplitudenwert in einen *digital codierten* Wert um, indem man ihn als ganzzahliges Vielfaches einer kleinsten Einheit ΔU (Abb. 92.4c, d) darstellt. Das Signal besteht nur noch aus den zwei logischen Zuständen 0 und 1. Damit kann auch bei gestörter Übertragung der richtige Wert häufig noch erkannt werden. Das Signal nach Abb. 92.4c tritt am Ausgang eines Digital-Analog-Umsetzers, also auf der Empfängerseite, auf. Abb. 92.4d zeigt ein analoges Signal, das auf der Senderseite zur Übertragung aufbereitet worden ist, indem es abgetastet und anschließend in einem Analog-Digital-Umsetzer in digital codierte Werte umgesetzt wurde. Dabei können der codierte und der tatsächliche Wert maximal um $\pm\Delta U/2$ voneinander abweichen:

größte Abweichung bei digitaler Codierung

$$\pm\frac{\Delta U}{2} = \pm\frac{1}{2} \cdot \frac{1}{2^n - 1} \cdot U \qquad (92.8)$$

n Anzahl der Bits zur Codierung, U Maximalwert von $s(t)$ in V, $\Delta U = U/(2^n - 1)$ in V.

92.4.9 Signale im Frequenzbereich

Für die rechnerische Erfassung von Übertragungskanälen hat es sich bewährt, das *Amplitudenspektrum* der Signale, d. h. die Amplituden der im Signal vorkommenden Frequenzanteile in Abhängigkeit von der Frequenz darzustellen. Man benutzt

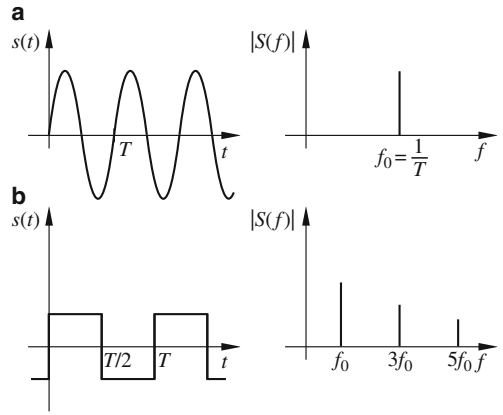


Abb. 92.5 Signaldarstellung periodischer Funktionen im Zeit- und Frequenzbereich. **a** Sinusförmiges Signal, **b** rechteckförmiges Signal

dazu die Beziehung zwischen Frequenz und Periodendauer, $f = 1/T$, die in dieser Form nur für periodische Signale gilt und die deshalb auf die in der Nachrichtentechnik verwendeten stochastischen Signale erweitert werden muss.

92.4.9.1 Periodische sinusförmige Signale

In Abb. 92.5a ist sowohl der Verlauf in Abhängigkeit von der Zeit, $s(t)$, als auch der Verlauf in Abhängigkeit von der Frequenz, $|\underline{S}(f)|$, dargestellt. Der Zusammenhang zwischen beiden Darstellungen ist durch $f_0 = 1/T$ gegeben. Hinweis: $\underline{S}(f)$ kann nach (92.14) eine komplexe Funktion sein. Für die Darstellung wählt man dann entweder $\underline{S}(f)$ oder $|\underline{S}(f)|$.

92.4.9.2 Periodische nichtsinusförmige Signale

Abb. 92.5b zeigt als Beispiel einen periodischen rechteckförmigen Verlauf $s(t)$. Zur Berechnung der im Signal enthaltenen Frequenzanteile wird die Fourierreihe verwendet, mit der das Signal $s(t)$ auch in folgender Form darstellbar ist:

Fourierreihe

$$s(t) = a_0 + \sum_{n=1}^{\infty}[a_n \cdot \cos(n\omega_0 t) + b_n \cdot \sin(n\omega_0 t)]$$

$$a_0 = \frac{1}{T}\int_0^T s(t)dt; \quad a_n = \frac{2}{T}\int_0^T s(t) \cdot \cos(n\omega_0 t)dt;$$

$$b_n = \frac{2}{T}\int_0^T s(t) \cdot \sin(n\omega_0 t)dt \qquad (92.9a)$$

Eine für die Nachrichtentechnik vorteilhafte Darstellung hat die Form:

$$s(t) = s_0 + \sum_{n=1}^{\infty} \hat{s}_n \cdot \cos(n \cdot \omega_0 \cdot t + \varphi_n) \qquad (92.9b)$$

Der Zusammenhang zwischen beiden Darstellungen ist gegeben durch:

$$a_0 = s_0, \quad a_n = \hat{s}_n \cdot \cos\varphi_n;$$

$$b_n = -\hat{s}_n \cdot \sin\varphi_n; \quad \hat{s}_n = \sqrt{a_n^2 + b_n^2} \qquad (92.9c)$$

Wird $s(t)$ entsprechend dem Verlauf in Abb. 92.5b in (92.9a) mit $s(t) = U_0$ und den Integralgrenzen 0 bis $T/2$ sowie $s(t) = -U_0$ und den Integralgrenzen $T/2$ bis T eingesetzt, ergibt sich:

$$s(t) = \frac{2 \cdot U_0}{\pi}\left(\frac{1}{1} \cdot \sin(1\omega_0 t) + \frac{1}{3} \cdot \sin(3\omega_0 t) + \frac{1}{5} \cdot \sin(5\omega_0 t) + \ldots\right) \qquad (92.10)$$

Die Scheitelwerte der Amplituden bei den diskreten Frequenzen $n \cdot \omega_0/2p = n \cdot f_0$ sind in Abb. 92.5b in das Diagramm $|\underline{S}(f)| = f(f)$ eingetragen. (Hinweis: Es können auch die Effektivwerte der Amplituden eingetragen werden. Die Scheitelwerte nach (92.10) sind dann durch $\sqrt{2}$ zu dividieren.)

92.4.9.3 Nichtperiodische Signale

Die Informationsübertragung geschieht mit stochastischen Signalen. In der Praxis reicht es aber häufig aus, die Eigenschaften eines Übertragungskanals oder eines Vierpols mit nichtperiodischen deterministischen Signalen zu erfassen, denn man kann daraus auf das näherungsweise Verhalten bei stochastischen Signalen schließen. Um die in einem nichtperiodischen Signal enthaltenen Frequenzanteile zu bestimmen, wird zunächst eine für die Anwendung in der Nachrichtentechnik besonders geeignete Darstellung für periodische Signale abgeleitet und anschließend auf nichtperiodische Signale erweitert.

Fouriertransformation

Euler'sche Formel: $\cos x = \dfrac{e^{jx} + e^{-jx}}{2}$

Für die in (92.9b) dargestellten Summanden gilt damit:

$$s(t) = s_0 + \sum_{n=1}^{\infty} \hat{s}_n \cdot \frac{e^{j(n\omega_0 t + \varphi_n)} + e^{-j(n\omega_0 t + \varphi_n)}}{2}$$

$$= s_0 + \sum_{n=1}^{\infty} \hat{s}_n e^{j\varphi_n} \frac{1}{2} e^{jn\omega_0 t} + \sum_{n=1}^{\infty} \hat{s}_n e^{-j\varphi_n} \frac{1}{2} e^{-jn\omega_0 t} \qquad (92.11)$$

Definitionen:

$$\underline{C}_n = \frac{\hat{s}_n \cdot e^{j\varphi_n}}{2}; \quad \underline{C}_{-n} = \frac{\hat{s}_n \cdot e^{-j\varphi_n}}{2}$$

Für den zweiten Summanden in (92.11) kann man schreiben:

$$\sum_{n=1}^{\infty} \hat{s}_n e^{-j\varphi_n} \frac{1}{2} e^{-jn\omega_0 t} = \sum_{n=-1}^{-\infty} \hat{s}_n e^{j\varphi_n} \frac{1}{2} e^{jn\omega_0 t} \qquad (92.12)$$

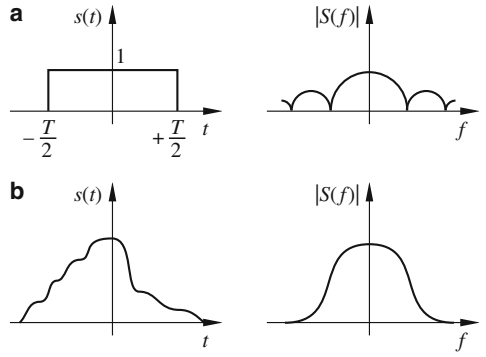


Abb. 92.6 Signaldarstellung nichtperiodischer Funktionen im Zeit- und Frequenzbereich. **a** Rechteckimpuls, **b** stochastisches Signal

Setzt man (92.12) in (92.11) ein, folgt:

Komplexe Fourierreihe:

$$s(t) = \sum_{-\infty}^{+\infty} \underline{C}_n \cdot e^{jn\omega_0 t} \quad \text{mit } \underline{C}_n = \frac{1}{T}\int_0^T s(t)e^{-jn\omega_0 t}dt \qquad (92.13)$$

Formal ist eine negative Frequenz eingeführt worden, wobei zu jeder negativen Frequenz $-n\cdot f_0$ eine entsprechende positive Frequenz $+n\cdot f_0$ gehört. Die Darstellung in (92.13) heißt *komplexe Fourierreihe*.

Das in einem nichtperiodischen Signal enthaltene Spektrum erhält man durch den Grenzübergang $T \to \infty$ und damit $n\cdot f_0 \to f$, $f_0 = 1/T \to df$. Die Summe geht in ein Integral über, die entstehende Funktion $\underline{S}(f)$ heißt *(komplexe) Spektraldichte der Zeitfunktion* $s(t)$, die mathematische Operation ist die *Fouriertransformation*.

Fouriertransformation:

$$\underline{S}(f) = \int_{-\infty}^{+\infty} s(t)\cdot e^{-j2\pi ft}dt;$$
$$s(t) = \int_{-\infty}^{+\infty} \underline{S}(f)\cdot e^{j2\pi ft}df \qquad (92.14)$$

$\underline{S}(f)$ Spektraldichte eines gegebenen Zeitsignals $s(t)$.

In Abb. 92.6a ist ein Rechteckimpuls mit der dazugehörigen Spektraldichte dargestellt. Es wurde gemäß Hinweis im Abschn. 92.4.9.1 der Betrag $|\underline{S}(f)|$ aufgetragen. Es sind unendlich viele Frequenzanteile vorhanden, was durch den unendlich steilen Anstieg und Abfall hervorgerufen wird. Da ein Übertragungskanal niemals unendlich viele Frequenzen übertragen kann, ist der dargestellte Rechteckimpuls in dieser Form nicht übertragbar. Das ist andererseits auch nicht erforderlich, denn mit technischen Mitteln kann er nicht erzeugt werden: Zu den Zeitpunkten $-T/2$ und $+T/2$ müssten unendlich viele Funktionswerte zwischen 0 und 1 gleichzeitig existieren, was im Widerspruch zur Eindeutigkeit physikalischer Gesetze steht.

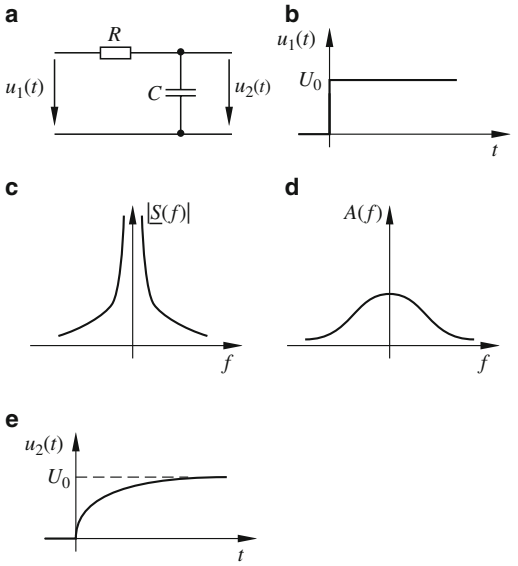


Abb. 92.7 Sprungantwort eines Tiefpasses. **a** Schaltbild, **b** Einheitssprung als Anregung, **c** Spektraldichte des Einheitssprunges, **d** Übertragungsfunktion des Tiefpasses, **e** Ausgangsspannung des Tiefpasses

Abb. 92.6b zeigt ein stochastisches Signal (z. B. Sprache) mit zugehöriger Spektraldichte.

Zwei weitere Größen zur Beschreibung eines Systems sind die *Übertragungsfunktion* $\underline{H}(f)$ („Spektraldichte" eines Übertragungssystems) als das Verhältnis von Ausgangsgröße $\underline{S}_2(f)$ zu Eingangsgröße $\underline{S}_1(f)$ sowie deren Betrag $A(f)$. Die zugeordneten Zeitfunktionen werden mit $h(t)$ und $a(t)$ bezeichnet:

Übertragungsfunktion:

$$\underline{H}(f) = \frac{\underline{S}_2(f)}{\underline{S}_1(f)} = |\underline{H}(f)|e^{-jb(f)};$$
$$|\underline{H}(f)| = A(f) \qquad (92.15)$$

$\underline{S}_1(f)$ Eingangsgröße, $\underline{S}_2(f)$ Ausgangsgröße, $\underline{H}(f)$ auch mit „Spektraldichte" eines Übertragungssystems bezeichnet; Beispiel: Verhältnis von Ausgangs- zu Eingangsspannung an einem Tiefpass, siehe Beispiel 5.

Zugeordnete Zeitfunktionen: $h(t)$, $a(t)$ Dämpfungsmaß und $b(t)$ Phasenmaß.

Beispiel 5

Die Eingangsspannung des in Abb. 92.7 gezeigten Tiefpasses ist ein Einheitssprung.

Die Übertragungsfunktion ergibt sich zu:

$$\underline{H}(f) = \frac{\underline{S}_2(f)}{\underline{S}_1(f)} = \frac{\underline{U}_2(j\omega)}{\underline{U}_1(j\omega)} = \frac{1}{1+j\omega T}; \quad T = RC$$
$$A(f) = \frac{1}{\sqrt{1+(\omega T)^2}}$$

Der nach (92.14) berechnete Betrag der Spektraldichte, $|\underline{S}(f)|$, ist in Abb. 92.7c eingezeichnet, $A(f)$ in Abb. 92.7d. Formt man (92.15) um und setzt für $\underline{S}_1(f)$ die Spektraldichte des Einheitssprunges ein, ergibt sich: $\underline{S}_2(f) = \underline{S}_1(f) \cdot \underline{H}(f)$ und daraus mit (92.14) durch Rücktransformation der in Abb. 92.7e eingezeichnete Verlauf der Spannung am Ausgang des Tiefpasses. ◀

Diskrete Fouriertransformation (DFT)

Es wird von einem zeitbegrenzten Signal ausgegangen, das zu den Zeitpunkten $n \cdot T_a$ abgetastet wird. Dabei hat n den Bereich von 0 bis $N-1$. T_a ist das Abtastintervall. Die gesamte Messzeit ist dann $N \cdot T_a$. Damit setzt man in (92.14): $t \to n \cdot T_a$, $s(t) \to s(n \cdot T_a)$, $e^{-j\omega t} \to e^{-j\omega nTa}$. Es ergibt sich:

Diskrete Fouriertransformation:

$$S_d(f) = \sum_{n=0}^{N-1} f(nT_a) \cdot e^{-j2\pi kn/N} \quad \text{mit } k = 0,1,\ldots,N-1 \tag{92.16}$$

Durch geeignete mathematische Algorithmen kann man auch bei einer großen Anzahl von Abtastwerten die Transformation ausreichend schnell durchführen.

Schnelle Fouriertransformation (Fast Fourier Transform, FFT)

Der Grundgedanke ist, die Folge $f(n)$ eines gegebenen Signals mit N Werten so zu zerlegen, dass mehrere kleinere Teilfolgen entstehen, die getrennt transformiert und anschließend zur Gesamtfunktion überlagert werden können. Diese Aufteilung ist dann besonders günstig, wenn sich N aus 2^k mit $k = 2,3,4,\ldots$ ergibt. Die Eingangsfolge wird zunächst in zwei Teilfolgen zerlegt, deren Elemente den Abstand $2 \cdot T_a$ haben. Diese 2 Teilfolgen werden wiederum in je 2 Teilfolgen zerlegt mit dem Abstand $4 \cdot T_a$ usw., bis die einzelnen Folgen nur noch aus 2 Elementen bestehen. Auf diese wird dann der DFT-Algorithmus angewendet.

Laplacetransformation

Hier folgt aus Gründen der Systematik eine kurze Einführung in die Laplacetransformation. Eine ausführlichere Darstellung gibt es in Teil XIV „Signal- und Systemtheorie“, Kap. 112 „Nichtperiodische zeitkontinuierliche Signale“.

Die Laplacetransformation unterscheidet sich von der Fouriertransformation u. a. dadurch, dass die zeitlichen Vorgänge für $t < 0$ null sind. Der betrachtete Vorgang setzt im Zeit- und Frequenzbereich erst ab $t = 0$ ein (keine negativen Frequenzen) und steht damit in anschaulicherer Beziehung zur Praxis als bei der Fouriertransformation. Um Verwechslungen mit dem Laplace-Operator s zu vermeiden, wurde die Zeitfunktion mit $f(t)$ anstatt $s(t)$ bezeichnet. Die Transformationsformeln lauten:

Laplacetransformation:

$$L[f(t)] = F(s) = \int_{+0}^{\infty} f(t) \cdot e^{-st} dt;$$

$$L^{-1}[F(s)] = f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s) \cdot e^{st} ds \tag{92.17}$$

L^{-1} inverse Laplacetransformation (Rücktransformation); $F(s)$ Laplacetransformierte der Zeitfunktion $f(t)$.

s ist der Laplace-Operator (in der Literatur auch mit p bezeichnet), es gilt $s = \sigma + j\omega$. Die Einheit ist $1/\mathrm{s}$, die Analogie zum Frequenzspektrum ist damit gegeben. Der Realteil σ musste eingeführt werden, damit die Integrale nach (92.17) endliche Werte ergeben, er spielt bei den weiteren Betrachtungen keine Rolle.

Im Folgenden soll die Anwendung ohne mathematische Ableitungen gezeigt werden.

Berechnung eines Zeitvorganges mit der Laplacetransformation

Das gegebene Netzwerk als Verkörperung einer Leitung, eines Vierpols oder eines Übertragungskanals wird in die s-Ebene transformiert: Widerstände R bleiben unverändert, Kapazitäten erhalten die „Impedanz“ $1/sC$, Induktivitäten und Gegeninduktivitäten erhalten die „Impedanz“ sL bzw. sM. Hier wird die Analogie zur komplexen Rechnung erkennbar, die ja nur für eingeschwungene sinusförmige Vorgänge gilt. Ist ein Kondensator zum Zeitpunkt $t = 0$ auf die Spannung U_0 aufgeladen (diese Spannung kann sich zum Zeitpunkt $t = 0$ nicht sprunghaft ändern), wird das in der s-Ebene durch Reihenschaltung einer Spannungsquelle mit der Spannung U_0/s zum Kondensator berücksichtigt. Die Richtung dieser Spannungsquelle ist die gleiche wie die der Spannung am Kondensator zum Zeitpunkt $t = 0$. Hinweis: Auch wenn die Spannung dieser Spannungsquelle für $t > 0$ als konstanter Wert in Erscheinung tritt, obwohl sich der Kondensator in der Regel umlädt, ist das Ergebnis der Berechnung richtig, denn die am Kondensator in der Zeitebene auftretende Spannung entspricht der Summe (vorzeichenrichtig addiert) aus der eventuell vorhandenen Spannung U_0/s und der Spannung am Kondensator $1/sC$ in der s-Ebene! Fließt durch eine Spule oder Gegeninduktivität zum Zeitpunkt $t = 0$ ein Strom I_0 (dieser Strom kann sich zum Zeitpunkt $t = 0$ ebenfalls nicht sprunghaft ändern), wird dies durch eine Spannungsquelle mit der Spannung $L \cdot I_0$ in Reihe zu sL berücksichtigt. Die Richtung dieser Spannungsquelle ist *entgegengesetzt* zur Richtung des zum Zeitpunkt $t = 0$ fließenden Stromes I_0 (die Spule wird zur Spannungsquelle). Auch hier gilt: Die in der Zeitebene an der Spule auftretende Spannung entspricht der Summe aus der Spannungsquelle mit der Spannung $L \cdot I_0$ und der an der Spule mit sL anstehenden Spannung (vorzeichenrichtig addiert). Wird das Netzwerk von Spannungs- und/oder Stromquellen ab $t = 0$ gespeist, ist deren Verlauf in die s-Ebene zu transformieren und entsprechend einzusetzen. Häufig vorkommende Verläufe kann man Tabellen entnehmen, siehe Tab. 92.2.

Tab. 92.2 Korrespondenzen zur Laplacetransformation (Auswahl)

$F(s)$	$f(t)$
$\frac{1}{s}$	1
$\frac{1}{s^2}$	t
$\frac{1}{s+a}$	e^{-at}
$\frac{1}{s(s+a)}$	$\frac{1}{a}\cdot(1-e^{-at})$
$\frac{1}{(s+a)^2}$	$t\cdot e^{-at}$
$\frac{1}{(s+a)(s+b)}, a\neq b$	$\frac{1}{b-a}(e^{-at}-e^{-bt})$
$\frac{1}{s(s+a)(s+b)}, a\neq b$	$\frac{1}{ab}+\frac{b\cdot e^{-at}-a\cdot e^{-bt}}{a\cdot b(a-b)}$
$\frac{1}{s(s+a)^2}$	$\frac{1-e^{-at}-a\cdot t\cdot e^{-at}}{a^2}$

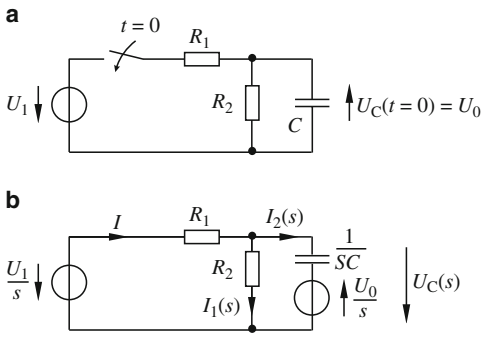


Abb. 92.8 Berechnung eines nichtperiodischen Vorganges mit der Laplace-Transformation. **a** Netzwerk in der Zeitebene, **b** Netzwerk in der s-Ebene

Die zu berechnende Größe wird zunächst in der s-Ebene nach den Regeln einer Netzwerkberechnung mit den dort gültigen Gesetzen (Ohm'sches Gesetz, Kirchhoff'sche Gesetze, Spannungsteiler, Überlagerungssatz) bestimmt. Dann wird sie unter Anwendung einer Tabelle in die Zeitebene zurücktransformiert. Die in Tab. 92.2 aufgelisteten Äquivalenzen sind für die Hin- und Rücktransformation anwendbar.

Zwei Rechenregeln sind wichtig. Regel 1: Besteht die Zeitfunktion aus mehreren Summanden, werden diese einzeln transformiert und in der s-Ebene wieder zu einer Summe vereinigt; das gilt entsprechend für eine Summe in der s-Ebene bei Rücktransformation in die Zeitebene. Regel 2: Eine Konstante, mit der die Zeitfunktion multipliziert wird, bleibt als multiplikative Konstante in der s-Ebene erhalten und umgekehrt. Die Spannungen und Ströme haben in der s-Ebene die Einheit Vs bzw. As, was wiederum auf eine Spektralverteilung hindeutet.

Beispiel 6

In Abb. 92.8 ist ein Netzwerk gezeigt, in dem der Kondensator C zum Zeitpunkt $t = 0$ auf die Spannung U_0 in der eingezeichneten Richtung aufgeladen ist. Zum Zeitpunkt $t = 0$ wird eine Gleichspannung $U = U_1$ auf das Netzwerk geschaltet. Berechnet werden soll der Verlauf der Spannung am Kondensator C.

Der Schalter in der Zeitebene wird in der s-Ebene durch eine Sprungfunktion der Höhe U_1 ersetzt: U_1/s (Tab. 92.2, Sprungfunktion $1/s$, multipliziert mit der Konstanten U_1, Regel 2). Die Widerstände und Kondensatoren gehen gemäß den dargestellten Regeln in die s-Ebene über; der auf die Spannung U_0 aufgeladene Kondensator erhält in Reihe eine Spannungsquelle mit der Spannung U_0/s. Die Richtung ist die gleiche wie die der Spannung an C für $t = 0$. Die gesuchte Spannung U_C ist die Reihenschaltung aus U_0/s und der Spannung an $1/sC$. Für das Netzwerk können die 3 Gleichungen angesetzt werden:

$$\frac{U_1}{s} = I\cdot R_1 + I_1\cdot R_2;$$
$$I = I_1 + I_2; \quad I_1\cdot R_2 = I_2\cdot\frac{1}{sC} - \frac{U_0}{s}$$

Daraus kann I_2 berechnet werden:

$$I_2 = \frac{1}{R_1}\cdot\frac{U_1 + U_0\frac{R_1+R_2}{R_2}}{s+\frac{R_1+R_2}{R_2R_1C}}$$

für U_C folgt:

$$U_C = I_2\cdot\frac{1}{sC} - \frac{U_0}{s}$$

Der erste Summand von U_C muss so umgeformt werden, dass ein Ausdruck $\frac{A}{Bs(s+a)}$ entsteht, der dann mit Regel 2 und Tab. 92.2 zurücktransformiert werden kann (A, B und a sind Konstante, die kein „s" enthalten):

$$U_C = \frac{1}{R_1C}\cdot\frac{1}{s}\cdot\frac{U_1 + U_0\frac{R_1+R_2}{R_2}}{s+\frac{R_1+R_2}{R_2R_1C}} - \frac{U_0}{s}$$

Der zweite Summand, $-U_0/s$, wird nach Regel 1 zu $-U_0$ direkt zurücktransformiert.

Die am Kondensator stehende Spannung für $t \geq 0$ ist damit:

$$u_C = \left(U_1\frac{R_2}{R_1+R_2} + U_0\right)\cdot(1-e^{\frac{-t}{\tau}}) - U_0$$

mit $\tau = \frac{R_1R_2C}{R_1+R_2}$.

Ein einmaliger *rechteckförmiger* Spannungsverlauf kann dadurch auf das Netzwerk gegeben werden, dass die Spannung U_1 zum Zeitpunkt t_1 ($0 < t_1 < \infty$) zu null gemacht wird. Weiterhin wird der Wert der Kondensatorspannung zum Zeitpunkt t_1 für die in Reihe zum Kondensator liegende Spannungsquelle vorzeichenrichtig eingesetzt (also den Wert U_0/s aus dem ersten Teil durch $u_C(t_1)/s$ ersetzen). Die weitere Berechnung erfolgt analog zu dem oben gezeigten Rechengang. ◀

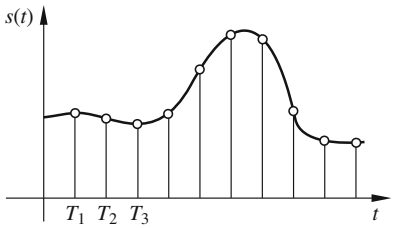


Abb. 92.9 Signalabtastung

92.4.10 Abtasttheorem von Shannon

Abtasttheorem im Zeitbereich

Enthält ein analoges kontinuierliches Signal nur Frequenzanteile bis zu einer maximalen Frequenz f_g, so ist der Signalverlauf eindeutig dadurch gegeben, dass nach Abb. 92.9 das Signal zu diskreten Zeitpunkten T_1, T_2, T_3 usw. abgetastet wird. Für die konstanten zeitlichen Abstände der Abtastimpulse, $T_{n+1} - T_n$, und damit für die Abtastfrequenz f_{ab} gilt:

Abtastfrequenz

$$f_{ab} = \frac{1}{T_{n+1} - T_n} \geq 2 \cdot f_g \quad \text{(theoret.)};$$
$$f_{ab} \geq 2{,}2 \cdot f_g \quad \text{(prakt.)} \qquad (92.18)$$

f_g höchste im Signal vorkommende Frequenz in Hz.

Es genügt, diese abgetasteten Amplitudenwerte zu übertragen. Der Empfänger kann daraus das Originalsignal durch Zwischenschalten eines Tiefpasses zurückgewinnen. Weitere Einzelheiten dazu folgen bei der Behandlung der *Pulsamplitudenmodulation PAM* im Abschn. 96.4.1. Die meisten vorkommenden Signale sind vom Prinzip her frequenzbandbegrenzt (Sprachsignal enthält aufgrund des Kehlkopfaufbaues nur Frequenzen bis etwa 15 kHz). Aus praktischen Erwägungen heraus könnte es aber erforderlich sein, den im Originalsignal enthaltenen Frequenzbereich einzuschränken, weil dann die Abtastfrequenz verringert werden kann. Die sich hieraus ergebenden Vorteile werden bei den *Zeitmultiplexverfahren* im Kap. 100, Mehrfachübertragung, beschrieben. Allerdings bedeutet eine zusätzliche Frequenzbandbegrenzung auch eine Signalverfälschung (z. B. werden sehr schnelle Signaländerungen verlangsamt). Da es keine Filter mit unendlich steiler Flanke bei der maximal zulässigen Frequenz f_g gibt und bei der Abtastung eines sinusförmigen Signals mit $f = f_g$ möglicherweise immer genau in den Nulldurchgängen abgetastet wird, wählt man die Abtastfrequenz zu $f_{ab} \geq 2{,}2 \cdot f_g$.

Abtasttheorem im Frequenzbereich

Ein Signal mit der begrenzten Zeitdauer T_g lässt sich vollständig durch das Frequenzspektrum bei den diskreten Frequenzen im Abstand Δf darstellen:

Frequenzabstand

$$\Delta f \leq \frac{1}{2 \cdot T_g} \quad \text{(theoretisch)};$$
$$\Delta f \leq \frac{1}{2{,}2 \cdot T_g} \quad \text{(praktisch)} \qquad (92.19)$$

T_g begrenzte Zeitdauer des Signals in s.

Gleichung (92.19) folgt aus der Verknüpfung von Zeit und Frequenz über die Fouriertransformation. Für den Faktor 2,2 gilt die gleiche Begründung wie oben.

92.4.11 Zufällige (stochastische) Signale

Zufällige (stochastische) Signale haben keine erkennbaren zeitabhängigen Gesetzmäßigkeiten, trotzdem besitzen sie in den meisten Fällen sogenannte statistische Kenngrößen, die nicht von der Zeit abhängen, sofern man diese Kenngrößen über einen genügend großen Zeitraum betrachtet. Für die Nachrichtentechnik besitzen sie eine große Bedeutung (siehe Abschn. 92.4.11.3): Zum einen als Nutzsignal zur Nachrichtenübertragung, zum anderen als Störsignal in Form von Rauschen.

92.4.11.1 Rauschen

Rauschen ist ein statistischer Vorgang, der durch Wärmebewegung von Ladungsträgern hervorgerufen wird und deshalb ein stochastisches Signal darstellt. Nach der Herkunft des Rauschens unterscheidet man äußere und innere Quellen.

a) Äußere Quellen

- *Atmosphärisches Rauschen*: Es entsteht im Wesentlichen durch Blitzentladungen von Gewittern. Das Rauschspektrum nimmt oberhalb etwa 10 MHz deutlich ab.
- *Kosmisches Rauschen*: Es handelt sich um die Radiostrahlung entfernter Sterne. Es tritt ab etwa 50 MHz störend in Erscheinung, wenn das atmosphärische Rauschen abgeklungen ist.

b) Innere Quellen

- *Widerstandsrauschen oder thermisches Rauschen*: Es entsteht in Wirkwiderständen. Der Mittelwert der Rauschspannung $\bar{u}_R$ berechnet sich zu:
  Widerstandsrauschen: *Rauschspannung*

$$\bar{u}_R = \sqrt{4kTBR} \qquad (92.20a)$$

k Boltzmann-Konstante $1{,}38 \cdot 10^{-23}$ W s/K; T absolute (thermodynamische) Temperatur in K; B Bandbreite (ausgewerteter oder berücksichtigter Frequenzbereich) in Hz; R Widerstand in Ω.

Der Rauschstrom ergibt sich zu:
Widerstandsrauschen: *Rauschstrom*

$$\bar{i}_R = \sqrt{\frac{4kTB}{R}} \tag{92.20b}$$

Formelgrößen siehe (92.20a).
Ein rauschender Widerstand erzeugt an einem gleich großen nichtrauschenden (d. h. idealisierten) Widerstand bei Leistungsanpassung die Rauschleistung
Rauschleistung

$$\bar{p}_R = \frac{\bar{u}_R}{2} \cdot \frac{\bar{i}_R}{2} = kTB \quad \text{bei Leistungsanpassung} \tag{92.20c}$$

Formelgrößen siehe (92.20a).

Beispiel 7

Bei Zimmertemperatur (20 °C = 293 K) beträgt die Rauschspannung eines Widerstandes $R = 10\,\text{k}\Omega$ bei einer Bandbreite von 1 MHz:

$$\bar{u}_R = \sqrt{4 \cdot 1{,}38 \cdot 10^{-23}\,\frac{\text{W s}}{\text{K}} \cdot 293\,\text{K} \cdot 10^6\,\frac{1}{\text{s}} \cdot 10^4\,\frac{\text{V}}{\text{A}}} = 12{,}7\,\mu\text{V}$$

Die Rauschleistung errechnet sich zu:

$$\bar{p}_R = 1{,}38 \cdot 10^{-23}\,\frac{\text{W s}}{\text{K}} \cdot 293\ \text{K} \cdot 10^6\,\frac{1}{\text{s}} = 4 \cdot 10^{-15}\,\text{W}$$

Sie ist unabhängig vom Wert des Widerstandes. Da die Rauschspannung von der Temperatur abhängt, werden Verstärker für besondere Anwendungen (Empfänger für Signale von Weltraumsonden) im Temperaturbereich von einigen Kelvin betrieben (aber oberhalb der Supraleitung). Die Rauschspannung des 10-kΩ-Widerstandes würde dann bei 10 K Betriebstemperatur von 12,7 µV auf etwa 2,3 µV absinken. ◀

- $1/f$-*Rauschen* oder *Funkelrauschen*: Rauschen, das bei Frequenzen unterhalb 1 kHz auftritt und zu niedrigeren Frequenzen hin zunimmt (Amplitude $\sim 1/f$). Es macht sich besonders bei Halbleitern (Transistoren) unangenehm bemerkbar.
- „*Popcorn*"-*Rauschen*: Es handelt sich um sporadisch auftretendes Rauschen mit kurzer Dauer (µs-Bereich) und Amplituden, die um ein Vielfaches über denen des „normalen" Rauschens liegen. Es führt in Verstärkerschaltungen häufig zu Übersteuerungen.
- *Schrotrauschen*: Es wird verursacht durch ungleichmäßige Ladungsträgerinjektion bzw. -emission in Halbleitern und Elektronenröhren. Das mittlere Stromschwankungsquadrat ist gegeben zu:

Schrotrauschen als Stromschwankungsquadrat:

$$\bar{i}_s^2 = 2q_eIB$$

q_e Elementarladung des Elektrons $1{,}6 \cdot 10^{-19}$ A s; B Bandbreite in Hz; I Anoden- bzw. Halbleiterstrom in A.

- *Stromverteilungsrauschen*: Es entsteht durch statistische Schwankungen der Stromaufteilung auf Basis und Kollektor bei Transistoren bzw. auf Gitter und Anode bei Elektronenröhren.

c) Weißes Rauschen

Im betrachteten Frequenzbereich sind, über einen größeren Zeitbereich betrachtet, alle Frequenzen mit gleicher Amplitude vorhanden.

d) Kennzeichnung der Rauscheigenschaften von Verstärkern und Empfängern

Mit *Störabstand* wird das Verhältnis von Signalleistung P_S zu Störsignalleistung P_N bezeichnet:

Störabstand

$$S = \frac{P_S}{P_N} \quad \text{oder} \quad S = 10\,\text{lg}\,\frac{P_S}{P_N} \quad \text{in dB} \tag{92.21}$$

P_S Signalleistung (engl. *s*ignal); P_N Rauschleistung (engl. *n*oise).

In der Rundfunk-Empfangstechnik wird ein Störabstand von 26 dB entsprechend $P_S/P_N \approx 400$ gefordert.

Zum Vergleich der Rauscheigenschaften von Verstärkern oder Empfängern wird die *Rauschzahl* F verwendet. Sie ist definiert als das Verhältnis von Signal- zu Rauschleistung am Eingang bezogen auf das entsprechende Verhältnis am Ausgang:

Rauschzahl

$$F = \frac{(P_S/P_N)_e}{(P_S/P_N)_a} \tag{92.22}$$

P_S Signalleistung; P_N Rauschleistung; Indizes: e Eingang; a Ausgang.

Rauschmaß

$$F = 10 \cdot \lg(F) \quad \text{in dB} \tag{92.23}$$

Die Rauschzahl am Ausgang eines Vierpols (Verstärkers) hängt auch vom Rauschen der Signalquelle ab, deshalb verwendet man bevorzugt die *Rauschtemperatur* Q. Der Vierpol hat die Leistungsverstärkung G und wird als rauschfrei angenommen. Am Eingang liegt ein Widerstand R, der die gleiche Rauschleistung P_N am Ausgang des Vierpols erzeugt wie der Vierpol selber: $P_N = G \cdot k \cdot B \cdot \vartheta_V$ ((92.20c), Leistungsanpassung; Formelgrößen siehe (92.20a)). ϑ_V ist die Rauschtemperatur des Vierpols. Entsprechend rauscht auch die den Vierpol speisende Signalquelle mit der Rauschtemperatur ϑ_{SS}: $P_{SS} = G \cdot k \cdot B \cdot \vartheta_{SS}$.

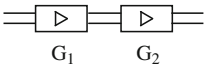


Abb. 92.10 Rauschen eines zweistufigen Verstärkers

Die effektive Rauschleistung am Ausgang ergibt sich durch Addition: $P_{\text{eff}} = G \cdot k \cdot B(\vartheta_{\text{SS}} + \vartheta_V)$. Der Ausdruck $(\vartheta_{\text{SS}} + \vartheta_V)$ wird *effektive Rauschtemperatur* ϑ_{eff} genannt. Damit folgt für die Rauschzahl:

Rauschzahl

$$F = \frac{P_{\text{SS}} + P_N}{P_{\text{SS}}} = 1 + \frac{P_N}{P_{\text{SS}}} = 1 + \frac{GkB\vartheta_V}{GkB\vartheta_{\text{SS}}}$$
$$= 1 + \frac{\vartheta_V}{\vartheta_{\text{SS}}} = 1 + F_Z \qquad (92.24)$$

F_Z ist die *zusätzliche Rauschzahl*, die das (zusätzliche) Rauschen des Vierpols oder Verstärkers angibt; $F_Z = 0$ heißt, dass er rauschfrei ist.

Rauschmaß $a_F = 10 \lg F$ in dB (92.25)

Rauschabstand $a_r = 10 \lg \frac{P_S}{P_N}$ in dB (92.26)

P_S Signalleistung (engl. *s*ignal); P_N Rauschleistung (engl. *n*oise).

Sind zwei Verstärkerstufen hintereinandergeschaltet, wie in Abb. 92.10 gezeigt, ergibt sich als Gesamtrauschzahl:

Gesamtrauschzahl zweistufiger Verstärker

$$F = F_1 + \frac{F_2 - 1}{G_1} \qquad (92.27)$$

F_1 Rauschzahl des Verstärkers 1; F_2 Rauschzahl des Verstärkers 2; G_1 Leistungsverstärkung des Verstärkers 1.

Die erste Verstärkerstufe sollte deshalb ein möglichst niedriges Rauschen (F_1) und eine hohe Verstärkung (G_1) besitzen.

92.4.11.2 Kenngrößen von stochastischen Signalen

a) Mittelwerte

Es werden Definitionen wie in der Messtechnik verwendet, die auf die Anwendungen in der Nachrichtentechnik zugeschnitten sind. Deshalb wird das Integrationsintervall häufig von $-T$ bis $+T$ gewählt, sodass sich der Faktor $1/2$ vor dem Integral ergibt.

Arithmetischer Mittelwert

$$\bar{s} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s(t)dt$$

Quadratischer Mittelwert

$$\bar{s}^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \{s(t)\}^2 dt$$

Effektivwert

$$S_{\text{eff}} = \sqrt{\bar{s}^2}$$

Streuung

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \{s(t) - \bar{s}\}^2 dt$$

Standardabweichung

$$\sigma = \sqrt{\sigma^2} \qquad (92.28)$$

In der Praxis kann der Grenzübergang $T \to \infty$ nicht verwirklicht werden. Es hat sich aber gezeigt, dass der jeweilige Mittelwert schon nach kurzer Zeit (Erfahrungswert) erreicht wird und sich bei weiterer Erhöhung von T nicht mehr wesentlich ändert.

b) Korrelationsfunktionen

Korrelation heißt Ähnlichkeit bzw. Übereinstimmung, auch Verknüpfung.

Autokorrelationsfunktion (AKF):

$$\Phi_{\text{SS}}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s(t) \cdot s(t-\tau)dt \qquad (92.29)$$

Anschaulich beschreibt sie die „innere Ähnlichkeit" eines Signals *mit sich selbst*, wenn man beide um die Zeit t gegeneinander verschiebt.

Beispiel 8

a) Die Autokorrelationsfunktion für eine Gleichspannung U_0 berechnet sich zu: $\Phi_{SS}(t) = U_0^2 \neq f(t)$. Die Ähnlichkeit ist stets gleich groß und maximal.
b) Für einen sinusförmigen Verlauf $s(t) = \hat{s}(\cos\omega t + \varphi)$ folgt: $\Phi_{\text{SS}}(\tau) = \frac{\hat{s}^2}{2}\cos\omega\tau$. Wenn $t = 0$ ist, stimmen beide Signale überein, und die Ähnlichkeit ist maximal. Für $\omega t = \pm 90°$ bzw. $\pm 270°$ weichen beide Signale maximal voneinander ab, die Ähnlichkeit ist null.
c) Das Signal ist Rauschen. Da Rauschen ein statistischer Vorgang ist, besteht keinerlei Ähnlichkeit zwischen benachbarten Augenblickswerten, und damit ergibt sich: $\Phi_{\text{SS}}(t \neq 0) = 0$. ◀

Kreuzkorrelationsfunktion (KKF):

$$\Phi_{s_1 s_2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s_1(t) \cdot s_2(\tau - t)dt \qquad (92.30)$$

Sie gibt die Ähnlichkeit *zweier unterschiedlicher* Signale an.

c) Spektrale Leistungsdichte

Sie ist die Leistung P als Funktion der Frequenz, bezogen auf Δf:

Spektrale Leistungsdichte

$$\Phi_S(f) = \lim_{\Delta f \to 0} \frac{\Delta P}{\Delta f} \tag{92.31}$$

Theorem von Wiener-Khintchine: *Die spektrale Leistungsdichte ist die Fouriertransformierte der Autokorrelationsfunktion.*

92.4.11.3 Anwendungen der Kenngrößen von stochastischen Signalen

Die im vorigen Kapitel erläuterten Kenngrößen finden Anwendung in der Korrelations-Mess- und -Übertragungstechnik. Das Prinzip soll an zwei Beispielen gezeigt werden.

a) Die Bilder von Raumsonden sind beim Empfang auf der Erdoberfläche sehr stark durch Rauschen gestört, wobei die Amplituden des Nutzsignales nur ein Bruchteil derjenigen des Störsignales sind. Man veranlasst deshalb die Raumsonde, dasselbe Bild viele tausendmal zu senden. Mit jedem neu eintreffenden Bild wird die Autokorrelationsfunktion, bestehend aus der Überlagerung von Nutz- und Rauschsignal, mit allen vorhergehenden Bildern gebildet. Das Nutzsignal ist durch das wiederholte Senden ein quasi periodisches Signal, dessen Autokorrelationsfunktion ein Maximum bei $\tau = 0$ hat ($\tau = 0$ gilt bei periodischen Signalen mit der Periodendauer T auch für $T, 2T, 3T, \ldots$). Die Autokorrelationsfunktion des Rauschens geht nach Beispiel 8 gegen null, sodass nach einer ausreichenden Anzahl von Bildwiederholungen das Nutzsignal nahezu ungestört übrig bleibt. (Siehe dazu Abschn. 108.3).

b) Das Rauschen eines Verstärkers kann in der Praxis dadurch verringert werden, dass das Nutzsignal in zwei oder mehr getrennten rauschenden Verstärkern parallel verstärkt und von den Ausgangssignalen die Kreuzkorrelationsfunktion gebildet wird. Das Rauschen der Verstärker ist nicht korreliert, sodass es sich teilweise heraushebt. Übrig bleibt das korrelierte Nutzsignal, das eine maximale Übereinstimmung für $\tau = 0$ aufweist.

92.4.12 Verzerrungen

Der (*komplexe*) *Frequenzgang* oder *der* (*komplexe*) *Übertragungsfaktor* eines Vierpols oder eines Verstärkers ist das Verhältnis einer beliebigen Ausgangsgröße zu einer beliebigen Eingangsgröße. In der Praxis wird es sich meist um gleichartige Größen handeln, z. B. Spannungen. Der *komplexe Frequenzgang* wird mit $\underline{H}(j\omega)$ bezeichnet und kann in Form von Betrag und Phase oder als Summe von Realteil $R(\omega)$ und Imaginärteil $X(\omega)$ dargestellt werden:

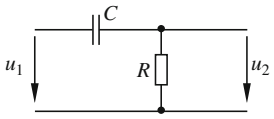


Abb. 92.11 RC-Hochpass

(*Komplexer*) *Frequenzgang oder* (*komplexer*) *Übertragungsfaktor*:

$$\begin{aligned}\underline{H}(j\omega) &= |\underline{H}(j\omega)| \cdot e^{jb(\omega)} = C \cdot e^{a(\omega)} \cdot e^{jb(\omega)} \\ &= R(\omega) + jX(\omega)\end{aligned} \tag{92.32}$$

a Dämpfungsmaß; b Phasenmaß; C Konstante, falls Eingangs- und Ausgangsgröße unterschiedliche Dimension besitzen, $R(\omega)$ Realteil, $X(\omega)$ Imaginärteil.

a wird als *Dämpfungsmaß*, b als *Phasenmaß* bezeichnet;

$$a = \lg D = \lg \frac{P_1}{P_2}$$

D Dämpfungsfaktor; P_1 Leistung am Eingang; P_2 Leistung am Ausgang; siehe auch Abschn. 92.5 „Kenngrößen der Übertragungsstrecke“.

Hängen a bzw. b von der Frequenz ab, spricht man von *Dämpfungsverzerrungen* bzw. *Phasenverzerrungen*.

Dämpfungsverzerrungen	$a = f(\omega)$
Phasenverzerrungen	$b = f(\omega)$

Bei Vierpolen verwendet man auch den aus der Regelungstechnik stammenden Begriff „*Übertragungsfunktion*“ für das Verhältnis von Ausgangs- zu Eingangsspannung bei Leerlauf am Ausgang.

$$\textit{Übertragungsfunktion} \quad \underline{H}(j\omega) = \frac{\underline{U}_2}{\underline{U}_1} \tag{92.33}$$

bei Vierpolen, Leerlauf am Ausgang.

$\underline{U}_1$ Eingangsspannung, $\underline{U}_2$ Ausgangsspannung

Beispiel 9

Ein RC-Hochpass nach Abb. 92.11 hat den Übertragungsfaktor (Übertragungsfunktion) nach (92.32) bzw. (92.33):

$$\begin{aligned}\underline{H}(j\omega) = \frac{\underline{U}_2}{\underline{U}_1} &= \frac{j\omega RC}{1 + j\omega RC} = \sqrt{\frac{(\omega RC)^2}{1 + (\omega RC)^2}} \cdot e^{j \arctan \frac{1}{\omega RC}} \\ &= \frac{(\omega RC)^2}{1 + (\omega RC)^2} + j\frac{\omega RC}{1 + (\omega RC)^2}\end{aligned}$$

Ein nicht sinusförmiges Signal am Eingang hat am Ausgang eine Phasenverzerrung, da $b(\omega) = \arctan(1/\omega RC) \neq \text{const}$. Ebenso treten Dämpfungsverzerrungen auf, da $a = f(\omega)$. ◄

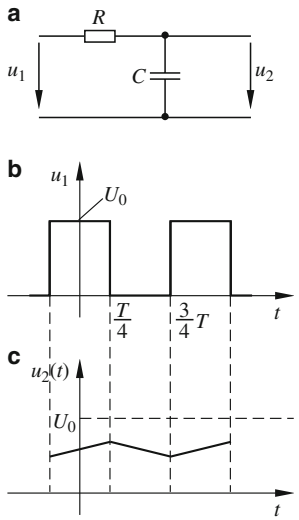


Abb. 92.12 Lineare Verzerrungen. **a** RC-Tiefpass, **b** Eingangssignal, **c** linear verzerrtes Ausgangssignal

92.4.12.1 Lineare Verzerrungen

Von *linearen Verzerrungen* spricht man, wenn Eingangs- und Ausgangssignal eines Vierpols oder Verstärkers *nicht* den gleichen zeitlichen Verlauf haben, aber keine *zusätzlichen* Frequenzen im Ausgangssignal gegenüber dem Eingangssignal enthalten sind (Fourierzerlegung beider Signale). Sie entstehen häufig dadurch, dass der Vierpol oder Verstärker Tiefpassverhalten besitzt und damit den Frequenzbereich einschränkt (siehe Beispiel 10). Diese Verzerrungen sind prinzipiell korrigierbar.

Beispiel 10

Ein periodisches Rechtecksignal durchläuft einen RC-Tiefpass nach Abb. 92.12a. Der Frequenzgang ergibt sich bei unbelastetem Ausgang zu: $\underline{H}(j\omega) = \frac{\underline{U}_2}{\underline{U}_1} = \frac{1}{1+j\omega RC}$. Die Oberschwingungen werden mit wachsender Frequenz zunehmend gedämpft, sodass das Signal am Ausgang in seinem zeitlichen Verlauf verändert ist (Abb. 92.12c). Durch Reihenschaltung eines weiteren Vierpols mit dem inversen Frequenzgang $\underline{H}_{\text{inv}} = (\frac{1}{1+j\omega RC})^{-1} = (1+j\omega RC)$ kann das verzerrte Signal wieder in den Originalzustand versetzt werden. ◀

92.4.12.2 Nichtlineare Verzerrungen

Sie entstehen an gekrümmten Kennlinien, an denen der Quotient aus Spannung und Strom nicht mehr konstant ist, sondern von der Spannung bzw. dem Strom abhängt (z. B. bei Dioden, spannungsabhängigen Widerständen). Es entstehen *zusätzliche neue* Frequenzen im verzerrten Ausgangssignal, wie sie sich z. B. ergeben, wenn man als Eingangssignal eine sinusförmige Größe wählt und für die verzerrte Ausgangsgröße die Fourierzerlegung durchführt. Abb. 92.13 zeigt das Entstehen nichtlinearer Verzerrungen an einer gekrümmten Kennlinie.

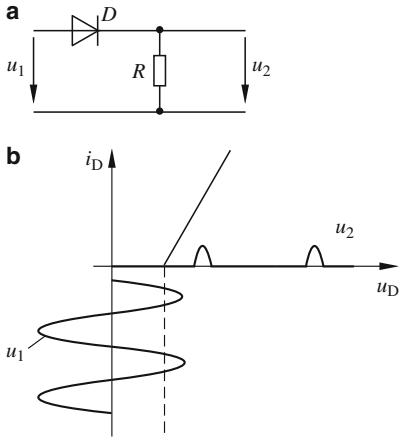


Abb. 92.13 Entstehung von nichtlinearen Verzerrungen an einer idealisierten Diodenkennlinie

92.4.12.3 Klirrfaktor

Der Klirrfaktor dient zur Beurteilung von nichtlinearen Verzerrungen. Zur Bestimmung wird folgendes Verfahren eingesetzt: Der Prüfling (z. B. Verstärker) erhält ein sinusförmiges Eingangssignal mit der Frequenz f_0 (bei NF-Verstärkern in der Regel 1 kHz). Aus dem Ausgangssignal wird die Grundschwingung mit der Frequenz f_0 ausgefiltert. Ein eventuell verbleibendes Restsignal ist durch nichtlineare Verzerrungen des Verstärkers entstanden, dessen Effektivwert gleich dem Effektivwert aller Oberschwingungen ist. Dieser Wert ist um so größer, je mehr das Ausgangssignal von der Sinusform abweicht. Er heißt *Klirrfaktor* und ist wie folgt definiert (DIN 40110 und 40148):

Klirrfaktor

$$k = \sqrt{\frac{U_2^2 + U_3^2 + U_4^2 + \ldots}{U_1^2 + U_2^2 + U_3^2 + \ldots}} = \sqrt{\frac{\sum_{i=2}^{\infty} U_i^2}{\sum_{i=1}^{\infty} U_i^2}} \qquad (92.34)$$

U_1 Effektivwert der Grundschwingung oder ersten Harmonischen; U_2 Effektivwert der ersten Oberschwingung oder zweiten Harmonischen usw.

Er wird meist in % angegeben. Daneben gibt es noch die folgende Darstellung:

Klirrfaktor

$$k' = \frac{\sqrt{U_2^2 + U_3^2 + U_4^2 + \ldots}}{U_1} \qquad (92.35)$$

Der Zusammenhang zwischen k' und k ist gegeben durch:

$$k = \frac{k'}{\sqrt{1 + k'^2}}$$

Je kleiner der Klirrfaktor ist, um so weniger unterscheiden sich k' und k.

Für besondere Anwendungen werden *Teilklirrfaktoren* verwendet:

Teilklirrfaktoren:

$$k_2 \approx k_2' = \frac{U_2}{U_1}; \quad k_3 \approx k_3' = \frac{U_3}{U_1}; \quad \ldots \tag{92.36}$$

Das „≈"-Zeichen gilt, wenn die Teilklirrfaktoren viel kleiner als 1 sind.

$$\textit{Klirrdämpfung} \quad a_k = 20\,\lg\frac{1}{k} \tag{92.37}$$

Beispiel 11

Für den in Abb. 92.5 dargestellten rechteckförmigen Signalverlauf gibt (92.10) die Fourierreihe an. Setzt man die Effektivwerte in (92.34) ein, folgt für den Klirrfaktor:

$$k = \sqrt{\frac{(\frac{2U_0}{\pi\sqrt{2}})^2[(\frac{1}{3})^2 + (\frac{1}{5})^2 + (\frac{1}{7})^2 + \ldots]}{(\frac{2U_0}{\pi\sqrt{2}})^2[(\frac{1}{1})^2 + (\frac{1}{3})^2 + (\frac{1}{5})^2 + (\frac{1}{7})^2 + \ldots]}}$$

$$= \sqrt{\frac{[(\frac{1}{1})^2 + (\frac{1}{3})^2 + (\frac{1}{5})^2 + (\frac{1}{7})^2 + \ldots] - 1}{[(\frac{1}{1})^2 + (\frac{1}{3})^2 + (\frac{1}{5})^2 + (\frac{1}{7})^2 + \ldots]}}$$

Der Ausdruck $1/\sqrt{2}$ im Zähler und im Nenner ist wegen des Überganges von den Scheitel- zu den Effektivwerten erforderlich.

Der Faktor $2U_0/\pi\sqrt{2}$ lässt sich im Zähler und im Nenner als gemeinsamer Faktor ausklammern und kürzen. Im Zähler wurde dann in der eckigen Klammer zunächst der Wert $(1/1)^2$ addiert und anschließend wieder abgezogen. Für die eckige Klammer im Zähler und im Nenner gilt:

$$\left[\left(\frac{1}{1}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{7}\right)^2 + \ldots\right] = \frac{\pi^2}{8}$$

Damit erhält man:

$$k = \sqrt{\frac{\frac{\pi^2}{8} - 1}{\frac{\pi^2}{8}}} = 0{,}435 \quad \text{oder} \quad 43{,}5\,\% \qquad \blacktriangleleft$$

92.5 Kenngrößen der Übertragungsstrecke

Unter „Übertragungsstrecke" werden hier bevorzugt entweder einzelne Vierpole, Verstärker und elektrische Leitungen oder aber Kombinationen aus ihnen verstanden (Hinweis 1: Natürlich lassen sich auch Verstärker und Leitungen unter dem Begriff „Vierpol" einordnen. Hinweis 2: In Abb. 92.14 ist ein komplexer Vierpol als Kettenschaltung von Leitung 1 → Verstärker 1 → Leitung 2 → Verstärker 2 → Leitung 3 → Tiefpassfilter → Leitung 4 dargestellt). Um die Verknüpfung von Eingangs- und Ausgangsgrößen sowohl von Einzelvierpolen als auch von Gesamtvierpolen zu erfassen, wurden die folgenden Kenngrößen definiert (Indizes: 1 Eingang, 2 Ausgang):

92.5.1 Dämpfungsfaktor

$$\textit{Leistungsdämpfungsfaktor} \qquad D_P = \frac{P_1}{P_2} \tag{92.38a}$$

$$\textit{Spannungsdämpfungsfaktor} \qquad D_U = \frac{U_1}{U_2} \tag{92.38b}$$

$$\textit{Stromdämpfungsfaktor} \qquad D_I = \frac{I_1}{I_2} \tag{92.38c}$$

Die Dämpfungsfaktoren sind ≥ 1, wenn die Größen am Ausgang höchstens gleich denen am Eingang sind. Ein Dämpfungsfaktor $D < 1$ bedeutet damit eine Verstärkung. Da man bei einer Verstärkung aber von der Vorstellung eines Zahlenwertes > 1 ausgeht, hat man *Übertragungsfaktoren* definiert:

92.5.2 Übertragungsfaktor

Er wird auch mit *Verstärkungsfaktor* bezeichnet.

$$\textit{Leistungsübertragungsfaktor} \qquad T_P = \frac{P_2}{P_1} \tag{92.39a}$$

$$\textit{Spannungsübertragungsfaktor} \qquad T_U = \frac{U_2}{U_1} \tag{92.39b}$$

$$\textit{Stromübertragungsfaktor} \qquad T_I = \frac{I_2}{I_1} \tag{92.39c}$$

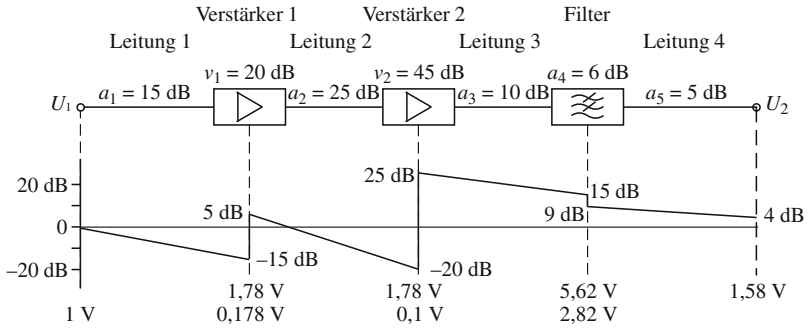


Abb. 92.14 Pegeldiagramm

Da in der Nachrichtentechnik Dämpfungs- und Übertragungsfaktoren (Verstärkungsfaktoren) im Bereich von mehreren Zehnerpotenzen vorkommen und in einer Übertragungskette mehrere dieser Werte miteinander multipliziert werden müssen, hat man logarithmische Maße definiert, bei denen die Multiplikation in eine Addition übergeht. Als Basis dient die Zahl 10 (verwendet bei Vierpolen und Verstärkern) oder die Zahl $e = 2{,}718\ldots$ (verwendet bei elektrischen Leitungen).

92.5.3 Dämpfungsmaß

Es wird in Dezibel (dB) angegeben. *Leistungsdämpfungsmaß*

$$a_P = 10\,\lg\frac{P_1}{P_2} \quad \text{in dB} \tag{92.40a}$$

Spannungsdämpfungsmaß

$$a_U = 10\cdot\lg\frac{U_1^2}{Z_1}\frac{Z_2}{U_2^2} \quad \text{in dB}$$

$$a_U = 10\cdot\lg\frac{U_1^2}{U_2^2} = 20\cdot\lg\frac{U_1}{U_2} \quad \text{in dB} \quad \text{für } Z_1 = Z_2 \tag{92.40b}$$

Stromdämpfungsmaß

$$a_I = 10\cdot\lg\frac{I_1^2 Z_1}{I_2^2 Z_2} \quad \text{in dB}$$

$$a_I = 10\cdot\lg\frac{I_1^2}{I_2^2} = 20\cdot\lg\frac{I_1}{I_2} \quad \text{in dB} \quad \text{für } Z_1 = Z_2 \tag{92.40c}$$

Z_1 Eingangswiderstand des Vierpols, reell, Z_2 Lastwiderstand am Ausgang des Vierpols, reell.

Die Bedingung $Z_1 = Z_2$ spielt bei der Leistungsanpassung (Verstärker) und der Leitungsanpassung (elektrische Leitungen) eine Rolle. Mit der Beziehung $\lg[x^2] = 2\cdot\lg[x]$ ergeben sich dann die dargestellten Formeln. Das Dämpfungsmaß ist positiv, wenn die Ausgangsgröße kleiner ist als die Eingangsgröße. Die Übertragungsmaße (Verstärkungsmaße) ergeben sich analog zu (92.40a), (92.40b) und (92.40c).

92.5.4 Übertragungsmaß, Verstärkungsmaß

Beide Bezeichnungen werden verwendet, und der Wert wird in Dezibel (dB) angegeben. Sie unterscheiden sich vom Dämpfungsmaß durch das Vorzeichen. Das Verstärkungsmaß ist positiv, wenn die Ausgangsgröße größer ist als die Eingangsgröße. Ein positives Verstärkungsmaß ist damit ein negatives Dämpfungsmaß und umgekehrt.

Leistungsübertragungsmaß

$$v_P = -a_P = 10\cdot\lg\frac{P_2}{P_1} \quad \text{in dB} \tag{92.41a}$$

Spannungsübertragungsmaß

$$v_U = -a_U = 10\cdot\lg\frac{U_2^2}{Z_2}\frac{Z_1}{U_1^2} \quad \text{in dB}$$

$$v_U = -a_U = 20\cdot\lg\frac{U_2}{U_1} \quad \text{in dB} \quad \text{für } Z_1 = Z_2 \tag{92.41b}$$

Stromübertragungsmaß

$$v_I = -a_I = 10\cdot\lg\frac{I_2^2}{I_1^2}\frac{Z_2}{Z_1} \quad \text{in dB}$$

$$v_I = -a_I = 20\cdot\lg\frac{I_2}{I_1} \quad \text{in dB} \quad \text{für } Z_1 = Z_2 \tag{92.41c}$$

Z_1 Eingangswiderstand des Vierpols, reell, Z_2 Lastwiderstand am Ausgang des Vierpols, reell.

Besteht eine Übertragungsstrecke aus der Hintereinanderschaltung („Kettenschaltung“) von mehreren Einzelvierpolen, folgt das Gesamtdämpfungsmaß aus der Summe der Teildämpfungsmaße:

Gesamt-Dämpfungsmaß(-Verstärkungsmaß):

$$a_{\text{ges}} = a_1 + a_2 + a_3 + \ldots \quad \text{mit } a_n = -v_n \tag{92.42}$$

Beispiel 12

Gegeben: $a_{U1} = -12\,\text{dB}$; $a_{U2} = 25\,\text{dB}$; $a_{U3} = -40\,\text{dB}$. $U_1 = 1\,\text{mV}$; $U_3 = ?\,\text{mV}$.

$a_{U\,\text{ges}} = a_{U1} + a_{U2} + a_{U3} = (-12 + 25 - 40)\,\text{dB} = -27\,\text{dB}$. Da Z_1 und Z_2 nicht gegeben sind, wird $Z_1 = Z_2$ gesetzt. $a_{U\,\text{ges}} = 20\,\lg(U_1/U_3) = -27\,\text{dB}$. $a_{U\,\text{ges}}$ ist negativ, und damit wird U_1 verstärkt. Für U_3 folgt: $U_1/U_3 = 10^{-27/20}$, $U_3 = U_1/10^{-27/20} = U_1/10^{27/20} = 1\,\text{mV}\cdot 22{,}4 = 22{,}4\,\text{mV}$. ◀

92.5.5 Pegel

Häufig interessiert das Verhältnis einer Leistung, einer Spannung oder eines Stromes zu einem beliebigen Bezugswert. Man spricht dann von einem *Pegel*. Wichtig ist, dass zu einer Pegelangabe der Bezugswert angegeben wird oder dieser Wert bekannt ist.

92.5.5.1 Absoluter Pegel

Absoluter Leistungspegel

$$L_{P\,\text{abs}} = 10\cdot\lg\frac{P_x}{1\,\text{mW}} \quad \text{in dBm} \tag{92.43a}$$

Absoluter Spannungspegel

$$L_{U\,\text{abs}} = 20\cdot\lg\frac{U_x}{775\,\text{mV}} \quad \text{in dB für } Z_1 = Z_2 \tag{92.43b}$$

Absoluter Strompegel

$$L_{I\,\text{abs}} = 20\cdot\lg\frac{I_x}{1{,}29\,\text{mA}} \quad \text{in dB} \quad \text{für } Z_1 = Z_2 \tag{92.43c}$$

Tab. 92.3 Beispiele für Pegelangaben

Pegelart	Definition	Bezeichnung
Absoluter Leistungspegel	$10 \cdot \lg\left(\frac{P}{1\,\mathrm{mW}}\right)$	dBm
	$10 \cdot \lg\left(\frac{P}{1\,\mathrm{W}}\right)$	dBW
Absoluter Spannungspegel	$20 \cdot \lg\left(\frac{\lvert U\rvert}{775\,\mathrm{mV}}\right)$	dB
	$20 \cdot \lg\left(\frac{\lvert U\rvert}{1\,\mathrm{V}}\right)$	dBV
Relativer Pegel	$10 \cdot \lg\left(\frac{P}{P_0}\right)$	dBr
Absoluter Pegel in Antennenschaltungen	$20 \cdot \lg\left(\frac{\lvert U\rvert}{1\,\mu\mathrm{V}}\right)$	dB(μV)
Absoluter Pegel, reduziert auf 0-dBr-Punkt	Es gilt: $\frac{P}{\mathrm{dBm}} = \frac{P}{\mathrm{dBm0}} + \frac{P}{\mathrm{dBr}}$	dBm0

Die Bezugswerte (Zusatz „m“ bei der Leistungsangabe in dB) sind vom Telefon abgeleitet, bei dem an der Hörkapsel eine mittlere elektrische Leistung von 1 mW an 600 Ω (Widerstand des Telefons und Leitungs- oder Wellenwiderstand der Telefonleitung) zur Verständigung ausreicht.

92.5.5.2 Relativer Pegel

Relativer Leistungspegel

$$L_P = 10 \cdot \lg \frac{P_x}{P_0} \quad \text{in dBr} \tag{92.44a}$$

Relativer Spannungspegel

$$L_U = 20 \cdot \lg \frac{U_x}{U_0} \quad \text{in dBr} \quad \text{mit } Z_1 = Z_2 \tag{92.44b}$$

Relativer Strompegel

$$L_I = 20 \cdot \lg \frac{I_x}{I_0} \quad \text{in dBr} \quad \text{mit } Z_1 = Z_2 \tag{92.44c}$$

P_0, U_0, I_0 sind geeignet gewählte Bezugsgrößen.

In der Antennen-Messtechnik wird beispielsweise $U_0 = 1\,\mu\mathrm{V}$ gewählt, weil bei Antennenspannungen in dieser Größenordnung Mono-Rundfunkempfang möglich wird. Die Angabe erfolgt dann in dB (μV).

In der Tab. 92.3 sind einige Pegelangaben aufgelistet.

92.5.5.3 Dämpfungsmaß, Übertragungsmaß

Sie sind folgendermaßen definiert:

- Dämpfungsmaß (Pegel) $a = L_1 - L_2$ in dB, positiv,
- Übertragungsmaß (Pegel) $a = L_1 - L_2$ in dB, negativ.

Es gilt:

$$\begin{aligned} &L_1 \text{ Pegel am Punkt 1 in dB } (L_P \text{ oder } L_U \text{ oder } L_I) \\ &L_2 \text{ Pegel am Punkt 2 in dB } (L_P \text{ oder } L_U \text{ oder } L_I) \end{aligned} \tag{92.45}$$

Ein negatives Dämpfungsmaß ist damit ein Übertragungsmaß und bedeutet, dass die Ausgangsgröße größer ist als die Eingangsgröße.

92.5.5.4 Pegeldiagramm

Um bei komplexen Übertragungsstrecken einen schnellen Überblick über die an jedem Ort herrschenden Pegel bzw. Absolutwerte zu erhalten, verwendet man das Pegeldiagramm. Abb. 92.14 zeigt ein Beispiel.

Zu jedem Ort ist der vorhandene Pegel (hier Spannungspegel) sofort angebbar. Durch die willkürliche aber praktische Festlegung der Eingangsgröße U_1 auf einen Pegel von 0 dB kann man außerdem erkennen, ob der Pegel irgendwo einen kritischen Wert unterschreitet, sodass ein Mindest-Störabstand nicht mehr gewährleistet ist.

Zweitore, Vierpole

93

Wilfried Plaßmann



93.1 Zweitor allgemein 1148
93.2 Spezielle Zweitore 1155
93.3 Wellenparameter passiver Zweitore 1157

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_93

93.1 Zweitor allgemein

93.1.1 Grundlagen

Zweitore (nach DIN 4899; alte Bezeichnung „Vierpole“) sind Funktionseinheiten mit zwei Toren, wobei zu jedem Tor ein Klemmenpaar gehört. An eines der Tore wird das Eingangssignal angeschlossen, die zugehörigen Anschlüsse werden mit Eingangsklemmen bezeichnet. Entsprechend wird an den zwei Ausgangsklemmen das Ausgangssignal abgegriffen. Eventuell erforderliche Anschlüsse für eine Stromversorgung werden nicht mitgezählt. Der interne Aufbau besteht aus einer beliebigen Zusammenschaltung von Widerständen, Kondensatoren, Spulen, Transformatoren, Dioden, Transistoren usw. Es wird vorausgesetzt, dass die Zweitore zeitinvariant sind, d. h. die Eigenschaften hängen nicht von der Zeit ab.

Zweitore werden in folgender Weise charakterisiert:

Passive Zweitore: Die am Ausgang entnehmbare Leistung ist kleiner oder gleich der am Eingang zugeführten Leistung.

Aktive Zweitore: Die am Ausgang entnehmbare Leistung ist größer als die am Eingang zugeführte Leistung.

Lineare Zweitore: Innerhalb des Zweitors gilt, dass in jedem Zweig das Verhältnis von Spannung zu Strom nicht vom Strom bzw. der Spannung abhängt. Enthält ein Zweitor nichtlineare Bauelemente wie Dioden, Transistoren oder spannungsabhängige Widerstände, werden deren Kennlinien häufig linearisiert. Dadurch entsteht näherungsweise ein lineares Zweitor.

Nichtlineare Zweitore: Die nichtlinearen Kennlinien von z. B. Dioden, Transistoren, spannungsabhängigen Widerständen und Spulen mit Eisen werden nicht linearisiert. Damit ist die funktionale Abhängigkeit zwischen den Eingangs- und Ausgangsgrößen des Zweitors nicht so einfach darstellbar wie bei linearen Zweitoren.

Die folgenden Ausführungen setzen lineare, zeitinvariante Zweitore voraus. Die elektrische Leitung ist ein Zweitor besonderer Art. Sie wird deshalb im nächsten Kapitel gesondert behandelt.

Abb. 93.1 zeigt die für die Ein- und Ausgangsgrößen verwendeten Bezeichnungen (DIN 1344, 40148). Das Zweitor nach Abb. 93.1b besitzt die Besonderheit, dass ein Eingangs- und ein Ausgangsanschluss miteinander verbunden sind. Zweitore dieser Art werden bevorzugt verwendet, da die durchgehende Verbindung häufig als Anschlusspunkt für Abschirmungen, für das Gehäusepotential und den Netz-Schutzleiter dient. Auf diese Weise lassen sich Störungen in Signalkreisen klein halten. Die Richtung des Stromes $\underline{I}_2$ wird in der Regel positiv genommen, wenn er in das Zweitor hineinfließt. Diese Betrachtungsweise ist z. B. für Zweitore in Telefon-Signalkreisen sinnvoll, da Generator (Sprecher) und Verbraucher (Hörer) ihre Rollen tauschen und damit der Signalfluss in beiden Richtungen erfolgt.

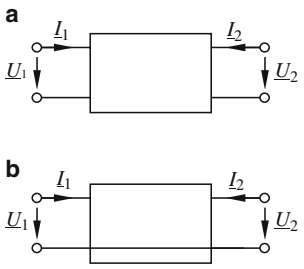


Abb. 93.1 Zweitor. **a** Allgemein, **b** gemeinsamer Anschluss für Ein- und Ausgang

Die im folgenden dargestellten Beziehungen für Zweitore sind wesentliche Ergebnisse der Zweitortheorie und sind aus zwei Gründen von besonderer Bedeutung:

1. Umfangreiche Zweitore lassen sich auf wenige sogenannte „Elementarzweitore“ zurückführen bzw. als Kombination aus diesen darstellen. Die Parameter dieser Elementarzweitore sind in Tab. 93.4 zusammengestellt.
2. Mehrere Zweitore können durch geeignet gewählte Parameter der Einzelzweitore auf einfache Weise zu einem Gesamtzweitor zusammengefügt werden (Tab. 93.1 und Abb. 93.2).

93.1.2 Zweitorgleichungen, Zusammenschaltung von Zweitoren

Das Zweitor wird in seinen Eigenschaften vollständig durch *zwei* Gleichungen beschrieben, die die vier Größen $\underline{U}_1$, $\underline{U}_2$, $\underline{I}_1$ und $\underline{I}_2$ nach Abb. 93.1 miteinander verknüpfen. Je nach Art dieser Verknüpfung ergeben sich unterschiedliche Parameter (Faktoren bei den Spannungen und Strömen, siehe Tab. 93.1, linke Spalte). Die Bedeutung dieser unterschiedlichen Parameter liegt darin, dass eine bestimmte von den 5 möglichen Arten der Zusammenschaltung zweier Einzelzweitore nach Abb. 93.2a–e nur mit den zugeordneten Parametern erfolgen kann. Die mittlere Spalte von Tab. 93.1 listet die Arten der Zusammenschaltung auf, während in der rechten Spalte die erforderlichen Parameter stehen und welche Operationen für eine Zusammenschaltung erforderlich sind. Liegen die Parameter in einer anderen als der geforderten Form vor, müssen sie anhand von Tab. 93.2 umgewandelt werden.

Hinweise zu Tab. 93.1:

1. Bei den $\underline{H}$-Parametern ist alternativ auch die Schreibweise mit kleinen Buchstaben zulässig, um die Verbindung zu den bei Transistoren verwendeten $\underline{h}$-Parametern herzustellen. So wird beispielsweise die Stromverstärkung nicht mehr mit B bzw. b, sondern mit h_{21} bezeichnet.
2. Bei der Reihen-Parallel- und der Parallel-Reihen-Schaltung (Abb. 93.2c, d) muss das zweite Zweitor zwischen Ein- und

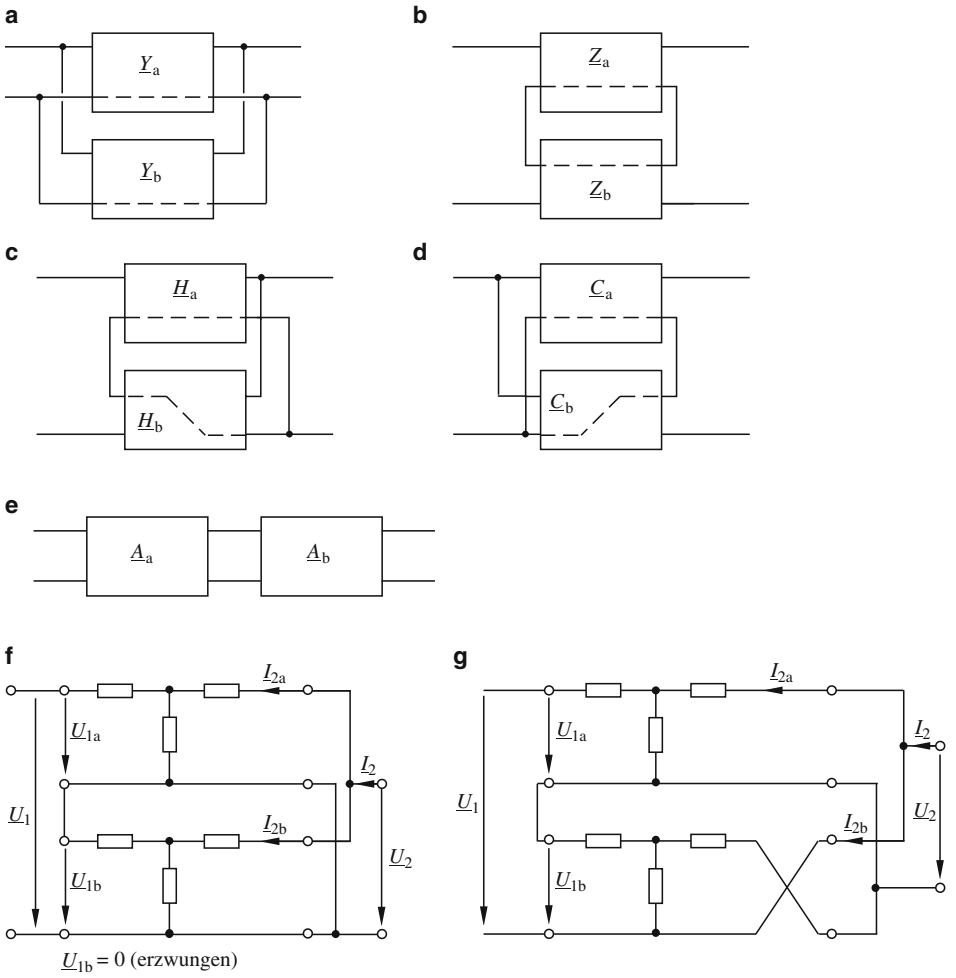


Abb. 93.2 Zusammenschaltung von Zweitoren. **a** Parallel-Parallel-Schaltung, **b** Reihen-Reihen-Schaltung, **c** Reihen-Parallel-Schaltung, **d** Parallel-Reihen-Schaltung, **e** Kettenschaltung, **f** Reihen-Parallel-Schaltung (unteres Zweitor *nicht* gedreht), **g** Reihen-Parallel-Schaltung (unteres Zweitor gedreht)

Ausgang in der dargestellten Weise „gedreht" werden. Prinzipiell kann die Zusammenschaltung zweier Zweitore nach Abb. 93.2c oder Abb. 93.2d auf 4 Arten erfolgen. Abb. 93.2f zeigt am Beispiel der Reihen-Parallel-Schaltung, wie bei einer von diesen vier Möglichkeiten der Zusammenschaltung $\underline{U}_{1b} = 0$ erzwungen wird.

Auch zwei weitere, hier nicht dargestellte Möglichkeiten, führen dazu, dass eine Spannung zu Null erzwungen wird. Nur die „gedrehte" Version von Zweitor 2 nach Abb. 93.2g erfüllt die Bedingungen $\underline{U}_1 = \underline{U}_{1a} + \underline{U}_{1b}$ für die Reihenschaltung am Eingang und $\underline{I}_2 = \underline{I}_{2a} + \underline{I}_{2b}$ für die Parallelschaltung am Ausgang. In analoger Weise lässt sich die Drehung für die Parallel-Reihen-Schaltung nach Abb. 93.2d begründen.

3. Bei der Kettenschaltung zweier Zweitore müssen die $\underline{A}$-Parameter des links angeordneten Zweitors auch bei der Matrizenmultiplikation links stehen.

Da zwei Gleichungen unabhängig von der Art der Parameter das Zweitor vollständig und eindeutig beschreiben, können die Parameter mit Tab. 93.2 ineinander umgerechnet werden.

93.1.3 Bestimmung der Zweitorparameter

Die Bestimmung der Zweitorparameter wird am Beispiel von $\underline{Z}_{11}$ gezeigt. In der Gleichung $\underline{U}_1 = \underline{Z}_{11} \cdot \underline{I}_1 + \underline{Z}_{12} \cdot \underline{I}_2$ wird $\underline{I}_2$ durch Leerlauf am Ausgang zu Null erzwungen. Korrekterweise spricht man bei $\underline{Z}_{11}$ von der Leerlauf-Eingangsimpedanz. Dann ergibt sich $\underline{Z}_{11}$ aus dem Verhältnis $\underline{U}_1/\underline{I}_1$. Die Bestimmung von $\underline{Z}_{12}$ wird über $\underline{I}_1 = 0$ erreicht. Diese Impedanz heißt Leerlauf-Übertragungsimpedanz. Obwohl $\underline{Z}_{11}$ ohne Berücksichtigung von $\underline{Z}_{12}$ ($\underline{I}_2 = 0$ gesetzt) und $\underline{Z}_{12}$ ohne Berücksichtigung von $\underline{Z}_{11}$ ($\underline{I}_1 = 0$ gesetzt) bestimmt worden ist, erfüllen sie die Gesamtgleichung $\underline{U}_1 = \underline{Z}_{11} \cdot \underline{I}_1 + \underline{Z}_{12} \cdot \underline{I}_2$. Das liegt an der Gültigkeit des Überlagerungssatzes. (Hinweis: Eine Gleichung mit 2 Unbekannten lässt sich so *nicht* lösen.)

Allgemein gilt: $\underline{I}_2 = 0$ wird durch Leerlauf am Ausgang, $\underline{I}_1 = 0$ durch Leerlauf am Eingang erzwungen. Analog wird $\underline{U}_2 = 0$ durch Kurzschluss am Ausgang, $\underline{U}_1 = 0$ durch Kurzschluss am Eingang erreicht. In Tab. 93.3 sind alle Parameter dargestellt. Da Zweitore im allgemeinen auch Induktivitäten und Kapa-

Tab. 93.1 Zweitorgleichungen und Zusammenschaltung von zwei Zweitoren (siehe Hinweise im Text)

Name der Parameter, Zweitorgleichungen	Art der Zusammenschaltung zweier Zweitore (siehe Abb. 93.2)	Zugehörige Parameteroperationen; Zweitor 1: Index a, Zweitor 2: Index b
Zweitorgleichungen in Leitwertform: $\underline{Y}$-Parameter $\underline{I}_1 = \underline{Y}_{11} \cdot \underline{U}_1 + \underline{Y}_{12} \cdot \underline{U}_2$ $\underline{I}_2 = \underline{Y}_{21} \cdot \underline{U}_1 + \underline{Y}_{22} \cdot \underline{U}_2$	Eingänge parallel, Ausgänge parallel (Abb. 93.2a)	$\underline{Y}_{11} = \underline{Y}_{11a} + \underline{Y}_{11b}$ $\underline{Y}_{12} = \underline{Y}_{12a} + \underline{Y}_{12b}$ $\underline{Y}_{21} = \underline{Y}_{21a} + \underline{Y}_{21b}$ $\underline{Y}_{22} = \underline{Y}_{22a} + \underline{Y}_{22b}$
Zweitorgleichungen in Impedanzform: $\underline{Z}$-Parameter $\underline{U}_1 = \underline{Z}_{11} \cdot \underline{I}_1 + \underline{Z}_{12} \cdot \underline{I}_2$ $\underline{U}_2 = \underline{Z}_{21} \cdot \underline{I}_1 + \underline{Z}_{22} \cdot \underline{I}_2$	Eingänge in Reihe, Ausgänge in Reihe (Abb. 93.2b)	$\underline{Z}_{11} = \underline{Z}_{11a} + \underline{Z}_{11b}$ $\underline{Z}_{12} = \underline{Z}_{12a} + \underline{Z}_{12b}$ $\underline{Z}_{21} = \underline{Z}_{21a} + \underline{Z}_{21b}$ $\underline{Z}_{22} = \underline{Z}_{22a} + \underline{Z}_{22b}$
Zweitorgleichungen in Reihen-Parallel- oder Hybrid-Form: $\underline{H}$-Parameter $\underline{U}_1 = \underline{H}_{11} \cdot \underline{I}_1 + \underline{H}_{12} \cdot \underline{U}_2$ $\underline{I}_2 = \underline{H}_{21} \cdot \underline{I}_1 + \underline{H}_{22} \cdot \underline{U}_2$ alternative Schreibweise: $\underline{h}$-Parameter $\underline{U}_1 = \underline{h}_{11} \cdot \underline{I}_1 + \underline{h}_{12} \cdot \underline{U}_2$ $\underline{I}_2 = \underline{h}_{21} \cdot \underline{I}_1 + \underline{h}_{22} \cdot \underline{U}_2$	Eingänge in Reihe, Ausgänge parallel (Abb. 93.2c) Achtung: Zweites Zweitor muss nach Abb. 93.2c am Ausgang gedreht werden. Erläuterung siehe Abb. 93.2f	$\underline{H}_{11} = \underline{H}_{11a} + \underline{H}_{11b}$ $\underline{H}_{12} = \underline{H}_{12a} + \underline{H}_{12b}$ $\underline{H}_{21} = \underline{H}_{21a} + \underline{H}_{21b}$ $\underline{H}_{22} = \underline{H}_{22a} + \underline{H}_{22b}$ $\underline{h}_{11} = \underline{h}_{11a} + \underline{h}_{11b}$ $\underline{h}_{12} = \underline{h}_{12a} + \underline{h}_{12b}$ $\underline{h}_{21} = \underline{h}_{21a} + \underline{h}_{21b}$ $\underline{h}_{22} = \underline{h}_{22a} + \underline{h}_{22b}$
Zweitorgleichungen in Parallel-Reihen-Form: $\underline{C}$-Parameter $\underline{I}_1 = \underline{C}_{11} \cdot \underline{U}_1 + \underline{C}_{12} \cdot \underline{I}_2$ $\underline{U}_2 = \underline{C}_{21} \cdot \underline{U}_1 + \underline{C}_{22} \cdot \underline{I}_2$	Eingänge parallel, Ausgänge in Reihe (Abb. 93.2d) Achtung: Zweites Zweitor muss nach Abb. 93.2d am Ausgang gedreht werden	$\underline{C}_{11} = \underline{C}_{11a} + \underline{C}_{11b}$ $\underline{C}_{12} = \underline{C}_{12a} + \underline{C}_{12b}$ $\underline{C}_{21} = \underline{C}_{21a} + \underline{C}_{21b}$ $\underline{C}_{22} = \underline{C}_{22a} + \underline{C}_{22b}$
Zweitorgleichungen in Kettenform: $\underline{A}$-Parameter $\underline{U}_1 = \underline{A}_{11} \cdot \underline{U}_2 + \underline{A}_{12} \cdot (-\underline{I}_2)$ $\underline{I}_1 = \underline{A}_{21} \cdot \underline{U}_2 + \underline{A}_{22} \cdot (-\underline{I}_2)$	Kettenschaltung (Abb. 93.2e)	$\begin{bmatrix} \underline{A}_{11}\underline{A}_{12} \\ \underline{A}_{21}\underline{A}_{22} \end{bmatrix} = \begin{bmatrix} \underline{A}_{11a}\underline{A}_{12a} \\ \underline{A}_{21a}\underline{A}_{22a} \end{bmatrix} \cdot \begin{bmatrix} \underline{A}_{11b}\underline{A}_{12b} \\ \underline{A}_{21b}\underline{A}_{22b} \end{bmatrix}$ Matrizenmultiplikation; Index a: linkes Zweitor, Index b: rechtes Zweitor

Tab. 93.2 Umrechnung von Zweitorparametern

	$\underline{Y}$	$\underline{Z}$	$\underline{H}$	$\underline{C}$	$\underline{A}$
$\underline{Y}$	$\underline{Y}_{11}\underline{Y}_{12}$ $\underline{Y}_{21}\underline{Y}_{22}$	$\frac{\underline{Z}_{22}}{\det \underline{Z}} \frac{-\underline{Z}_{12}}{\det \underline{Z}}$ $\frac{-\underline{Z}_{21}}{\det \underline{Z}} \frac{\underline{Z}_{11}}{\det \underline{Z}}$	$\frac{1}{\underline{H}_{11}} \frac{-\underline{H}_{12}}{\underline{H}_{11}}$ $\frac{\underline{H}_{21}}{\underline{H}_{11}} \frac{\det \underline{H}}{\underline{H}_{11}}$	$\frac{\det \underline{C}}{\underline{C}_{22}} \frac{\underline{C}_{12}}{\underline{C}_{22}}$ $\frac{-\underline{C}_{21}}{\underline{C}_{22}} \frac{1}{\underline{C}_{22}}$	$\frac{\underline{A}_{22}}{\underline{A}_{12}} \frac{-\det \underline{A}}{\underline{A}_{12}}$ $\frac{-1}{\underline{A}_{12}} \frac{\underline{A}_{11}}{\underline{A}_{12}}$
$\underline{Z}$	$\frac{\underline{Y}_{22}}{\det \underline{Y}} \frac{-\underline{Y}_{12}}{\det \underline{Y}}$ $\frac{-\underline{Y}_{21}}{\det \underline{Y}} \frac{\underline{Y}_{11}}{\det \underline{Y}}$	$\underline{Z}_{11}\underline{Z}_{12}$ $\underline{Z}_{21}\underline{Z}_{22}$	$\frac{\det \underline{H}}{\underline{H}_{22}} \frac{\underline{H}_{12}}{\underline{H}_{22}}$ $\frac{-\underline{H}_{21}}{\underline{H}_{22}} \frac{1}{\underline{H}_{22}}$	$\frac{1}{\underline{C}_{11}} \frac{-\underline{C}_{12}}{\underline{C}_{11}}$ $\frac{\underline{C}_{21}}{\underline{C}_{11}} \frac{\det \underline{C}}{\underline{C}_{11}}$	$\frac{\underline{A}_{11}}{\underline{A}_{21}} \frac{\det \underline{A}}{\underline{A}_{21}}$ $\frac{1}{\underline{A}_{21}} \frac{\underline{A}_{22}}{\underline{A}_{21}}$
$\underline{H}$	$\frac{1}{\underline{Y}_{11}} \frac{-\underline{Y}_{12}}{\underline{Y}_{11}}$ $\frac{\underline{Y}_{21}}{\underline{Y}_{11}} \frac{\det \underline{Y}}{\underline{Y}_{11}}$	$\frac{\det \underline{Z}}{\underline{Z}_{22}} \frac{\underline{Z}_{12}}{\underline{Z}_{22}}$ $\frac{-\underline{Z}_{21}}{\underline{Z}_{22}} \frac{1}{\underline{Z}_{22}}$	$\underline{H}_{11}\underline{H}_{12}$ $\underline{H}_{21}\underline{H}_{22}$	$\frac{\underline{C}_{22}}{\det \underline{C}} \frac{-\underline{C}_{12}}{\det \underline{C}}$ $\frac{-\underline{C}_{21}}{\det \underline{C}} \frac{\underline{C}_{11}}{\det \underline{C}}$	$\frac{\underline{A}_{12}}{\underline{A}_{22}} \frac{\det \underline{A}}{\underline{A}_{22}}$ $\frac{-1}{\underline{A}_{22}} \frac{\underline{A}_{21}}{\underline{A}_{22}}$
$\underline{C}$	$\frac{\det \underline{Y}}{\underline{Y}_{22}} \frac{\underline{Y}_{12}}{\underline{Y}_{22}}$ $\frac{-\underline{Y}_{21}}{\underline{Y}_{22}} \frac{1}{\underline{Y}_{22}}$	$\frac{1}{\underline{Z}_{11}} \frac{-\underline{Z}_{12}}{\underline{Z}_{11}}$ $\frac{\underline{Z}_{21}}{\underline{Z}_{11}} \frac{\det \underline{Z}}{\underline{Z}_{11}}$	$\frac{\underline{H}_{22}}{\det \underline{H}} \frac{-\underline{H}_{12}}{\det \underline{H}}$ $\frac{-\underline{H}_{21}}{\det 1} \frac{\underline{H}_{11}}{\det \underline{H}}$	$\underline{C}_{11}\underline{C}_{12}$ $\underline{C}_{21}\underline{C}_{22}$	$\frac{\underline{A}_{21}}{\underline{A}_{11}} \frac{-\det \underline{A}}{\underline{A}_{11}}$ $\frac{1}{\underline{A}_{11}} \frac{\underline{A}_{12}}{\underline{A}_{11}}$
$\underline{A}$	$\frac{-\underline{Y}_{22}}{\underline{Y}_{21}} \frac{-1}{\underline{Y}_{21}}$ $\frac{-\det \underline{Y}}{\underline{Y}_{21}} \frac{-\underline{Y}_{11}}{\underline{Y}_{21}}$	$\frac{\underline{Z}_{11}}{\underline{Z}_{21}} \frac{\det \underline{Z}}{\underline{Z}_{21}}$ $\frac{1}{\underline{Z}_{21}} \frac{\underline{Z}_{22}}{\underline{Z}_{21}}$	$\frac{-\det \underline{H}}{\underline{H}_{21}} \frac{-\underline{H}_{11}}{\underline{H}_{21}}$ $\frac{-\underline{H}_{22}}{\underline{H}_{21}} \frac{-1}{\underline{H}_{21}}$	$\frac{1}{\underline{C}_{21}} \frac{\underline{C}_{22}}{\underline{C}_{21}}$ $\frac{\underline{C}_{11}}{\underline{C}_{21}} \frac{\det \underline{C}}{\underline{C}_{21}}$	$\underline{A}_{11}\underline{A}_{12}$ $\underline{A}_{21}\underline{A}_{22}$

Tab. 93.3 Definitionen und Bezeichnungen von Zweitorparametern

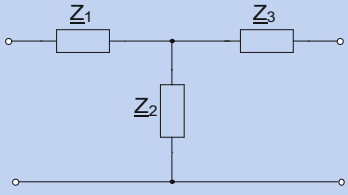
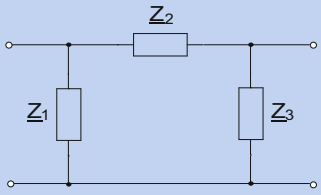
$\underline{Z}$-Parameter	$\underline{Z}_{11} = \frac{\underline{U}_1}{\underline{I}_1}\Big\vert_{\underline{I}_2=0}$	Leerlauf-Eingangsimpedanz
	$\underline{Z}_{12} = \frac{\underline{U}_1}{\underline{I}_2}\Big\vert_{\underline{I}_1=0}$	Leerlauf-Übertragungsimpedanz rückwärts
	$\underline{Z}_{21} = \frac{\underline{U}_2}{\underline{I}_1}\Big\vert_{\underline{I}_2=0}$	Leerlauf-Übertragungsimpedanz vorwärts
	$\underline{Z}_{22} = \frac{\underline{U}_2}{\underline{I}_2}\Big\vert_{\underline{I}_1=0}$	Leerlauf-Ausgangsimpedanz
$\underline{Y}$-Parameter	$\underline{Y}_{11} = \frac{\underline{I}_1}{\underline{U}_1}\Big\vert_{\underline{U}_2=0}$	Kurzschluss-Eingangsadmittanz
	$\underline{Y}_{12} = \frac{\underline{I}_1}{\underline{U}_2}\Big\vert_{\underline{U}_1=0}$	Kurzschluss-Übertragungsadmittanz rückwärts
	$\underline{Y}_{21} = \frac{\underline{I}_2}{\underline{U}_1}\Big\vert_{\underline{U}_2=0}$	Kurzschluss-Übertragungsadmittanz vorwärts
	$\underline{Y}_{22} = \frac{\underline{I}_2}{\underline{U}_2}\Big\vert_{\underline{U}_1=0}$	Kurzschluss-Ausgangsadmittanz
$\underline{H}$-Parameter	$\underline{H}_{11} = \frac{\underline{U}_1}{\underline{I}_1}\Big\vert_{\underline{U}_2=0}$	Kurzschluss-Eingangsimpedanz
	$\underline{H}_{12} = \frac{\underline{U}_1}{\underline{U}_2}\Big\vert_{\underline{I}_1=0}$	Leerlauf-Spannungsrückwirkung
	$\underline{H}_{21} = \frac{\underline{I}_2}{\underline{I}_1}\Big\vert_{\underline{U}_2=0}$	Kurzschluss-Stromübersetzung vorwärts
	$\underline{H}_{22} = \frac{\underline{I}_2}{\underline{U}_2}\Big\vert_{\underline{I}_1=0}$	Leerlauf-Ausgangsadmittanz
$\underline{C}$-Parameter	$\underline{C}_{11} = \frac{\underline{I}_1}{\underline{U}_1}\Big\vert_{\underline{I}_2=0}$	Leerlauf-Eingangsadmittanz
	$\underline{C}_{12} = \frac{\underline{I}_1}{\underline{I}_2}\Big\vert_{\underline{U}_1=0}$	Kurzschluss-Stromrückwirkung
	$\underline{C}_{21} = \frac{\underline{U}_2}{\underline{U}_1}\Big\vert_{\underline{I}_2=0}$	Leerlauf-Spannungsübersetzung
	$\underline{C}_{22} = \frac{\underline{U}_2}{\underline{I}_2}\Big\vert_{\underline{U}_1=0}$	Kurzschluss-Ausgangsimpedanz
$\underline{A}$-Parameter	$\underline{A}_{11} = \frac{\underline{U}_1}{\underline{U}_2}\Big\vert_{\underline{I}_2=0}$	Reziproke Leerlauf-Spannungsübersetzung vorwärts
	$\underline{A}_{12} = \frac{\underline{U}_1}{-\underline{I}_2}\Big\vert_{\underline{U}_2=0}$	Negative reziproke Kurzschluss-Übertragungsadmittanz vorwärts
	$\underline{A}_{21} = \frac{\underline{I}_1}{\underline{U}_2}\Big\vert_{\underline{I}_2=0}$	Reziproke Leerlauf-Übertragungsimpedanz vorwärts
	$\underline{A}_{22} = \frac{\underline{I}_1}{-\underline{I}_2}\Big\vert_{\underline{U}_2=0}$	Negative reziproke Kurzschluss-Stromübersetzung vorwärts

zitäten enthalten, spricht man korrekterweise von Impedanzen bzw. Admittanzen. Werden die Bezeichnungen Widerstände und Leitwerte bei den Zweitorparametern verwendet, setzt man voraus, dass der Anwender diese Bezeichnungen gegebenenfalls durch die korrekten Begriffe ersetzt.

93.1.4 Elementarzweitore

Häufig lassen sich Zweitore, aufgebaut aus einer Vielzahl von Einzelelementen, auf wenige sogenannte *Elementarzweitore* zurückführen, deren Zweitorparameter in Tabellen abgelesen werden können und die zur Bildung des Gesamtzweitors entsprechend zusammengefügt werden müssen. In Tab. 93.4

Tab. 93.4 Parameter der Elementarzweitore „T-Zweitor" und „π-Zweitor"

	T-Zweitor	Π-Zweitor
		
$(\underline{Y})$	$\underline{Y}_{11} = \frac{\underline{Z}_2+\underline{Z}_3}{\underline{Z}_1\underline{Z}_2+\underline{Z}_1\underline{Z}_3+\underline{Z}_2\underline{Z}_3}$ $\underline{Y}_{12} = \frac{-\underline{Z}_2}{\underline{Z}_1\underline{Z}_2+\underline{Z}_1\underline{Z}_3+\underline{Z}_2\underline{Z}_3}$ $\underline{Y}_{21} = \frac{-\underline{Z}_2}{\underline{Z}_1\underline{Z}_2+\underline{Z}_1\underline{Z}_3+\underline{Z}_2\underline{Z}_3}$ $\underline{Y}_{22} = \frac{\underline{Z}_1+\underline{Z}_2}{\underline{Z}_1\underline{Z}_2+\underline{Z}_1\underline{Z}_3+\underline{Z}_2\underline{Z}_3}$	$\underline{Y}_{11} = \frac{1}{\underline{Z}_1} + \frac{1}{\underline{Z}_2}$ $\underline{Y}_{12} = -\frac{1}{\underline{Z}_2}$ $\underline{Y}_{21} = -\frac{1}{\underline{Z}_2}$ $\underline{Y}_{22} = \frac{1}{\underline{Z}_2} + \frac{1}{\underline{Z}_3}$
$(\underline{Z})$	$\underline{Z}_{11} = \underline{Z}_1 + \underline{Z}_2$ $\underline{Z}_{12} = \underline{Z}_2$ $\underline{Z}_{21} = \underline{Z}_2$ $\underline{Z}_{22} = \underline{Z}_2 + \underline{Z}_3$	$\underline{Z}_{11} = \frac{\underline{Z}_1(\underline{Z}_2+\underline{Z}_3)}{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3)}$ $\underline{Z}_{12} = \frac{\underline{Z}_1\underline{Z}_3}{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3)}$ $\underline{Z}_{21} = \frac{\underline{Z}_1\underline{Z}_3}{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3)}$ $\underline{Z}_{22} = \frac{\underline{Z}_3(\underline{Z}_1+\underline{Z}_2)}{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3)}$
$(\underline{H})$	$\underline{H}_{11} = \frac{\underline{Z}_1\underline{Z}_2+\underline{Z}_1\underline{Z}_3+\underline{Z}_2\underline{Z}_3}{\underline{Z}_2+\underline{Z}_3}$ $\underline{H}_{12} = \frac{\underline{Z}_2}{\underline{Z}_2+\underline{Z}_3}$ $\underline{H}_{21} = \frac{-\underline{Z}_2}{\underline{Z}_2+\underline{Z}_3}$ $\underline{H}_{22} = \frac{1}{\underline{Z}_2+\underline{Z}_3}$	$\underline{H}_{11} = \frac{\underline{Z}_1\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{12} = \frac{\underline{Z}_1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{21} = \frac{-\underline{Z}_1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{22} = \frac{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3}{\underline{Z}_3(\underline{Z}_1+\underline{Z}_2)}$
$(\underline{C})$	$\underline{C}_{11} = \frac{1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{12} = \frac{-\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{21} = \frac{\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{22} = \underline{Z}_3 + \frac{\underline{Z}_1\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$	$\underline{C}_{11} = \frac{\underline{Z}_1+\underline{Z}_2+\underline{Z}_3}{\underline{Z}_1(\underline{Z}_2+\underline{Z}_3)}$ $\underline{C}_{12} = \frac{-\underline{Z}_1}{\underline{Z}_2+\underline{Z}_3}$ $\underline{C}_{21} = \frac{\underline{Z}_1}{\underline{Z}_2+\underline{Z}_3}$ $\underline{C}_{22} = \frac{\underline{Z}_2\underline{Z}_3}{\underline{Z}_2+\underline{Z}_3}$
$(\underline{A})$	$\underline{A}_{11} = 1 + \frac{\underline{Z}_1}{\underline{Z}_2}$ $\underline{A}_{12} = \underline{Z}_1 + \underline{Z}_3 + \frac{\underline{Z}_1\underline{Z}_3}{\underline{Z}_2}$ $\underline{A}_{21} = \frac{1}{\underline{Z}_2}$ $\underline{A}_{22} = 1 + \frac{\underline{Z}_3}{\underline{Z}_2}$	$\underline{A}_{11} = 1 + \frac{\underline{Z}_2}{\underline{Z}_3}$ $\underline{A}_{12} = \underline{Z}_2$ $\underline{A}_{21} = \frac{1}{\underline{Z}_1} + \frac{1}{\underline{Z}_3} + \frac{\underline{Z}_2}{\underline{Z}_1\underline{Z}_3}$ $\underline{A}_{22} = 1 + \frac{\underline{Z}_2}{\underline{Z}_1}$

sind zunächst die Zweitorparameter der Elementarzweitore „T-Zweitor" und „π-Zweitor" dargestellt. Die anschließend aufgelisteten Zweitore „Längswiderstand", „Querwiderstand" und die „Γ-Zweitore 1 und 2" in Tab. 93.5 sind aus Tab. 93.4 abgeleitet, indem die entsprechenden Impedanzen mit $\underline{Z}_n = 0$ oder $\underline{Z}_n \to \infty$ eingesetzt werden.

Die $\underline{Z}$-Parameter existieren beim Längswiderstand nicht, weil entweder $\underline{I}_1$ oder $\underline{I}_2$ zu Null erzwungen wird und damit der jeweils andere Strom ebenfalls Null ist ($\underline{I}_1 = \underline{I}_2$). Dieser steht aber in der Definitionsgleichung im Nenner, sodass man einen unbestimmten Ausdruck erhält. Beim Querwiderstand führen die zu Null erzwungenen Spannungen bei den Y-Parametern zu unbestimmten Ausdrücken.

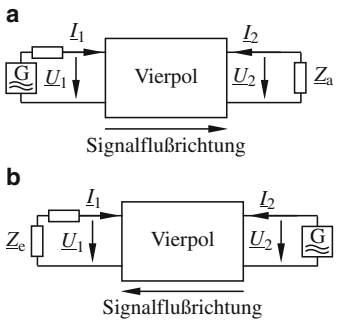


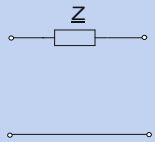
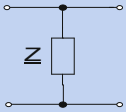
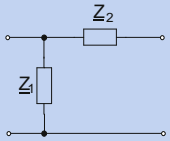
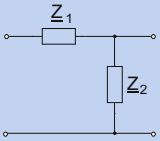
Abb. 93.3 Signalflussrichtung bei Zweitoren. **a** Vorwärtsrichtung, **b** Rückwärtsrichtung

93.1.5 Betriebskenngrößen

Beim Einsatz von Zweitoren gibt es sechs Verknüpfungen zwischen den Größen $\underline{U}_1$, $\underline{U}_2$, $\underline{I}_1$ und $\underline{I}_2$. Man bezeichnet sie als *Betriebskenngrößen in Vorwärtsrichtung*, wenn die Signalflussrichtung im Zweitor von links nach rechts verläuft (Abb. 93.3a). Entsprechend handelt es sich um *Betriebskenngrößen in Rückwärtsrichtung* bei einer Signalflussrichtung von rechts nach links (Abb. 93.3b). Die Lage des Zweitors bleibt dabei unverändert. Diese Kenngrößen sind in Tab. 93.6 dargestellt.

Da im allgemeinen Fall an den Ausgang eines in Vorwärtsrichtung betriebenen Zweitors eine Lastimpedanz $\underline{Z}_a$ nach Abb. 93.3 angeschlossen ist, sind die Betriebskenngrößen in Tab. 93.7 für diesen Fall aufgelistet. Für Zweitore in Rückwärtsrichtung ist der Eingang mit einer Lastimpedanz $\underline{Z}_e$ beschaltet.

Tab. 93.5 Parameter von vereinfachten Elementarzweitoren aus Tab. 93.4

	Längswiderstand	Querwiderstand	Γ-Zweitor 1	Γ-Zweitor 2
				
$(\underline{Y})$	$\underline{Y}_{11} = \frac{1}{\underline{Z}}$ $\underline{Y}_{12} = -\frac{1}{\underline{Z}}$ $\underline{Y}_{21} = -\frac{1}{\underline{Z}}$ $\underline{Y}_{22} = \frac{1}{\underline{Z}}$	Existieren nicht	$\underline{Y}_{11} = \frac{1}{\underline{Z}_1} + \frac{1}{\underline{Z}_2}$ $\underline{Y}_{12} = -\frac{1}{\underline{Z}_2}$ $\underline{Y}_{21} = -\frac{1}{\underline{Z}_2}$ $\underline{Y}_{22} = \frac{1}{\underline{Z}_2}$	$\underline{Y}_{11} = \frac{1}{\underline{Z}_1}$ $\underline{Y}_{12} = -\frac{1}{\underline{Z}_1}$ $\underline{Y}_{21} = -\frac{1}{\underline{Z}_1}$ $\underline{Y}_{22} = \frac{1}{\underline{Z}_1} + \frac{1}{\underline{Z}_2}$
$(\underline{Z})$	Existieren nicht	$\underline{Z}_{11} = \underline{Z}$ $\underline{Z}_{12} = \underline{Z}$ $\underline{Z}_{21} = \underline{Z}$ $\underline{Z}_{22} = \underline{Z}$	$\underline{Z}_{11} = \underline{Z}_1$ $\underline{Z}_{12} = \underline{Z}_1$ $\underline{Z}_{21} = \underline{Z}_1$ $\underline{Z}_{22} = \underline{Z}_1 + \underline{Z}_2$	$\underline{Z}_{11} = \underline{Z}_1 + \underline{Z}_2$ $\underline{Z}_{12} = \underline{Z}_2$ $\underline{Z}_{21} = \underline{Z}_2$ $\underline{Z}_{22} = \underline{Z}_2$
$(\underline{H})$	$\underline{H}_{11} = \underline{Z}$ $\underline{H}_{12} = 1$ $\underline{H}_{21} = -1$ $\underline{H}_{22} = 0$	$\underline{H}_{11} = 0$ $\underline{H}_{12} = 1$ $\underline{H}_{21} = -1$ $\underline{H}_{22} = \frac{1}{\underline{Z}}$	$\underline{H}_{11} = \frac{\underline{Z}_1\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{12} = \frac{\underline{Z}_1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{21} = -\frac{\underline{Z}_1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{H}_{22} = \frac{1}{\underline{Z}_1+\underline{Z}_2}$	$\underline{H}_{11} = \underline{Z}_1$ $\underline{H}_{12} = 1$ $\underline{H}_{21} = -1$ $\underline{H}_{22} = \frac{1}{\underline{Z}_2}$
$(\underline{C})$	$\underline{C}_{11} = 0$ $\underline{C}_{12} = -1$ $\underline{C}_{21} = 1$ $\underline{C}_{22} = \underline{Z}$	$\underline{C}_{11} = \frac{1}{\underline{Z}}$ $\underline{C}_{12} = -1$ $\underline{C}_{21} = 1$ $\underline{C}_{22} = 0$	$\underline{C}_{11} = \frac{1}{\underline{Z}_1}$ $\underline{C}_{12} = -1$ $\underline{C}_{21} = 1$ $\underline{C}_{22} = \underline{Z}_1 + \underline{Z}_2$	$\underline{C}_{11} = \frac{1}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{12} = -\frac{\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{21} = \frac{\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$ $\underline{C}_{22} = \frac{\underline{Z}_1\underline{Z}_2}{\underline{Z}_1+\underline{Z}_2}$
$(\underline{A})$	$\underline{A}_{11} = 1$ $\underline{A}_{12} = \underline{Z}$ $\underline{A}_{21} = 0$ $\underline{A}_{22} = 1$	$\underline{A}_{11} = 1$ $\underline{A}_{12} = 0$ $\underline{A}_{21} = \frac{1}{\underline{Z}}$ $\underline{A}_{22} = 1$	$\underline{A}_{11} = 1$ $\underline{A}_{12} = \underline{Z}_2$ $\underline{A}_{21} = \frac{1}{\underline{Z}_1}$ $\underline{A}_{22} = 1 + \frac{\underline{Z}_2}{\underline{Z}_1}$	$\underline{A}_{11} = 1 + \frac{\underline{Z}_1}{\underline{Z}_2}$ $\underline{A}_{12} = \underline{Z}_1$ $\underline{A}_{21} = \frac{1}{\underline{Z}_2}$ $\underline{A}_{22} = 1$

Tab. 93.6 Definitionen der Betriebskenngrößen (nach [II.1])

Betriebskenngröße im Vorwärtsbetrieb	Definition
Eingangsadmittanz	$\underline{Y}_{\text{in}} = \frac{\underline{I}_1}{\underline{U}_1}$
Eingangsimpedanz	$\underline{Z}_{\text{in}} = \frac{\underline{U}_1}{\underline{I}_1}$
Übertragungsadmittanz vorwärts	$\underline{Y}_{\text{üf}} = \frac{\underline{I}_2}{\underline{U}_1}$
Übertragungsimpedanz vorwärts	$\underline{Z}_{\text{üf}} = \frac{\underline{U}_2}{\underline{I}_1}$
Spannungsübersetzung vorwärts	$\underline{v}_{\text{uf}} = \frac{\underline{U}_2}{\underline{U}_1}$
Stromübersetzung vorwärts	$\underline{v}_{\text{if}} = \frac{\underline{I}_2}{\underline{I}_1}$

Betriebskenngröße im Rückwärtsbetrieb	Definition
Ausgangsadmittanz	$\underline{Y}_{\text{out}} = \frac{\underline{I}_2}{\underline{U}_2}$
Ausgangsimpedanz	$\underline{Z}_{\text{out}} = \frac{\underline{U}_2}{\underline{I}_2}$
Übertragungsadmittanz rückwärts	$\underline{Y}_{\text{ür}} = \frac{\underline{I}_1}{\underline{U}_2}$
Übertragungsimpedanz rückwärts	$\underline{Z}_{\text{ür}} = \frac{\underline{U}_1}{\underline{I}_2}$
Spannungsübersetzung rückwärts	$\underline{v}_{\text{ur}} = \frac{\underline{U}_1}{\underline{U}_2}$
Stromübersetzung rückwärts	$\underline{v}_{\text{ir}} = \frac{\underline{I}_1}{\underline{I}_2}$

Die in Tab. 93.7 aufgelisteten Betriebskenngrößen vereinfachen sich dann, wenn die Lastimpedanz entweder Null ist (Kurzschluss am Ein- bzw. Ausgang, entspricht einer unendlich großen Lastadmittanz) oder unendlich groß ist (Leerlauf am Ein- bzw. Ausgang, entspricht einer Lastadmittanz Null). In Tab. 93.8 sind die entsprechenden Größen eingetragen.

Beispiel 1

In Abb. 93.4 ist ein Netzwerk gezeigt, mit dem RC-Oszillatoren aufgebaut werden können. Es dient dazu, bei *einer* Frequenz $f = f_0$ einen Phasenwinkel von 180° zwischen $\underline{U}_1$ und $\underline{U}_2$ zu erzeugen (Gegenphase, deshalb ist das Vorzeichen ohne Bedeutung). Die für eine Gesamtphasenverschiebung von 360° erforderlichen weiteren 180° werden mit einem Transistor bzw. einer Operationsverstärkerschaltung erreicht.

Berechnet werden soll diese Frequenz in Abhängigkeit von R und C und das Verhältnis von Ausgangsspannung zu Eingangsspannung bei dieser Frequenz.

Das Netzwerk wird zerlegt in die Elementar-Zweitore „T-Zweitor“ und „π-Zweitor“. Da beide in Kettenschaltung zusammengeschaltet sind, werden nach Tab. 93.1 die $\underline{A}$-Parameter beider Zweitore benötigt. Gesucht ist $\underline{v}_{\text{uf}}$. Der

Tab. 93.7 Betriebskenngrößen mit Lastadmittanz $\underline{Y}$

Betriebskenngrößen in Vorwärtsrichtung					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{in}}$	$\frac{\det \underline{Y}+\underline{Y}_{11}\cdot\underline{Y}_a}{\underline{Y}_{22}+\underline{Y}_a}$	$\frac{1+\underline{Z}_{22}\cdot\underline{Y}_a}{\underline{Z}_{11}+\underline{Y}_a\cdot\det\underline{Z}}$	$\frac{\underline{H}_{22}+\underline{Y}_a}{\det\underline{H}+\underline{H}_{11}\cdot\underline{Y}_a}$	$\frac{\underline{C}_{11}+\underline{Y}_a\cdot\det\underline{C}}{1+\underline{C}_{22}\cdot\underline{Y}_a}$	$\frac{\underline{A}_{21}+\underline{A}_{22}\cdot\underline{Y}_a}{\underline{A}_{11}+\underline{A}_{12}\cdot\underline{Y}_a}$
$\underline{Z}_{\text{in}}$	$\frac{\underline{Y}_{22}+\underline{Y}_a}{\det\underline{Y}+\underline{Y}_{11}\cdot\underline{Y}_a}$	$\frac{\underline{Z}_{11}+\underline{Y}_a\cdot\det\underline{Z}}{1+\underline{Z}_{22}\cdot\underline{Y}_a}$	$\frac{\det\underline{H}+\underline{H}_{11}\cdot\underline{Y}_a}{\underline{H}_{22}+\underline{Y}_a}$	$\frac{1+\underline{C}_{22}\cdot\underline{Y}_a}{\underline{C}_{11}+\underline{Y}_a\cdot\det\underline{C}}$	$\frac{\underline{A}_{11}+\underline{A}_{12}\cdot\underline{Y}_a}{\underline{A}_{21}+\underline{A}_{22}\cdot\underline{Y}_a}$
$\underline{Y}_{\text{üf}}$	$\frac{\underline{Y}_{21}\cdot\underline{Y}_a}{\underline{Y}_{22}+\underline{Y}_a}$	$\frac{-\underline{Z}_{21}\cdot\underline{Y}_a}{\underline{Z}_{11}+\underline{Y}_a\cdot\det\underline{Z}}$	$\frac{\underline{H}_{21}\cdot\underline{Y}_a}{\det\underline{H}+\underline{H}_{11}\cdot\underline{Y}_a}$	$\frac{-\underline{C}_{21}\cdot\underline{Y}_a}{1+\underline{C}_{22}\cdot\underline{Y}_a}$	$\frac{-\underline{Y}_a}{\underline{A}_{11}+\underline{A}_{12}\cdot\underline{Y}_a}$
$\underline{Z}_{\text{üf}}$	$\frac{-\underline{Y}_{21}}{\det\underline{Y}+\underline{Y}_{11}\cdot\underline{Y}_a}$	$\frac{\underline{Z}_{21}}{1+\underline{Z}_{22}\cdot\underline{Y}_a}$	$\frac{-\underline{H}_{21}}{\underline{H}_{22}+\underline{Y}_a}$	$\frac{\underline{C}_{21}}{\underline{C}_{11}+\underline{Y}_a\cdot\det\underline{C}}$	$\frac{1}{\underline{A}_{21}+\underline{A}_{22}\cdot\underline{Y}_a}$
$\underline{v}_{\text{uf}}$	$\frac{-\underline{Y}_{21}}{\underline{Y}_{22}+\underline{Y}_a}$	$\frac{\underline{Z}_{21}}{\underline{Z}_{11}+\underline{Y}_a\cdot\det\underline{Z}}$	$\frac{-\underline{H}_{21}}{\det\underline{H}+\underline{H}_{11}\cdot\underline{Y}_a}$	$\frac{\underline{C}_{21}}{1+\underline{C}_{22}\cdot\underline{Y}_a}$	$\frac{1}{\underline{A}_{11}+\underline{A}_{12}\cdot\underline{Y}_a}$
$\underline{v}_{\text{if}}$	$\frac{\underline{Y}_{21}\cdot\underline{Y}_a}{\det\underline{Y}+\underline{Y}_{11}\cdot\underline{Y}_a}$	$\frac{-\underline{Z}_{21}\cdot\underline{Y}_a}{1+\underline{Z}_{22}\cdot\underline{Y}_a}$	$\frac{\underline{H}_{21}\cdot\underline{Y}_a}{\underline{H}_{22}+\underline{Y}_a}$	$\frac{-\underline{C}_{21}\cdot\underline{Y}_a}{\underline{C}_{11}+\underline{Y}_a\cdot\det\underline{C}}$	$\frac{-\underline{Y}_a}{\underline{A}_{21}+\underline{A}_{22}\cdot\underline{Y}_a}$

Betriebskenngrößen in Rückwärtsrichtung					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{out}}$	$\frac{\det\underline{Y}+\underline{Y}_{22}\cdot\underline{Y}_e}{\underline{Y}_{11}+\underline{Y}_e}$	$\frac{1+\underline{Z}_{11}\cdot\underline{Y}_e}{\underline{Z}_{22}+\underline{Y}_e\cdot\det\underline{Z}}$	$\frac{\underline{H}_{22}+\underline{Y}_e\cdot\det\underline{H}}{1+\underline{H}_{11}\cdot\underline{Y}_e}$	$\frac{\underline{C}_{11}+\underline{Y}_e}{\det\underline{C}+\underline{C}_{22}\cdot\underline{Y}_e}$	$\frac{\underline{A}_{21}+\underline{A}_{11}\cdot\underline{Y}_e}{\underline{A}_{22}+\underline{A}_{12}\cdot\underline{Y}_e}$
$\underline{Z}_{\text{out}}$	$\frac{\underline{Y}_{11}+\underline{Y}_e}{\det\underline{Y}+\underline{Y}_{22}\cdot\underline{Y}_e}$	$\frac{\underline{Z}_{22}+\underline{Y}_e\cdot\det\underline{Z}}{1+\underline{Z}_{11}\cdot\underline{Y}_e}$	$\frac{1+\underline{H}_{11}\cdot\underline{Y}_e}{\underline{H}_{22}+\underline{Y}_e\cdot\det\underline{H}}$	$\frac{\det\underline{C}+\underline{C}_{22}\cdot\underline{Y}_e}{\underline{C}_{11}+\underline{Y}_e}$	$\frac{\underline{A}_{22}+\underline{A}_{12}\cdot\underline{Y}_e}{\underline{A}_{21}+\underline{A}_{11}\cdot\underline{Y}_e}$
$\underline{Y}_{\text{ür}}$	$\frac{\underline{Y}_{12}\cdot\underline{Y}_e}{\underline{Y}_{11}+\underline{Y}_e}$	$\frac{-\underline{Z}_{12}\cdot\underline{Y}_e}{\underline{Z}_{22}+\underline{Y}_e\cdot\det\underline{Z}}$	$\frac{-\underline{H}_{12}\cdot\underline{Y}_e}{1+\underline{H}_{11}\cdot\underline{Y}_e}$	$\frac{\underline{C}_{12}\cdot\underline{Y}_e}{\det\underline{C}+\underline{C}_{22}\cdot\underline{Y}_e}$	$\frac{-\underline{Y}_e\cdot\det\underline{A}}{\underline{A}_{22}+\underline{A}_{12}\cdot\underline{Y}_e}$
$\underline{Z}_{\text{ür}}$	$\frac{-\underline{Y}_{12}}{\det\underline{Y}+\underline{Y}_{22}\cdot\underline{Y}_e}$	$\frac{\underline{Z}_{12}}{1+\underline{Z}_{11}\cdot\underline{Y}_e}$	$\frac{\underline{H}_{12}}{\underline{H}_{22}+\underline{Y}_e\cdot\det\underline{H}}$	$\frac{-\underline{C}_{12}}{\underline{C}_{11}+\underline{Y}_e}$	$\frac{\det\underline{A}}{\underline{A}_{21}+\underline{A}_{11}\cdot\underline{Y}_e}$
$\underline{v}_{\text{ur}}$	$\frac{-\underline{Y}_{12}}{\underline{Y}_{11}+\underline{Y}_e}$	$\frac{\underline{Z}_{12}}{\underline{Z}_{22}+\underline{Y}_e\cdot\det\underline{Z}}$	$\frac{\underline{H}_{12}}{1+\underline{H}_{11}\cdot\underline{Y}_e}$	$\frac{-\underline{C}_{12}}{\det\underline{C}+\underline{C}_{22}\cdot\underline{Y}_e}$	$\frac{\det\underline{A}}{\underline{A}_{22}+\underline{A}_{12}\cdot\underline{Y}_e}$
$\underline{v}_{\text{ir}}$	$\frac{\underline{Y}_{12}\cdot\underline{Y}_e}{\det\underline{Y}+\underline{Y}_{22}\cdot\underline{Y}_e}$	$\frac{-\underline{Z}_{12}\cdot\underline{Y}_e}{1+\underline{Z}_{11}\cdot\underline{Y}_e}$	$\frac{-\underline{H}_{12}\cdot\underline{Y}_e}{\underline{H}_{22}+\underline{Y}_e\cdot\det\underline{H}}$	$\frac{\underline{C}_{12}\cdot\underline{Y}_e}{\underline{C}_{11}+\underline{Y}_e}$	$\frac{-\underline{Y}_e\cdot\det\underline{A}}{\underline{A}_{21}+\underline{A}_{11}\cdot\underline{Y}_e}$

Hinweise:
$\det\underline{Z}=\underline{Z}_{11}\cdot\underline{Z}_{22}-\underline{Z}_{12}\cdot\underline{Z}_{21}$, $\det\underline{Y}=\underline{Y}_{11}\cdot\underline{Y}_{22}-\underline{Y}_{12}\cdot\underline{Y}_{21}$, $\det\underline{C}=\underline{C}_{11}\cdot\underline{C}_{22}-\underline{C}_{12}\cdot\underline{C}_{21}$, $\det\underline{H}=\underline{H}_{11}\cdot\underline{H}_{22}-\underline{H}_{12}\cdot\underline{H}_{21}$, $\det\underline{A}=\underline{A}_{11}\cdot\underline{A}_{22}-\underline{A}_{12}\cdot\underline{A}_{21}$
Es wurde die Admittanz $\underline{Y}_a$ bzw. $\underline{Y}_e$ an Stelle der Impedanz $\underline{Z}_a$ bzw. $\underline{Z}_e$ eingesetzt, weil sich dadurch eine übersichtlichere Schreibweise ergibt. Die Umrechnung geschieht über $\underline{Y}_a=1/\underline{Z}_a$; $\underline{Y}_e=1/\underline{Z}_e$

Ausgang ist nicht belastet, sodass sich nach Tab. 93.8 $\underline{v}_{\text{uf}}=1/\underline{A}_{11}$ ergibt. $\underline{A}_{11}$ ist das Ergebnis der Matrizenmultiplikation der $\underline{A}$-Parameter beider Zweitore. Da hier nur *ein* Parameter zu berechnen ist, genügt es, von dem links angeordneten T-Zweitor die obere Zeile, also $\underline{A}_{11\text{TVP}}$ und $\underline{A}_{12\text{TVP}}$ zu bestimmen, von dem rechts stehenden π-Zweitor die erste Spalte, also $\underline{A}_{11\pi\text{VP}}$ und $\underline{A}_{21\pi\text{VP}}$. Diese Parameter werden Tab. 93.4 entnommen, in die die Bauelemente aus Abb. 93.4 eingesetzt werden:

$$\underline{A}_{11\text{TVP}}=1+Rj\omega C;$$
$$\underline{A}_{12\text{TVP}}=R+R+R^2j\omega C$$
$$\underline{A}_{11\pi\text{VP}}=1+Rj\omega C;$$
$$\underline{A}_{21\pi\text{VP}}=j\omega C+j\omega C+Rj\omega Cj\omega C=2j\omega C-\omega^2C^2R$$

Damit ergibt sich für $\underline{A}_{11}$:

$$\begin{aligned}\underline{A}_{11}&=\underline{A}_{11\text{TVP}}\cdot\underline{A}_{11\pi\text{VP}}+\underline{A}_{12\text{TVP}}\cdot\underline{A}_{21\pi\text{VP}}\\&=(1+j\omega RC)(1+j\omega RC)\\&\quad+(2R+R^2j\omega C)(2j\omega C-\omega^2C^2R)\\&=1+2j\omega RC-\omega^2C^2R^2+4j\omega CR\\&\quad-2\omega^2C^2R^2-2\omega^2C^2R^2-j\omega^3C^3R^3\\&=1+6j\omega RC-5\omega^2C^2R^2-j\omega^3C^3R^3\\&=1-5\omega^2C^2R^2+j\omega RC(6-\omega^2R^2C^2)\end{aligned}$$

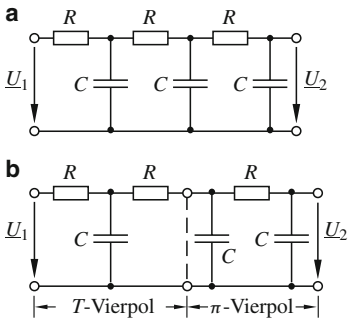


Abb. 93.4 Phasenschieber-Netzwerk (zu Beispiel 1). **a** Netzwerk, **b** Zerlegung in zwei Elementar-Zweitore

Für einen Phasenwinkel 0° oder 180° muss der Imaginärteil von $1/\underline{A}_{11}$ und damit von $\underline{A}_{11}$ Null sein, d. h. es muss gelten: $6-\omega_0^2R^2C^2=0$. Für die Frequenz folgt: $\omega_0=\frac{\sqrt{6}}{RC}$. Setzt man diesen Wert in die Gleichung für $1/\underline{A}_{11}$ ein, ergibt sich:

$$\left.\frac{1}{\underline{A}_{11}}\right|_{\omega_0}=\frac{1}{1-5C^2R^2\omega_0^2}=\frac{1}{1-5C^2R^2(\frac{\sqrt{6}}{RC})^2}=-\frac{1}{29}$$

Da der Wert negativ ist, beträgt die Phasenverschiebung bei ω_0 zwischen Ein- und Ausgangsspannung 180°, die Amplitude ist am Ausgang auf 1/29 abgesunken. ◀

Tab. 93.8 Betriebskenngrößen für die Sonderfälle „Kurzschluss" und „Leerlauf"

Betriebskenngrößen in Vorwärtsrichtung bei Leerlauf am Ausgang ($\underline{Z}_a \to \infty$)					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{in}}$	$\frac{\det \underline{Y}}{\underline{Y}_{22}}$	$\frac{1}{\underline{Z}_{11}}$	$\frac{\underline{H}_{22}}{\det \underline{H}}$	$\underline{C}_{11}$	$\frac{\underline{A}_{21}}{\underline{A}_{11}}$
$\underline{Z}_{\text{in}}$	$\frac{\underline{Y}_{22}}{\det \underline{Y}}$	$\underline{Z}_{11}$	$\frac{\det \underline{H}}{\underline{H}_{22}}$	$\frac{1}{\underline{C}_{11}}$	$\frac{\underline{A}_{11}}{\underline{A}_{21}}$
$\underline{Y}_{\text{üf}}$	0	0	0	0	0
$\underline{Z}_{\text{üf}}$	$\frac{-\underline{Y}_{21}}{\det \underline{Y}}$	$\underline{Z}_{21}$	$\frac{-\underline{H}_{21}}{\underline{H}_{22}}$	$\frac{\underline{C}_{21}}{\underline{C}_{11}}$	$\frac{1}{\underline{A}_{21}}$
$\underline{v}_{\text{uf}}$	$\frac{-\underline{Y}_{21}}{\underline{Y}_{22}}$	$\frac{\underline{Z}_{21}}{\underline{Z}_{11}}$	$\frac{-\underline{H}_{21}}{\det \underline{H}}$	$\underline{C}_{21}$	$\frac{1}{\underline{A}_{11}}$
$\underline{v}_{\text{if}}$	0	0	0	0	0

Betriebskenngrößen in Rückwärtsrichtung bei Leerlauf am Eingang ($\underline{Z}_e \to \infty$)					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{out}}$	$\frac{\det \underline{Y}}{\underline{Y}_{11}}$	$\frac{1}{\underline{Z}_{22}}$	$\underline{H}_{22}$	$\frac{\underline{C}_{11}}{\det \underline{C}}$	$\frac{\underline{A}_{21}}{\underline{A}_{22}}$
$\underline{Z}_{\text{out}}$	$\frac{\underline{Y}_{11}}{\det \underline{Y}}$	$\underline{Z}_{22}$	$\frac{1}{\underline{H}_{22}}$	$\frac{\det \underline{C}}{\underline{C}_{11}}$	$\frac{\underline{A}_{22}}{\underline{A}_{21}}$
$\underline{Y}_{\text{ür}}$	0	0	0	0	0
$\underline{Z}_{\text{ür}}$	$\frac{-\underline{Y}_{12}}{\det \underline{Y}}$	$\underline{Z}_{12}$	$\frac{\underline{H}_{12}}{\underline{H}_{22}}$	$\frac{-\underline{C}_{12}}{\underline{C}_{11}}$	$\frac{\det \underline{A}}{\underline{A}_{21}}$
$\underline{v}_{\text{ur}}$	$\frac{-\underline{Y}_{12}}{\underline{Y}_{11}}$	$\frac{\underline{Z}_{12}}{\underline{Z}_{22}}$	$\underline{H}_{12}$	$\frac{-\underline{C}_{12}}{\det \underline{C}}$	$\frac{\det \underline{A}}{\underline{A}_{22}}$
$\underline{v}_{\text{ir}}$	0	0	0	0	0

Betriebskenngrößen in Vorwärtsrichtung bei Kurzschluss am Ausgang ($\underline{Z}_a = 0$)					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{in}}$	$\underline{Y}_{11}$	$\frac{\underline{Z}_{22}}{\det \underline{Z}}$	$\frac{1}{\underline{H}_{11}}$	$\frac{\det \underline{C}}{\underline{C}_{22}}$	$\frac{\underline{A}_{22}}{\underline{A}_{12}}$
$\underline{Z}_{\text{in}}$	$\frac{1}{\underline{Y}_{11}}$	$\frac{\det \underline{Z}}{\underline{Z}_{22}}$	$\underline{H}_{11}$	$\frac{\underline{C}_{22}}{\det \underline{C}}$	$\frac{\underline{A}_{12}}{\underline{A}_{22}}$
$\underline{Y}_{\text{üf}}$	$\underline{Y}_{21}$	$\frac{-\underline{Z}_{21}}{\det \underline{Z}}$	$\frac{\underline{H}_{21}}{\underline{H}_{11}}$	$\frac{-\underline{C}_{21}}{\underline{C}_{22}}$	$\frac{1}{\underline{A}_{12}}$
$\underline{Z}_{\text{üf}}$	0	0	0	0	0
$\underline{v}_{\text{uf}}$	0	0	0	0	0
$\underline{v}_{\text{if}}$	$\frac{\underline{Y}_{21}}{\underline{Y}_{11}}$	$\frac{\det \underline{Z}}{\underline{Z}_{22}}$	$\underline{H}_{21}$	$\frac{-\underline{C}_{21}}{\det \underline{C}}$	$\frac{1}{\underline{A}_{22}}$

Betriebskenngrößen in Rückwärtsrichtung bei Kurzschluss am Eingang ($\underline{Z}_e = 0$)					
	$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{\text{out}}$	$\underline{Y}_{22}$	$\frac{\underline{Z}_{11}}{\det \underline{Z}}$	$\frac{\det \underline{H}}{\underline{H}_{11}}$	$\frac{1}{\underline{C}_{22}}$	$\frac{\underline{A}_{11}}{\underline{A}_{12}}$
$\underline{Z}_{\text{out}}$	$\frac{1}{\underline{Y}_{22}}$	$\frac{\det \underline{Z}}{\underline{Z}_{11}}$	$\frac{\underline{H}_{11}}{\det \underline{H}}$	$\underline{C}_{22}$	$\frac{\underline{A}_{12}}{\underline{A}_{11}}$
$\underline{Y}_{\text{ür}}$	$\underline{Y}_{12}$	$\frac{-\underline{Z}_{12}}{\det \underline{Z}}$	$\frac{-\underline{H}_{12}}{\underline{H}_{11}}$	$\frac{\underline{C}_{12}}{\underline{C}_{22}}$	$\frac{\det \underline{A}}{\underline{A}_{12}}$
$\underline{Z}_{\text{ür}}$	0	0	0	0	0
$\underline{v}_{\text{ur}}$	0	0	0	0	0
$\underline{v}_{\text{ir}}$	$\frac{\underline{Y}_{12}}{\underline{Y}_{22}}$	$\frac{-\underline{Z}_{12}}{\underline{Z}_{11}}$	$\frac{-\underline{H}_{12}}{\det \underline{H}}$	$\underline{C}_{12}$	$\frac{\det \underline{A}}{\underline{A}_{11}}$

Beispiel 2

Das Phasenschieber-Netzwerk aus Abb. 93.4 ist am Ausgang mit einem Widerstand $\underline{R}_a$ belastet, Abb. 93.5.

Zur Berechnung der Spannungsverstärkung vorwärts wird entweder die Formel $\underline{v}_{uf} = \frac{1}{\underline{A}_{11}+\underline{A}_{12}\cdot\underline{Y}_a}$ verwendet, oder der Widerstand $\underline{R}_a$ wird mit dem parallel liegenden Kondensator zusammengefasst zu: $\underline{Z}_{ges} = \frac{R_a\cdot\frac{1}{j\omega C}}{R_a+\frac{1}{j\omega C}} = \frac{R_a}{1+R_a j\omega C}$. Dieser Wert wird in die Formel aus Beispiel 1 an entsprechender Stelle eingesetzt:

$$\begin{aligned}
\underline{A}_{11TVP} &= 1 + Rj\omega C;\\
\underline{A}_{12TVP} &= R + R + R^2 j\omega C\\
[1mm]\underline{A}_{11\pi VP} &= 1 + \frac{R}{\underline{Z}_{ges}}\\
&= \frac{\underline{Z}_{ges}+R}{\underline{Z}_{ges}};\\
\underline{A}_{21\pi VP} &= j\omega C + \frac{1}{\underline{Z}_{ges}} + \frac{Rj\omega C}{\underline{Z}_{ges}}\\
&= \frac{\underline{Z}_{ges}j\omega C + 1 + Rj\omega C}{\underline{Z}_{ges}}
\end{aligned}$$

Nach einigen Umformungen folgt schließlich:

$$\begin{aligned}
\frac{1}{\underline{A}_{11}} &= \frac{\underline{U}_2}{\underline{U}_1}\\
&= \frac{R_a}{\left[\begin{array}{l} R_a(1-5\omega^2R^2C^2) + R(3-R^2\omega^2C^2)\\ +j\omega RC(6R_a+4R-\omega^2R^2C^2R_a)\end{array}\right]}
\end{aligned}$$

Für $R_a \to \infty$ folgt daraus der Wert aus Beispiel 1. Der Phasenwinkel beträgt Null oder 180° für $6R_a + 4R - \omega_0^2R^2C^2R_a = 0$; d. h.

$$\omega_0 = \sqrt{\frac{6R_a+4R}{R^2C^2R_a}} = \frac{1}{RC}\sqrt{6+\frac{4R}{R_a}}$$

Setzt man diesen Wert ein, ergibt sich nach einiger Rechnung:

$$\begin{aligned}
\left.\frac{1}{\underline{A}_{11}}\right|_{\omega_0} &= \left.\frac{\underline{U}_2}{\underline{U}_1}\right|_{\omega_0}\\
&= \frac{1}{-29-23\frac{R}{R_a}-4\frac{R^2}{R_a^2}}
\end{aligned}$$

Der Wert ist negativ, demnach handelt es sich um den Phasenwinkel 180°. Auch hier bekommt man das Ergebnis aus Beispiel 1 mit $R_a \to \infty$. ◄

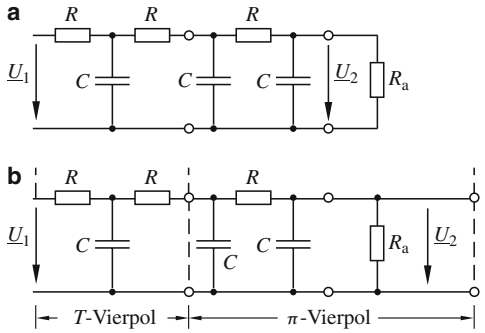


Abb. 93.5 Phasenschieber-Netzwerk (zu Beispiel 2), am Ausgang belastet. **a** Netzwerk, **b** Einbeziehen von R_a in das π-Zweitor

93.2 Spezielle Zweitore

In der Vermittlungstechnik (Telefon) werden Zweitore benötigt, bei denen der Signalfluss in Vorwärts- und Rückwärtsrichtung gleich sein muss. Hat ein Zweitor diese und andere besondere Eigenschaften, lassen sie sich an den Zweitorparametern ablesen.

93.2.1 Übertragungssymmetrische (reziproke) Zweitore

Ein Zweitor ist übertragungssymmetrisch (kopplungssymmetrisch, reziprok), wenn für ihn der Kirchhoff'sche Umkehrungssatz gilt (Abb. 93.6): Bei Stromeinspeisung mit $\underline{I}_0$ am Eingang und anschließend am Ausgang muss gelten: $\underline{U}_1 = \underline{U}_2$. Bei Spannungseinspeisung lautet die Bedingung: $\underline{I}_1 = \underline{I}_2$.

Passive Zweitore sind übertragungssymmetrisch.

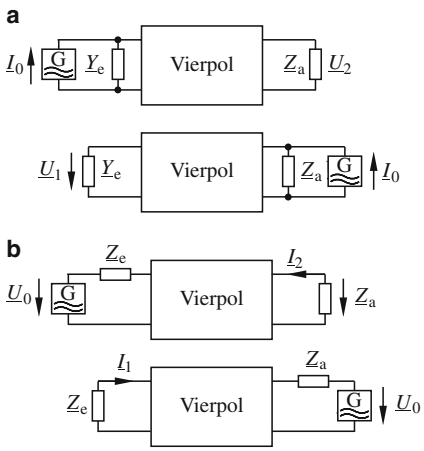


Abb. 93.6 Übertragungssymmetrische Zweitore. **a** Stromeinspeisung, **b** Spannungseinspeisung

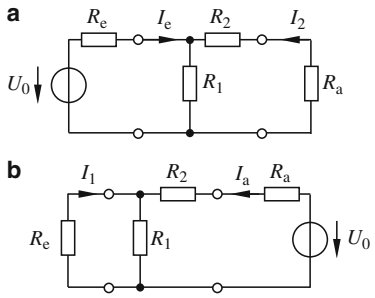


Abb. 93.7 Zur Ableitung der Übertragungssymmetrie (zu Beispiel 3). **a** Speisung am Eingang mit U_0, **b** Speisung am Ausgang mit U_0

Beispiel 3

Abb. 93.7 zeigt einen Gamma-Zweitor, an dem die Übertragungssymmetrie gezeigt werden soll.

Der Einfachheit halber werden ohmsche Widerstände und eine Gleichspannungsquelle eingesetzt. Für die Speisung am Eingang (Abb. 93.7a) folgt:

$$\frac{-I_2}{I_e} = \frac{R_1}{R_1 + R_2 + R_a};$$

$$I_e = \frac{U_0}{R_e + \frac{R_1(R_2+R_a)}{R_1+R_2+R_a}} = \frac{U_0(R_1 + R_2 + R_a)}{(R_1 + R_2 + R_a)R_e + R_1(R_2 + R_a)};$$

$$-I_2 = \frac{U_0 \cdot R_1}{R_e(R_1 + R_2 + R_a) + R_1(R_2 + R_a)}$$

In analoger Weise folgt bei Speisung am Ausgang (Abb. 93.7b):

$$\frac{I_1}{-I_a} = \frac{R_1}{R_1 + R_e};$$

$$I_a = \frac{U_0}{R_a + R_2 + \frac{R_1 \cdot R_e}{R_1+R_e}} = \frac{U_0(R_1 + R_e)}{(R_1 + R_e)(R_a + R_2) + R_1 \cdot R_e};$$

$$-I_1 = \frac{U_0 \cdot R_1}{R_e(R_1 + R_2 + R_a) + R_1(R_2 + R_a)}$$

Es gilt: $I_1 = I_2$. Damit ist das Zweitor übertragungssymmetrisch. ◀

Allgemein kann man die in Tab. 93.9 aufgeführten Bedingungen für übertragungssymmetrische Zweitore ableiten.

Tab. 93.9 Bedingungen für übertragungssymmetrische Zweitore

$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{12} = \underline{Y}_{21}$	$\underline{Z}_{12} = \underline{Z}_{21}$	$\underline{H}_{12} = -\underline{H}_{21}$	$\underline{C}_{12} = -\underline{C}_{21}$	$\det \underline{A} = 1$

Tab. 93.10 Bedingungen für widerstandssymmetrische Zweitore

$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{11} = \underline{Y}_{22}$	$\underline{Z}_{11} = \underline{Z}_{22}$	$\det \underline{H} = 1$	$\det \underline{C} = 1$	$\underline{A}_{11} = \underline{A}_{22}$

Tab. 93.11 Bedingungen für längssymmetrische Zweitore

$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{11} = \underline{Y}_{22}$	$\underline{Z}_{11} = \underline{Z}_{22}$	$\det \underline{H} = 1$	$\det \underline{C} = 1$	$\underline{A}_{11} = \underline{A}_{22}$
$\underline{Y}_{12} = \underline{Y}_{21}$	$\underline{Z}_{12} = \underline{Z}_{21}$	$\underline{H}_{12} = -\underline{H}_{21}$	$\underline{C}_{12} = -\underline{C}_{21}$	$\det \underline{A} = 1$

Tab. 93.12 Bedingungen für rückwirkungsfreie Zweitore

$\underline{Y}$-Parameter	$\underline{Z}$-Parameter	$\underline{H}$-Parameter	$\underline{C}$-Parameter	$\underline{A}$-Parameter
$\underline{Y}_{12} = 0$	$\underline{Z}_{12} = 0$	$\underline{H}_{12} = 0$	$\underline{C}_{12} = 0$	$\det \underline{A} = 0$

93.2.2 Widerstandssymmetrische Zweitore

Bei widerstandssymmetrischen Zweitoren sind die Eingangsimpedanz bei beliebiger Abschlussimpedanz in Vorwärtsrichtung und die Ausgangsimpedanz (Blick vom Ausgang in das Zweitor hinein) bei gleicher Abschlussimpedanz am Eingang gleich. Die Bedingungen sind in Tab. 93.10 aufgeführt.

Ein Beispiel für einen nahezu rückwirkungsfreien Zweitor ist ein von einer Sendeantenne ausgesandtes elektromagnetisches Feld. Senderseitig lässt sich nicht feststellen, wie viele Lastimpedanzen in Form von Empfängerantennen vorhanden sind, sofern sie sich im Fernfeld der Senderantenne befinden (siehe Abschn. 95.1).

93.2.3 Längssymmetrische Zweitore

Bei längssymmetrischen Zweitoren sind alle Betriebseigenschaften unabhängig von der Betriebsrichtung, d. h. sie sind sowohl übertragungssymmetrisch als auch widerstandssymmetrisch. Damit ergeben sich die in Tab. 93.11 aufgeführten Bedingungen für die Parameter.

Beispiel 4

Der in Tab. 93.4 dargestellte T-Zweitor ist dann längssymmetrisch, wenn nach Tab. 93.11 gilt: $\underline{Z}_{11} = \underline{Z}_{22}$ und $\underline{Z}_{12} = \underline{Z}_{21}$. Das ist der Fall, wenn gilt: $\underline{Z}_1 = \underline{Z}_3$. Dieses Ergebnis entspricht auch der Anschauung, da man bei längssymmetrischen Zweitoren nicht unterscheiden kann, ob man ihn in Vorwärts- oder Rückwärtsrichtung betreibt. ◀

93.2.4 Rückwirkungsfreie Zweitore

Wirkt die Ausgangsgröße nicht auf den Eingang zurück, spricht man von einem rückwirkungsfreien Zweitor. Passive Zweitore besitzen diese Eigenschaft nicht; in der Praxis kann man sie nur näherungsweise erfüllen. Bei Transistoren bemüht man sich, die Rückwirkung auf den Eingang so klein wie möglich zu machen. Tab. 93.12 stellt die notwendigen Bedingungen dar.

93.3 Wellenparameter passiver Zweitore

Die Leitungstheorie ist ein Teil der Zweitortheorie und arbeitet mit Wellenparametern anstelle von Zweitorparametern. Da elektrische Leitungen als passive Zweitore übertragungssymmetrisch sind, hat man hierfür *Wellenparameter* festgelegt.

93.3.1 Allgemeine passive Zweitore

Wird ein Zweitor in Vorwärtsrichtung betrieben und mit einer Lastimpedanz $\underline{Z}_a$ belastet (Abb. 93.8a), ergibt sich am Eingang die Eingangsimpedanz $\underline{Z}_{\text{in}}$. Sie ist abhängig vom Zweitor und der Lastimpedanz $\underline{Z}_a$ und lässt sich wie folgt berechnen:

$$\underline{U}_1 = \underline{Z}_{11} \cdot \underline{I}_1 + \underline{Z}_{12} \cdot \underline{I}_2 \quad \text{(1. Zweitorgleichung);}$$

$$\text{Division durch } \underline{I}_1\text{:} \quad \frac{\underline{U}_1}{\underline{I}_1} = \underline{Z}_{11} + \frac{\underline{Z}_{12} \cdot \underline{I}_2}{\underline{I}_1}$$

$$\underline{U}_2 = \underline{Z}_{21} \cdot \underline{I}_1 + \underline{Z}_{22} \cdot \underline{I}_2 \quad \text{(2. Zweitorgleichung);}$$

$$\text{Division durch } \underline{I}_2\text{:} \quad \frac{\underline{U}_2}{\underline{I}_2} = \underline{Z}_{21} \cdot \frac{\underline{I}_1}{\underline{I}_2} + \underline{Z}_{22}$$

Es gilt:

$$\frac{-\underline{U}_2}{\underline{I}_2} = \underline{Z}_a; \quad \frac{\underline{I}_1}{\underline{I}_2} = (-\underline{Z}_a - \underline{Z}_{22}) \cdot \frac{1}{\underline{Z}_{21}};$$

eingesetzt in die oberste Gleichung: *Eingangsimpedanz*

$$\underline{Z}_{\text{in}} = \underline{Z}_{11} - \frac{\underline{Z}_{12} \cdot \underline{Z}_{21}}{\underline{Z}_a + \underline{Z}_{22}}$$

Sonderfall Kurzschluss am Ausgang ($\underline{Z}_a = 0$):

$$\underline{Z}_{\text{in}\,K} = \underline{Z}_{11} - \frac{\underline{Z}_{12} \cdot \underline{Z}_{21}}{\underline{Z}_{22}} = \frac{\det \underline{Z}}{\underline{Z}_{22}}$$

Sonderfall Leerlauf am Ausgang ($\underline{Z}_a \to \infty$):

$$\underline{Z}_{\text{in}\,L} = \underline{Z}_{11}$$

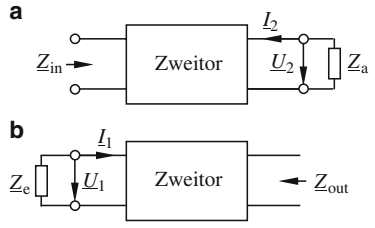


Abb. 93.8 Eingangs- und Ausgangsimpedanz eines belasteten Zweitors. **a** $\underline{Z}_{\text{in}}$ bei $\underline{Z}_a$ am Ausgang, **b** $\underline{Z}_{\text{out}}$ bei $\underline{Z}_e$ am Eingang

In entsprechender Weise erhält man beim Betrieb in Rückwärtsrichtung und Lastimpedanz $\underline{Z}_e$ am Eingang die Ausgangsimpedanz $\underline{Z}_{\text{out}}$ (am Ausgang in das Zweitor hineingesehen):

Ausgangsimpedanz

$$\underline{Z}_{\text{out}} = \underline{Z}_{22} - \frac{\underline{Z}_{12} \cdot \underline{Z}_{21}}{\underline{Z}_{11} + \underline{Z}_e}$$

Sonderfall Kurzschluss am Eingang ($\underline{Z}_e = 0$):

$$\underline{Z}_{\text{out}\,K} = \underline{Z}_{22} - \frac{\underline{Z}_{12} \cdot \underline{Z}_{21}}{\underline{Z}_{11}}$$

Sonderfall Leerlauf am Eingang ($\underline{Z}_e \to \infty$):

$$\underline{Z}_{\text{out}\,L} = \underline{Z}_{22}$$

Durch entsprechenden Ansatz weiterer Zweitorgleichungen können die Ein- und Ausgangsimpedanzen in Abhängigkeit von den $\underline{Y}$-, $\underline{C}$-, $\underline{H}$- und $\underline{A}$-Parametern dargestellt werden. Mit den $\underline{A}$-Parametern ergibt sich:

Eingangsimpedanz

$$\underline{Z}_{\text{in}} = \frac{\underline{A}_{11} \cdot \underline{Z}_a + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_a + \underline{A}_{22}}$$

Sonderfall Kurzschluss am Ausgang ($\underline{Z}_a = 0$):

$$\underline{Z}_{\text{in}\,K} = \frac{\underline{A}_{12}}{\underline{A}_{22}}$$

Sonderfall Leerlauf am Ausgang ($\underline{Z}_a \to \infty$):

$$\underline{Z}_{\text{in}\,L} = \frac{\underline{A}_{11}}{\underline{A}_{21}}$$

Ausgangsimpedanz

$$\underline{Z}_{\text{out}} = \frac{\underline{A}_{22} \cdot \underline{Z}_e + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_e + \underline{A}_{11}}$$

Sonderfall Kurzschluss am Eingang ($\underline{Z}_e = 0$):

$$\underline{Z}_{\text{out}\,K} = \frac{\underline{A}_{12}}{\underline{A}_{11}}$$

Sonderfall Leerlauf am Eingang ($\underline{Z}_e \to \infty$):

$$\underline{Z}_{\text{out}\,L} = \frac{\underline{A}_{22}}{\underline{A}_{21}}$$

93.3.2 Längssymmetrische passive Zweitore

Elektrische Leitungen sind längssymmetrische Zweitore, und deshalb sind die Wellenparameter besonders für diese Zweitore von Bedeutung. Für längssymmetrische Zweitore gilt nach

Tab. 93.11: $\underline{Z}_{11} = \underline{Z}_{22}$, $\underline{Z}_{12} = \underline{Z}_{21}$, $\underline{A}_{11} = \underline{A}_{22}$, $\det \underline{A} = 1$. Dann ergeben sich Eingangs- und Ausgangsimpedanz zu:

Eingangsimpedanz

$$\underline{Z}_{\text{in}} = \underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{11} + \underline{Z}_a} = \frac{\underline{A}_{11} \cdot \underline{Z}_a + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_a + \underline{A}_{11}} \tag{93.1a}$$

Ausgangsimpedanz

$$\underline{Z}_{\text{out}} = \underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{11} + \underline{Z}_e} = \frac{\underline{A}_{22} \cdot \underline{Z}_e + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_e + \underline{A}_{22}} \tag{93.1b}$$

Sie sind gleich, wenn Ein- und Ausgang mit der gleichen Impedanz abgeschlossen werden ($\underline{Z}_e = \underline{Z}_a$). Für die Sonderfälle Kurzschluss und Leerlauf ergibt sich:

Eingangsimpedanz bei Kurzschluss am Ausgang ($\underline{Z}_a = 0$):

$$\underline{Z}_{\text{in}\,K} = \underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{22}} = \frac{\underline{A}_{12}}{\underline{A}_{11}}$$

Eingangsimpedanz bei Leerlauf am Ausgang ($\underline{Z}_a \to \infty$):

$$\underline{Z}_{\text{in}\,L} = \underline{Z}_{11} = \frac{\underline{A}_{11}}{\underline{A}_{21}}$$

Ausgangsimpedanz bei Kurzschluss am Eingang ($\underline{Z}_e = 0$):

$$\underline{Z}_{\text{out}\,K} = \underline{Z}_{22} - \frac{\underline{Z}_{12} \cdot \underline{Z}_{21}}{\underline{Z}_{11}} = \frac{\underline{A}_{12}}{\underline{A}_{22}}$$

Ausgangsimpedanz bei Leerlauf am Eingang ($\underline{Z}_e \to \infty$):

$$\underline{Z}_{\text{out}\,L} = \underline{Z}_{22} = \frac{\underline{A}_{22}}{\underline{A}_{21}}$$

93.3.3 Wellenwiderstand bei passiven längssymmetrischen Zweitoren

In Analogie zu den elektrischen Leitungen bestimmt man diejenige Lastimpedanz $\underline{Z}_a$ am Ausgang, bei der die Eingangsimpedanz $\underline{Z}_{\text{in}}$ gleich $\underline{Z}_a$ ist. Dieser Wert wird *Wellenwiderstand* des Zweitors genannt. Aus (93.1a) folgt damit:

Eingangsimpedanz

$$\underline{Z}_{\text{in}} = \underline{Z}_a = \underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{11} + \underline{Z}_a} = \frac{\underline{A}_{11} \cdot \underline{Z}_a + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_a + \underline{A}_{11}}$$

Eingesetzt folgt:

$$\underline{Z}_{\text{in}} = \underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{11} + \underline{Z}_{\text{in}}} = \frac{\underline{A}_{11} \cdot \underline{Z}_{\text{in}} + \underline{A}_{12}}{\underline{A}_{21} \cdot \underline{Z}_{\text{in}} + \underline{A}_{11}}$$

Nach $\underline{Z}_{\text{in}}$ aufgelöst:

$$\underline{Z}_{\text{in}}^2 = \underline{Z}_{11}^2 - \underline{Z}_{12}^2 \quad \text{bzw.} \quad \underline{Z}_{\text{in}}^2 = \frac{\underline{A}_{12}}{\underline{A}_{21}}$$

Die Ausdrücke $\underline{Z}_{11}^2 - \underline{Z}_{12}^2$ bzw. $\frac{\underline{A}_{12}}{\underline{A}_{21}}$ kann man ersetzen durch das Produkt aus den Eingangsimpedanzen $\underline{Z}_{\text{in}\,L}$ und $\underline{Z}_{\text{in}\,K}$:

$$\begin{aligned} \underline{Z}_{11}^2 - \underline{Z}_{12}^2 &= \underline{Z}_{\text{in}\,L} \cdot \underline{Z}_{\text{in}\,K} \\ &= \underbrace{\frac{\underline{Z}_{11}^2}{\underline{Z}_{11}}}_{\underline{Z}_{\text{in}\,L}} \cdot \underbrace{\left(\underline{Z}_{11} - \frac{\underline{Z}_{12}^2}{\underline{Z}_{22}}\right)}_{\underline{Z}_{\text{in}\,K}} = \underline{Z}_{11}^2 - \underline{Z}_{12}^2 \end{aligned}$$

$$\frac{\underline{A}_{12}}{\underline{A}_{21}} = \underline{Z}_{\text{in}\,L} \cdot \underline{Z}_{\text{in}\,K} = \underbrace{\frac{\underline{A}_{11}}{\underline{A}_{21}}}_{\underline{Z}_{\text{in}\,L}} \cdot \underbrace{\frac{\underline{A}_{12}}{\underline{A}_{11}}}_{\underline{Z}_{\text{in}\,K}} = \frac{\underline{A}_{12}}{\underline{A}_{21}}$$

Damit folgt für den Wellenwiderstand eines passiven längssymmetrischen Zweitors:

$$\textit{Wellenwiderstand} \quad \underline{Z} = \sqrt{\underline{Z}_{\text{in}\,L} \cdot \underline{Z}_{\text{in}\,K}} \tag{93.2}$$

Da es sich um längssymmetrische Zweitore handelt, ist der Wellenwiderstand, vom Ausgang her gesehen, der gleiche.

93.3.4 Übertragungsmaß bei passiven längssymmetrischen Zweitoren

Die Übertragungseigenschaften eines passiven längssymmetrischen Zweitors, der mit dem Wellenwiderstand abgeschlossen ist, können durch die folgenden *Wellenübertragungsmaße* beschrieben werden:

$$\textit{Spannungsdämpfungsmaß} \quad g_u = \ln \frac{\underline{U}_1}{\underline{U}_2}$$

$$\textit{Stromdämpfungsmaß} \quad g_i = \ln \frac{\underline{I}_1}{\underline{I}_2}$$

komplexes Wellendämpfungsmaß

$$g = \frac{1}{2} \cdot (g_u + g_i) = \ln \sqrt{\frac{\underline{U}_1}{\underline{U}_2} \cdot \frac{\underline{I}_1}{\underline{I}_2}} = a + jb$$

a Dämpfungsmaß (wahlweise mit dem Kurzzeichen Np für Neper), b Phasenmaß

Für längssymmetrische Zweitore gilt (ohne Ableitung):

$$\begin{aligned} g = g_u = g_i &= \ln \frac{\underline{U}_1}{\underline{U}_2} = \ln \frac{\underline{I}_1}{\underline{I}_2} \\ &= \ln(\underline{A}_{11} + \sqrt{\underline{A}_{12} \cdot \underline{A}_{21}}) = a + jb \end{aligned} \tag{93.3}$$

Die drei Wellenübertragungsmaße sind gleich.

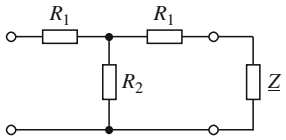


Abb. 93.9 Längssymmetrischer T-Zweitor (zu Beispiel 5)

Beispiel 5

Bei dem längssymmetrischen Zweitor nach Abb. 93.9 soll das Dämpfungsmaß berechnet werden. Er ist mit dem Wellenwiderstand

$$Z = \sqrt{Z_{\text{in}\,L} \cdot Z_{\text{in}\,K}}$$
$$= \sqrt{(R_1 + R_2) \cdot (R_1 + \frac{R_1 \cdot R_2}{R_1 + R_2})} = \sqrt{R_1(R_1 + 2 \cdot R_2)}$$

nach (93.2) abgeschlossen. Mit (93.3) folgt:

$$g = g_u = g_i = \ln \frac{\underline{U}_1}{\underline{U}_2} = \ln \frac{\underline{I}_1}{\underline{I}_2} = \ln(\underline{A}_{11} + \sqrt{\underline{A}_{12} \cdot \underline{A}_{21}})$$
$$= \ln \left(1 + \frac{R_1}{R_2} + \sqrt{\left(2 \cdot R_1 + \frac{R_1^2}{R_2}\right) \frac{1}{R_2}}\right)$$
$$= \ln \left(1 + \frac{R_1}{R_2} + \sqrt{\frac{R_1(R_1 + 2R_2)}{R_2^2}}\right)$$

Für $R_1 = R_2 = R$ erhält man:

$$Z = \sqrt{3 \cdot R^2} = \sqrt{3} \cdot R;$$
$$g = \ln \left(1 + 1 + \sqrt{\frac{R^2 + 2R^2}{R^2}}\right)$$
$$= \ln(2 + \sqrt{3}) = 1{,}32 \text{ (Np)}$$

Der Ausdruck in der Klammer von g gibt das Spannungsverhältnis $U_1/U_2 = 3{,}73$ an. Bei passiven Zweitoren ist stets $U_1 \geq U_2$ und $I_1 \geq I_2$, sodass sich für g ein Wert ≥ 0 ergibt. ◀

93.3.5 Spezielle Zweitore

Die im folgenden behandelten Zweitore werden in der Nachrichtentechnik häufiger angewendet.

93.3.5.1 Doppel-T-Filter

Abb. 93.10 zeigt die Schaltung. Es handelt sich um zwei T-Zweitore, die eingangs- und ausgangsseitig parallel geschaltet sind. Die Schaltung ist am Ausgang nicht belastet. Mit Hilfe

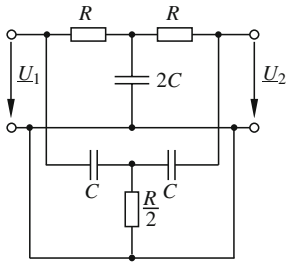


Abb. 93.10 Doppel-T-Filter

der für diese Art der Zusammenschaltung erforderlichen $\underline{Y}$-Parameter (Tab. 93.1) wird die Spannungsübersetzung vorwärts berechnet. Die angegebene Dimensionierung führt zu einem einfachen Ergebnis für $|\underline{v}_{\text{uf}}|$, das hier ohne Ableitung angegeben wird:

$$|\underline{v}_{\text{uf}}| = \frac{(1 - \omega^2 R^2 C^2)^2}{(1 - \omega^2 R^2 C^2)^2 + 16\omega^2 R^2 C^2}$$

Es handelt sich um eine Bandsperre, die bei der Frequenz $\omega_0 = 1/RC$ rechnerisch die Ausgangsspannung Null hat. In der Praxis wird man diesen Wert nicht erreichen können, da die Widerstände und Kondensatoren stets geringe Abweichungen voneinander haben. Bei entsprechender Auswahl der Bauelemente kann $|\underline{v}_{\text{uf}}|_{\omega_0}$ aber $< 10^{-3}$ gemacht werden.

93.3.5.2 Kreuzschaltung

Die Schaltung und die $\underline{Z}$-Parameter dieses Zweitors sind in Abb. 93.11a angegeben.

Die Spannungsübersetzung vorwärts ergibt sich bei Leerlauf am Ausgang nach Tab. 93.8 zu:

$$\underline{v}_{\text{uf}}|_{\underline{Z}_a \to \infty} = \frac{\underline{Z}_{21}}{\underline{Z}_{11}}$$

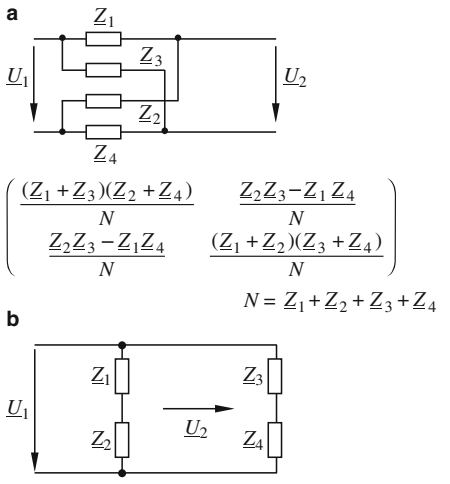


Abb. 93.11 Kreuzzweitor. **a** Schaltung mit $\underline{Z}$-Parametern, **b** umgezeichnet zur *Wheatstone'schen Brückenschaltung*

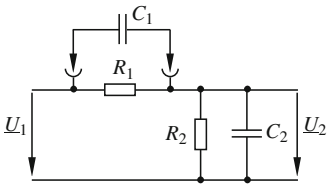


Abb. 93.12 Frequenzkompensierter Spannungsteiler

Sie wird Null für $\underline{Z}_2 \cdot \underline{Z}_3 = \underline{Z}_1 \cdot \underline{Z}_4$. Diese Bedingung lässt sich bei Impedanzen dann erfüllen, wenn sie sowohl für den Real- als auch für den Imaginärteil gilt. Wird das Zweitor nach Abb. 93.11b umgezeichnet, erkennt man die *Wheatstone'sche Brückenschaltung*. Die oben angeführte Bedingung ist die Abgleichbedingung dieser Brücke.

93.3.5.3 Frequenzkompensierter Spannungsteiler

In Abb. 93.12 ist der Fall gezeigt, dass mit den Widerständen R_1 und R_2 ein Spannungsteiler aufgebaut werden soll. Zu dem Widerstand R_2 liegt eine unerwünschte Kapazität C_2 parallel (Schaltungs- und Transistorkapazität wie beim Oszilloskop-Eingang: $R_2 = 1\,\mathrm{M\Omega}$; $C_2 = 50\,\mathrm{pF}$). Damit bildet das Zweitor einen Tiefpass und verfälscht den Signalverlauf des Eingangssignals. Durch Einsetzen und geeignete Dimensionierung des Kondensators C_1 wird die Frequenzabhängigkeit der Spannungsteilung aufgehoben. Der Ausgang des Zweitors ist unbelastet, sodass für die Spannungsübersetzung vorwärts gilt:

$$\underline{v}_{\mathrm{uf}} = 1/\underline{A}_{11};$$

$$\underline{A}_{11} = 1 + \frac{\underline{Z}_1}{\underline{Z}_2} = 1 + \frac{\frac{R_1 \cdot \frac{1}{j\omega C_1}}{R_1 + \frac{1}{j\omega C_1}}}{\frac{R_2 \cdot \frac{1}{j\omega C_2}}{R_2 + \frac{1}{j\omega C_2}}} = 1 + \frac{R_1}{R_2} \cdot \frac{1 + j\omega R_2 C_2}{1 + j\omega R_1 C_1}$$

$$\underline{v}_{\mathrm{uf}} = 1/\underline{A}_{11} = \frac{1}{1 + \frac{R_1}{R_2} \cdot \frac{1+j\omega R_2 C_2}{1+j\omega R_1 C_1}}$$

Wählt man $R_1 \cdot C_1 = R_2 \cdot C_2$, so kürzt sich der Ausdruck, der die Kreisfrequenz ω enthält, heraus. Die Spannungsübersetzung ist frequenzunabhängig und hat den Wert $\underline{v}_{\mathrm{uf}} = \frac{R_2}{R_1+R_2}$. Mit den zwei Widerständen stellt man das gewünschte Teilerverhältnis ein und dimensioniert dann den Kondensator C_1 zu: $C_1 = R_2 C_2 / R_1$.

Leitungen

94

Wilfried Plaßmann und Egon Döring



94.1 Leitungsbeläge und Leitungsgleichungen 1162
94.2 Leitung mit sinusförmigen Spannungen und Strömen 1163
94.3 Leitung mit sinusförmigen Spannungen und Strömen und beliebiger Abschlussimpedanz Z_e 1166
94.4 Verlustlose Leitung 1169
94.5 Daten von Leitungen 1177
94.6 Hochfrequenzleitungen 1177
94.7 s-Parameter 1184
94.8 Kreisdiagramm 1186

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_94

94.1 Leitungsbeläge und Leitungsgleichungen

Unter Leitung wird hier eine homogene Zweidrahtleitung (Doppelleitung) verstanden. Eine Leitung ist *homogen*, wenn die elektrischen Eigenschaften längs der Leitung konstant sind, was in der Regel dadurch erreicht wird, dass sich die mechanischen Abmessungen und das Material längs der Leitung nicht ändern. Zwischen den zwei Leitern liegt eine Spannung, und es fließt ein Strom in den einen Leiter hinein und im zweiten Leiter zurück. Die Anordnung der zwei Leiter zueinander kann unterschiedlich sein, siehe Abb. 94.1: a) Zwei Leiter parallel. Sind sie verdrillt, dann befindet sich Isoliermaterial zwischen den Leitern mit $e_r > 1$. b) Ein Leiter als Hinleiter gegenüber Erde als Rückleiter. c) Zwei konzentrisch angeordnete Leiter, von denen der innere als Hinleiter und der äußere als Rückleiter dient (Koaxialleitung). Abb. 94.1d zeigt die Verläufe des elektrischen und des magnetischen Feldes eines Koaxialkabels. Beide Felder stehen, und das gilt auch für die Leiteranordnungen Abb. 94.1a, b, senkrecht auf der Ausbreitungsrichtung der Welle. Man spricht deshalb bei der Ausbreitung von einem *TEM-Typ* (*t*ransversal *e*lektro*m*agnetisch).

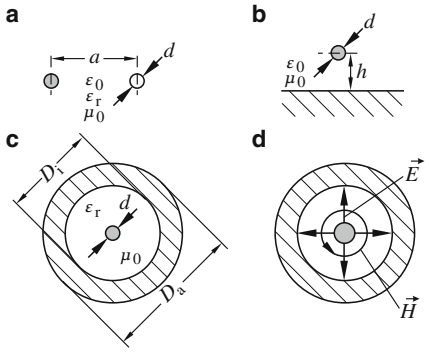


Abb. 94.1 Anordnungen von Zweidrahtleitungen. **a**–**c** Siehe Text, **d** Feldverläufe im Koaxialkabel

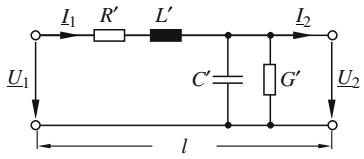


Abb. 94.2 Ersatzschaltbild einer elektrischen Leitung

Abb. 94.2 zeigt das Ersatzschaltbild einer elektrischen Leitung. Der im Leiter fließende Strom hat ein Magnetfeld zur Folge, das durch die Induktivität L' erfasst wird, und er verursacht an dem ohmschen Widerstand R' der Leitung einen Spannungsfall. Hin- und Rückleiter bilden einen Kondensator mit der Kapazität C', und das zwischen den beiden Leitern vorhandene verlustbehaftete Dielektrikum ergibt den Leitwert G'.

Dabei hat man die längs der Leitung kontinuierlich verteilten Größen Induktivität, Widerstand, Kapazität und Leitwert in jeweils einem diskreten Element (L', R', C', G' in Abb. 94.2) zusammengefasst. Der Einfachheit halber berücksichtigt man die Elemente der Rückleitung in denen der Hinleitung und nimmt die Rückleitung als idealen Leiter an. Bei den hier betrachteten homogenen Leitungen nimmt der Wert dieser diskreten Elemente linear mit der Länge der Leitung zu. Deshalb werden sie auf eine Längeneinheit, meist 1 km, bezogen und als *Leitungsbeläge* L', R', C' und G' bezeichnet. Die Werte einer Leitung mit der Länge l_0 erhält man durch Multiplikation der Leitungsbeläge mit l_0:

$$R_{l_0} = R' \cdot l_0; \quad L_{l_0} = L' \cdot l_0;$$
$$C_{l_0} = C' \cdot l_0; \quad G_{l_0} = G' \cdot l_0$$

Die Leitungsbeläge werden häufig in den Einheiten H/km, Ω/km, F/km und S/km oder, den Werten aus der Praxis angepasst, in Ω/km, mH/km, nF/km und mS/km angegeben.

Bei höheren Frequenzen ändern sich die Werte der Leitungsbeläge durch den *Skineffekt*. Tab. 94.1 enthält die Berechnungsformeln für die in Abb. 94.1 gezeigten Leiteranordnungen für niedrige und für hohe Frequenzen (nach [10]).

Es handelt sich bei den in Tab. 94.1 angegebenen Formeln um Näherungslösungen, mit denen die ungefähren Werte der Leitungsbeläge berechnet werden können. Für eine genauere Bestimmung müssen die Frequenzabhängigkeit der Leitungsbeläge und die Feldverzerrungen berücksichtigt werden. Bei dem Induktivitätsbelag wurde das Feld *in* den Leitern vernachlässigt und nur das Feld *zwischen* den Leitern eingesetzt.

Kapazitäts- und Leitwertbelag sind wie folgt verknüpft: $G' = C' \cdot (\gamma/\varepsilon)$. Da sich zwischen den beiden Leitern Isoliermaterial befindet, wird für die elektrische Leitfähigkeit der Ausdruck $\gamma = \gamma_{\text{isolier}}$ eingesetzt. Damit ergeben sich zwei *Kabel-Kenngrößen*:

Zeitkonstante für selbstständige Entladung

$$\tau = R \cdot C = \frac{C'}{G'} = \frac{\varepsilon}{\gamma_{\text{isolier}}} \tag{94.1}$$

Verlustfaktor

$$\tan\delta = \frac{P}{|Q|} = \frac{U^2 \cdot G'}{U^2 \cdot \omega \cdot C'} = \frac{\gamma_{\text{isolier}}}{\omega \cdot \varepsilon} \tag{94.2}$$

P Wirkleistung in W, Q Blindleistung in Var.

Setzt man in Abb. 94.2 nicht die diskreten Elemente ein, sondern die längs der Leitung kontinuierlich vorhandenen Leitungsbeläge, führt die Berechnung dieses Vierpols auf die *Leitungsgleichungen* in Form eines Systems von partiellen Differenzialgleichungen:

Leitungsgleichungen:

$$\frac{\partial^2 u}{\partial x^2} = R'G'u + (R'C' + L'G')\frac{\partial u}{\partial t} + L'C'\frac{\partial^2 u}{\partial t^2} \tag{94.3a}$$

$$\frac{\partial^2 i}{\partial x^2} = R' \cdot G' \cdot i + (R' \cdot C' + L' \cdot G')\frac{\partial i}{\partial t} + L' \cdot C'\frac{\partial^2 i}{\partial t^2} \tag{94.3b}$$

Die Lösung dieses Gleichungssystems ist für *beliebige* Verläufe von $u(t)$ und $i(t)$ bei Leitungen, bei denen alle vier Leitungsbeläge *nicht Null* sind, sehr umfangreich. In der Praxis sind

Tab. 94.1 Leitungsbeläge für verschiedene Leiteranordnungen

Anordnung	Niedrige Frequenzen	Hohe Frequenzen
Widerstandsbelag R'		
Zweidraht-Leitung, Abb. 94.1a	$R' \approx 2 \cdot \frac{\rho \cdot l}{A} \cdot \frac{1}{l} = 2 \cdot \frac{\rho}{A}$ $\approx 2 \cdot \frac{l}{\gamma \cdot A} \cdot \frac{1}{l} = 2 \cdot \frac{1}{\gamma \cdot A}$	$R' \approx 2 \cdot \frac{1}{\gamma \cdot A} \cdot \frac{d}{4} \cdot \sqrt{\pi \cdot f \cdot \gamma \cdot \mu}$ für $\frac{d}{4} \cdot \sqrt{\pi \cdot f \cdot \gamma \cdot \mu} \gg 1$
Eindraht-Leitung, Erde als idealer Rückleiter, Abb. 94.1b	$R' \approx \frac{\rho}{A}$ $\approx \frac{1}{\gamma \cdot A}$	$R' \approx \frac{1}{\gamma \cdot A} \cdot \frac{d}{4} \cdot \sqrt{\pi \cdot f \cdot \gamma \cdot \mu}$ für $\frac{d}{4} \cdot \sqrt{\pi \cdot f \cdot \gamma \cdot \mu} \gg 1$
Koaxial-Leitung, Abb. 94.1c	$R' \approx \frac{1}{\gamma \cdot \frac{d_i^2 \cdot \pi}{4}} + \frac{1}{\gamma \cdot \pi \cdot \frac{D_a^2 - D_i^2}{4}}$	$R' \approx \frac{\sqrt{\pi \cdot f \cdot \gamma \cdot \mu}}{2 \cdot \pi \cdot \gamma \cdot \frac{d_i}{2}} + \frac{\sqrt{\pi \cdot f \cdot \gamma \cdot \mu}}{2 \cdot \pi \cdot \gamma \cdot \frac{D_a}{2}}$
Induktivitätsbelag L'		
Zweidraht-Leitung, Abb. 94.1a	$L' \approx \frac{\mu_0}{\pi} \cdot \ln \frac{2 \cdot a}{d}$	$L' \approx 2{,}3 \cdot \frac{\mu_0}{\pi} \cdot \ln \frac{2 \cdot a}{d}$
Eindraht-Leitung, Erde als idealer Rückleiter, Abb. 94.1b	$L' \approx \frac{\mu_0}{2 \cdot \pi} \cdot \ln \frac{4 \cdot h}{d}$	$L' \approx 1{,}15 \cdot \frac{\mu_0}{\pi} \cdot \ln \frac{4 \cdot h}{d}$
Koaxial-Leitung, Abb. 94.1c	$L' \approx \frac{\mu_0}{2 \cdot \pi} \cdot \ln \frac{D_i}{d}$	$L' \approx \frac{\mu_0}{4 \cdot \pi} \cdot \ln \frac{D_i}{d}$
Kapazitätsbelag C'		
Zweidraht-Leitung, Abb. 94.1a	$C' \approx \frac{\pi \cdot \varepsilon_0 \cdot \varepsilon_r}{\ln \frac{2 \cdot a}{d}}$	$C' \approx \frac{\pi \cdot \varepsilon_0 \cdot \varepsilon_r}{\ln \frac{2 \cdot a}{d}}$
Eindraht-Leitung, Erde als idealer Rückleiter, Abb. 94.1b	$C' \approx \frac{2 \cdot \pi \cdot \varepsilon_0}{\ln \frac{4 \cdot h}{d}}$	$C' \approx \frac{2 \cdot \pi \cdot \varepsilon_0}{\ln \frac{4 \cdot h}{d}}$
Koaxial-Leitung, Abb. 94.1c	$C' \approx \frac{2 \cdot \pi \cdot \varepsilon_0 \cdot \varepsilon_r}{\ln \frac{D_i}{d}}$	$C' \approx \frac{2 \cdot \pi \cdot \varepsilon_0 \cdot \varepsilon_r}{\ln \frac{D_i}{d}}$
Leitwertbelag (Ableitungsbelag) G'		
Zweidraht-Leitung, Abb. 94.1a	$G' \approx \frac{\pi \cdot \gamma}{\ln \frac{2 \cdot a}{d}}$	$G' \approx \frac{\pi \cdot \gamma}{\ln \frac{2 \cdot a}{d}}$
Eindraht-Leitung, Erde als idealer Rückleiter, Abb. 94.1b	$G' \approx \frac{2 \cdot \pi \cdot \gamma}{\ln \frac{4 \cdot h}{d}}$	$G' \approx \frac{2 \cdot \pi \cdot \gamma}{\ln \frac{4 \cdot h}{d}}$
Koaxial-Leitung, Abb. 94.1c	$G' \approx \frac{2 \cdot \pi \cdot \gamma}{\ln \frac{D_i}{d}}$	$G' \approx \frac{2 \cdot \pi \cdot \gamma}{\ln \frac{D_i}{d}}$

Bedeutung der Formelzeichen
ρ spezifischer elektrischer Widerstand in $\Omega\,\mathrm{m}$
γ elektrische Leitfähigkeit in $1/(\Omega\,\mathrm{m})$; $\gamma = 1/\rho$
A Fläche des Leiters in m^2
f Frequenz in Hz
ε_0 elektrische Feldkonstante $\approx 8{,}86 \cdot 10^{-12}\ \frac{\mathrm{A\,s}}{\mathrm{V\,m}}$
μ_0 magnetische Feldkonstante $\approx 4 \cdot \pi \cdot 10^{-7}\ \frac{\mathrm{V\,s}}{\mathrm{A\,m}}$
ε_r Permittivitätszahl (Dielektrizitätszahl), materialabhängig
μ_r Permeabilitätszahl, materialabhängig
ε Permittivität (Dielektrizitätskonstante), $\varepsilon = \varepsilon_0 \cdot \varepsilon_r$
μ Permeabilität, $\mu = \mu_0 \cdot \mu_r$
d, a, h, D_i, D_a siehe Abb. 94.1

häufig die nachfolgend aufgeführten Voraussetzungen entweder ganz oder wenigstens näherungsweise erfüllt, sodass man zu einfachen und überschaubaren Lösungen des Gleichungssystems kommt:

1. Spannung und Strom haben sinusförmigen Verlauf; die Leitungsbeläge können beliebig sein (Abschn. 94.2).
2. Widerstands- und Leitwertbelag werden zu Null angenommen. Dann spricht man von einer *verlustlosen Leitung*. Spannungs- und Stromverlauf können beliebig sein (Abschn. 94.4).

94.2 Leitung mit sinusförmigen Spannungen und Strömen

Für sinusförmige Spannungen und Ströme lässt sich eine einfache Lösung der Differentialgleichungen (94.3a) und (94.3b) angeben.

94.2.1 Allgemeine Lösung

Die allgemeine Lösung lautet:

$$u(t,x) = \underbrace{\hat{u}_1 \cdot e^{-\alpha x} \cdot \cos(\omega t + \varphi_1 - \beta x)}_{u_h} + \underbrace{\hat{u}_2 \cdot e^{\alpha x} \cdot \cos(\omega t + \varphi_2 + \beta x)}_{u_r} \tag{94.4a}$$

$$i(t,x) = \underbrace{\hat{i}_1 \cdot e^{-\alpha x} \cdot \cos(\omega t + \varphi_1 - \beta x)}_{i_h} + \underbrace{\hat{i}_2 \cdot e^{\alpha x} \cdot \cos(\omega t + \varphi_2 + \beta x)}_{i_r} \tag{94.4b}$$

φ_1 beliebiger Anfangswinkel von $\hat{u}$ und $\hat{i}$, wird im folgenden zu Null gesetzt; x Längenkoordinate der Leitung: $x = 0$ Leitungsanfang, $x = l$ Leitungsende; a Dämpfungskoeffizient; β Phasenkoeffizient (siehe (94.6)); $\hat{u}_1$, $\hat{u}_2$, $\hat{i}_1$, $\hat{i}_2$ Integrationskon-

stanten, durch Randbedingungen zu bestimmen (siehe (94.9a), (94.9b), (94.10a), (94.10b)).

Bestimmt man die vier Konstanten $\hat{u}_1, \hat{u}_2, \hat{i}_1, \hat{i}_2$ in (94.4a), (94.4b) so, dass sie z. B. Spannung und Strom am Anfang der Leitung ($x = 0$, siehe (94.9a), (94.9b)) vorgeben, kann man an jedem Ort der Leitung zu jedem Zeitpunkt den Wert von Spannung und Strom berechnen. Beide Verläufe bestehen aus zwei Anteilen: Einer hinlaufenden Welle u_h bzw. i_h mit abnehmender Amplitude in Richtung Leitungsende und einer zurücklaufenden Welle u_r bzw. i_r mit abnehmender Amplitude in Richtung Leitungsanfang (Erklärung für die rücklaufende Welle siehe Abschn. 94.3.1 und 94.4.1).

94.2.2 Wellenwiderstand

Bildet man an beliebigen Orten längs der Leitung zu beliebigen Zeitpunkten den Quotienten aus Spannung und Strom der hinlaufenden *oder* rücklaufenden Welle, ergibt sich ein konstanter Wert, der mit *Wellenwiderstand* $\underline{Z}$ bezeichnet wird und sich mit den Leitungsbelägen berechnen lässt:

Wellenwiderstand (allgemeine Leitung)

$$\underline{Z} = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \tag{94.5}$$

94.2.3 Ausbreitungskoeffizient

Die in (94.4a), (94.4b) auftretenden Größen a und b ergeben sich aus der Lösung der Differenzialgleichungen und sind mit den Leitungsbelägen verknüpft:

Ausbreitungskoeffizient

$$\gamma = \sqrt{(R' + j\omega L') \cdot (G' + j\omega C')} = \alpha + j\beta \tag{94.6}$$

(Schreibweise γ anstatt $\underline{\gamma}$ gewählt), α *Dämpfungskoeffizient*, β *Phasenkoeffizient*.

Der in (94.6) auftretende Dämpfungskoeffizient α ist der Realteil, der Phasenkoeffizient β der Imaginärteil des Ausbreitungskoeffizienten γ.

94.2.4 Verlustlose Leitung

Bei einer verlustlosen Leitung gilt: $R' = G' = 0$, und Wellenwiderstand (94.5) und Ausbreitungskoeffizient γ (94.6) berechnen sich zu:

Wellenwiderstand

$$\underline{Z}_{vL} = \sqrt{\frac{L'}{C'}} \tag{94.7}$$

Ausbreitungskoeffizient

$$\gamma_{vL} = \sqrt{-\omega^2 L'C'} = j\omega\sqrt{L'C'} = j\beta; \quad \alpha = 0 \tag{94.8}$$

Weitere Eigenschaften der verlustlosen Leitung werden im Abschn. 94.4 behandelt.

94.2.5 Lösung mit Zeigerdarstellung

In der Praxis interessiert weniger der zeitliche Verlauf als vielmehr die Amplitude längs der Leitung. Man wählt deshalb eine Lösungsdarstellung mit Zeigern:

$$\underline{U}(x) = \underline{U}_1 \cdot e^{-\gamma x} + \underline{U}_2 \cdot e^{\gamma x}$$
$$\underline{I}(x) = \frac{1}{\underline{Z}}(\underline{U}_1 \cdot e^{-\gamma x} + \underline{U}_2 \cdot e^{\gamma x})$$

$\underline{U}_1$, $\underline{U}_2$ Integrationskonstanten, durch Randbedingungen bestimmbar, $\underline{Z}$ Wellenwiderstand, γ siehe (94.6).

Auch hier bestehen Spannungs- und Stromverlauf aus einer hinlaufenden und einer rücklaufenden Welle. Der in (94.4a), (94.4b) auftretende Dämpfungskoeffizient α ist der Realteil, der Phasenkoeffizient β der Imaginärteil des Ausbreitungskoeffizienten g (siehe (94.6)). Gibt man sich am Leitungsanfang ($x = 0$, Index „a“) die Spannung $\underline{U}_a$ und den Strom $\underline{I}_a$ vor, so kann man die Integrationskonstanten berechnen:

Bedingungen am Leitungsanfang vorgegeben ($x = 0$, $\underline{U}_a$, $\underline{I}_a$):

$$\begin{aligned}
\underline{U}_a &= \underline{U}_1 + \underline{U}_2; \\
\underline{I}_a &= \frac{1}{\underline{Z}}(\underline{U}_1 - \underline{U}_2) \\
\Rightarrow \underline{U}_1 &= \frac{\underline{U}_a + \underline{Z} \cdot \underline{I}_a}{2}; \quad \underline{U}_2 = \frac{\underline{U}_a - \underline{Z} \cdot \underline{I}_a}{2} \\
\Rightarrow \underline{U}(x) &= \underbrace{\frac{\underline{U}_a + \underline{Z} \cdot \underline{I}_a}{2} \cdot e^{-\gamma x}}_{\underline{U}_h} + \underbrace{\frac{\underline{U}_a - \underline{Z} \cdot \underline{I}_a}{2} \cdot e^{\gamma x}}_{\underline{U}_r} \qquad (94.9a) \\
\Rightarrow \underline{I}(x) &= \underbrace{\frac{1}{2} \cdot \left(\frac{\underline{U}_a}{\underline{Z}} + \underline{I}_a\right) \cdot e^{-\gamma x}}_{\underline{I}_h} - \underbrace{\frac{1}{2} \cdot \left(\frac{\underline{U}_a}{\underline{Z}} - \underline{I}_a\right) \cdot e^{\gamma x}}_{\underline{I}_r} \qquad (94.9b)
\end{aligned}$$

$\underline{U}_a$, $\underline{I}_a$ Spannung bzw. Strom am Anfang der Leitung ($x = 0$); $\underline{Z}$ Wellenwiderstand der Leitung; Index „h“: hinlaufende (auf das Leitungsende zulaufende) Welle, Index „r“: rücklaufende (vom Leitungsende zum -anfang zurücklaufende) Welle.

In gleicher Weise kann man Spannung und Strom am Ende der Leitung ($x = l$, Index „e“) vorgeben. Das führt zu den folgenden Gleichungen:

Bedingungen am Leitungsende vorgegeben ($x = l$, $\underline{U}_e$, $\underline{I}_e$):

$$\underline{U}(x = l) = \underline{U}_e = \underline{U}_1 \cdot e^{-\gamma l} + \underline{U}_2 \cdot e^{\gamma l}$$

$$\underline{I}(x = l) = \underline{I}_e = \frac{1}{\underline{Z}} \cdot (\underline{U}_1 \cdot e^{-\gamma l} - \underline{U}_2 \cdot e^{\gamma l})$$

$$\Rightarrow \underline{U}(x) = \underbrace{\frac{\underline{U}_e + \underline{Z} \cdot \underline{I}_e}{2} \cdot e^{\gamma(l-x)}}_{\underline{U}_h} + \underbrace{\frac{\underline{U}_e - \underline{Z} \cdot \underline{I}_e}{2} \cdot e^{-\gamma(l-x)}}_{\underline{U}_r} \tag{94.10a}$$

$$\Rightarrow \underline{I}(x) = \underbrace{\frac{1}{2} \cdot \left(\frac{\underline{U}_e}{\underline{Z}} + \underline{I}_e\right) \cdot e^{\gamma(l-x)}}_{\underline{I}_h} - \underbrace{\frac{1}{2} \cdot \left(\frac{\underline{U}_e}{\underline{Z}} - \underline{I}_e\right) \cdot e^{-\gamma(l-x)}}_{\underline{I}_r} \tag{94.10b}$$

$\underline{U}_e$, $\underline{I}_e$ Spannung bzw. Strom am Ende der Leitung ($x = l$); $\underline{Z}$ Wellenwiderstand der Leitung, Index „h“: hinlaufende (auf das Leitungsende zulaufende) Welle, Index „r“: rücklaufende (vom Leitungsende zum -anfang zurücklaufende) Welle.

Die Gleichungen (94.9a), (94.9b) und (94.10a), (94.10b) werden auch als *physikalische Form der Leitungsgleichungen* bezeichnet.

94.2.6 Unendlich lange Leitung

Für $l \to \infty$ verschwindet in (94.9a), (94.9b) die rücklaufende Welle, und es bleibt für die Verläufe von Strom und Spannung längs der Leitung:

$$\underline{U}(x) = \underbrace{\frac{\underline{U}_a + \underline{Z} \cdot \underline{I}_a}{2} \cdot e^{-\gamma x}}_{\underline{U}_h} = \underbrace{\underline{U}_a \cdot e^{-\gamma x}}_{\underline{U}_h} \tag{94.11a}$$

$$\underline{I}(x) = \underbrace{\frac{1}{2} \cdot \left(\frac{\underline{U}_a}{\underline{Z}} + \underline{I}_a\right) \cdot e^{-\gamma x}}_{\underline{I}_h} = \underbrace{\left(\frac{\underline{U}_a}{\underline{Z}}\right) \cdot e^{-\gamma x}}_{\underline{I}_h} \tag{94.11b}$$

mit $\underline{I}_a = \frac{\underline{U}_a}{\underline{Z}}$. $\underline{U}_a$, $\underline{I}_a$ Spannung bzw. Strom am Anfang der Leitung ($x = 0$), $\underline{Z}$ Wellenwiderstand.

94.2.7 Anpassung

Ein für die Praxis interessanter Fall ist der, dass am Ende einer *endlich langen Leitung* eine Impedanz $\underline{Z}_e$ mit dem gleichen Wert wie der Wellenwiderstand $\underline{Z}$ der Leitung angeschlossen ist (Abb. 94.3):

$$\textit{Anpassung} \quad \underline{Z}_e = \underline{Z}$$

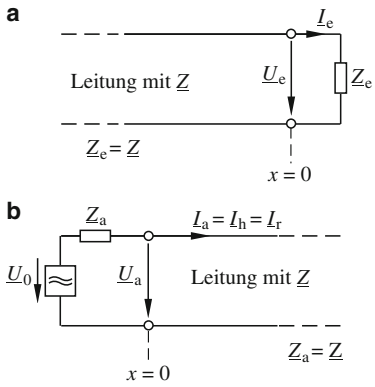


Abb. 94.3 Anpassung **a** am Ausgang, **b** am Eingang einer Leitung

Dann gilt am Ende der Leitung: $\underline{I}_e = \underline{U}_e/\underline{Z}$. Diese Bedingung gilt sowohl für das Leitungsende (aber noch in der Leitung) als auch für die Lastimpedanz $\underline{Z}_e$. Nach (94.9a), (94.9b) verschwindet die rücklaufende Welle. Dieser Fall wird *Anpassung* genannt und hat sehr große Bedeutung bei der Signalübertragung auf verlustlosen Leitungen. Auch am Leitungseingang spricht man von Anpassung, wenn der Innenwiderstand (Impedanz) des Generators gleich dem Wellenwiderstand der Leitung ist (Abb. 94.3b). Die Bedeutung der Anpassung am Ein- und Ausgang einer Leitung wird bei der Ausbreitung von Wanderwellen auf einer Leitung (Abschn. 94.4.2) gezeigt.

94.2.8 Phasengeschwindigkeit, Gruppengeschwindigkeit

Bei der Ausbreitung einer Welle auf der Leitung (hinlaufende *oder* zurücklaufende Welle betrachtet) findet man die gleiche Phasenlage nach (94.4a), (94.4b) bei konstant gehaltener Zeit nach $x_0 \cdot \beta = 2 \cdot \pi$ wieder. Andererseits ergibt sich an einem festen Ort die gleiche Phasenlage nach der Periodendauer $T = 1/f$. Ein Vergleich liefert:

$$\beta \cdot x_0 = 2\pi = \omega \cdot T \quad \Rightarrow \quad v = \frac{x_0}{T} = \frac{\omega}{\beta}$$

Phasengeschwindigkeit

$$v = \frac{\omega}{\beta} = \frac{2 \cdot \pi \cdot f}{\beta} \tag{94.12}$$

Mit dieser Geschwindigkeit breitet sich eine Welle auf der Leitung aus.

Die mittlere Leistung dagegen breitet sich auf der Leitung mit der *Gruppengeschwindigkeit*

$$v_{\mathrm{gr}} = \frac{d\omega}{d\beta} \tag{94.13}$$

aus. Ist die Phasengeschwindigkeit unabhängig von der Frequenz, sind Phasen- und Gruppengeschwindigkeit gleich: $\omega =$

$v \cdot \beta$ (aus (94.12)). Daraus folgt: $\frac{d\omega}{d\beta} = v = v_{gr}$ (siehe auch Abschn. 94.4 „Verlustlose Leitung“). Man spricht von *Dispersion*, wenn Phasen- und Gruppengeschwindigkeit *nicht* übereinstimmen:

$$\textit{Dispersion} \quad v \neq v_{gr}$$

Beispiel 1

Von einer verlustlosen Fernsprechleitung sind gegeben: $L' = 2{,}2\,\text{mH/km}$; $C' = 6\,\text{nF/km}$. Die Frequenz beträgt 1 kHz. Daraus berechnen sich folgende Größen:

$$\begin{aligned}
\underline{Z} = \underline{Z}_{vL} &= \sqrt{\frac{L'}{C'}} = \sqrt{\frac{2{,}2 \cdot 10^{-3}\,\text{V s} \cdot \text{V km}}{\text{A km} \cdot 6\,\text{A s} \cdot 10^{-9}}} \\
&= \sqrt{\frac{2{,}2}{6} \cdot 10^6\,\Omega} = 606\,\Omega \\
\gamma = \gamma_{vL} &= \sqrt{j\omega L' j\omega C'} = \sqrt{-\omega^2 L' C'} = j\omega\sqrt{L'C'} \\
&= j2 \cdot \pi \cdot 10^5\,\frac{1}{\text{s}}\sqrt{2{,}2 \cdot 6 \cdot 10^{-12}}\,\frac{\text{s}}{\text{km}} = j\frac{2{,}28}{\text{km}} \\
\Rightarrow \quad \alpha = 0; \quad \beta &= \frac{2{,}28}{\text{km}} \\
v = \frac{\omega}{\beta} &= \frac{2 \cdot \pi \cdot f}{\beta} \\
&= \frac{2 \cdot \pi \cdot 10^3\,\text{km}}{\text{s} \cdot 2{,}28} = 2{,}75 \cdot 10^3\,\frac{\text{km}}{\text{s}}
\end{aligned}$$

◄

94.3 Leitung mit sinusförmigen Spannungen und Strömen und beliebiger Abschlussimpedanz Z_e

94.3.1 Reflexionsfaktor, Übertragungsfaktor

Schließt man eine endlich lange Leitung am Ende mit einer beliebigen Impedanz $\underline{Z}_e$ ab, so gilt bei $x = l$: $\underline{U}_e = \underline{Z}_e \cdot \underline{I}_e$. In (94.10a) eingesetzt, erhält man:

$$\underline{U}(x) = \underbrace{\frac{\underline{Z}_e + \underline{Z}}{2} \cdot \underline{I}_e \cdot e^{\gamma(l-x)}}_{\underline{U}_h} + \underbrace{\frac{\underline{Z}_e - \underline{Z}}{2} \cdot \underline{I}_e \cdot e^{-\gamma(l-x)}}_{\underline{U}_r}$$

$\underline{Z}_e$ am Ende der Leitung angeschlossene Impedanz, $\underline{Z}$ Wellenwiderstand der Leitung.

Das Verhältnis von zurücklaufender Welle zu vorlaufender Welle ist:

$$\frac{\underline{U}_r}{\underline{U}_h} = \frac{\underline{Z}_e - \underline{Z}}{\underline{Z}_e + \underline{Z}} \cdot e^{-2\gamma(l-x)} \tag{94.14}$$

Dieses Verhältnis *am Ende* der Leitung nennt man *Reflexionsfaktor* $\underline{r}$:

Reflexionsfaktor

$$\underline{r} = \left.\frac{\underline{U}_r}{\underline{U}_h}\right|_{x=l} = \underbrace{\frac{\underline{Z}_e - \underline{Z}}{\underline{Z}_e + \underline{Z}}}_{\text{„1“}} = \frac{\frac{\underline{Z}_e}{\underline{Z}} - 1}{\frac{\underline{Z}_e}{\underline{Z}} + 1} = \underbrace{\frac{\underline{z} - 1}{\underline{z} + 1}}_{\text{„2“}} \tag{94.15}$$

mit $\frac{\underline{Z}_e}{\underline{Z}} = \underline{z}$. $\underline{Z}_e$ am Ende der Leitung angeschlossene Impedanz; $\underline{Z}$ Wellenwiderstand der Leitung.

Die Darstellung „2“ wird der Darstellung „1“ vorgezogen. Für das entsprechende Verhältnis der Ströme ergibt sich der gleiche Ausdruck. Am Leitungs*ende* gelten die zwei Beziehungen:

$$1) \quad \underline{U}_e/\underline{I}_e = \underline{Z} \tag{94.16a}$$

(noch innerhalb der Leitung gemäß Definition von $\underline{Z}$ nach Abschn. 94.2.2),

$$2) \quad \underline{U}_e/\underline{I}_e = \underline{Z}_e \quad \text{(Ohm’sches Gesetz für } \underline{Z}_e\text{)} \tag{94.16b}$$

Erst durch Einführen einer rücklaufenden Welle können beide Bedingungen erfüllt werden. Mit (94.15) lässt sich auch die Amplitude der in die Impedanz $\underline{Z}_e$ „hineinlaufenden“ Welle $\underline{U}_{h2}$ berechnen:

Ansatz:

$$\underline{U}_{h1} + \underline{U}_{r1} = \underline{U}_{h2};$$

mit $\underline{U}_{r1} = \underline{U}_{h1}\frac{\underline{Z}_e - \underline{Z}}{\underline{Z}_e + \underline{Z}}$ folgt:

$$\begin{aligned}
\underline{U}_{h2} &= \underline{U}_{h1} + \underline{U}_{h1}\frac{\underline{Z}_e - \underline{Z}}{\underline{Z}_e + \underline{Z}} = \underline{U}_{h1}\left(1 + \frac{\underline{Z}_e - \underline{Z}}{\underline{Z}_e + \underline{Z}}\right) \\
&= \underline{U}_{h1}\frac{2\underline{Z}_e}{\underline{Z}_e + \underline{Z}} = \underline{U}_{h1} \cdot \ddot{u}_u
\end{aligned}$$

in analoger Weise:

$$\underline{I}_{h2} = \underline{I}_{h1}\frac{2\underline{Z}}{\underline{Z}_e + \underline{Z}} = \underline{I}_{h1} \cdot \ddot{u}_i$$

Spannungsübertragungsfaktor

$$\ddot{u}_u = \frac{2\underline{Z}_e}{\underline{Z}_e + \underline{Z}} \tag{94.17a}$$

Stromübertragungsfaktor

$$\ddot{u}_i = \frac{2\underline{Z}}{\underline{Z}_e + \underline{Z}} \tag{94.17b}$$

$\underline{U}_{h1}, \underline{I}_{h1}$ vom Generator auf das Leitungsende zulaufende Welle; $\underline{U}_{r1}, \underline{I}_{r1}$ in Richtung Leitungsanfang reflektierte Welle; $\underline{U}_{h2}, \underline{I}_{h2}$ in die Impedanz $\underline{Z}_e$ hineinlaufende Welle; $\underline{Z}_e$ Impedanz am Ende der Leitung; $\underline{Z}$ Wellenwiderstand der Leitung.

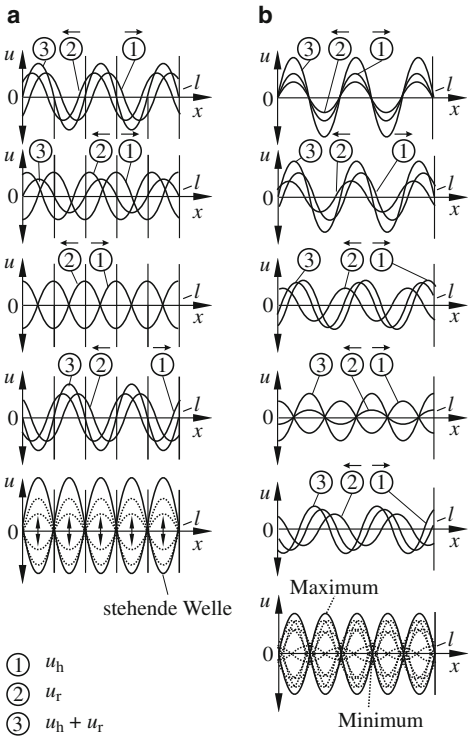


Abb. 94.4 Hin- und rücklaufende Welle am Ausgang. **a** Kurzschluss, **b** Fehlanpassung

Abb. 94.4 zeigt hin- und rücklaufende Welle der Spannung zu verschiedenen Zeitpunkten, wenn am Leitungsende Kurzschluss (a) oder Fehlanpassung (b) herrscht. Bei Kurzschluss hat die rücklaufende Welle stets die entgegengesetzt gleich große Amplitude wie die hinlaufende Welle, weil nur so die Bedingung erfüllbar ist, dass die Spannung am Ausgang wegen des Kurzschlusses zu jedem Zeitpunkt Null ist. Auch bei Fehlanpassung ermöglicht die rücklaufende Welle, dass die genannten zwei Bedingungen (94.16a), (94.16b) erfüllt werden können.

Bisher wurde am Ende der Leitung eine diskrete Impedanz $\underline{Z}_e$ eingesetzt. Dort kann aber auch eine zweite Leitung mit dem Wellenwiderstand $\underline{Z}_{el}$ angeschlossen werden (Stoßstelle zweier Leitungen). Die Berechnung geschieht mit den angegebenen Formeln, indem $\underline{Z}_e$ durch $\underline{Z}_{el}$ ersetzt wird.

94.3.2 Eingangsimpedanz

Die Gleichungen (94.10a), (94.10b) werden umgeformt, und es wird von der Beziehung

$$\sinh\gamma(l-x) = \frac{e^{\gamma(l-x)} - e^{-\gamma(l-x)}}{2};$$

$$\cosh\gamma(l-x) = \frac{e^{\gamma(l-x)} + e^{-\gamma(l-x)}}{2};$$

$$\tanh\gamma(l-x) = \frac{\sinh\gamma(l-x)}{\cosh\gamma(l-x)} = \frac{e^{\gamma(l-x)} - e^{-\gamma(l-x)}}{e^{\gamma(l-x)} + e^{-\gamma(l-x)}}$$

Gebrauch gemacht. Dann können die Leitungsgleichungen in der *mathematischen Form* dargestellt werden:

$$\underline{U}(x) = \underline{U}_e \cdot \cosh\gamma(l-x) + \underline{Z} \cdot \underline{I}_e \cdot \sinh\gamma(l-x)$$

$$\underline{I}(x) = \underline{I}_e \cdot \cosh\gamma(l-x) + \frac{\underline{U}_e}{\underline{Z}} \cdot \sinh\gamma(l-x) \qquad (94.18)$$

Benutzt man weiterhin

$$\underline{Z}_a = \frac{\underline{U}_a}{\underline{I}_a}; \quad \underline{Z}_e = \frac{\underline{U}_e}{\underline{I}_e};$$

$$\underline{U}(x=0) = \underline{U}_a = \underline{U}_e \cdot \cosh\gamma l + \underline{Z} \cdot \underline{I}_e \cdot \sinh\gamma l$$

$$\underline{I}(x=0) = \underline{I}_a = \underline{I}_e \cdot \cosh\gamma l + \frac{\underline{U}_e}{\underline{Z}} \cdot \sinh\gamma l$$

folgt:

Eingangsimpedanz am Anfang einer beliebig langen Leitung:

$$\underline{Z}_a = \underline{Z} \cdot \frac{\underline{Z}_e + \underline{Z} \cdot \tanh\gamma l}{\underline{Z} + \underline{Z}_e \cdot \tanh\gamma l} \qquad (94.19)$$

$\underline{Z}_e$ am Leitungsende angeschlossene Impedanz; $\underline{Z}$ Wellenwiderstand der Leitung.

Da die Funktion $\tanh x$ für $x > 3$ nahezu 1 ist, hat eine hinreichend lange Leitung, d. h. $\gamma \cdot l > 3$, eine Eingangsimpedanz, die, unabhängig von der Impedanz am Leitungsende, näherungsweise gleich dem Wellenwiderstand der Leitung ist.

Eingangsimpedanz am Anfang einer langen Leitung:

$$\underline{Z}_a \approx \underline{Z} \quad \text{für } \gamma \cdot l > 3 \qquad (94.20)$$

$\underline{Z}$ Wellenwiderstand der Leitung.

94.3.3 Verzerrungsfreie Leitung

Bei dieser Leitung sind Dämpfungskoeffizient und Phasengeschwindigkeit unabhängig von der Frequenz. Die Bedingung lautet:

Verzerrungsfreie Leitung:

$$\frac{R'}{L'} = \frac{G'}{C'}$$

$$\Rightarrow \quad \alpha = R'\sqrt{\frac{C'}{L'}}; \quad \underline{Z} = \sqrt{\frac{L'}{C'}}; \quad \beta = \omega\sqrt{L'C'} \qquad (94.21)$$

Damit werden Signale unabhängig von der Frequenz gedämpft und verzögert. Ein beliebiges Signal ändert also nicht seine Kurvenform. Es tritt am Leitungsende lediglich mit geringerer Amplitude und zeitverzögert auf.

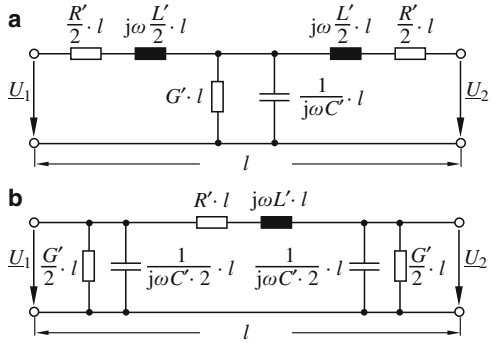


Abb. 94.5 Leitung als Vierpol. **a** T-Vierpol, **b** π-Vierpol

94.3.4 Leitung als Vierpol

94.3.4.1 Allgemeine Ersatzschaltung

Häufig liegen Leitungen gegebener Länge l vor, bei denen ein Generator ein Signal in den Eingang einspeist und bei denen nur das Signal am Ende der Leitung interessiert. Man betrachtet die Leitung deshalb als Vierpol und wendet die Vierpolgesetze an. Aus (94.18) kann man die $\underline{A}$-Parameter ableiten:

$$(\underline{A}) = \begin{pmatrix} \cosh\gamma l & \underline{Z}\cdot\sinh\gamma l \\ \frac{\sinh\gamma l}{\underline{Z}} & \cosh\gamma l \end{pmatrix} \tag{94.22}$$

Es gilt: $\det A = \cosh^2\gamma l - \sinh^2\gamma l = 1$; $\underline{A}_{11} = \underline{A}_{22}$. Der Vierpol „Leitung“ ist damit widerstands- und übertragungssymmetrisch, was sich unmittelbar aus der Anschauung bestätigen lässt (Abschn. 93.2.1 und 93.2.2). Es sind nur zwei Kenngrößen zur Beschreibung erforderlich: der Wellenwiderstand $\underline{Z}$ und das Produkt aus dem Ausbreitungskoeffizienten γ und der Leitungslänge l. Gleichung (94.19) stimmt mit dem Ausdruck für die Eingangsimpedanz eines am Ausgang belasteten Vierpols überein, siehe Tab. 93.7. Abb. 94.5 stellt die Leitung als Elementarvierpol dar.

Vergleicht man (94.22) mit den entsprechenden Matrizen des T- und π-Vierpols (Tab. 93.4), lässt sich eine elektrische Leitung der Länge l als Vierpol darstellen. Es gelten folgende Beziehungen:

π-Vierpol:

$$\begin{aligned}
\underline{A}_{11} &= \underline{A}_{22} = \cosh\gamma l = 1 + (R+j\omega L)\cdot(G+j\omega C)\cdot\frac{1}{2} \\
\underline{A}_{12} &= \underline{Z}\cdot\sinh\gamma l = R + j\omega L \\
\underline{A}_{21} &= \frac{\sinh\gamma l}{\underline{Z}} = (G+j\omega C) + (R+j\omega L)\cdot(G+j\omega C)^2\cdot\frac{1}{4} \\
&= (G+j\omega C)\cdot\left(1 + \frac{(R+j\omega L)\cdot(G+j\omega C)}{4}\right)
\end{aligned} \tag{94.23a}$$

T-Vierpol:

$$\begin{aligned}
\underline{A}_{11} &= \underline{A}_{22} = \cosh\gamma l = 1 + (R+j\omega L)\cdot(G+j\omega C)\cdot\frac{1}{2} \\
\underline{A}_{12} &= \underline{Z}\cdot\sinh\gamma = (R+j\omega L) + (G+j\omega C)\cdot(R+j\omega L)^2\cdot\frac{1}{4} \\
&= (R+j\omega L)\cdot\left(1 + \frac{(G+j\omega C)\cdot(R+j\omega L)}{4}\right) \\
\underline{A}_{21} &= \frac{\sinh\gamma l}{\underline{Z}} = G + j\omega C;
\end{aligned} \tag{94.23b}$$

$\underline{Z}$ Wellenwiderstand der Leitung.

Weiterhin gilt für (94.23a), (94.23b): $R = R'\cdot l$; $L = L'\cdot l$; $C = C'\cdot l$; $G = G'\cdot l$.

Vergleicht man die $\underline{A}$-Parameter beider Vierpole mit gleichem Index miteinander, fällt auf, dass $\underline{A}_{12}$ und $\underline{A}_{21}$ *nicht übereinstimmen*. Das ist die Folge davon, dass die *homogene Leitung* durch *diskrete Elemente* nachgebildet worden ist. Man sieht, dass das nur bedingt zulässig ist.

Durch Umformen der entsprechenden Gleichung (94.23a), (94.23b) und Zerlegung in Real- und Imaginärteil können die Elemente des π- bzw. T-Vierpols aus dem Wellenwiderstand $\underline{Z}$ und γl bestimmt werden:

π-Vierpol:

$$\begin{aligned}
R + j\omega L &= \underline{Z}\cdot\sinh(\gamma l), \\
G + j\omega C &= \frac{2}{\underline{Z}}\cdot\tanh\left(\frac{\gamma l}{2}\right),
\end{aligned}$$

T-Vierpol:

$$\begin{aligned}
R + j\omega L &= 2\cdot\underline{Z}\cdot\tanh\left(\frac{\gamma l}{2}\right), \\
G + j\omega C &= \frac{1}{\underline{Z}}\cdot\sinh(\gamma l)
\end{aligned} \tag{94.24}$$

Beide Vierpole erfassen die Leitung, wie oben begründet, nur näherungsweise.

94.3.4.2 Elektrisch kurze Leitung

Gilt für eine Leitung $|\gamma\cdot l| \ll 1$, können in (94.18) bzw. (94.22) Näherungen eingesetzt werden:

$$\begin{aligned}
&\sinh(\gamma\cdot l) \approx \gamma\cdot l, \\
&\cosh(\gamma\cdot l) \approx 1 + \frac{1}{2}(\gamma\cdot l)^2 \quad \text{für } |\gamma\cdot l| \ll 1
\end{aligned} \tag{94.25}$$

Damit ergeben sich die Elementarvierpole nach Abb. 94.5. Das sind aber auch die Ersatzschaltungen, die für die Ableitung der Leitungsgleichungen (94.3a), (94.3b) verwendet wurden und bei denen die Leitungslängen infinitesimal klein waren. Demnach darf das Ersatzschaltbild nach Abb. 94.5 nur dann verwendet werden, wenn die Bedingung nach (94.25) erfüllt ist.

Bei einer elektrisch kurzen Leitung ist längs der Leitung zu einem beliebig angenommenen Zeitpunkt überall die gleiche Spannung und der gleiche Strom vorhanden.

94.4 Verlustlose Leitung

94.4.1 Eigenschaften

Eine Leitung ist verlustlos, wenn gilt: $R' = G' = 0$. Damit folgen die Eigenschaften der Leitung zu:

Verlustlose Leitung:

$$\alpha = 0 \quad \Rightarrow \quad \underline{Z}_{vL} = \sqrt{\frac{L'}{C'}}; \quad \beta_{vL} = \omega\sqrt{L'C'};$$
$$v_{vL} = \frac{\omega}{\beta} = \frac{\omega}{\omega\sqrt{L'C'}} = \frac{1}{\sqrt{L'C'}} \qquad (94.26)$$

Sowohl der Wellenwiderstand $\underline{Z}_{vL}$ als auch die Phasengeschwindigkeit v_{vL} sind frequenzunabhängig. Der Dämpfungskoeffizient ist erwartungsgemäß Null, der Wellenwiderstand $\underline{Z}_{vL}$ ist reell. Die Lösung der Differentialgleichung (94.3a) lautet damit:

Lösung der Wellengleichung:

$$u(x,t) = u(x \pm v \cdot t) \qquad (94.27)$$

Die sich ausbreitenden Wellen heißen *Wanderwellen* und haben die Eigenschaft, dass beliebige Signalverläufe bei der Ausbreitung in der Leitung weder in der Amplitude noch in der Signalform geändert werden. Das gilt aber nur, wenn die hinlaufende bzw. rücklaufende Welle *für sich* betrachtet wird. Bei Reflexionen am Leitungsanfang oder -ende gibt es Signalverfälschungen durch Überlagerungen, siehe Abschn. 94.4.2 und Beispiel 2.

Die rücklaufende Welle lässt sich anschaulich erklären: Der Ausgang der Leitung sei offen. Da weder in der Leitung noch im Widerstand am Ausgang Leistung verbraucht wird, fließt die vom Generator zum Ausgang transportierte Leistung als rücklaufende Welle wieder zum Generator zurück.

Verlustarme Leitung: Auch wenn die Bedingung $R' = G' = 0$ in der Praxis nicht erfüllt ist, kann man doch häufig eine Leitung als „nahezu verlustlos" ansehen, wenn gilt: $\omega L' \gg R'$; $\omega C' \gg G'$. Sie wird dann als *verlustarme* Leitung bezeichnet.

94.4.2 Wanderwellen bei Reflexion am Leitungsein- und -ausgang

Abb. 94.6a zeigt eine Anordnung, bei der zum Zeitpunkt $t = 0$ der Schalter geschlossen wird. Es gilt: $R_i \neq Z \neq R_e$. u_a ist nach der Spannungsteilerregel:

$$u_a = U_0 \cdot \frac{Z}{R_i + Z} = u_{h1}^{(1)}$$

Eine Welle mit dieser Amplitude breitet sich auf der Leitung in Richtung Leitungsende aus und erreicht es nach der Laufzeit T.

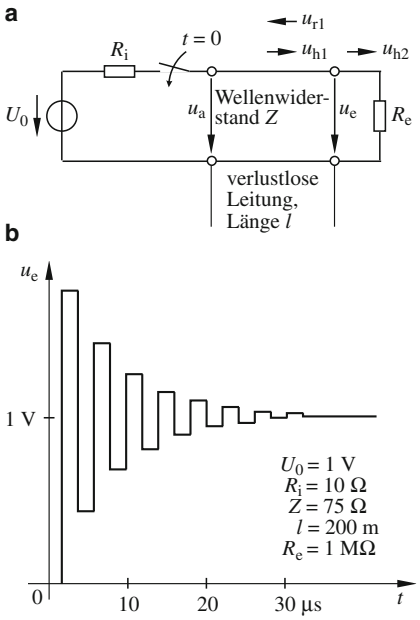


Abb. 94.6 Reflexion. **a** Modell, **b** Beispiel

Am Leitungsende muss gelten:

$$u_{h1}^{(1)} + u_{r1}^{(1)} = u_{h2}^{(1)};$$
$$u_{r1}^{(1)} = u_{h1}^{(1)} \frac{R_e - Z}{R_e + Z} \quad \text{(mit Gleichung (94.15))}$$
$$u_{h2}^{(1)} = u_{h1}^{(1)} \frac{2R_e}{R_e + Z} \quad \text{(mit Gleichung (94.17a))}$$

u_{h1} auf das Leitungsende zulaufende Welle, u_{r1} in Richtung Leitungsanfang reflektierte Welle, u_{h2} in den Widerstand R_e hineinlaufende Welle, R_e Widerstand am Ende der Leitung, Z Wellenwiderstand der Leitung (reell, da verlustlose Leitung).

Für $T < t < 3T$ ist die Spannung an R_e gleich $u_{h2}^{(1)} = \frac{Z}{R_i+Z} \cdot \frac{2R_e}{R_e+Z} \cdot U_0$. Die rücklaufende Welle $u_{r1}^{(1)}$ wird nach $2T$ am Leitungsanfang ebenfalls teilweise reflektiert, die Amplitude ist:

$$u_{h1}^{(2)} = u_{r1}^{(1)} \frac{R_i - Z}{R_i + Z} = \frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} \cdot \frac{Z}{R_i + Z} \cdot U_0$$

Erreicht diese Welle nach $3T$ das Leitungsende, gehen davon

$$u_{h2}^{(2)} \cdot \ddot{u}_u = \frac{2R_e}{R_e + Z} \cdot \frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} \cdot \frac{Z}{R_i + Z} \cdot U_0$$

in den Widerstand R_e und addieren sich zu der bereits seit T vorhandenen Spannung $u_{h2}^{(1)}$. Für $3T < t < 5T$ ist die Spannung an R_e:

$$u_{h2}^{(1)} + u_{h2}^{(2)} = \frac{Z}{R_i + Z} \cdot \frac{2R_e}{R_e + Z} \cdot U_0$$
$$+ \frac{Z}{R_i + Z} \cdot \frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} \cdot \frac{2R_e}{R_e + Z} \cdot U_0$$

Für $5T < t < 7T$ folgt für die Spannung am Ende der Leitung

$$u_{h2}^{(1)} + u_{h2}^{(2)} + u_{h2}^{(3)} = \frac{Z}{R_i + Z} \cdot \frac{2R_e}{R_e + Z} \cdot U_0 + \frac{Z}{R_i + Z} \cdot \frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} \cdot \frac{2R_e}{R_e + Z} \cdot U_0$$

$$u_{h2}^{(1)} + u_{h2}^{(2)} + u_{h2}^{(3)} = \left(\frac{Z}{R_i + Z} \cdot \frac{2R_e}{R_e + Z} \cdot U_0 \right) \cdot \left(1 + \frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} + \left[\frac{R_e - Z}{R_e + Z} \cdot \frac{R_i - Z}{R_i + Z} \right]^2 \right)$$

Auf diese Weise kann man den Spannungsverlauf am Ende der Leitung berechnen.

Beispiel 2

Ein Spannungssprung von 0 auf 1 V, den ein Generator mit dem Innenwiderstand $R_i = 10\,\Omega$ erzeugt, wird über eine 200 m lange Messleitung (Koaxialkabel) mit dem Wellenwiderstand $Z = 75\,\Omega$ auf den Eingang eines Oszilloskopes mit dem Eingangswiderstand $R_e = 1\,\mathrm{M}\Omega$ gegeben. Abb. 94.6b zeigt den dargestellten Verlauf. Es ist danach sogar möglich, dass die Spannung am Ausgang der Leitung zeitweise *über* dem stationären Endwert liegt. ◀

94.4.3 Elektrisch lange Leitung

Die im Abschn. 94.4.2, Wanderwellen, und Abb. 94.6b dargestellten Mehrfachreflexionen am Ein- und Ausgang der Leitung bei Fehlanpassung treten nur dann störend in Erscheinung, wenn die Leitung *„elektrisch lang"* ist. Die Bedingung lautet:

Elektrisch lange Leitung:

$$\frac{l}{\lambda} > 0{,}01 \quad \text{mit } \lambda = \frac{v}{f} \tag{94.28}$$

l Länge der Leitung in m; v Phasengeschwindigkeit in m/s (94.26); f Signalfrequenz in Hz.

Dabei wurden (94.6), (94.12) und (94.25) verwendet:

$$|\gamma \cdot l| < 0{,}1$$

$$|\gamma \cdot l| = |j \cdot \beta \cdot l| = \frac{\omega \cdot l}{v} \quad \Rightarrow \quad \frac{2 \cdot \pi \cdot f \cdot l}{v} > 0{,}1$$

$$\Rightarrow \quad \frac{f \cdot l}{v} > \frac{0{,}1}{2\pi}$$

oder $\frac{f \cdot l}{v} > 0{,}01$.

Alternativ spricht man von *elektrisch kurzer Leitung*:

Elektrisch kurze Leitung:

$$\frac{l}{\lambda} < 0{,}01 \quad \text{mit } \lambda = \frac{v}{f} \tag{94.29}$$

l Länge der Leitung in m; v Phasengeschwindigkeit in m/s (94.26); f Signalfrequenz in Hz.

Anschaulich bedeutet eine elektrisch lange Leitung, dass zu einem beliebig angenommenen Zeitpunkt die Spannungs- und Stromverteilung längs der Leitung nicht mehr konstant ist.

Beispiel 3

Zur Darstellung von unbekannten Spannungsverläufen auf dem Oszilloskopschirm werden Koaxialkabel als Signalleitung verwendet. Ein solches Kabel hat häufig eine Länge von 2 m und eine Phasengeschwindigkeit $v = 200.000\,\mathrm{km/s}$. Die Signalfrequenz soll 1 MHz betragen (zur Vereinfachung zunächst sinusförmiges Signal angenommen). Dann ergibt sich mit (94.28): $l \cdot f/v = ((2\,\mathrm{m}) \cdot (10^6)/\mathrm{s})/(2 \cdot 10^8\,\mathrm{m/s}) = 10^{-2}$. Diese Leitung befindet sich im Übergangsbereich zwischen elektrisch kurzer und elektrisch langer Leitung. Ist das Signal nicht sinusförmig, was häufig der Fall ist, gilt für die Oberschwingungen, dass die Leitung elektrisch lang ist. Es kommt zu Signalverfälschungen, sodass bei Frequenzen im MHz-Bereich mit Anpassung gearbeitet werden sollte. ◀

94.4.4 Leitung als Transformator

Der Wellenwiderstand einer verlustlosen Leitung ist reell (94.26). Dann geht (94.19) mit den Zusammenhängen

$$\tanh(\alpha \pm j\beta) = \frac{\sinh(2\alpha) \pm j \cdot \sin(2\beta)}{\cosh(2\alpha) + \cos(2\beta)},$$

$$\sin(2\beta) = 2 \cdot \sin(\beta) \cdot \cos(\beta),$$

$$\cos(2\beta) = 2 \cdot \cos^2(\beta) - 1,$$

$$\sinh(0) = 0, \quad \cosh(0) = 1$$

über in:

Eingangsimpedanz am Anfang der Leitung:

$$\underline{Z}_a = \underline{Z} \cdot \frac{\underline{Z}_e + j \cdot \underline{Z} \cdot \tan \beta l}{\underline{Z} + j \cdot \underline{Z}_e \cdot \tan \beta l} \tag{94.30}$$

$\underline{Z}_e$ am Leitungsende angeschlossene Impedanz; $\underline{Z}$ Wellenwiderstand der Leitung.

Wählt man die Leitungslänge l zu $\lambda/4$ ($\Rightarrow \beta \cdot l = \pi/2$) bzw. $\lambda/2$ ($\Rightarrow \beta \cdot l = \pi$), folgt daraus für die Eingangsimpedanz am Anfang der Leitung:

$$\underline{Z}_a\big|_{\beta l = \pi/2} = \frac{\underline{Z}^2}{\underline{Z}_e} \qquad \lambda/4\text{-Transformator}$$

$$\underline{Z}_a\big|_{\beta l = \pi} = \underline{Z}_e \qquad \lambda/2\text{-Transformator} \tag{94.31}$$

Ist der Wellenwiderstand $\underline{Z} = \sqrt{\underline{Z}_a \cdot \underline{Z}_e}$, kann beim $\lambda/4$-Transformator jeder Leitungsabschluss in jede andere Impedanz $\underline{Z}_a$ transformiert werden. Beim $\lambda/2$-Transformator erhält man am Eingang die am Ende der Leitung angeschlossene Impedanz $\underline{Z}_e$, und zwar unabhängig von den Leitungseigenschaften.

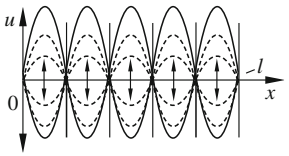


Abb. 94.7 Stehende Wellen (siehe auch Abb. 94.4)

94.4.5 Stehende Wellen

Betreibt man eine verlustlose Leitung am Ende im Leerlauf oder im Kurzschluss, ist der Reflexionsfaktor nach (94.15) $+1$ für den Leerlauf- bzw. -1 für den Kurzschlussfall. Spannung bzw. Strom am Ausgang werden zu Null erzwungen, sodass Strom- und Spannungsmaximum (oder entsprechende andere Werte) entlang der Leitung ortsfest sind. Beide haben eine Phasenverschiebung von 90° zueinander. Abb. 94.7 zeigt den Spannungsverlauf auf einer Leitung bei Kurzschluss am Leitungsende ($x = l$), vom Ende aus betrachtet. Bildet man den Quotienten $U_{\min}/U_{\max}$ aus dem Minimal- und dem Maximalwert der Spannung nach Abb. 94.4b unten, erhält man den *Anpassungsfaktor*, der mit dem Betrag des Reflexionsfaktors in der dargestellten Weise zusammenhängt:

Anpassungsfaktor

$$m = \frac{1-|\underline{r}|}{1+|\underline{r}|}, \quad \text{und daraus } |\underline{r}| = \frac{1-m}{1+m} \tag{94.32}$$

$\underline{r}$ Reflexionsfaktor nach (94.15).

Für Leerlauf bzw. Kurzschluss ist $m = 0$. Misst man das Spannungsmaximum $U_{\max}$ und das Spannungsminimum $U_{\min}$ mit einem Wechselspannungsmessgerät, bekommt man mit (94.15) auch eine Aussage über das Verhältnis von Abschlusswiderstand zu Leitungswiderstand. Herrscht am Ende der Leitung Anpassung, ist $m = 1$.

94.4.6 Kettenleiter

Schaltet man mehrere Vierpole in Kettenschaltung nach Abb. 94.8a zusammen, erhält man einen Kettenleiter. Die Berechnung erfolgt hier nach Abb. 94.8b für einen verlustlosen Kettenleiter, weil er in der Praxis die größte Bedeutung hat. Abhängig davon, ob für die Einzelvierpole die T- oder π-Schaltung nach Abb. 94.5 gewählt wird, erhält man für die Einzelvierpole der Länge l mit (94.23a), (94.23b):

T-Vierpol:

$$\cosh\gamma\cdot l = 1-\frac{\omega^2 LC}{2}; \quad \underline{Z} = \sqrt{\frac{L}{C}\cdot\frac{1}{1-\frac{\omega^2 LC}{4}}} \tag{94.33a}$$

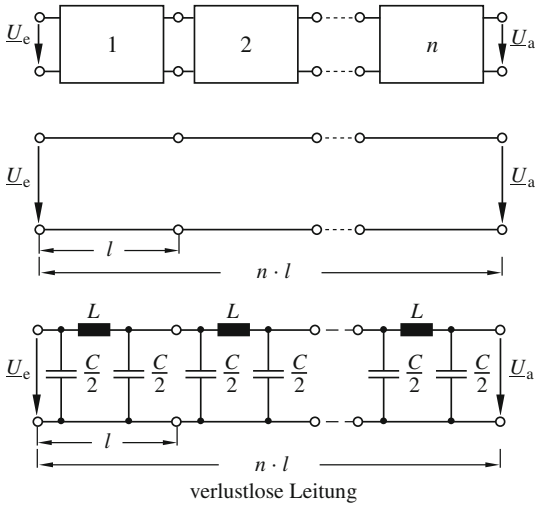


Abb. 94.8 Kettenleiter. **a** allgemeiner Aufbau, **b** elektrische Leitung als Kettenleiter, **c** Ersatzschaltbild der elektrischen Leitung

π-Vierpol:

$$\cosh\gamma\cdot l = 1-\frac{\omega^2 LC}{2}; \quad \underline{Z} = \sqrt{\frac{L}{C}\cdot\left(1-\frac{\omega^2 LC}{4}\right)} \tag{94.33b}$$

Sind die Größen L, R, G und C eines Elementarvierpols der Länge l gegeben, kann die Kettenschaltung dadurch berechnet werden, dass man bei n Kettengliedern die Größe $\gamma\cdot l$ durch $n\cdot\gamma\cdot l$ ersetzt (94.18). Damit werden L, R, G und C jeweils mit n multipliziert:

Kettenleiter mit n gleichen Elementen:

$$\begin{aligned}\underline{U}(x=0) &= \underline{U}_a = \underline{U}_e\cdot\cosh(n\cdot\gamma\cdot l) + \underline{Z}\cdot\underline{I}_e\cdot\sinh(n\cdot\gamma\cdot l)\\ \underline{I}(x=0) &= \underline{I}_a = \underline{I}_e\cdot\cosh(n\cdot\gamma\cdot l) + \frac{\underline{U}_e}{\underline{Z}}\cdot\sinh(n\cdot\gamma\cdot l)\end{aligned} \tag{94.34}$$

Für die hier betrachteten verlustlosen Kettenleiter wird gesetzt: $R' = G' = 0$.

94.4.7 Wellenfilter

Eine Anwendung für die Kettenleiter gibt es bei den Wellenfiltern. Es werden bei Bedarf mehrere Elementarvierpole in Kettenschaltung verbunden. Der dann entstehende Gesamtvierpol soll betrachtet werden. Nach (94.33a), (94.33b) gilt für T- und π-Vierpol: $\cosh(\gamma\cdot l) = 1-(\omega^2\cdot L\cdot C/2)$. Trennt man γ in Real- und Imaginärteil $a + jb$, so folgt:

$$\begin{aligned}\cosh(a+jb) &= \cosh(a)\cdot\cos(b) + j\cdot\sinh(a)\cdot\sin(b)\\ &= 1-\frac{\omega^2 LC}{2}\end{aligned}$$

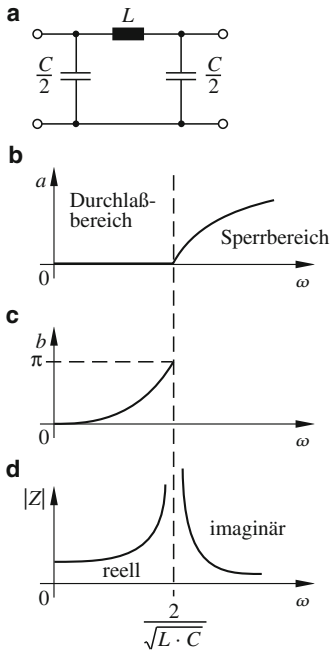


Abb. 94.9 Wellenfilter. **a** Schaltbild, **b** Dämpfungskoeffizient, **c** Phasenkoeffizient, **d** Wellenwiderstand $\underline{Z}$

Der Imaginärteil ist Null:

$$\sinh(a) \cdot \sin(b) = 0;$$

$$\cosh(a) \cdot \cos(b) = 1 - \frac{\omega^2 LC}{2}$$

Diese Gleichungen werden erfüllt durch:

1. $a = 0$ und $\cos(b) = 1 - \frac{\omega^2 \cdot L \cdot C}{2}$ oder
2. $b = 0, \pi, 2\pi, \ldots$ und $\cosh(a) = |1 - \frac{\omega^2 \cdot L \cdot C}{2}|$.

Die Mathematik liefert:

$$|\cos(b)| \leq 1; \quad |\cosh(a)| \geq 1.$$

Für $|1 - \frac{\omega^2 \cdot L \cdot C}{2}| < 1$ gilt deshalb:

$$a = 0 \quad \text{und} \quad \cos(b) = 1 - \frac{\omega^2 \cdot L \cdot C}{2};$$

für $|1 - \frac{\omega^2 \cdot L \cdot C}{2}| > 1$ gilt:

$$b = 0, \pi, 2\pi, \ldots \quad \text{und} \quad \cosh(a) = \left|1 - \frac{\omega^2 \cdot L \cdot C}{2}\right|$$

Für $0 < \omega < 2/\sqrt{LC}$ist die Dämpfung Null, das Filter ist im Durchlassbereich. Für $\omega > 2/\sqrt{LC}$ nimmt die Dämpfung mit wachsender Frequenz zu. Abb. 94.9 zeigt Dämpfungs- und Phasenkoeffizient als Funktion der Frequenz.

94.4.8 Filterentwurf

Im vorhergehenden Abschnitt war durch Kettenschaltung die erweiterte Wirkung eines Tiefpasses dargestellt. Will man die Filterwirkung verändern, sind zwei Wege von besonderer Bedeutung:

1. Konvertierung der Filtercharakteristik
2. Modifikation der (Tiefpass-)Filtercharakteristik

Zur Konvertierung wird zunächst die realistischere Darstellung der Tiefpasscharakteristik (TP) von Abb. 94.9b in Abb. 94.10a idealisiert. Die Frequenzachse ist mit der auf die Grenzfrequenz ω_G normierten Frequenz Ω beschriftet. Die erste Konversion ist die Inversion, Abb. 94.10b. Es entsteht ein Hochpass. Wird der Tiefpass an der Ordinate gespiegelt, Abb. 94.10c, ergibt sich der Bandpass. Die beim Bandpass in Abb. 94.10c auftretenden negativen Frequenzen haben zunächst nur theoretisch Bedeutung. Das negative Vorzeichen absolut verschwindet bei der Transformation, in der der Nullpunkt des System zu einer realen, höheren Frequenz (mindestens $> |\omega_G|$) transformiert wird. Das negative Vorzeichen bezieht sich dann zunächst noch auf die – normierte – Mittenfrequenz Ω_0 des Systems. Dies ist in Abb. 93.10d dargestellt. Der Bezug ist jetzt die normierte Mittenfrequenz Ω_0 und die Eckfrequenzen des Bandpasses ergeben sich zu $\Omega_0 + \Omega_G$ und $\Omega_0 - \Omega_G$. Für Ω existiert immer noch eine negative Achse.

In Abb. 94.10e ist Ω entnormiert und es erscheinen neben der realen Mittenfrequenz ω_0 die obere Grenzfrequenz mit $\omega_0 = \omega_0 + \omega_G$ und die untere Grenzfrequenz $\omega_u = \omega_0 - \omega_G$ ohne dass Probleme mit dem negativen Vorzeichen auftreten.

Die zunächst für die vereinfachende Betrachtung idealisiert angenommenen Verläufe werden nach bestimmten Gesetzmäßigkeiten realisiert. Für die Charakteristiken der Tiefpässe sind Frequenzgänge nach Butterworth (Potenzfilter) Abb. 94.11a, Bessel Abb. 94.11b, Tschebyscheff I Abb. 94.11c bzw. Tschebyscheff II Abb. 94.11d oder Cauer-Parameter (elliptische) Filter Abb. 94.11e üblich.

Butterworth und Besselfilter zeichnen sich durch einen „glatten“ Verlauf des Amplitudenganges im Durchlassbereich aus. Der Übergang zum Sperrbereich, die Flankensteilheit ist nur mäßig ausgeprägt. Als Grenzfrequenz gilt der Punkt, bei dem der Amplitudengang den 3 dB Wert überschreitet. Tschebyscheff I- und Cauer-Charakteristiken weisen Welligkeiten im Durchlassbereich auf. Die Grenzfrequenz des Durchlassbereiches ist erreicht, wenn der Betrag der Welligkeit dauerhaft überschritten wird. Tschebyscheff II- und Cauer-Filter weisen auch Welligkeiten im Sperrbereich auf. Die Welligkeit im Sperrbereich muss eine gewisse Mindestdämpfung sicherstellen. Die Sperrgrenzfrequenz ist dann die Frequenz im Übergangsbereich bei der erstmalig diese Mindestdämpfung erreicht wird.

Da die Berechnung der Filterelemente nach der Betriebsgrößentheorie zwar zweckmäßig aber sehr aufwändig ist, werden dafür Rechnerprogramme eingesetzt. Bevor Rechner mit entsprechender Rechenleistung allgemein verfügbar waren wurden die notwendigen Unterlagen in „Filterkatalogen“ z. B. von R. Saal

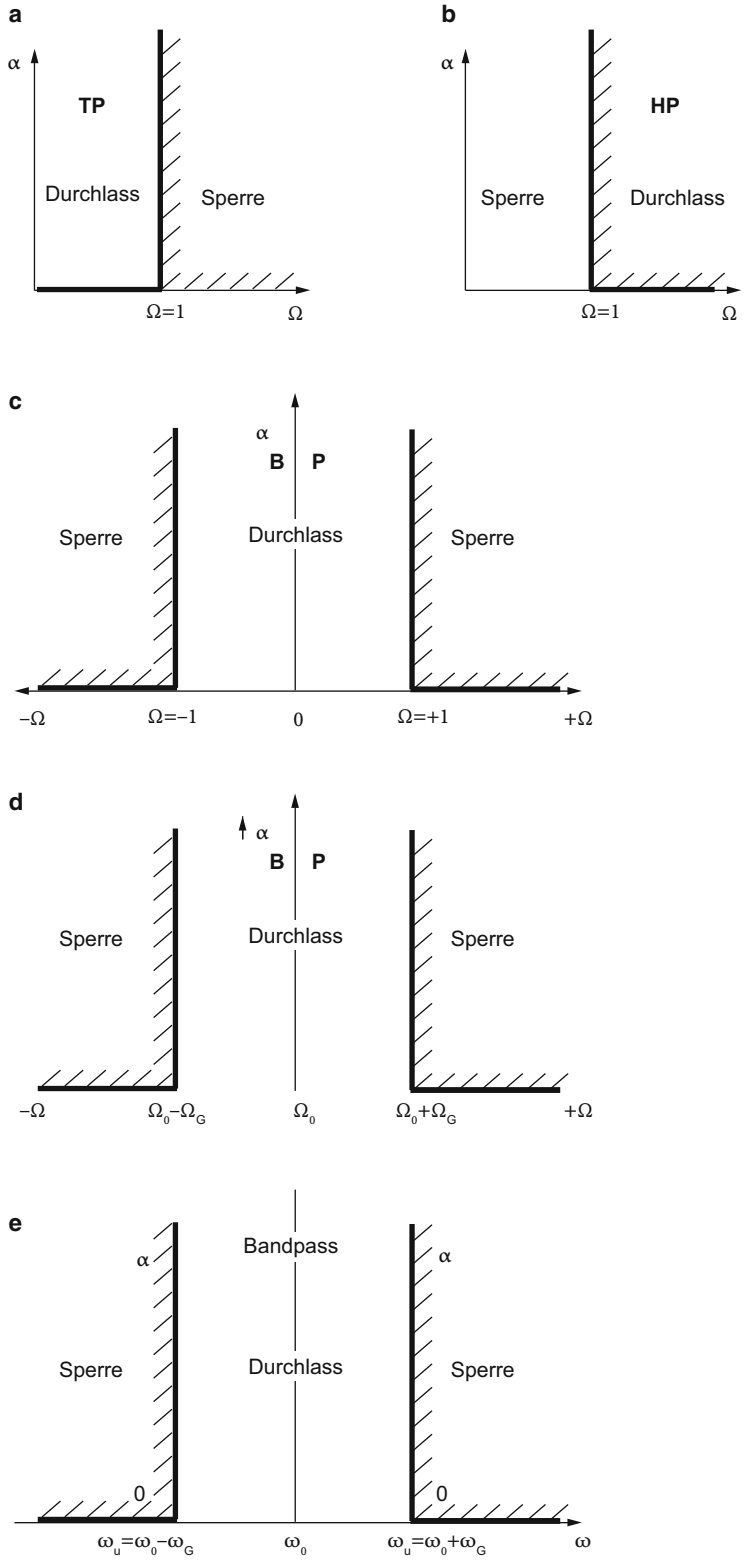


Abb. 94.10 Konversion und Transformation eines Tiefpasses. **a** Idealisierter Tiefpass (TP) mit unendlicher Dämpfung im Sperrbereich und unendlicher Flankensteilheit mit normierten Freuqenzen Ω, **b** Konversion zum Hochpass (HP) durch Vertauschung von Sperr- und Durchlassbereich ebenfalls mit normierten Freuqenzen Ω, **c** Konversion des Tiefpasses zum Bandpass (BP) durch Spiegelung des Tiefpasses an der Ordinate und normierten Frequenzen, die jetzt auch negative Werte aufweisen, **d** Transformierter Bandpass (BP) im Bereich der normierten Frequenzen, wobei die negative Ω-Achse erhalten bleibt, **e** Transformierter Bandpass (BP) mit realen Frequenzen ω, wobei jetzt lediglich das negative Vorzeichen von ω_G die Subtraktion von der Nennfrequenz ω_0 bedeutet und so die untere Grenzfrequenz ω_u liefert

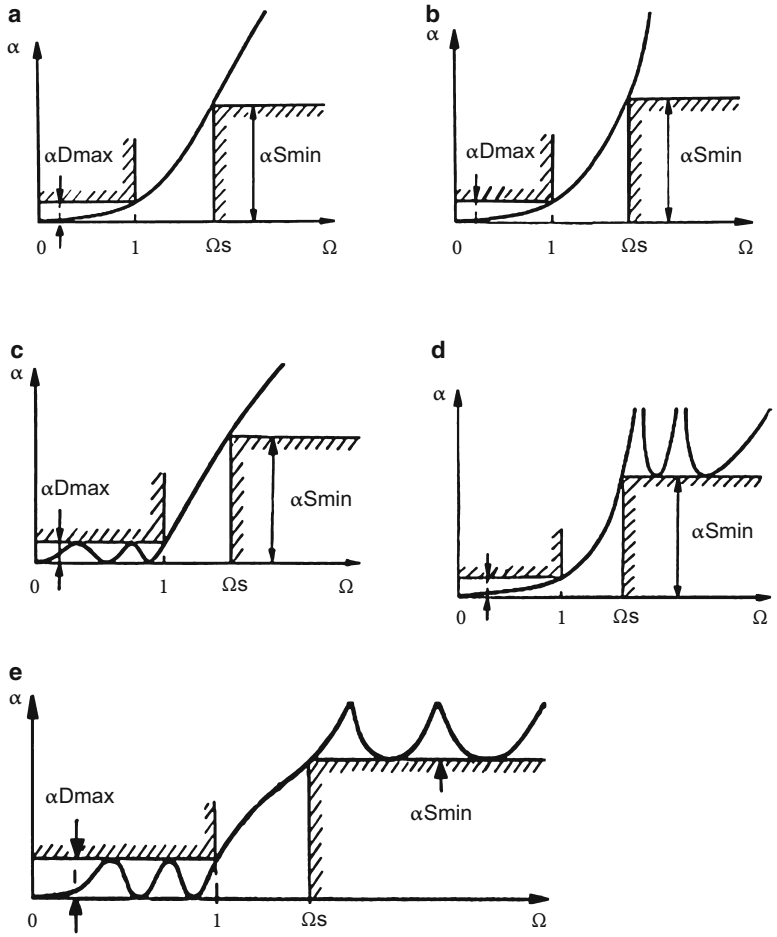


Abb. 94.11 Tiefpässe. α_{Dmax} = maximal zulässige Dämpfung im Durchlassbereich, α_{Smin} = minimal zulässige Dämpfung im Sperrbereich, **a** Potenz oder Butterworth-Filter, **b** Bessel-Filter, **c** Tschebyscheff-I-Filter, **d** Tschebyscheff-II-Filter, **e** Cauer-Parameter oder elliptischer Filter

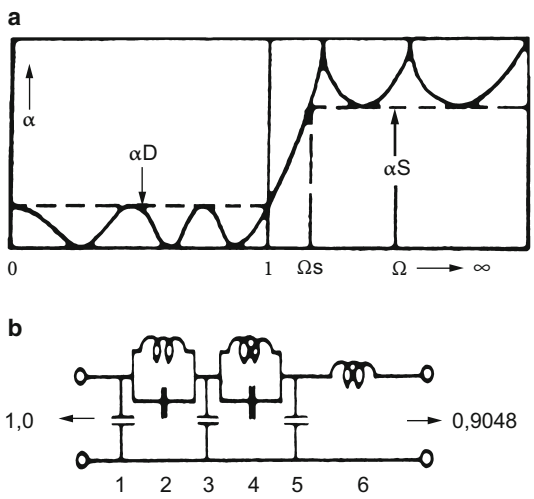


Abb. 94.12 Cauer-Parameter Tiefpass-Filter 6. Ordnung (nach Saal „Der Entwurf von Filtern mit Hilfe des Kataloges normierter Tiefpässe"). **a** Normierter Frequenzgang, **b** Schaltung

(Telefunken) und G. Pfitzenmaier (Siemens) veröffentlicht. Der Vorteil der Filterkataloge besteht in der sofort erkennbaren Charakteristik des gesuchten Filters, da die Kataloge den grundsätzlichen Frequenzgang und auch den Bauelementeaufwand mit drucken. Als Beispiel hierzu ist in Abb. 94.12a der Frequenzgang eines Cauerfilters 6. Ordnung und in Abb. 94.12b die Schaltung dazu gezeigt (n. R. Saal „Der Entwurf von Filtern mit Hilfe des Kataloges normierter Tiefpässe"). Zu beachten ist noch, dass bei der Transformation vom Tiefpass zum Bandpass auch die Blindelemente transformieren. In Abb. 94.13 sind die Konsequenzen zusammengestellt.

Neben der Welligkeit im Durchlassbereich, die ja auch die Lage der Grenzfrequenz bestimmt, sind noch die Flankensteilheit im Übergangsbereich und die Gleichmäßigkeit der Gruppenlaufzeit wichtige Dimensionierungsgrößen.

Abschließend sei noch bemerkt, dass durch die Extremstellen in den Sperrbereichen von Tschebyscheff II- und Cauer Parameter-Filtern durch geeignete Dimensionierung einzelne Frequenzen gezielt ausgeblendet werden können.

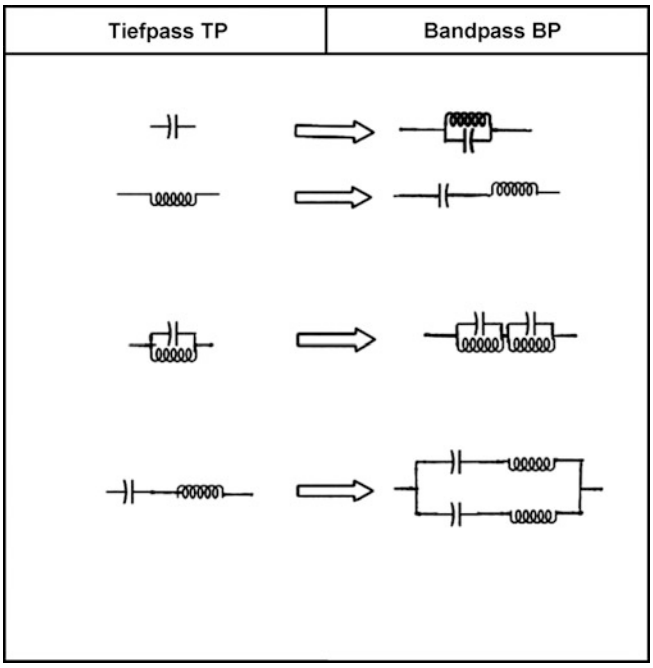


Abb. 94.13 Schema der Transformation der Tiefpasselemente zu Bandpasselementen (nach Saal „Der Entwurf von Filtern mit Hilfe des Kataloges normierter Tiefpässe")

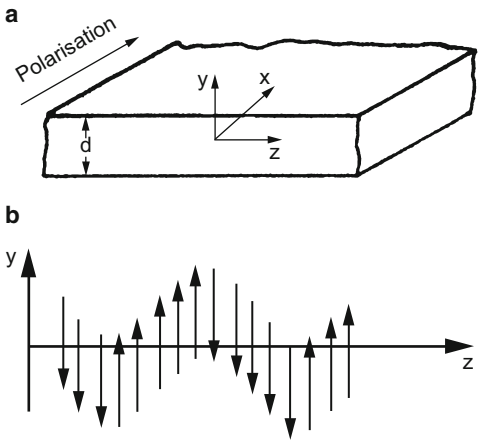


Abb. 94.14 Dielektrikum für Filter. **a** Polarisiertes Dielektrikum, **b** Scherschwingungen bei Wechselspannungsanregung

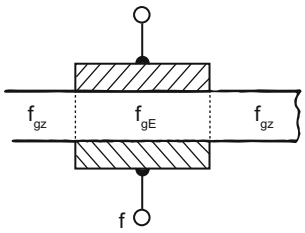


Abb. 94.15 Mit Elektroden belastetes piezoelektrisches Dielektrikum

94.4.9 Dielektrische Filter

Spulenfilter sind in der Herstellung aufwendig und benötigen im Endeinbau einen individuellen Abgleich. Dielektrische Filter nutzen mechanische Effekte aus die bei der Herstellung abgeglichen werden und dann dauerhaft konstant sind.

In einem dünnen unendlich ausgebreiteten Dielektrikum mit piezoelektrischen Eigenschaften werden bei Längspolarisation und elektrischer Anregung von den Oberflächen Scherschwingungen in z-Richtung angeregt, Abb. 94.14. Die Ausbreitungskonstante β ist dabei:

$$\beta = \frac{\pi}{d}\sqrt{\left(\frac{f}{f_{gz}}\right)^2 - 1} \qquad (94.35)$$

d ist die Substratdicke (Abb. 94.14a) und f_{gz} die Grenzfrequenz oberhalb der keine Wellenausbreitung in z-Richtung stattfindet. Wird das Dielektrikum mit Elektroden versehen, so gilt wegen der erhöhten Masse für diesen Bereich eine niedrigere Grenzfrequenz f_{gE} und damit

$$f_{gE} < f < f_{gz} \qquad (94.36)$$

Beim Betrieb mit einer Frequenz f zwischen f_{gE} und f_{gz} bleibt die Energie im Elektrodenbereich „gefangen" (energy trapping) was zu ausgeprägten mechanischen Resonanzen führt.

Bei Teilung einer Elektrode, Abb. 94.15 können verschiedene Resonanzen durch Vielfachkopplung in solche Phasenbeziehungen zueinander gelangen, dass einzelne Frequenzen optimal angeregt werden. Es stellen sich sogenannte symmetrische und antisymmetrische Schwingmoden ein, Abb. 94.16a.

Den einfachsten Fall stellt der „Kondensator" als Quarzfilter dar, das durch die hohe Güte des äquivalenten Schwingkreises eine Bandbreite unterhalb von 1 kHz aufweist (s. a. Abschn. 97.2.1).

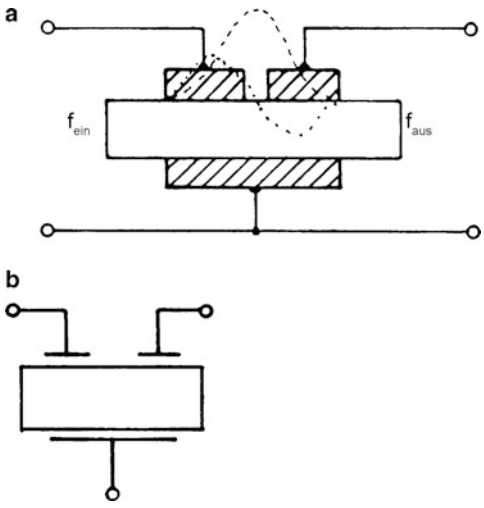


Abb. 94.16 Dielektrisches Filter. **a** Piezoelektrisches Dielektrikum mit geteilter Elektrode und mögliche Schwingungsanregungen, **b** Das Schaltbild eines derartigen dreipoligen Filters

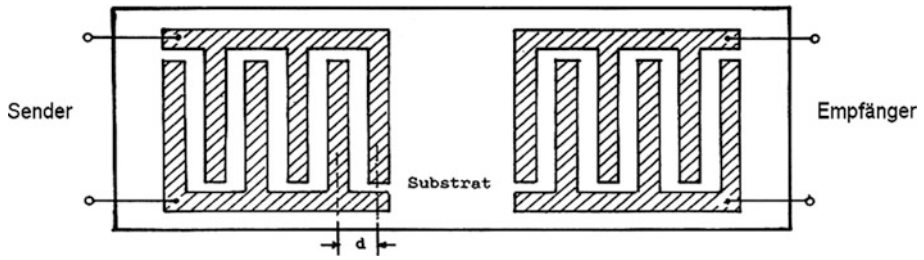


Abb. 94.17 Interdigitalwandler. Prinzip der Fingerstrukturen eines Interdigitalwandler (ID-Wandler) zur Erzeugung von Oberflächenwellen

Größere Bandbreiten lassen sich mit Teilung der Elektroden gemäß Abb. 94.16 erreichen. Diese Anordnung ermöglicht Bandbreiten von kHz im 10-MHz-Bereich.

Nach einem weiter modifizierten Prinzip zur Frequenzselektion arbeiten Oberflächenwellenfilter OFW oder englisch SAW (surface acoustic waves). Grundlage sind wieder mechanische Schwingungen die zur Interferenz von Wellenzügen an der Oberfläche eines geeigneten Substrats, z. B. Lithium-Niobat $LiNbO_3$, führen. Zur Erzeugung der Oberflächenwellen dienen fingerartig ineinandergreifende Interdigitalwandler (ID-Wandler) Abb. 94.17. Die sogenannte Synchronfrequenz f_0 eines solchen Systems ist:

$$f_0 = \frac{v_w}{2d} \tag{94.37}$$

v_w = Ausbreitungsgeschwindigkeit der Oberflächenwelle. Sie beträgt z. B. für Lithiumniobat ($LiNbO_3$) 3488 m/s. Der Bandpassfrequenzgang einer solchen Struktur folgt im einfachsten Fall – gleichmäßig, symmetrisch – der Spaltfunktion $(\sin x)/x$ mit ihren Periodizitäten. Durch Wichtung der Fingerstrukturen kann der Frequenzgang beeinflusst werden.

1) Abstandsvariation (d) → Laufzeitwichtung (Frequenz),
2) Breitenvariation → Feldstärkewichtung,
3) Längenvariation → Intensitätswichtung (der Frequenzen).

Wird z. B. die Länge nach $(\sin x)/x$ gewichtet ergibt sich eine erhöhte Bandbreite gegenüber der einfachen Struktur.

Phasengang und Gruppenlaufzeit sind durch die „Abstandsgrößen“ festgelegt und damit konstruktiv definierbar. Die Konstanz dieser Größen ist durch die mechanische Konstanz des Systems gegeben.

Grundsätzlich problematisch ist bei allen dielektrischen Filtern ihre „kapazitive Grundcharakteristik“, d. h. außerhalb des eigentlichen Betriebsbereiches existiert immer noch ein gewisses „kapazitives Übersprechen“. Für die Selektion weiter außerhalb der genutzten Frequenzen ist daher u. U. ein gewöhnliches LC-Filter sinnvoll. Durch entsprechende „Antennenstrukturen“ der Finger der ID-Wandler sind auch komplizierte Frequenzgänge wie etwa Stufenstrukturen bei Fernseh ZF-Filtern realisierbar. Zur Unterdrückung von Reflexionen sind OFW-Filter durchweg niederohmig anzusteuern, wohingegen der Abschluss hochohmig auszuführen ist.

In Abb. 94.18 sind die drei mit dielektrischen Filtern möglichen Kurvenformen, sehr schmal, schmal und breit dargestellt. In Abb. 94.18a) ist quasi der Quarz-Kondensator mit verbreiterter Resonanz zu sehen. Es ist offensichtlich, dass außerhalb

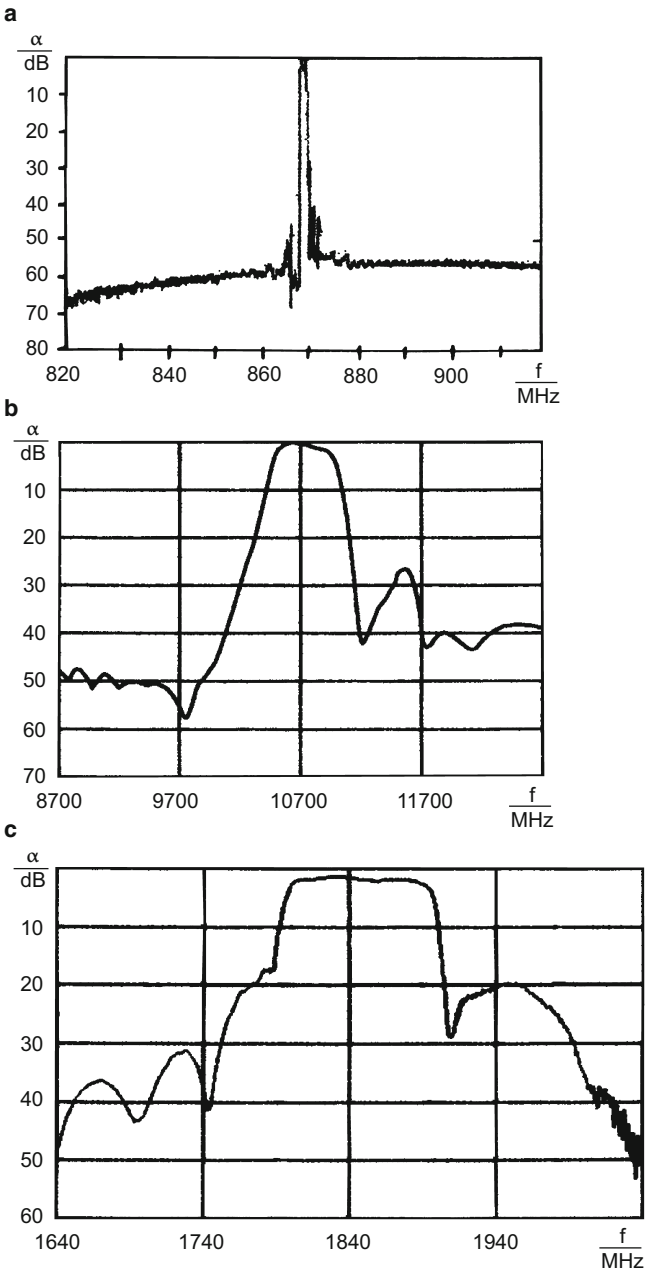


Abb. 94.18 Filtercharakteristiken. α = Dämpfung, **a** Quarzfilter, Bandbreite im kHz-Bereich, Beispiel: EPCOS B3575, **b** keramisches Filter mit Bandbreite im zig-kHz-Bereich, Beispiel: Murata CERAFIL SFLA10M7 mit ca. 250 kHz Bandbreite, **c** Oberflächenwellenfilter mit Bandbreite im MHz-Bereich, Beispiel: Murata SAWEN1G84

der Resonanz auch noch Signale kapazitiv Übersprechen. Dieser Effekte ist auch bei den beiden anderen Kurven erkennbar. Abb. 94.18b) liegt der Aufbau nach Abb. 94.16 zugrunde. Abb. 94.18c) zeigt die typische Durchlasscharakteristik eines Handy-Eingangsfilters.

94.5 Daten von Leitungen

Tab. 94.2 gibt Beispiele für Leitungsbeläge häufig vorkommender Leitungen (Näherungswerte).

Häufig kann man für die Leitungsbeläge Vernachlässigungen oder Näherungen angeben, sodass sich für den Dämpfungs- und Phasenkoeffizienten sowie für den Wellenwiderstand vereinfachte Ausdrücke ergeben. Sie sind in Tab. 94.3 zusammengestellt.

94.6 Hochfrequenzleitungen

94.6.1 Hochfrequenz-Koaxialkabel

In der Hochfrequenztechnik werden für die allgemeine Signalübertragung weitgehend Koaxialkabel mit geringen Verlusten eingesetzt. Für den in der Praxis wichtigen Wellenwiderstand $\underline{Z}$ gilt für geringe Verluste α_R, mit den mechanischen Daten $d =$ Durchmesser des Innenleiters und $D =$ (Innen-)Durchmesser des Außenleiters und der Permittivitätszahl ε_r:

$$\underline{Z} = \frac{60}{\sqrt{\varepsilon_r}} \cdot \ln \frac{D}{d} \tag{94.38}$$

Die zugehörige – geringe – Dämpfung wird mit dem Flächenwiderstand $R_F = \rho/\delta$, d. h. spezifischer Widerstand ρ durch

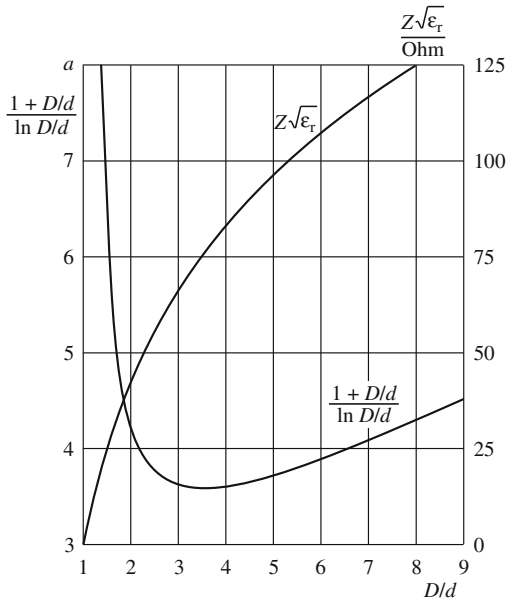


Abb. 94.19 Wellenwiderstand $\underline{Z}\sqrt{\varepsilon_r}$ und Hilfsfunktion $[1 + D/d]/[\ln(D/d]$ als Funktion von D/d

Eindringtiefe δ (s. a. (94.39)), zu:

$$\alpha_R = \frac{R_F \cdot \sqrt{\varepsilon_r}}{2 \cdot \pi \cdot 60 \cdot D} \cdot \frac{1 + \frac{D}{d}}{\ln \frac{D}{d}} \tag{94.38a}$$

Wie das Diagramm in Abb. 94.19 zeigt, weist der zweite Term in (94.38a), nämlich $[1 + D/d]/[\ln(D/d)]$, minimale Werte im Bereich zwischen $D/d = 3 \ldots 4$ auf. Für $D/d = 3{,}6$ wird die Kabeldämpfung minimal. Die höchste Spannungsfestigkeit erreichen Koaxialkabel bei $D/d = 2{,}7$. Aus den typischen Werten von $2 \ldots 3$ für ε_r folgen die Standardwerte von 50, 60 und 75 Ω für Koaxialkabel.

Für einfache Kabel weist D Werte von einigen mm auf, und der Außenmantel besteht aus ein- oder mehrlagigem Geflecht. Der Innenleiter kann aus Draht oder Litze aufgebaut sein. Die Isolation zwischen Innen- und Außenleiter ist geschäumtes PE oder PTFE. Kabel dieser Art mit 75 Ω Wellenwiderstand sind typisch für die in Häusern verlegten Breitbandkabel. Die üblichen 50-Ω-Messkabel weisen wegen der notwendigen hohen Flexibilität gleichfalls diesen einfachen Aufbau auf. Die dann notwendige gute Abschirmung wird oft durch doppeltes Außengeflecht

Tab. 94.2 Beispiele für Leitungsbeläge

Leitungstyp	R' in Ω/km	L' in mH/km	G' in µS/km	C' in nF/km
Freileitungen (Energieversorgung, $f = 50$ Hz)	0,2	1,5	0,5	5
Fernsprechleitung ($f = 1$ kHz)	5	2,2	0,8	6
Fernsprechkabel ($f = 1$ kHz)	60	0,6	1,0	50

Tab. 94.3 Leitungs-Kenngrößen für besondere Leitungen

Bedingung	α	β	$\underline{Z}$	Beispiel
$R' = G' = 0$	0	$\omega\sqrt{L'C'}$	$\sqrt{\frac{L'}{C'}}$	Verlustlose Leitung
$G' \ll \omega C'$ $R' \ll \omega L'$	$\frac{R'}{2}\sqrt{\frac{C'}{L'}} + \frac{G'}{2}\sqrt{\frac{L'}{C'}}$ mit $G' \approx 0$ folgt $\frac{R'}{2}\sqrt{\frac{C'}{L'}}$	$\omega\sqrt{L'C'}$	$\sqrt{\frac{L'}{C'}}$	Freileitung allgemein
$\frac{R'}{L'} = \frac{G'}{C'}$	$R'\sqrt{\frac{C'}{L'}}$	$\omega\sqrt{L'C'}$	$\sqrt{\frac{L'}{C'}}$	Verzerrungsfreie Leitung
$\omega L' \ll R'$ $G' \approx 0$	$\sqrt{\frac{\omega C' R'}{2}}$	$\sqrt{\frac{\omega C' R'}{2}}$	$\sqrt{\frac{R'}{j\omega C'}}$	Leitung mit geringem Querschnitt bei tiefen Frequenzen

Tab. 94.4 Daten einiger einfacher Koaxialleitungen nach RFS Radio Frequency Systems GmbH

Typ	Wellenwiderstand Z (Ω)	Durchmesser d (mm)	Durchmesser D (mm)	Außendurchmesser (mm)	Abschirmung einfach oder 1. + 2. doppelt	Isolation	C' (pF/m)	Dämpfung (dB/100 m) 100 MHz	Dämpfung (dB/100 m) 1 GHz
RF 174	50 ± 2	7 × 0,16 Cu Litze	1,91 Geflecht Cu/Sn	2,7	Geflecht Cu/Sn 85 %	PE-Schaum	100	26,0	93,0
RG 58	50 ± 2	19 × 0,18 Cu/Sn Litze	3,55 Geflecht Cu/Sn	5,0	Geflecht Cu/Sn 96 %	PE-Schaum	101	15,6	61,0
RGC58RGFLEX	50 ± 2	1,06 Cu	3,60 Geflecht Cu/Sn	5,0	1. Al/PET Folie 100 % 2. Geflecht Cu/Sn 75 %	PE-Schaum	82	10,4	35,4
RGC213RGFLEX	50 ± 2	2,55 Cu	8,14 Geflecht Cu/Sn	10,34	1. Al/PET Folie 100 % 2. Geflecht Cu/Sn 78 %	PE-Schaum	82	4,15	14,9
RG213	50 ± 2	7 × 0,75 Cu/Sn Litze	8,11 Geflecht Cu/Sn	10,31	Geflecht Cu/Sn 96 %	PE-Schaum	101	6,90	27,3
RGC8RGFLEX	50 ± 2	2,74 Cu	8,13 Geflecht Cu/Sn	10,24	1. Al/PET Folie 100 % 2. Geflecht Cu/Sn 77 %	PE-Schaum	78	4,08	13,2

Cu/Sn = verzinntes Kupfer, Abschirmung % = Überdeckungsgrad durch das jeweilige Material

Tab. 94.5 Daten einiger CELLFLEX-Kabel von RFS Radio Frequency Systems GmbH

Typ	Wellenwiderstand Z (Ω)	Nenngröße	Isolation	C' (pF/m)	P_{mittel} bei 100 MHz (kW)	P_{mittel} bei 1 GHz (kW)	Dämpfung (dB/100 m) 100 MHz	Dämpfung (dB/100 m) 1 GHz
LCF 14-50	50 ± 1	1/4″	PE	80	4,20	13,9	4,20	13,9
SCF 14-50	50 ± 1	1/4″	PE	82	5,82	19,5	5,82	19,5
LCF 12-50	50 ± 1	1/2″	PE	76	2,16	7,20	2,16	7,20
SCF 12-50	50 ± 1	1/2″	PE	82	3,33	11,2	3,33	11,2
LCF 38-50	50 ± 1,5	3/8″	PE	76	3,43	11,3	3,43	11,3
LCF 114-50	50 ± 1	1 1/4″	PE	75	3,14	3,14	0,889	3,14
LCF 214-50A	50 ± 1	2 1/4″	PE	75	2,11	2,11	0,566	2,11

LCF = Low Level Foam Coax, SCF = Superflexible Foam Coax, PE = Polyethylen, C' = Kapazitätsbelag, P_{mittel} = zulässige mittlere Leistung (die Spitzenleistungen liegen wesentlich höher)

realisiert. Die Tab. 94.4 zeigt die Typennamen und die grundsätzlichen Daten einiger einfacher Koaxialkabel.

Durch den *Skin-Effekt* und die damit verbundene geringe Eindringtiefe δ verschiebt sich der Stromfluss im Innenleiter mit zunehmender Frequenz immer mehr in die äußere Randschicht, sodass der Innenleiter hohl ausgebildet werden kann.

Für die Praxis ergibt sich die durch den Skin-Effekt verminderte Eindringtiefe δ zu

$$\frac{\delta_{\text{Cu}}}{\mu\text{m}} = \frac{66}{\sqrt{\frac{f}{\text{MHz}}}} \quad \text{für Kupfer;}$$
$$\frac{\delta_{\text{Ag}}}{\mu\text{m}} = \frac{64}{\sqrt{\frac{f}{\text{MHz}}}} \quad \text{für Silber} \qquad (94.39)$$

Die kommerzielle Technik stellt sowohl erhöhte Anforderungen an die Toleranzen der Kabeldaten als auch an die Langzeitstabilität. Für geringere Leistungen – ca. 10 kW im MHz- und 0,5 kW im GHz-Bereich – wird die Isolation noch raumfüllend



Abb. 94.20 Koaxialkabel mit geschäumter Isolation und gewelltem Außenmantel, CELLFLEX-Kabel von RFS kabelmetal

aus geschäumtem verlustarmem PE ausgeführt. Der Außenleiter wird sowohl aus Stabilitätsgründen für die elektrischen Daten als auch aus Gründen der mechanischen Flexibilität gewellt ausgeführt. Der Innenleiter kann sowohl noch massiv als auch schon hohl sein. Abb. 94.20 zeigt als Beispiel den Aufbau von CELLFLEX-Kabeln der Fa. RFS Radio Frequency Systems GmbH. Die typischen Daten einiger gewellter Kabel mit geschäumter Isolation sind in der Tab. 94.5 zusammengestellt.

Für noch höhere Leistungen bis in den MW-Bereich hinein werden Außen- und teilweise auch die Innenleiter als Wellrohre ausgebildet. Die Isolation erfolgt nur noch durch verlustarme Wendel oder Stützen aus Niederdruck-PE oder -PTFE. Abb. 94.21

Tab. 94.6 Daten einiger Koaxialkabel. Heliflex von RFS Radio Frequency Systems GmbH

Typ allgemein (RFS)	Wellenwiderstand Z (Ω)	Durchmesser d (mm)	Durchmesser D (mm)	Isolation	C' (pF/m)	P_{mittel} 100 MHz (kW)	P_{mittel} 1 GHz (kW)	Dämpfung (dB/100 m) 100 MHz	Dämpfung (dB/100 m) 1 GHz
3/8″ (HCA38-50JT)	50 ± 1	4,0 Drt	12,4 Wlr	PE Wendel	74,0	P3,67	1,13	2,76	9,31
7/8″ (HCA78-50JT)	50 ± 0,5	9,0 Ro	20,2 Wlr	PE Wendel	71,0	12,1	3,62	1,21	4,25
1-1/8″ (HCA118-50J)	50 ± 0,5	12,0 Ro	33,2 Wlr	PE Wendel	73,0	12,1	3,72	0,910	3,09
2-1/4″ (HCA214-50J)	50 ± 0,5	22,6 Wlr	56,6 Wlr	PE Wendel	66,6	23,0	6,64	0,553	1,93
3-1/8″ (HCA318-50J)	50 ± 0,5	34,8 Wlr	85,5 Wlr	PE Wendel	70,0	54,2	17,0	0,362	1,26
5″ (HCA500-50JT)	50 ± 0,5	58,0 Wlr	140,5 Wlr	PE Wendel	70,0	189	62,6 800 MHz	0,223	0,699 800 MHz
8″ (HCA800-50J)	50 ± 0,5	88,5 Wlr	215,0 Wlr	PE Wendel	69,0	279	111 600 MHz	0,149	0,397 600 MHz

Drt = Draht, Ro = Rohr, Wlr = Wellrohr, PE = Polyethylen (Wendel s. a. Abb. 94.21), C' = Kapazitätsbelag, P_{mittel} = zulässige mittlere Leistung (die Spitzenleistungen liegen wesentlich höher)

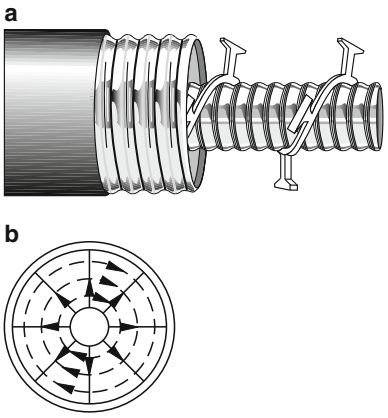


Abb. 94.21 **a** FLEXWELL-Kabel nach RFS kabelmetal mit gewelltem Außen- und Innenleiter sowie PE-Stützwendel, **b** Feldlinienbild eines Koaxialkabels nach **a** (—— = E-Feldlinien, – – – = H-Feldlinien)

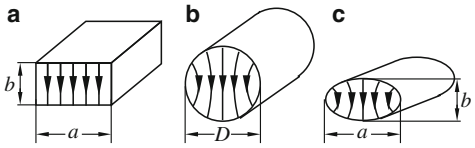


Abb. 94.22 Querschnittsformen von Hohlleitern mit Verlauf des elektrischen Feldes. **a** rechteckig, **b** rund, **c** elliptisch

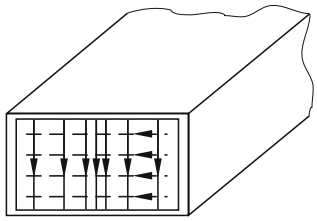


Abb. 94.23 Feldverläufe in einem Rechteck-Hohlleiter mit TE- oder H-Welle (—— = E-Feldlinien, – – – = H-Feldlinien)

zeigt in a) einen solchen Aufbau für ein FLEXWELL-Kabel von RFS kabelmetal. Die prinzipiellen Verläufe der elektrischen und magnetischen Feldlinien sind zum besseren Verständnis in Abb. 94.21b eingetragen. Das Innenleiterwellrohr ist im Inneren selbstverständlich feldfrei. In der Tab. 94.6 sind die Daten einiger FLEXWELL-Kabel zusammengestellt. Gegen äußere Einflüsse sind die Kabel in praktisch allen Fällen durch zusätzliche strapazierfähige Kunststoffummantelungen geschützt.

94.6.2 Hohlleiter

Mit weiter zunehmender Frequenz ist es sogar sinnvoll, den Innenleiter ganz wegzulassen. Man kommt dann zum eigentlichen *Hohlleiter*. Abb. 94.22 zeigt die am meisten verwendeten Querschnittsformen.

In Abb. 94.23 ist ein Beispiel für den Momentanwert der Verläufe des elektrischen und des magnetischen Feldes dargestellt. Da bei Hohlleitern entweder das elektrische oder das magnetische Feld Komponenten in Ausbreitungsrichtung hat, gelten die Bezeichnungen nach Tab. 94.7.

Rechteckhohlleiter: 1. Index: Anzahl längs der Breitseite, 2. Index: Anzahl längs der Schmalseite. Der Index ist Null, falls keine Änderung längs der entsprechenden Seite auftritt.

Tab. 94.7 Bezeichnung der Wellen in Hohlleitern

Typ	Komponente in Ausbreitungsrichtung	Bezeichnung (siehe auch Abb. 94.24)
E-Welle	Elektrisches Feld	TM-Welle: transversal-magnetische Welle
H-Welle	Magnetisches Feld	TE-Welle: transversal-elektrische Welle

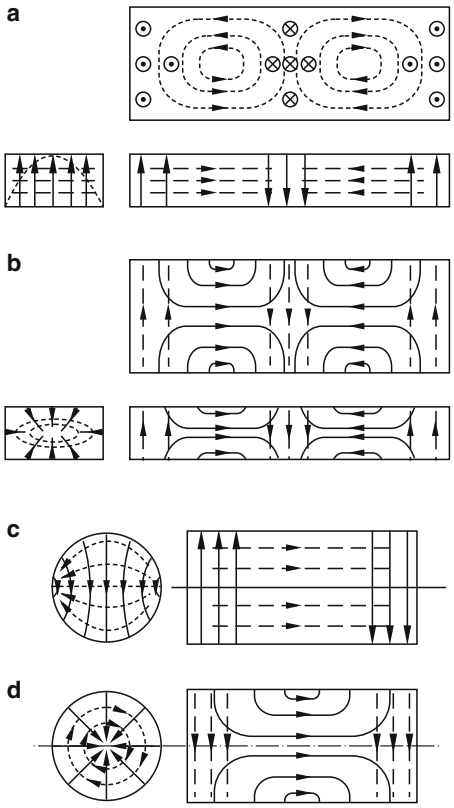


Abb. 94.24 Beispiele für Schwingungstypen in Hohlleitern. **a** H_{10}-Welle im Rechteckhohlleiter, **b** E_{11}-Welle im Rechteckhohlleiter, **c** H_{11}-Welle im Rundhohlleiter, **d** E_{01}-Welle im Rundhohlleiter (—— = E-Feldlinien, – – – = H-Feldlinien)

Rundhohlleiter, *Elliptische Hohlleiter*: 1. Index: Halbe Anzahl längs des Umfanges, 2. Index: Anzahl in radialer Richtung (Maximum in der Achse wird mitgezählt). Für elliptische Hohlleiter werden evtl. noch Zusatzindizierungen verwendet, siehe dazu die Erläuterungen zu Abb. 94.27 weiter unten. Abb. 94.24 erläutert die Bezeichnungsweise an einigen Beispielen. Am häufigsten werden angewendet:

Rechteckhohlleiter: H_{10}- und E_{11}-Welle

Rundhohlleiter, *Elliptische Hohlleiter*: H_{11}- und E_{01}-Welle

Man kann sich die Ausbreitung als fortwährende Reflexion an den Wänden des Hohlleiters vorstellen. Das Material muss deshalb elektrisch gut leitend sein, damit in den Wandungen möglichst wenig Verluste entstehen. (Es gibt noch kein Material, das eine gute elektrische und eine gute magnetische Leitfähigkeit gleichzeitig besitzt.)

Die elektromagnetischen Wellen breiten sich im Hohlleiter mit Lichtgeschwindigkeit c aus. Abb. 94.25 zeigt eine entsprechende Wellenfront. Folgende Größen lassen sich unmittelbar entnehmen:

$$\textit{Gruppengeschwindigkeit} \qquad v_g = c \cdot \sin\alpha \tag{94.40a}$$

$$\textit{Phasengeschwindigkeit} \qquad v_p = \frac{c}{\sin\alpha} \tag{94.40b}$$

c Lichtgeschwindigkeit $\approx 3 \cdot 10^8$ m/s.

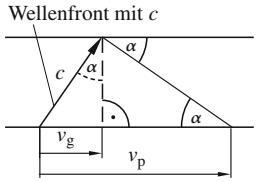


Abb. 94.25 Ebene Welle im Rechteck-Hohlleiter

Nach DIN 47301 werden die sich ausbreitenden Wellen durch zwei Indizes gekennzeichnet, die die Zahl der Maxima der elektrischen Feldstärke $\vec{E}$ angeben:

Die Energieausbreitung geschieht mit der Gruppengeschwindigkeit. Die Phasengeschwindigkeit ist stets größer oder gleich der Lichtgeschwindigkeit.

Die Wellenausbreitung wird erst oberhalb einer kritischen Wellenlänge λ_{kr} technisch nutzbar. Den Ausbreitungskoeffizienten kann man berechnen, er ergibt sich zu:

Ausbreitungskoeffizient:

$$\gamma = \alpha + j\beta = 2\pi\sqrt{\frac{1}{\lambda_{kr}^2} - \frac{1}{\lambda^2}} \tag{94.41}$$

λ_{kr} kritische Wellenlänge in m; λ Wellenlänge der sich im Hohlleiter ausbreitenden Welle.

Für $\lambda_{kr} < \lambda$ ist γ reell, d. h. die Welle wird gedämpft. Für $\lambda_{kr} > \lambda$ dagegen ist die Dämpfung Null, sodass eine Übertragung sinnvoll ist. Die Abmessungen des Hohlleiters (Durchmesser bzw. Länge und Breite nach Abb. 94.22) und λ_{kr} liegen in der gleichen Größenordnung. Einer Frequenz 3 GHz entspricht eine Wellenlänge von 10 cm, sodass die Hohlleiter für Frequenzen oberhalb dieses Wertes wirtschaftlich vertretbare Abmessungen bekommen. Hohlleiterabmessungen und zu übertragende Frequenz werden so aufeinander abgestimmt, dass sich *nur* der gewünschte Wellentyp ausbreiten kann: Die Frequenz der Welle ist größer als die kritische Frequenz des Wellentyps, aber kleiner als die aller anderen möglichen Wellentypen. Tab. 94.8 gibt Beispiele für die kritische Wellenlänge und die daraus folgende *kritische Frequenz* f_{kr}, die aus der Beziehung $f_{kr} = c/\lambda_{kr}$ (c ist die Lichtgeschwindigkeit) folgt. Erst für Frequenzen oberhalb etwa 5 GHz ergeben sich für Hohlleiter handliche mechanische Abmessungen.

Tab. 94.8 Beispiele für kritische Frequenzen und Wellenlängen bei Hohlleitern

	λ_{kr}	f_{kr} für $D = 200$ mm	f_{kr} für $D = 10$ mm
Rechteckhohlleiter			
H_{10}	$2a$	750 MHz	15 GHz
E_{11}	$\frac{2ab}{\sqrt{a^2+b^2}}$		
E_{11} ($a = b$)	$\frac{2a}{\sqrt{2}}$	1,06 GHz	21,2 GHz
E_{11} ($a = 2b$)	$\frac{2a}{\sqrt{5}}$	1,68 GHz	33,5 GHz
Rundhohlleiter			
H_{11}	$1{,}71 \cdot D$	877 MHz	17,6 GHz
E_{01}	$1{,}31 \cdot D$	1,15 GHz	22,9 GHz

a, b siehe Abb. 94.22

Tab. 94.9 Daten einiger genormter Rechteck- und elliptischer Hohlleiter

Rechteckhohlleiter					Elliptische Hohlleiter					
Typ nach DIN 47302 bzw. IEC 153	Frequenz-bereich	Dämpfung bei		Band	Typ nach DIN 47302 bzw. IEC 153	Frequenz-bereich	Dämpfung bei		Dämpfung bei	
	GHz	f/GHz	dB/m			GHz	f/GHz	dB/100 m	f/GHz	dB/100 m
R 48	3,94 … 5,99	4,73	0,0345	G	E 60	5,6…6,4	5,6	4,15	6,4	3,8
R 70	5,38 … 8,17	6,45	0,0575	J	E 78	7,1…8,5	7,1	6,2	8,5	5,6
R 100	8,20 … 12,5	9,84	0,110	X	E 105	10,0…11,7	10,0	9,6	11,7	8,9
R 140	11,9 … 18,0	14,2	0,176	P	E 130	10,7…13,3	10,7	12,6	13,3	11,2
R 220	17,6 … 26,7	21,1	0,368	K	E 185	17,3…19,7	17,3	20,3	19,7	18,9
R 320	26,3 … 40,0	31,6	0,583	R	E 220	21,2…23,6	21,2	28,8	23,6	28,09

Wegen der in der Praxis unterschiedlichen Nutzungslängen sind die Dämpfungen für Rechteckhohlleiter in dB/m angegeben, während sie für elliptische Hohlleiter in dB/100 m angegeben sind

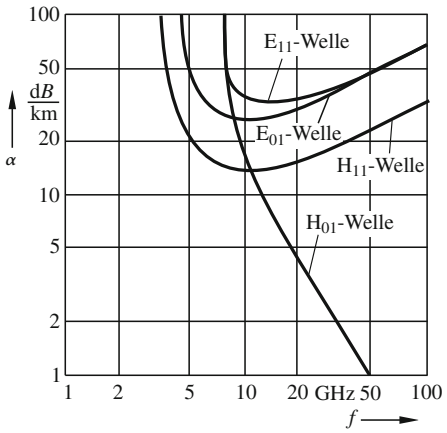


Abb. 94.26 Dämpfungskoeffizienten im Cu-Rundhohlleiter mit 50 mm Innendurchmesser (nach Frommer, Funktechnik)

Abb. 94.26 gibt Beispiele für den Verlauf des Dämpfungskoeffizienten in einem Kupfer-Rundhohlleiter. Zu erkennen ist die sprunghafte Zunahme der Dämpfung unterhalb der kritischen Frequenzen, was nach (94.41) auch zu erwarten ist.

Zu beachten ist noch, dass der (Feld-)Wellenwiderstand der Hohlleiter frequenzabhängig ist. Mit Z_0, dem Wellenwiderstand des freien Raumes (s. a. Kap. 95 „Antennen“, (95.4)), und der kritischen Frequenz f_{kr} des Hohlleiters gilt:

$$Z_F = \frac{Z_0}{\sqrt{1-(f_{kr}/f)^2}} \qquad (94.42)$$

Die aus (94.41) folgenden kritischen Frequenzen führen dazu, dass jeder Hohlleiter nur eine gewisse Bandbreite übertragen kann. Zwar sind im Prinzip auch höhere Schwingungstypen in Hohlleitern möglich, die jedoch in der Praxis vermieden werden. Ein Rechteckhohlleiter hat zunächst ein Frequenzverhältnis aus Frequenz der Grundwelle (1/2 Wellenlänge über a) zur Frequenz mit dem nächsthöheren Mode (2 · 1/2 Wellenlänge über a) von 1 zu 2. Praktisch genutzt wird aber nur ein Verhältnis von 1,2 zu 1,9. Um alle technischen Frequenzen übertragen zu können, gibt es standardisierte Typen mit einem Seitenverhältnis $a : b = 2 : 1$. In der Tab. 94.9 ist eine Auswahl von Rechteckhohlleitern aus IEC 153 zusammengestellt. Die R-Nummer entspricht dabei dem Zehnfachen der mittleren Betriebsfrequenz, gemessen in GHz. Beispiel: R 100 → mittlere Betriebsfrequenz 10 GHz · 10 → 100.

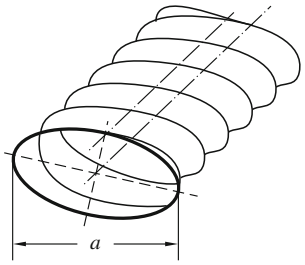


Abb. 94.27 Elliptischer Wellrohrhohlleiter

Rundhohlleiter haben zwar den Vorteil, dass Wellen vom Typ H_{01} sich mit zunehmender Frequenz immer stärker im Zentrum konzentrieren und damit eine mit der Frequenz abnehmende Dämpfung aufweisen (s. a. Abb. 94.26), sie zeigen aber ein instabiles Verhalten bei höheren Frequenzen mit Energieentzug für den Grundmode.

Nachteilig für Rechteck- und Rundhohlleiter ist in jedem Fall die geringe Flexibilität. Mobile Systeme sind damit nicht aufzubauen. Eine Lösung bieten hier die aus Wellrohren entwickelten elliptischen (Wellrohr-)Hohlleiter (Abb. 94.27). Sie sind, fast wie gewöhnliche Kabel, trommelbar und weisen Fertigungslängen von ca. 250 m auf. Wie Abb. 94.22 bereits zeigt, lehnen sie stark an die Rundhohlleiter an, und auch die Indizierung der Wellenformen ist entsprechend. Allerdings unterscheidet man noch cosinuselliptische Wellen mit der Zusatzindizierung „c“, z. B. H_{c11} (s. a. Abb. 94.28a) und sinuselliptische Wellen mit dem Zusatz „s“, z. B. H_{s11} (s. a. Abb. 94.28b).

Wegen der großen praktischen Bedeutung der elliptischen Hohlleiter sind auch diese in ihren Daten genormt und mit E-Num-

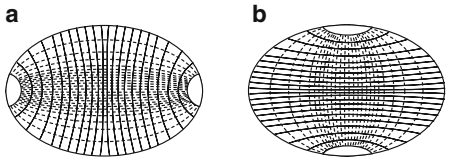


Abb. 94.28 Wellenformen in elliptischen Hohlleitern (nach W. Krank). **a** H_{c11}, **b** H_{s11} (—— = E-Feldlinien, – – – = H-Feldlinien)

mern nach der gleichen Methode wie die Rechteckhohlleiter gekennzeichnet. In der Tab. 94.9 sind auch Daten von einigen elliptischen Hohlleitern unter den Kennbuchstaben E_{xx} mit aufgeführt.

94.6.3 Streifenleitungen

Die Technik der gedruckten Schaltung hat bei hohen Frequenzen zu einem eigenen Leitungstyp, der *Streifenleitung*, geführt. Darunter ist ein flacher leitender Streifen zu verstehen, der durch ein Dielektrikum von einer oder mehreren großflächigen Gegenelektroden getrennt ist. Die wesentlichen Arten sind in Abb. 94.29 dargestellt.

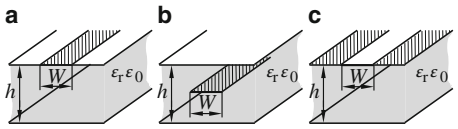


Abb. 94.29 Streifenleitungen. **a** Microstrip-Leitung, **b** Triplate-Leitung, **c** Schlitzleitung

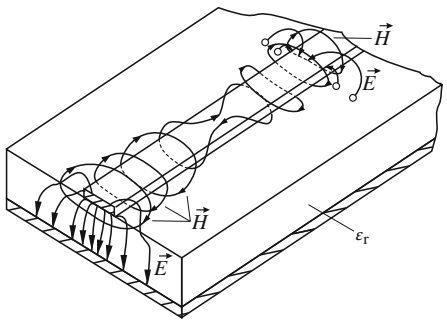


Abb. 94.30 Feldverlauf bei einer Streifenleitung

Mikrostreifenleitung, *Microstrip-* oder *Streifenleitung* im engeren Sinne ist dabei die Ausführung in Abb. 94.29a. Die in Abb. 94.30 dargestellte Feldverteilung an einer derartigen Leitung zeigt einen komplizierten Verlauf, was zu aufwändigen Berechnungen insbesondere des für die Praxis notwendigen Wellenwiderstandes von Streifenleitungen führt. Da die Feldkomponenten $\vec{E}$ und $\vec{H}$ zwar nicht exakt wie bei TEM-Wellen, aber doch weitgehend senkrecht zur Ausbreitungsrichtung verlaufen, spricht man auch von Quasi-TEM-Wellen. Wichtig ist bei der Untersuchung derartiger Systeme neben der relativen Permittivität ε_r des Substrates ganz besonders das Verhältnis w/h, d. h. Streifenbreite zu Substratdicke. Nach *Hammerstad* ergibt sich der Wellenwiderstand Z von Streifenleitungen mit den Bezeichnungen gemäß Abb. 94.29a zu:

$$\frac{Z}{\Omega} = \frac{60 \cdot \ln\left(\frac{8 \cdot h}{w} + \frac{w}{4 \cdot h}\right)}{\sqrt{\varepsilon_{\text{eff}}}} \quad \text{für } \frac{w}{h} < 1 \tag{94.43a}$$

$$\frac{Z}{\Omega} = \frac{120 \cdot \pi}{\sqrt{\varepsilon_{\text{eff}}}\left(\frac{w}{h} + 1{,}393 + 0{,}667 \cdot \ln\left(\frac{w}{h} + 1{,}44\right)\right)} \quad \text{für } \frac{w}{h} > 1 \tag{94.43b}$$

ε_{eff} ergibt sich aus der relativen Permittivität ε_r des Substrates zu:

$$\varepsilon_{\text{eff}} = 0{,}5 \cdot (\varepsilon_r + 1) + 0{,}5 \cdot (\varepsilon_r - 1) \cdot F \tag{94.43c}$$

$$F = \frac{1}{\sqrt{1 + 12 \cdot \frac{h}{w}}} + 0{,}04 \cdot \left(1 - \frac{w}{h}\right)^2 \quad \text{für } \frac{w}{h} < 1 \tag{94.43d}$$

$$F = \frac{1}{\sqrt{1 + 12 \cdot \frac{h}{w}}} \quad \text{für } \frac{w}{h} > 1 \tag{94.43e}$$

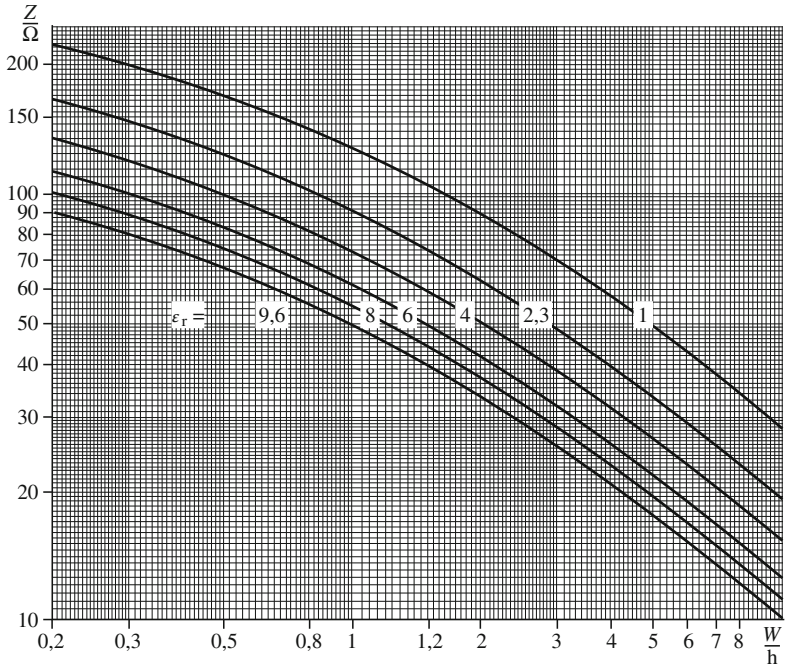


Abb. 94.31 Wellenwiderstand Z/Ω als Funktion von w/h, d. h. der Streifenbreite zur Substratdicke einer Streifenleitung mit der Permittivität ε_r als Parameter

Tab. 94.10 Verhältnis Streifenbreite zu Substratdicke (w/h) von Streifenleitungen nach Hammerstad für ausgewählte Werte von Z und $\varepsilon_r = 1 \ldots 9{,}9$, w = Streifenbreite in mm für eine Substratdicke $h = 1{,}5$ mm

$Z = 50\,\Omega$			$Z = 60\,\Omega$			$Z = 75\,\Omega$		
ε_r	w/h	w	ε_r	w/h	w	ε_r	w/h	w
1,00	4,91	7,36	1,00	3,78	5,67	1,00	2,68	4,03
1,50	3,91	5,87	1,50	2,99	4,49	1,50	2,09	3,14
2,00	3,29	4,94	2,00	2,49	3,74	2,00	1,71	2,57
2,10	3,19	4,79	2,10	2,41	3,62	2,10	1,65	2,48
2,20	3,10	4,65	2,20	2,34	3,51	2,20	1,60	2,40
2,25	3,06	4,59	2,25	2,31	3,46	2,25	1,57	2,36
2,30	3,01	4,52	2,30	2,27	3,41	2,30	1,55	2,32
2,35	2,97	4,46	2,35	2,24	3,36	2,35	1,52	2,28
2,40	2,93	4,40	2,40	2,21	3,31	2,40	1,50	2,25
2,50	2,86	4,29	2,50	2,14	3,22	2,50	1,45	2,18
2,60	2,78	4,18	2,60	2,09	3,13	2,60	1,41	2,11
2,70	2,72	4,08	2,70	2,03	3,05	2,70	1,36	2,05
2,80	2,65	3,98	2,80	1,98	2,97	2,80	1,32	1,99
2,90	2,59	3,89	2,90	1,93	2,90	2,90	1,29	1,93
3,00	2,53	3,80	3,00	1,88	2,83	3,00	1,25	1,88
3,20	2,42	3,63	3,20	1,79	2,69	3,20	1,18	1,78
3,40	2,32	3,48	3,40	1,71	2,57	3,40	1,12	1,68
3,60	2,23	3,34	3,60	1,64	2,46	3,60	1,06	1,60
3,80	2,14	3,21	3,80	1,57	2,35	3,80	1,01	1,52
4,00	2,06	3,10	4,00	1,51	2,26	4,00	0,97	1,46
4,20	1,99	2,99	4,20	1,45	2,17	4,20	0,93	1,39
4,40	1,92	2,89	4,40	1,39	2,09	4,40	0,89	1,33
4,50	1,89	2,84	4,50	1,36	2,05	4,50	0,87	1,30
4,60	1,86	2,79	4,60	1,34	2,01	4,60	0,85	1,28
4,70	1,83	2,74	4,70	1,32	1,98	4,70	0,83	1,25
4,80	1,80	2,70	4,80	1,29	1,94	4,80	0,81	1,22
4,90	1,77	2,65	4,90	1,27	1,90	4,90	0,80	1,20
5,00	1,74	2,61	5,00	1,25	1,87	5,00	0,78	1,17
5,20	1,69	2,54	5,20	1,20	1,81	5,20	0,75	1,13
5,50	1,61	2,42	5,50	1,14	1,72	5,50	0,71	1,06
6,00	1,50	2,25	6,00	1,05	1,58	6,00	0,64	0,97
6,50	1,40	2,11	6,50	0,98	1,47	6,50	0,59	0,88
7,00	1,32	1,98	7,00	0,91	1,37	7,00	0,54	0,81
7,50	1,24	1,86	7,50	0,85	1,28	7,50	0,50	0,75
8,00	1,16	1,75	8,00	0,80	1,20	8,00	0,46	0,69
8,50	1,10	1,65	8,50	0,75	1,12	8,50	0,42	0,64
9,00	1,04	1,56	9,00	0,70	1,06	9,00	0,39	0,59
9,10	1,03	1,54	9,10	0,69	1,04	9,10	0,39	0,58
9,20	1,02	1,52	9,20	0,68	1,03	9,20	0,38	0,57
9,30	1,01	1,51	9,30	0,68	1,02	9,30	0,38	0,57
9,40	1,00	1,50	9,40	0,67	1,01	9,40	0,37	0,56
9,50	0,99	1,49	9,50	0,66	0,99	9,50	0,37	0,55
9,60	0,98	1,47	9,60	0,65	0,98	9,60	0,36	0,54
9,70	0,97	1,46	9,70	0,65	0,97	9,70	0,36	0,54
9,80	0,96	1,45	9,80	0,64	0,96	9,80	0,35	0,53
9,90	0,95	1,43	9,90	0,63	0,95	9,90	0,35	0,52

Die Arbeit mit diesen Formeln ist recht aufwendig. Für geringere Genauigkeiten reicht das Diagramm in Abb. 94.31. Genauere Werte für die gängigen Wellenwiderstände von $Z = 50$, 60 und 75 Ω liefert die Tab. 94.10. Zu den üblichen Substratdicken von 1,5 mm gibt die Tab. 94.10 dann auch sofort die Streifenbreite w. Die Werte um $\varepsilon_r = 2{,}3$ sind fein unterteilt, da für genauere Arbeiten jeweils die Streuungen der ε_r-Werte des Trägermaterials beachtet werden müssen.

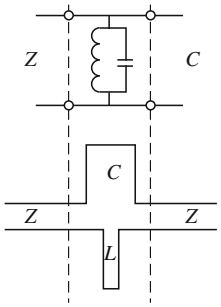


Abb. 94.32 Parallelresonanz an einer Streifenleitung durch eine niederohmige kapazitive und eine hochohmige induktive Stichleitung

Allgemein gilt, dass niederohmige Leitungen breit und hochohmige Leitungen schmal (dünn) ausfallen. Das Ätzen muss sehr sorgfältig geschehen, damit sowohl die vorgegebenen Streifenbreiten (w) ohne Schwankungen eingehalten werden und auch die Flanken der Streifen senkrecht sind und kein „Bahndammprofil" aufweisen.

Aus Abschn. 94.4.4, (94.30) folgt für $\underline{Z}_e = 0$ der Kurzschlusswiderstand $\underline{Z}_k$ bzw. für $\underline{Z}_e \to \infty$ der Leerlaufwiderstand $\underline{Z}_l$. Mit $\underline{Z} = Z$, d. h. reellem Wellenwiderstand, werden:

$$\underline{Z}_k = j \cdot Z \cdot \tan \beta \cdot l, \quad \underline{Z}_l = -j \cdot Z \cdot \cot \beta \cdot l \quad (94.30)$$

Für Werte von $l/\lambda \leq 0{,}25$ bedeutet dies induktives Verhalten der kurzgeschlossenen Leitung und kapazitives Verhalten bei Leerlauf. Diese Tatsache ist für Streifenleitungen besonders interessant, da somit in einfacher Weise Leitungen durch Stichleitungen für Transformationszwecke induktiv oder kapazitiv belastet werden können. Nachteilig ist u. U. die Länge der Stichleitung.

Gegenüber der Wellenlänge kurze ($l < 0{,}1 \cdot \lambda$) Leitungsabschnitte können niederohmig kapazitiv und hochohmig induktiv genutzt werden. Nieder- bzw. hochohmig ist immer in Bezug auf den Wellenwiderstand des restlichen Systems – meist 50 Ω – zu verstehen.

Für niederohmige Leitungsabschnitte gilt näherungsweise

$$\omega \cdot C \approx \frac{1}{Z} \cdot \sin \beta \cdot l \quad (94.44a)$$

und für hochohmige Leitungsabschnitte gilt näherungsweise

$$\omega \cdot L \approx Z \cdot \sin \beta \cdot l \quad (94.44b)$$

$\beta = 2 \cdot \pi \cdot \lambda$ und $l =$ Leitungslänge. Für den Wellenwiderstand Z ist hier der gewählte hochohmige bzw. niederohmige Wert des geplanten Blindelementes einzusetzen. Dabei ist auf den verfügbaren Platz für breite niederohmige Leitungsabschnitte zu achten. Bei hochohmigen Strecken wird die Leitung u. U. unrealisierbar schmal.

In Abb. 94.32 ist die Realisierung einer Parallelresonanz an einer Streifenleitung durch eine niederohmige (breite) kapazitive und hochohmige (schmale) induktive Stichleitung dargestellt.

Für Frequenzen im hunderter THz (Tera [10^{15}] Hertz) Bereich, (oder zehntel PHz (Peta [10^{18}] Hz) also Lichtwellen dienen Glasfaser-Lichtwellenleiter (LWL) als Leiter. Da dieser Frequenzbereich im Allgemeinen digital bearbeitet wird, sind die weiteren Betrachtungen hierzu in Kap. 103 „Nachrichtenübertragung über Lichtwellenleiter“ zu finden.

94.7 *s*-Parameter

Die Vierpolparameter (Kap. 93 „Zweitore, Vierpole“) betrachteten Systeme in der Widerstands-, Leitwert- oder Kettenform. Die Probleme, die sich auf Leitungen durch Reflexionen ergeben, blieben dabei unberücksichtigt. Wie die Leitungsgleichungen (94.9a), (94.9b) aber zeigen, sind allgemein eine hin- und eine rücklaufende Welle auf Leitungen wirksam. Dies macht sich besonders bei hohen Frequenzen bemerkbar, was durch eine veränderte Betrachtungsweise mit Hilfe der *Scatter-Parameter*, auch *Streu-Parameter* oder kurz *s-Parameter*, berücksichtigt werden kann. Diese zunächst für Leitungsbetrachtungen entwickelten Systemgrößen haben mit dem Übergang zu höheren Frequenzen erheblich an Bedeutung gewonnen.

Ausgang ist nicht mehr eine Spannung oder ein Strom, sondern eine normierte Leistungswelle, was gleichbedeutend mit der Wurzel aus einer Leistung ist. Allgemein: $a = U/\sqrt{Z}$ bzw. $I \cdot \sqrt{Z}$. Auf (94.9a) und (94.9b) angewandt bedeutet dies, dass $U = \underline{U}_h$ bzw. $\underline{U}_r$ und $R = \underline{Z}$ werden. In der Praxis kann dabei stets der reelle Wellenwiderstand Z statt des komplexen Wellenwiderstandes $\underline{Z}$ eingesetzt werden.

$$\frac{\underline{U}(x)}{\sqrt{\underline{Z}}} = \frac{\underline{U}_h}{\sqrt{\underline{Z}}} + \frac{\underline{U}_r}{\sqrt{\underline{Z}}} = a_x + b_x \qquad (94.45)$$

mit hinlaufender Welle $a = A \cdot e^{-\gamma \cdot x}$ und rücklaufender Welle $b = B \cdot e^{\gamma \cdot x}$. A und B stellen dann jeweils die normierten Leistungswellen dar.

Am Anfang der Leitung, Index 1, gilt dann:

$$\frac{\underline{U}_1}{\sqrt{\underline{Z}}} = a_1 + b_1 \qquad (94.46a)$$

und am Ende der Leitung, Index 2, gilt:

$$\frac{\underline{U}_2}{\sqrt{\underline{Z}}} = a_2 + b_2 \qquad (94.46b)$$

Das heißt, es tritt am Eingang eine rücklaufende Welle b_1 auf, die einerseits aus reflektierten Anteilen von a_1 am Anfang und andererseits aus Anteilen von a_2 am Ende der Leitung besteht. Entsprechend existiert eine zusammengesetzte Welle b_2 am Ende des Systems aus direkt am Ende von a_2 reflektierten Anteilen und von a_1 Anteilen, die vom Anfang der Leitung stammen. Der Zusammenhang zwischen diesen verschiedenen Komponenten wird durch. die *s*-Parameter wie folgt beschrieben:

$$b_1 = s_{11} \cdot a_1 + s_{12} \cdot a_2 \qquad (94.47a)$$

$$b_2 = s_{21} \cdot a_1 + s_{22} \cdot a_2 \qquad (94.47b)$$

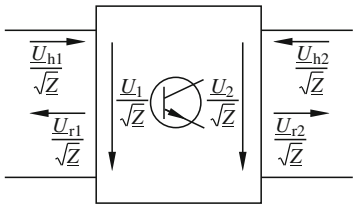


Abb. 94.33 Grundgrößen der *s*-Parameter, dargestellt an einem Transistor

Diese zunächst für Leitungen entwickelte Betrachtungsweise wird in der Höchstfrequenztechnik praktisch ausschließlich angewandt. Da Transistoren oberhalb von etwa 100 MHz weitgehend durch ihre *s*-Parameter beschrieben werden, seien an Hand von Abb. 94.33 die Bedeutungen der verschiedenen *s*-Parameter näher erläutert.

s_{11} $\quad \underline{U}_{h2}/\sqrt{\underline{Z}} = 0$

$$s_{11} = \frac{\underline{U}_{r1}}{\underline{U}_{h1}} \quad \text{Eingangsreflexionsfaktor}$$

s_{12} $\quad \underline{U}_{h1}/\sqrt{\underline{Z}} = 0$

$$s_{12} = \frac{\underline{U}_{r1}}{\underline{U}_{h2}} \quad \text{Rückübertragungsfaktor}$$

s_{21} $\quad \underline{U}_{h2}/\sqrt{\underline{Z}} = 0$

$$s_{21} = \frac{\underline{U}_{r2}}{\underline{U}_{h1}} \quad \text{Vorwärtsübertragungsfaktor}$$

s_{22} $\quad \underline{U}_{h1}/\sqrt{\underline{Z}} = 0$

$$s_{22} = \frac{\underline{U}_{r2}}{\underline{U}_{h2}} \quad \text{Ausgangsreflexionsfaktor}$$

Wie aus den Definitionen hervorgeht, können alle *s*-Parameter, obwohl zu Leistungswellen gehörig, als Spannungsverhältnisse gemessen werden. Das Problem besteht allerdings darin, dass die Wellen flussrichtungsabhängig zu messen sind. Dies ist mit sog. Richtkopplern möglich. Als Reflexionsfaktoren sind die *s*-Parameter dimensionslos und werden i. Allg. als komplexe Größen mit Betrag und Phase dargestellt. Zum Beispiel:

$$s_{21} = 3{,}3 \cdot e^{j \cdot 61^\circ}$$

Die grafische Darstellung benutzt Polarkoordinaten. Wirkliche Reflexionsfaktoren im klassischen Sinne sind aber nur die Größen s_{11} und s_{22}, die grafisch meist im Smith- oder Kreis-Diagramm (s. d.) dargestellt werden. Mitunter findet man auch eine Komponentendarstellung nach Real- und Imaginärteil.

94.7.1 Signalflussdiagramm

Die praktische Anwendung der *s*-Parameter führt zu einer veränderten Betrachtung von Schaltungen in Form von Signalflussdiagrammen. Hierzu gehören folgende Regeln:

1. Jede Variable, a_1, a_2, b_1 und b_2 wird durch einen Knoten dargestellt.

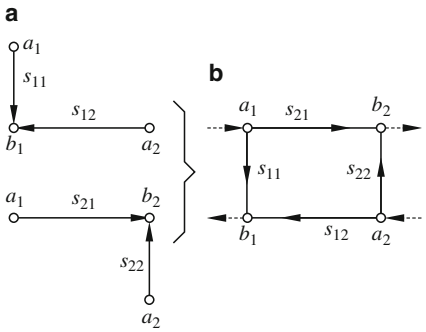


Abb. 94.34 Signalflussdiagramm eines Vierpols. **a** getrennte Darstellung, **b** zusammengefasste Darstellung, a: hineinfließende Wellen, b: rückfließende Wellen, s_{xx}: s-Parameter

2. Jeder s-Parameter, s_{11}, s_{12}, s_{21} und s_{22} wird durch einen Zweig dargestellt.
3. Die Zweige gehen von den Knoten der unabhängigen Variablen aus und enden bei den Knoten der abhängigen Variablen.
4. In den s-Parameter-Gleichungen sind a_1 und a_2 die unabhängigen Variablen und b_1 und b_2 die abhängigen Variablen.
5. „Jeder Knoten ist gleich der Summe der darin endenden Zweige."

In Abb. 94.34 ist das Signalflussdiagramm eines einfachen Vierpols gezeigt. In Abb. 94.34a sind die „reflektierten" Größen mit ihren Knoten b_1 und b_2 zunächst zur „Gewöhnung" getrennt dargestellt. In Abb. 94.34b sind die Signalflüsse dann zusammengefasst. Aus den Diagrammen ergibt sich sofort:

$$b_1 = s_{11} \cdot a_1 + s_{12} \cdot a_2 \tag{94.47c}$$

$$b_2 = s_{21} \cdot a_1 + s_{22} \cdot a_2 \tag{94.47d}$$

Diese Gleichungen sind, wie nicht anders zu erwarten war, mit (94.47a) und (94.47b) voll identisch.

An Abb. 94.34a wird jetzt auch leicht ersichtlich, was zunächst etwas willkürlich erschien, nämlich die Entstehung der abfließenden Größen b_1 und b_2. a_1 ist die dem Eingang des Vierpols als Steuersignal zufließende Leistungswelle. Davon wird wegen fehlerhafter Anpassung ein Bruchteil, nämlich $s_{11} \cdot a_1$ reflektiert. Dem Ausgang des Vierpols fließt eine Leistungswelle a_2 zu, welche die gleiche (!) Flussrichtung wie b_1 hat und durch „Übersprechen" mit einem Anteil $s_{12} \cdot a_2$ an den Eingang gelangt. Die im Ausgang abfließende Leistungswelle b_2 ist das dem Verbraucher zufließende Nutzsignal. Es besteht aus dem mit der gleichen Flussrichtung wie b_2 behafteten Eingangssignal a_1 verstärkt durch s_{21} also $s_{21} \cdot a_1$, und einem am Ausgang reflektierten Anteil $s_{22} \cdot a_2$ von einer auf den Ausgang aufgetroffenen Leistungswelle a_2, die in irgendeiner Form vom angeschlossenen Verbraucher kommt.

Die Zusammenhänge ändern sich, wenn Belastungen wirksam werden. Hierzu zunächst Abb. 94.35 mit dem Signalflussdiagramm des belasteten Generators. Warum die abfließenden Wellen jetzt die Bezeichnung b führen und die rück-(zu-)fließenden Wellen jetzt auf a lauten, wird sofort klar, wenn man noch

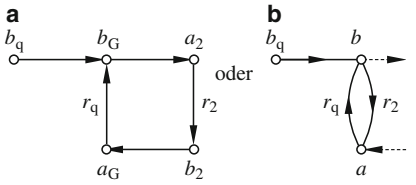


Abb. 94.35 Signalflussdiagramm des belasteten Generators. **a** getrennte Darstellung, **b** zusammengesetzte Darstellung, b: abfließende Welle, a: rückfließende Welle, b_q = normierte Leistungswelle der Quelle, r_q = Reflexionsfaktor der Quelle und r_2 = Reflexionsfaktor der Last

einmal Abb. 94.34b ganz rechts mit den gestrichelten Pfeilen betrachtet. b_q ist die normierte Leistungswelle des Generators und r_q sein Reflexionsfaktor. Aus Abb. 94.35b kann sofort abgelesen werden:

$$b = b_q + r_q \cdot a \quad \text{und} \tag{94.48a}$$

$$a = r_2 \cdot b \tag{94.48b}$$

Wechselweises Einsetzen von a bzw. b liefert:

$$b = \frac{b_q}{1 - r_q \cdot r_2} \quad \text{und} \tag{94.48c}$$

$$a = \frac{b_q}{1 - r_q \cdot r_2} r_2 \tag{94.48d}$$

Die maximal der Quelle entnehmbare Wirkleistung $P_{W\,\max}$ folgt – mit den konjugiert komplexen Größen a^* und b^* – aus der von der Last aufgenommenen Wirkleistung P_W aus zugeführten Leistung $a \cdot a^*$ minus reflektierter Leistung $b \cdot b^*$ zu:

$$P_W = a \cdot a^* - b \cdot b^* = |a|^2 \cdot (1 - |r_2|^2) \tag{94.49}$$

Im optimalen Falle soll die von der Last aufgenommene Leistungswelle a gleich der von der Quelle (Generator) abgegebenen Leistungswelle b nach (94.45) sein. Daraus folgt für die Leistung P_W:

$$P_W = |b_q|^2 \cdot \frac{1 - |r_2|^2}{|1 - r_q \cdot r_2|^2} \tag{94.50}$$

Das Maximum hieraus folgt für das Minimum des Nenners. Die Differenziation der Beträge von (94.50) liefert ein Leistungsmaximum für $|r_q| = |r_2|$ und aus der Phasenbedingung ergibt sich $r_2 = r_q^*$. Dies ist die allgemeine Anpassung auf einen konjugiert komplexen Abschlusswiderstand für die Entnahme einer maximalen Leistung aus einem Generator mit komplexem Innenwiderstand. Somit wird $P_{W\,\max}$:

$$P_{W\,\max} = |b_q|^2 \cdot \frac{1}{1 - |r_q|^2} \tag{94.51}$$

Die Zusammenfassung der bisherigen Betrachtungen liefert das Signalflussdiagramm des beschalteten Vierpols (Abb. 94.36). Die im Ausgang, Index 2, umgesetzte Leistung P_{W2} ergibt sich entsprechend (94.49) zu:

$$P_{W2} = b_2 \cdot b_2^* - a_2 \cdot a_2^* = |b_2|^2 \cdot (1 - |r_2|^2) \quad (94.49)$$

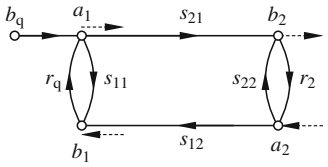


Abb. 94.36 Signalflussdiagramm des beschalteten Vierpols

b_q = normierte Leistungswelle der Quelle, a: hinlaufende Wellen, b: ablaufende Wellen, s_{xx} = s-Parameter des Vierpols (z. B. Transistor), r_q = Reflexionsfaktor der Quelle und r_2 = Reflexionsfaktor der Last.

94.7.2 Leistungsverstärkung

Mit den Grundgleichungen der s-Parameter und den Signalflüssen der äußeren Reflexionsfaktoren ergibt sich die Wirkleistung P_{W2} nach einigen Umformungen zu:

$$P_{W2} = |b_q|^2 \cdot \frac{|s_{21}|^2 \cdot (1 - |r_2|^2)}{|(1 - s_{11} \cdot r_q) \cdot (1 - s_{22} \cdot r_2) - s_{21} \cdot s_{12} \cdot r_q \cdot r_2|^2} \tag{94.52}$$

Bezogen auf die maximal dem Generator entnehmbare Wirkleistung $P_{W\max}$ (94.51) und mit der Randbedingung $s_{12} \to 0$ folgt daraus die wichtige Leistungsverstärkung G zu:

$$\frac{P_{W2}}{P_{W\max}} = G = \frac{|s_{21}|^2 \cdot (1 - |r_2|^2) \cdot (1 - |r_q|^2)}{|(1 - s_{11} \cdot r_q) \cdot (1 - s_{22} \cdot r_2)|^2} \tag{94.53}$$

Diesen Ausdruck kann man in 3 Komponenten aufspalten:

$$G = G_1 \cdot G_0 \cdot G_2 = \frac{1 - |r_q|^2}{|1 - s_{11} \cdot r_q|^2} \cdot |s_{21}|^2 \cdot \frac{1 - |r_2|^2}{|1 - s_{22} \cdot r_2|^2} \tag{94.53a}$$

Diese einzelnen Komponenten haben folgende Bedeutung:

G_1 = Wirkleistungsverstärkung am Eingangstor, maximal für $r_q = s_{11}^*$:

$$G_{1\max} = \frac{1}{1 - |s_{11}|^2}$$

$G_0 = |s_{21}|^2$ = Wirkleistungsverstärkung des Vierpols (Zweitors)

G_2 = Wirkleistungsverstärkung am Ausgangstor, maximal für $r_2 = s_{22}^*$:

$$G_{2\max} = \frac{1}{1 - |s_{22}|^2}$$

Die maximale Leistungsverstärkung des im Eingang und Ausgang konjugiert komplex angepassten Vierpols, $G_{\max}$, oder MAG = maximum available gain wird dann:

$$\text{MAG} = G_{1\max} \cdot G_0 \cdot G_{2\max} = \frac{|s_{21}|^2}{(1 - |s_{11}|^2) \cdot (1 - |s_{22}|^2)} \tag{94.54}$$

Dieser MAG-Wert ist für die Beurteilung von HF-Transistoren bedeutsam, da er die (theoretisch) maximal mögliche Leistungsverstärkung des Bauelementes angibt. Wenngleich dieser Wert in der Praxis nicht erreicht werden kann, so sind die Randbedingungen der konjugiert komplexen Anpassung in Schaltungen anzustreben und Sorge zu tragen, dass s_{12} nicht durch den Schaltungsaufbau vergrößert wird.

Weitere Folgerungen hierzu ergeben sich aus dem Arbeiten mit dem Kreisdiagramm.

94.8 Kreisdiagramm

Komplexe Widerstände werden i. Allg. in der komplexen Ebene dargestellt (Abb. 94.28a). Werden darin Kreise für konstante Anpassungsfaktoren (Stehwellenverhältnisse) m (s. a. (94.32)) und Kurven für konstante l/λ-Werte aufgetragen, so gelangt man zum *Buschbeck*-Diagramm. Wegen der unendlich großen komplexen Ebene ist auch das Diagramm unendlich groß. Da der Reflexionsfaktor $\underline{r}$ für passive Systeme nur Beträge zwischen Null und Eins annehmen kann, gehört dazu lediglich eine Kreisebene mit dem Radius 1 (Abb. 94.37b). Die Transformationsgleichung hierzu ist (94.15) mit

$$\underline{r} = \frac{\underline{z} - 1}{\underline{z} + 1} \quad (94.15) \tag{94.55}$$

$\underline{r}$ = komplexer Reflexionsfaktor und $\underline{z}$ = auf den Wellenwiderstand $\underline{Z}$ normierter komplexer Widerstand. Das so entstandene Diagramm ist vollständig in Abb. 94.38 dargestellt (vergrößerte Kopie zweckmäßig). Der äußere Kreis mit der Gradeinteilung ist der Kreis für den Reflexionsfaktor 1 und die verschiedenen Phasenwinkel von 0° … 180°. Kleinere Reflexionsfaktoren sind dann Kreise mit kleineren Radien. Der Mittelpunkt entspricht dem Reflexionsfaktor 0. Obwohl das Kreisdiagramm eigentlich

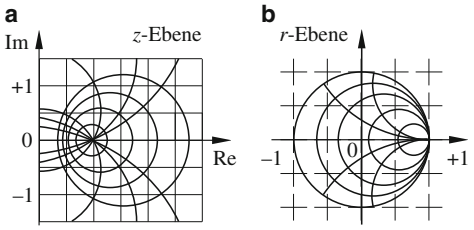


Abb. 94.37 Kreisdiagramm. **a** komplexe Widerstandsebene mit Kreisen für m = const und Kurven für l/λ = const (Buschbeck-Diagramm), **b** komplexe Reflexionsfaktorebene mit dem Radius 1 und Kreisen für konstante Wirkanteile und Kurven für konstante Blindanteile

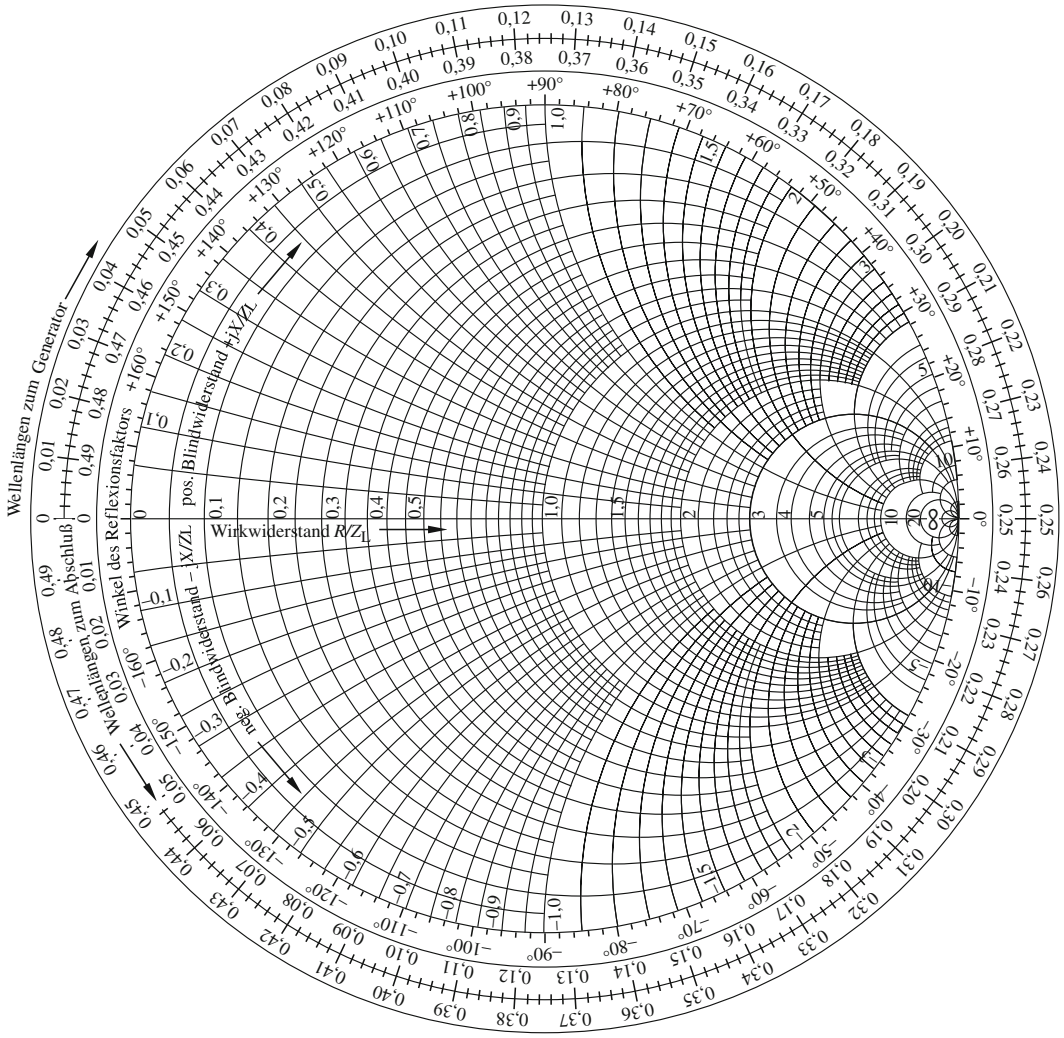


Abb. 94.38 Kreis- oder Smith-Diagramm

nur Reflexionsfaktoren darstellt, werden diese konzentrischen Kreise nicht berücksichtigt, da sie jederzeit leicht mit einem Zirkel einzutragen sind. Die dazu notwendige Skala befindet sich i. Allg. am Rande des Diagramms. Ins Diagramm eingetragen sind normierte Wirk- und Blindwiderstände. Die nach rechts verschobenen Kreise sind Kreise konstanten normierten Realteils (meist auf Wellenwiderstand Z normiert). Die nach oben gerichteten Kurven (Teile von apollonischen Kreisen) entsprechen konstanten normierten positiven Blindwiderständen, und die Kurven in der unteren Halbebene gehören zu konstanten normierten negativen Blindwiderständen. Ein Punkt in dem Diagramm entspricht der Serienschaltung normierter Wirk- und Blindwiderstände. Eine Spiegelung am Mittelpunkt oder eine Drehung um 180°, eine sog. Inversion, wandelt die normierte Serienschaltung in eine normierte Parallelschaltung um.

Obwohl das Kreisdiagramm zunächst nur für Widerstände entworfen wurde, ermöglicht die Inversion, d. h. die Drehung um 180°, den Übergang zu Leitwerten. Diese Drehung entspricht einer $\lambda/4$-Transformation auf einer Leitung. In der Praxis bedeutet dies aber auch, dass ein um 180° gedrehtes Kreisdiagramm ein Leitwertdiagramm darstellt (Abb. 94.40b). Die obere Halbebene weist jetzt negative normierte Blindleitwerte und die untere Halbebene positive normierte Blindleitwerte aus. In beiden Fällen bedeutet dies aber, dass die obere Halbebene (normierte) Induktivitäten ausweist und die untere Halbebene (normierte) Kapazitäten.

Beispiel 4 (Abb. 94.39)

1. $\underline{Z} = (200 + j100)\,\Omega$, Normierungswiderstand $Z = 100\,\Omega$: $\underline{z} = 2 + j1{,}0$; in Abb. 94.30 die Serienschaltung einer Induktivität mit einem Widerstand. Die Inversion liefert dazu $\underline{y} = 0{,}4 - j0{,}2$. (Achtung: Untere Halbebene liefert negative Vorzeichen! $1/j\omega \cdot L = -j/\omega \cdot L$.) Zum Normierungswiderstand $100\,\Omega$ gehört der Normierungsleitwert $1/100\,\Omega = 10\,\text{mS}$. Das heißt, die normierte äquivalente Parallelschaltung aus Induktivität und Widerstand hat $(4 - j2)\,\text{mS}$, was für den

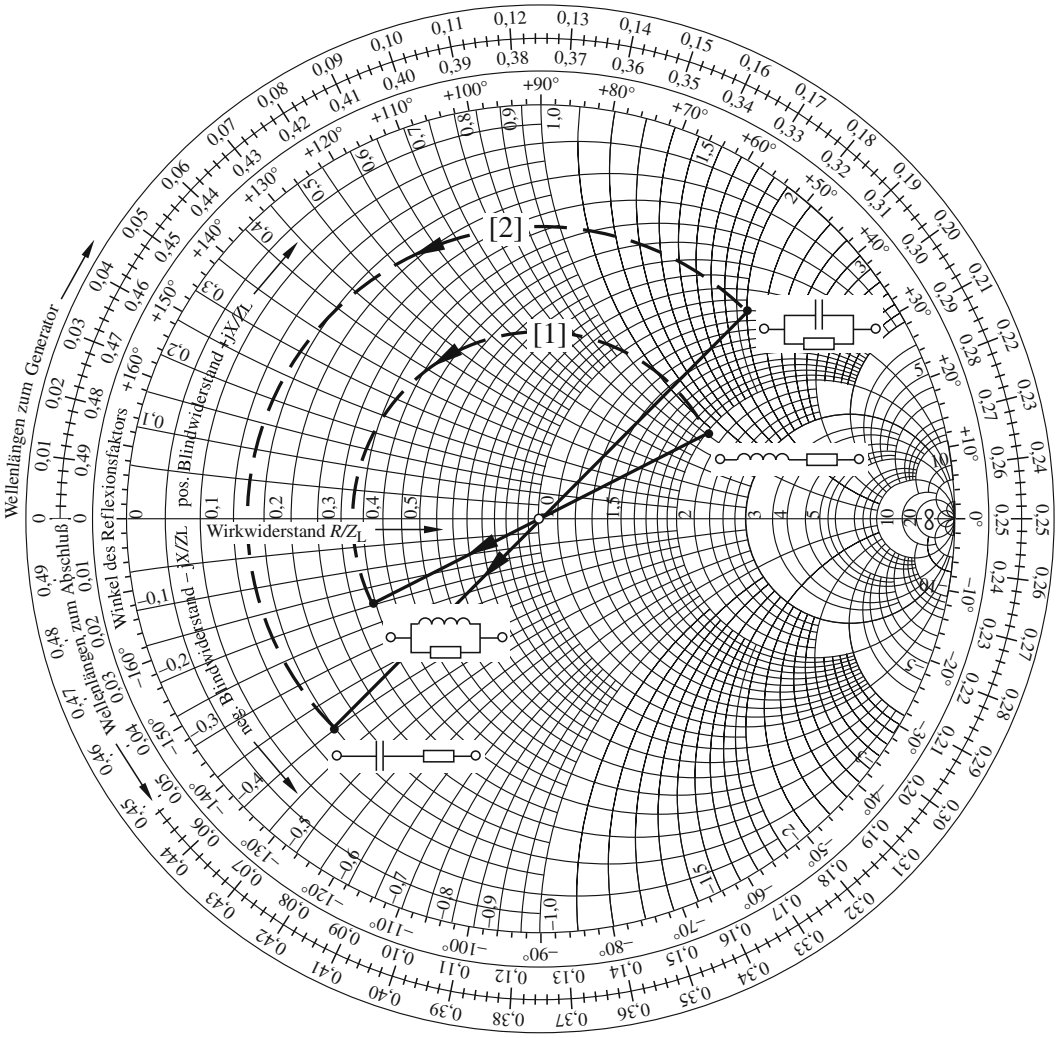


Abb. 94.39 Komplexe Widerstände und Leitwerte im Kreisdiagramm

Realteil 250 Ω und den Blindanteil $j500\,\Omega$ bedeutet. (Für die Bestimmung der Induktivität wird noch die Frequenz benötigt.)

2. $\underline{Y} = (0{,}01 + j0{,}02)$ S die Parallelschaltung eines Widerstandes mit einem Kondensator [2]. Abb. 94.30. Normierungsleitwert 10 mS: $\underline{y} = 1 + j2$. Die Inversion liefert jetzt $\underline{z} = 0{,}2 - j0{,}4$ und daraus $\underline{Z}$ $(40 - j20)\,\Omega$. (Genauere Ergebnisse als mit Abb. 94.31 lassen sich mit den Original(großen) Kreisdiagrammen erzielen.) ◀

Nachdem eine Drehung um 180° im Kreisdiagramm $\lambda/4$, also der Strecke einer Viertelwellenlänge entspricht, bedeutet eine vollständige Drehung um 360° eine halbe Wellenlänge, d. h. $\lambda/2$. Womit gemäß Leitungstransformation wieder die Ausgangswerte der Belastung einer Leitung erreicht wären (s. a. (94.31)).

In Abb. 94.40 ist gezeigt, welche Konsequenzen sich bei Bewegungen längs der verschiedenen Kurven für die Bauelemente einer Schaltung ergeben. In Abb. 94.40a liegt die Widerstandsebene vor. Bewegt man sich von einem Fixpunkt A auf einem Kreis, d. h. bei konstantem Widerstand, nach rechts, so bedeutet dies eine Zunahme der induktiven Serienkomponente. Das würde einer Schaltung mit vergrößerter Serieninduktivität entsprechen. Bewegt man sich auf dem gleichen Kreis vom Fixpunkt A nach links, so entspricht dies der Zunahme einer kapazitiven Serienkomponente. Bewegt man sich vom Punkt A auf einer Kurve für konstante Blindkomponente nach innen, so bedeutet dies die Erhöhung eines reellen Serienwiderstandes, in der umgekehrten Richtung die Verminderung einer Serienwiderstandskomponente.

Im Leitwertdiagramm (Abb. 94.40b) bedeutet eine Rechtsdrehung vom Punkt B eine Zunahme der kapazitiven Parallelkapazität und eine Linksdrehung eine Zunahme einer Parallelinduktivität. Bewegungen bei konstanten Blindanteilen führen

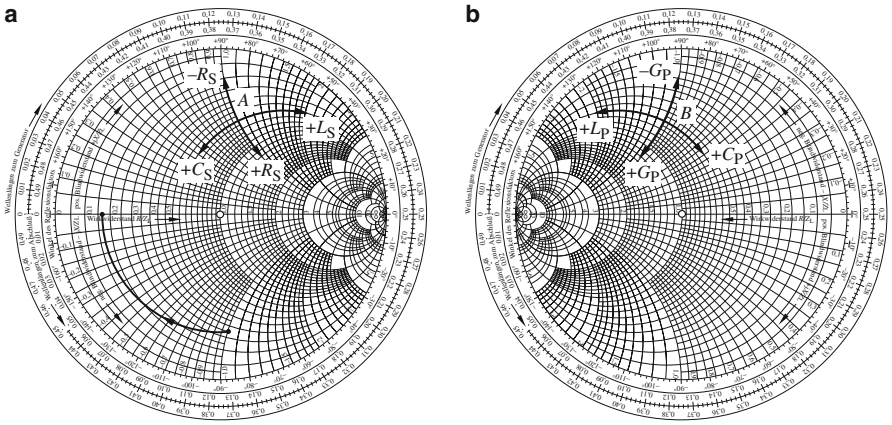


Abb. 94.40 Widerstands- und Leitwertdiagramme. **a** Kreisdiagramm für Widerstände, **b** Kreisdiagramm für Leitwerte (Erläuterungen im Text)

zur Vergrößerung des parallelen Wirkleitwertes, wenn sie nach innen gerichtet sind, und zur Verminderung bei einer Bewegungsrichtung nach außen.

Derartige Bewegungen oder Transformationen liefern zunächst einmal lediglich veränderte Schaltungen. Im Kreisdiagramm kann man aber sofort noch etwas anderes ablesen. Der Endpunkt einer Bewegung von A oder B hinweg liegt auf einem Kreis mit einem anderen Reflexionsfaktor. Es kann aber noch mehr erreicht werden. Wird in Abb. 94.40a der Pfeil bei C_s soweit verlängert, dass er die reelle Achse erreicht, so hat man einen komplexen Widerstand gemäß Punkt A ($R_s + j\omega L_s$) durch Zuschalten einer Serienkapazität reell gemacht. Allerdings wurde lediglich die induktive Komponente kompensiert und als Realteil die schon vorhandene Wirkkomponente erhalten. Wenn man allerdings vom Punkt A Parallelkomponenten hinzufügt, könnte man auch andere reelle Werte erreichen. Um das im Kreisdiagramm auszuführen, muss eine Inversion durchgeführt werden, d. h., man muss ins Leitwertdiagramm übergehen. Führt das noch nicht zum gewünschten Erfolg, muss u. U. wieder ins Widerstandsdiagramm rückinvertiert werden: Ein offensichtlich recht mühsames Verfahren.

94.8.1 Doppel-Kreisdiagramm

Diese und viele andere Aufgaben lassen sich mit Hilfe eines Doppel-Kreisdiagrammes lösen. (z. B. von H. Götting sen., Celler Str. 5, 31275 Röddensen zu beziehen).

Das zweifarbige Doppel-Kreisdiagramm hat die Inversion direkt eingearbeitet. Eine Serien-Parallel-Umwandlung geschieht jetzt so, dass lediglich die einen Koordinaten in das eine Diagramm (z. B. blau) eingetragen werden und danach die Koordinaten der anderen Farbe abgelesen werden. Die oben angeführten Beispiele können mit den Lösungen im Doppeldiagramm direkt abgelesen werden.

Eine weitere Transformationsaufgabe ist mit dem Kreisdiagramm elegant zu lösen. In Abschn. 94.4.4 folgt aus (94.30) die Veränderung von $\underline{Z}_a$ über eine Änderung der Leitunglänge l. Die Lösung dieser Aufgabe ist im Kreisdiagramm leicht zu bewerkstelligen. Es ist lediglich zu berücksichtigen, dass eine Drehung im Uhrzeigersinn in Richtung zum Generator wirkt, eine Drehung gegen den Uhrzeigersinn in Richtung zur Last transformiert.

Beispiel 5 (Abb. 94.40a)

In Abb. 94.31a ist ein Punkt L eingetragen, zu dem ein normierter komplexer Widerstand $\underline{z} = 0{,}35 - j1{,}0$ gehört. Wird diesem Widerstand in Richtung zum Generator ein dem (Normierungs-)Wellenwiderstand entsprechendes Leitungsstück vorgeschaltet, so ändert sich der am Leitungsanfang auftretende Widerstand mit der Leitungslänge gemäß (94.30). Im Kreisdiagramm bedeutet dies eine Drehung im Uhrzeigersinn auf einem Kreisbogen mit konstantem Radius, d. h. konstantem Betrag des Reflexionsfaktors. In Abb. 94.40a ist der Kreisbogen bis zur reellen Achse geführt und trifft dort auf den (normierten) Widerstandswert $r = 0{,}17$. Die zugehörige Leitungslänge (am inneren l/λ-Kreis abgelesen) beträgt $l/\lambda = 0{,}132$. Dies bedeutet, dass eine wellenwiderstandsrichtige Leitung mit der Länge $l = 0{,}132 \cdot \lambda$ den komplexen Anfangswiderstand $\underline{z}$ in einen reellen Widerstand $R = 0{,}17 \cdot Z\,\Omega$ transformiert hat. ◄

In der Praxis werden für Transformationen alle Möglichkeiten kombiniert, wobei gerade die Streifenleitungen besonders vielfältige Möglichkeiten bieten.

Das Doppel-Kreisdiagramm zeigt vor allem bei zusammengesetzten und mehrgliedrigen Transformationen seine Vorteile. Es ermöglicht sowohl den Schaltungsentwurf als auch seine Dimensionierung.

Beispiel 6 (Abb. 94.41)

Ein Widerstand von 50 Ω soll mit (verlustfreien) Blindelementen auf einen Wert von 200 Ω transformiert werden. Wie sieht die Schaltung aus und wie ist sie zu dimensionieren?

Zunächst ist für den Eintrag ins Diagramm zu normieren. Da keine speziellen Leitungs(wellen)widerstände vorliegen, bieten sich 100 Ω an. Somit $r_1 = 0{,}5$ und $r_2 = 2{,}0$. Diese Werte sind bei Punkt *A* bzw. *D* im Widerstandsdiagramm zu finden. (Im Leitwertdiagramm sind dort die zugehörigen Leitwerte mit 2,0 und 0,5 abzulesen.) Es ist jetzt ein geschlossener Transformationsweg im Diagramm zu entwickeln! Um eventuellen kapazitiven Komponenten (durch die Schaltung) bei r_1 vorzubeugen, soll die Schaltung mit einer Parallel-Kapazität beginnen. Dies bedeutet eine Rechtsdrehung auf einem Kreis mit konstantem Leitwert im Leitwertdiagramm. (Kurvenzug von *A* nach *B* in Abb. 94.41a.) Von Punkt *B* aus ist jetzt bei konstanter Wirkkomponente im Widerstandsdiagramm soweit nach rechts zu drehen (Punkt *C*) bis zu einem Kurvenzug für konstanten Wirkleitwert, der bei weiterer Rechtsdrehung zum Punkt *D* führt.

Die Schaltung dieses Kurvenzuges sieht wie in Abb. 94.41b aus. Die Strecken *AB* und *CD* waren Rechtsdrehungen im Leitwertdiagramm, also Parallel-Kapazitäten. Die Strecke *BC* gehörte zu einer Rechtsdrehung im Widerstandsdiagramm und bedeutet damit eine Serien-Induktivität.

Zur Dimensionierung müssen die normierten Werte aus dem Diagramm ermittelt werden. Der Bogen *AB* verläuft im Leitwertdiagramm von 0 … 2,0, d. h. $b_1 = 2{,}0$. Der Bogen *BC* verläuft im Widerstandsdiagramm von −0,25 (kapazitive Halbebene, daher minus) bis 0,65. Die Länge des Bogens ist dann *C* minus *B*, also $0{,}65 - (-0{,}25) = 0{,}9 = x$. Der Bogen *CD* liegt in der induktiven Halbebene des Leitwertdiagramms und beginnt demgemäß bei −1,3 und reicht bis 0 bei *D*. Die Bogenlänge ist *D* minus *C*, also $0-(-1{,}3) = 1{,}3 = b_2$. Die Konduktanzen bzw. Reaktanzen der Bauelemente betragen $B_1 = b_1 \cdot 10\,\text{mS} = 20\,\text{mS}$, $B_2 = b_2 \cdot 10\,\text{mS} = 9\,\text{mS}$ und $X = x \cdot 100\,\Omega = 90\,\Omega$. Um daraus Kondensatoren und Induktivitäten zu ermitteln, muss noch die Frequenz bekannt sein. ◀

Das Beispiel 6 hat die grundsätzliche Arbeit im Doppel-Kreisdiagramm erläutert. Die benutzten Bögen waren zunächst willkürlich in ihren Längen gewählt. Durch Variation oder Vorgabe bestimmter Elemente lässt sich sehr leicht ein passender geschlossener Transformationsweg finden. Ein Beginn im Punkt *A*, mit einem Bogen nach links oben, hätte eine Parallel-Induktivität als Eingang zur Folge gehabt, ein Beginn des Transformationsweges nach rechts oben eine Serien-Induktivität und der Beginn nach rechts unten eine Serien-Kapazität.

Ändert sich die Frequenz, so ändern sich die Längen der Bögen proportional. Mit steigender Frequenz steigen die Leitwerte,

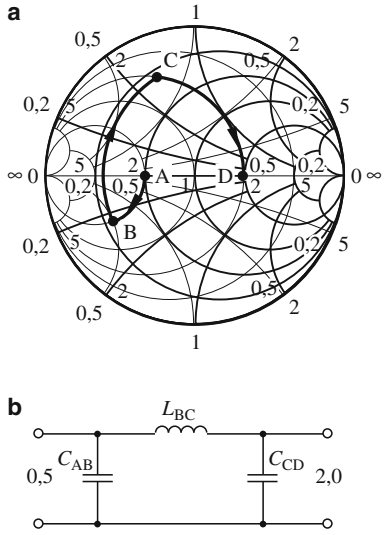


Abb. 94.41 Mehrgliedrige Transformation im Doppel-Kreisdiagramm. **a** Transformationsweg, **b** Schaltung (Erläuterung im Text)

d. h. größere Bogenlängen, und sinken die Widerstände, d. h. kürzere Bogenlängen, der Blindelemente und umgekehrt.

94.8.2 *s*-Parameter im Kreisdiagramm

Die s_{11}- und s_{22}-Parameter lassen sich als Reflexionsfaktoren sehr gut im Kreisdiagramm darstellen und werden in den Prospekten auch häufig so veröffentlicht.

Bei den Überlegungen zum MAG-Wert (94.54) war die Notwendigkeit des konjugiert komplexen Reflexionsfaktors aufgetaucht. Im Kreisdiagramm ist ein solcher Wert leicht darstellbar, da für einen nach Betrag und Phase gegebenen *s*-Parameter nur das Vorzeichen des Phasenwinkels umzukehren und im Kreisdiagramm einzutragen ist. Die Größen $G_{1\,max}$ und $G_{2\,max}$ in (94.54) verlangten aber diese konjugiert komplexen Reflexionsfaktoren an den Eingangs- bzw. Ausgangsklemmen eines Transistors. Wie gezeigt bietet das Kreisdiagramm die Möglichkeit, Transformationsschaltungen fast jeder beliebigen Art zu entwerfen. Es gibt aber noch eine Möglichkeit zu zeigen (nach Unterlagen von HewlettPackard), welche Verstärkungen sich ergeben, wenn die Anpassbedingungen nicht exakt erfüllt sind. Dies ist in der Praxis besonders für breitbandige Verstärker wichtig, da sowohl die Transistorendaten als auch die Schaltungen i. Allg. stark frequenzabhängig sind.

Die Ortskurven für konstante Werte von G_1 und G_2 sind Kreise im Kreisdiagramm, deren Mittelpunkte auf einer Geraden vom Zentrum des Diagramms zum Punkt $s{*}_{11}$ bzw. $s{*}_{22}$, d. h. den konjugiert komplexen *s*-Parametern, liegen.

Mit der normierten Verstärkung g_1

$$g_1 = \frac{G_1}{G_{1\,\max}} = \frac{G_1}{1 - |s_{11}|^2} \qquad (94.56a)$$

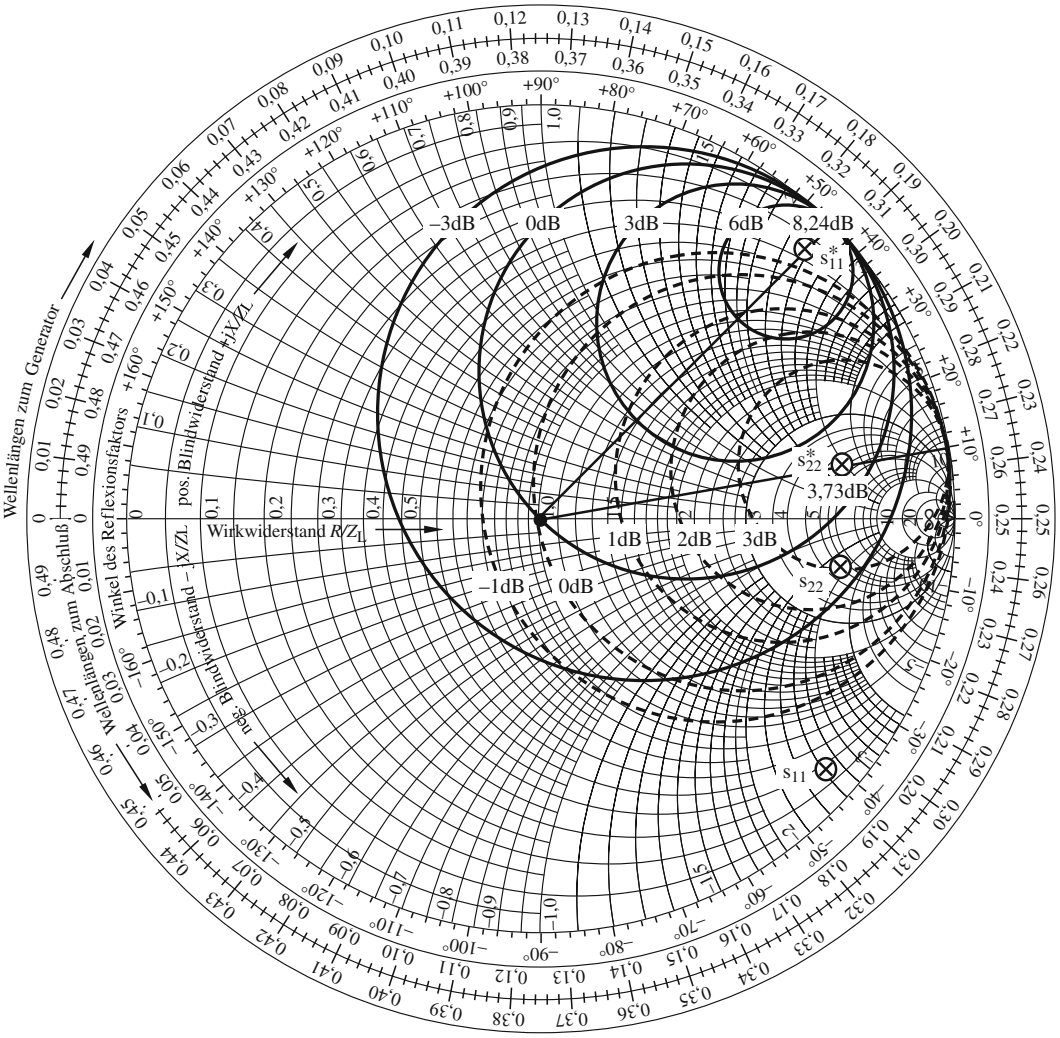


Abb. 94.42 Kreise konstanter Verstärkung im Kreisdiagramm, Lösung zu Beispiel 7 (Erläuterungen im Text)

beträgt der Radius R_G des Kreises für eine vorgegebene Verstärkung G_1:

Kreisradius für konstante Verstärkung

$$R_G = \frac{\sqrt{1-g_1}\cdot(1-|s_{11}|^2)}{1-(1-g_1)\cdot|s_{11}|^2} \qquad (94.56b)$$

Distanz D_G des Kreismittelpunktes vom Zentrum des Diagramms in Beträgen des Reflexionsfaktors (gemäß Diagramm-Maßstab):

Mittelpunktabstand für konstante Verstärkung

$$D_G = \frac{g_1\cdot|s_{11}|}{1-(1-g_1)\cdot|s_{11}|^2} \qquad (94.56c)$$

Die für die Eingangsseite mit s_{11} angeschriebenen Gleichungen gelten sinngemäß mit s_{22} für die Ausgangsseite eines Transistors. Die Verstärkungen G_1 bzw. $G_{1\,\max}$ sind in (94.56a) in absoluten Zahlen einzusetzen, auch wenn sie allgemein in *dB* angegeben werden.

Beispiel 7 (Abb. 94.42)

Für den Transistor CFY 11 von Siemens gelten bei 2 GHz und $Z = 50\,\Omega$ folgende s-Parameter:

$$s_{11} = 0{,}922\,\angle{-41°};\quad s_{12} = 0{,}026\,\angle 69° \approx 0;$$
$$s_{21} = 3{,}062\,\angle 144°;\quad s_{22} = 0{,}759\angle{-11°}$$

Damit ergibt sich auf der Eingangsseite

$$s_{11}^* = 0{,}922\,\angle 41° \quad \text{und} \quad G_{1\,\max} = 6{,}67 = 8{,}24\,\text{dB}$$

und für die Ausgangsseite

$$s_{22}^* = 0{,}759\,\angle 11° \quad \text{und} \quad G_{2\,\max} = 2{,}36 = 3{,}73\,\text{dB}$$

MAG $= G_1 + G_0 + G_2 = 8{,}24\,\text{dB} + 9{,}72dB + 3{,}73\,\text{dB} =$ 21,7 dB.

Tab. 94.11 Werte für Kreise konstanter Verstärkung im Eingang und Ausgang eines Transistors, dargestellt im Kreisdiagramm

Eingang			Ausgang		
G (dB)	D_G	R_G	G (dB)	D_G	R_G
8,24	0,992	0 (Lage von s_{11}^*)	3,73	0,759	0 (Lage von s_{22}^*)
6,0	0,873	0,145	3,0	0,705	0,193
3,0	0,682	0,310	2,0	0,629	0,299
0	0,498	0,498	1,0	0,554	0,369
−3,0	0,324	0,674	0	0,482	0,482
−6,0	0,190	0,808	−1,0	0,414	0,559

Die Rechenwerte für R_G und D_G sind in der Tab. 94.11 zusammengestellt, und Abb. 94.42 zeigt die Auswertung im (Widerstands-)Kreisdiagramm. ◀

Ähnlich wie die Verstärkung lassen sich auch die Rauschzahlen als Kreise im Kreisdiagramm oder in der komplexen Ebene darstellen. Diese Angaben werden i. Allg. von den Transistorherstellern in den Datenblättern gemacht.

Antennen

Wilfried Plaßmann



95.1 Grundlagen 1194
95.2 Kenngrößen 1196
95.3 Ausführungsformen von Antennen 1198
95.4 Wellenausbreitung 1200

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_95

95.1 Grundlagen

Die theoretischen Grundlagen und Ableitungen zu den Antennen hätten den Rahmen dieses Kapitels gesprengt, deshalb werden die wichtigsten Ergebnisse verwendet und, soweit möglich, anschaulich dargestellt. Für weitergehende Einzelheiten wird auf die Literatur verwiesen (z. B. [18]). Die im folgenden dargestellten Verhältnisse gelten wegen des *Reziprozitätssatzes* für Sende- und Empfängerantennen gleichermaßen.

Obwohl die Antenne im Allgemeinen einen komplexen Widerstand darstellt, hat sich die Bezeichnung Widerstand mit dem Formelzeichen Z durchgesetzt und soll auch hier verwendet werden (Ausnahme: Abschn. 95.4.3).

- *Antenne als Strahler*, *Poynting-Vektor*, *Strahlungsdichte*: Antennen sind Strahler, die die über Kabel zugeführte elektrische Energie in Strahlungsenergie umsetzen. Man spricht hier von Strahlung, weil der Generator (Sendeantenne) nicht feststellen kann, in welchem Maße die abgestrahlte Energie durch Gebäude oder Empfangsantennen (Verbraucher) absorbiert wird. Es gibt also keine Rückwirkung vom Verbraucher auf den Generator. Das durch die Antenne erzeugte Feld ist ein elektromagnetisches Feld, d. h. elektrisches Feld und magnetisches Feld treten zusammen auf und stehen senkrecht aufeinander, wenn man sich im Fernfeld (s. u.) der Antenne befindet. Der Energietransport geschieht senkrecht zu der aus den Vektoren $\vec{E}$ (elektrische Feldstärke) und $\vec{H}$ (magnetische Feldstärke) aufgespannten Ebene und wird mit dem *Poyntingvektor* $\vec{S}$ beschrieben:

Poynting-Vektor, *Strahlungsdichte*:

$$\vec{S} = \vec{E} \times \vec{H} \quad \text{bzw.} \quad |\vec{S}| = |\vec{E}| \cdot |\vec{H}| \cdot \sin\alpha \quad \text{in V A/m}^2 \tag{95.1}$$

α Winkel zwischen $\vec{E}$ und $\vec{H}$, $\vec{E} \times \vec{H}$ Vektorprodukt, $\vec{E}$ elektrische Feldstärke in V/m, $\vec{H}$ magnetische Feldstärke in A/m.

Der Poynting-Vektor gibt die Strahlungsdichte in der Einheit Leistung pro Fläche an.

- *Frequenz*, *Wellenlänge*: Der Energietransport geschieht mit Lichtgeschwindigkeit c, wobei im Vakuum und näherungsweise auch in Luft gilt: $\varepsilon_r = \mu_r = 1$

Lichtgeschwindigkeit:

$$c = \frac{1}{\sqrt{\varepsilon_0 \cdot \mu_0}} \approx 3 \cdot 10^5 \, \frac{\text{km}}{\text{s}} \tag{95.2}$$

ε_0 elektrische Feldkonstante, μ_0 magnetische Feldkonstante.

Zwischen der Wellenlänge λ und der Frequenz f einer elektromagnetischen Welle besteht die Beziehung:

$$\lambda \cdot f = c \tag{95.3}$$

c Lichtgeschwindigkeit $\approx 300.000\,\text{km/s}$.

- *Wellenwiderstand des leeren Raumes*: Das Verhältnis der Beträge von elektrischer und magnetischer Feldstärke ergibt unabhängig vom Ort einen konstanten reellen Wert und wird mit *Wellenwiderstand des leeren Raumes* bezeichnet:

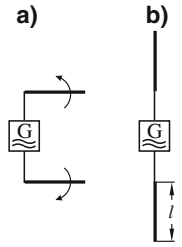


Abb. 95.1 Übergang Doppelleitung zur Dipolantenne. **a** Am Ende offene Leitung, **b** Dipolantenne

Wellenwiderstand des leeren Raumes:

$$\frac{|\vec{E}|}{|\vec{H}|} = Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 377\,\Omega \tag{95.4}$$

- *Dipol*: Die Abstrahlung des elektromagnetischen Feldes geschieht über eine Antenne, die man sich als eine am Ende leerlaufende Leitung vorstellen kann, die aufgeklappt wird, Abb. 95.1. Antennen dieser Art heißen *Dipol*, weil sie aus zwei strahlenden Elementen bestehen.
- *Hertz'scher Elementardipol*: Die Länge l der Leitung (Abb. 95.1) ist viel kleiner als die Wellenlänge λ, und es ergibt sich längs der Leitung eine nahezu konstante Stromverteilung.

Abb. 95.2 zeigt das in der Umgebung der Antenne vorhandene elektromagnetische Feld zu den Zeitpunkten $T/4$, $T/2$ und $3T/4$ (T Periodendauer, $T = 1/f$).

- *Nahfeld–Fernfeld*: Der Übergang vom Nah- zum Fernfeld hängt von der Form der Antenne ab und liegt in der Größenordnung $\lambda/2$. Erst im Fernfeld stehen die Feldvektoren $\vec{E}$ und $\vec{H}$ senkrecht aufeinander. Elektrische und magnetische Feldstärke nehmen im Fernfeld gemäß $1/r$ (r Entfernung zwischen Sendeantenne und dem betrachteten Ort) ab.
- *Polarisation*: Die Lage des elektrischen Feldes gibt die Polarisation an. Allgemein unterscheidet man:
  - *Elliptische* bzw. *kreisförmige Polarisation*: Die Feldvektoren $\vec{E}$ und $\vec{H}$ drehen sich ständig in Ellipsen- bzw. Kreisform, letztere nennt man *Zirkularpolarisation*. Je nach Drehrichtung spricht man von *rechts- bzw. linksdrehender Polarisation*.

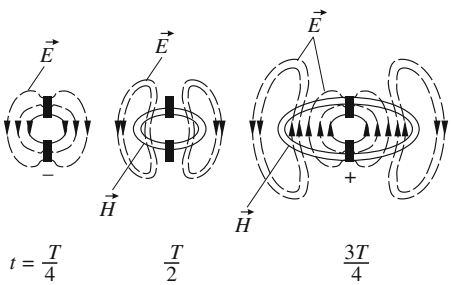


Abb. 95.2 Entstehung des elektromagnetischen Feldes einer Dipolantenne

– *Lineare Polarisation*: Die Feldvektoren haben eine festgelegte Richtung. Liegt der elektrische Feldvektor $\underline{\vec{E}}$ und $\underline{\vec{H}}$ parallel zur Erdoberfläche, handelt es sich um eine *horizontale Polarisation*, bei $\underline{\vec{E}}$ und $\underline{\vec{H}}$ senkrecht zur Erdoberfläche um eine *vertikale Polarisation* (Abb. 95.2).

Beispiele für horizontale Polarisation: UKW-Rundfunk, Fernsehen im Band III, IV und V. Beispiele für vertikale Polarisation: Lang-, Mittel- und Kurzwellenrundfunk, Fernsehen im Band I. Satelliten-Fernsehen verwendet horizontale und vertikale Polarisation.

- *Ausbreitungsrichtung des elektromagnetischen Feldes*: Die Ausbreitung des elektromagnetischen Feldes geschieht senkrecht zur Dipolachse in allen Richtungen, nicht aber in Achsrichtung des Dipols.
- *Einseitig geerdete Antenne*: In der in Abb. 95.1 dargestellten Dipolantenne kann anstelle des unteren Teils des Dipols eine *unendlich gut leitende waagerechte Fläche* eingesetzt werden. Die Erdoberfläche stellt diese Fläche, wenn auch nicht ideal, so aber doch in ausreichender Näherung, dar. An den Abstrahlverhältnissen ändert sich nichts, wie man mit der *Methode der Spiegelung* zeigen kann. Für die Praxis hat dieses Modell den Vorteil, dass man den einen Anschluss des Generators erden und eine durchgehende stabförmige Antennenkonstruktion verwenden kann. Diese Antennen heißen *Vertikalantennen* oder *fußpunktgespeiste Antennen*.
- $\lambda/4$-*Antenne* bzw. $\lambda/2$-*Dipol*: Bei dieser Antenne ist die Länge l nach Abb. 95.1 zu $\lambda/4$ gewählt. Sie hat deshalb besondere Bedeutung, weil nach Abschn. 94.4.4 „Leitung als Transformator“, (94.31), eine Leitung dieser Länge einen reellen Eingangswiderstand hat und sich der Innenwiderstand des Generators an den Wellenwiderstand des leeren Raumes anpassen lässt. Damit wird eine optimale Leistungsübertragung vom Generator auf das elektromagnetische Feld möglich.
- *Antennenverlängerung*: Vertikalantennen lassen sich bis zu einer Höhe h von etwa 150 m errichten. Das entspricht bei $h = \lambda/4$ einer Wellenlänge von 600 m und damit einer Frequenz $f = 500\,\text{kHz}$. Die Rundfunkübertragung auf Mittelwelle (Hektometerwellen) liegt im Bereich 500 ... 1500 kHz, sodass die erforderlichen Antennenabmessungen realisiert werden können. Für eine Übertragung auf Langwelle (Kilometerwellen) mit einem Frequenzbereich von etwa 80 ... 150 kHz und einer Wellenlänge von 3750 ... 2000 m müssten Antennen mit einer Höhe h von 937,5 ... 500 m errichtet werden, was technisch nicht möglich ist. Man hat deshalb am oberen Ende der Antennen zusätzliche *Endkapazitäten* C_e nach Abb. 95.3a erzeugt, indem man die Antennen entsprechend geformt hat. Bei den in Abb. 95.3b, c gezeigten häufig verwendeten Formen spricht man von $T-$ bzw. *Schirm-Antennen*. Die waagerechten (T-Antenne) bzw. die nach unten geneigten Flächen (Schirm-Antenne) auf der Antennenspitze bilden mit der Erdoberfläche eine zusätzliche Kapazität am Leitungsende. Dadurch erscheint die Antenne verlängert um den Wert

Antennenverlängerung:

$$l_v = \frac{\lambda \cdot \arctan(\omega C_e Z_A)}{2\pi} \qquad (95.5)$$

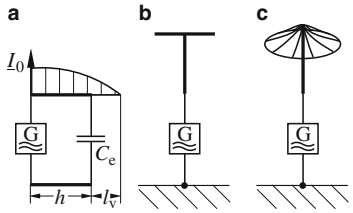


Abb. 95.3 Verkürzung von Vertikalantennen. **a** zusätzliche Endkapazität, **b** T-Antenne, **c** Schirm-Antenne

λ Wellenlänge und ω Kreisfrequenz der elektromagnetischen Welle in m bzw. 1/s, C_e Endkapazität in F, Z_A Antennenwiderstand in Ω.

Die Antennenhöhe kann damit verkürzt werden:

Verkürzte Antennenhöhe:

$$h_v = h - l_v = \frac{\lambda}{4} - l_v = \frac{\lambda}{4} - \frac{\lambda \cdot \arctan(\omega C_e Z_A)}{2\pi} \qquad (95.6)$$

h Antennenhöhe für $h = \lambda/4$.

Die Schwierigkeit bei der Auswertung der Gleichungen (95.5) und (95.6) liegt darin, dass der Antennenwiderstand Z_A durch die aufgeklappte Leitung nur näherungsweise bestimmt werden kann und von den geometrischen Abmessungen des Antennenstabes abhängt. Man kann aber erreichen, dass sich eine Antenne für Langwelle mit $\lambda/4 = 500\,\text{m}$ auf eine Länge von 250 m verkürzen lässt.

Beispiel 1

Eine Antenne hat eine Länge von 200 m und soll für die Frequenz $f = 150\,\text{kHz}$ als $\lambda/4$-Antenne eingesetzt werden. Sie hat den Antennenwiderstand $Z_A = 200\,\Omega$. Es gilt mit (95.3):

$$\lambda = c/f = (3 \cdot 10^8\,\text{m/s})/(1{,}5 \cdot 10^5/\text{s}) = 2000\,\text{m}$$
$$\Rightarrow \quad \lambda/4 = 500\,\text{m}$$

Gleichung (95.5) wird nach C_e umgeformt:

$$C_e = \frac{1}{\omega \cdot Z_A} \cdot \tan\frac{l_v \cdot 2 \cdot \pi}{\lambda}$$
$$= \frac{\text{s} \cdot \text{A}}{2 \cdot \pi \cdot 1{,}5 \cdot 10^5 \cdot 2 \cdot 10^2\,\text{V}} \cdot \tan\frac{3 \cdot 10^2\,\text{m} \cdot 2 \cdot \pi}{2 \cdot 10^3\,\text{m}}$$

Aus (95.6) wurde $l_v = h - h_v = (500 - 200)\,\text{m} = 300\,\text{m}$ eingesetzt. ◀

95.2 Kenngrößen

Die wesentlichen Eigenschaften von Antennen lassen sich den Kenngrößen entnehmen. Sie werden im Folgenden dargestellt, und es werden, sofern möglich, einfache Berechnungsformeln angegeben.

- *Richtdiagramm*: Im Fernfeld der Antenne stehen die Feldvektoren $\vec{E}$ und $\vec{H}$ senkrecht aufeinander. Der Energiefluss, erfasst durch den *Poynting-Vektor* (95.1), erfolgt wiederum senkrecht zu der durch die Feldvektoren aufgespannten Ebene. Nach Abb. 95.4 ist die Strahlungsdichte S abhängig vom Winkel ϑ und maximal für $\vartheta = 90°$. Vom Winkel φ dagegen hängt S wegen der Rotationssymmetrie nicht ab. Da $\vec{E}$ *und* $\vec{H}$ jeweils mit zunehmender Entfernung r gemäß $1/r$ abnehmen, folgt: $S \sim 1/r^2$. Gleichung (95.7a) gibt die Strahlungsdichte an.

Strahlungsdichte

$$S_r = \frac{Z_0 \cdot |\underline{I}_0|^2}{4\pi r^2} \cdot \left(\frac{l}{\lambda} \cdot \sin\vartheta\right) \tag{95.7a}$$

Z_0 Wellenwiderstand des leeren Raumes in Ω (95.4); $\underline{I}_0$ Strom in der Antenne in A, r Abstand von der Antenne in m, l Länge der Antenne in m (siehe Abb. 95.1, λ Wellenlänge in m, ϑ siehe Abb. 95.4.

Ersetzt man den Strom $\underline{I}_0$ durch die vom Strahlungswiderstand R_s aufgenommene Leistung P_s gemäß $|\underline{I}_0| = \sqrt{P_s/R_s}$ nach (95.11a) und (95.11b), ergibt sich:

Strahlungsdichte

$$S_r = \frac{3 \cdot P_s}{8 \cdot \pi \cdot r^2} \cdot \sin^2\vartheta \quad \text{für } l \ll \lambda \tag{95.7b}$$

r Abstand von der Antenne in m, ϑ siehe Abb. 95.4, P_s im Strahlungswiderstand R_s umgesetzte Leistung in W.

Hinweis: Die Abhängigkeit des Strahlungswiderstandes R_s vom Verhältnis l/λ ((95.11a) und (95.11b)) führt dazu, dass sich für die Kenngrößen der kurzen linearen Antenne und des $\lambda/2$-Dipols geringfügig voneinander abweichende Werte ergeben können.

Durch Normierung auf den Maximalwert für $\vartheta = 90°$ folgt der in Abb. 95.4 gezeigte Verlauf.

$$\textit{Richtdiagramm} \quad \frac{S_r}{S_{r\,\max}} = \sin^2\vartheta \tag{95.8}$$

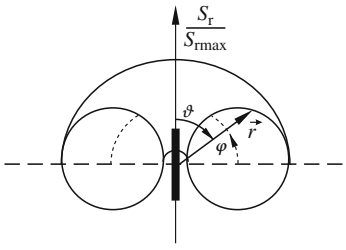


Abb. 95.4 Richtdiagramm eines Dipols (Dipol als durchgezogener Stab dargestellt)

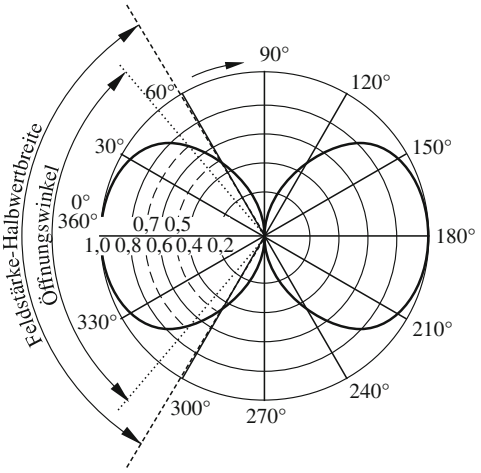


Abb. 95.5 Richtdiagramm in horizontaler Darstellung

Beispiel 2

Ein Dipol strahlt eine Leistung von $P_s = 1\,\text{kW}$ an R_s ab. Dann ergibt sich in einer Entfernung von $r = 10\,\text{km}$ parallel zur Dipolachse nach (IV.7b) eine Strahlungsdichte $S_r = \frac{3 \cdot 10^3\,\text{W}}{8 \cdot \pi \cdot 10^8\,\text{m}^2} \cdot \sin^2 90° = 1{,}2 \cdot 10^{-6}\,\frac{\text{W}}{\text{m}^2}$. Mit $S = E \cdot H$ und $H = E/Z_0$ folgt für die Feldstärke:

$$E = \sqrt{S \cdot Z_0} = \sqrt{\frac{3 \cdot P_s \cdot \sin^2\vartheta}{8 \cdot \pi \cdot r^2} \cdot Z_0}$$

$$= \sqrt{\frac{3 \cdot 10^3\,\text{V A}}{8 \cdot \pi \cdot 10^8\,\text{m}^2} \cdot 377\,\frac{\text{V}}{\text{A}}} = 21{,}2\,\frac{\text{m V}}{\text{m}} \quad \blacktriangleleft$$

- *Öffnungswinkel*: Er ist in Abb. 95.5 eingetragen und ergibt sich dadurch, dass innerhalb des Öffnungswinkels die relative Strahlungsdichte nach (95.8) $\geq 0{,}707$ ist.
- *Antennengewinn*, *Antennengewinnfaktor*, *Antennengewinnmaß*: Durch geeignete Antennenkonstruktionen kann man die Strahlungsdichte in bestimmten Richtungen im Raum konzentrieren. Ein Maß für die Richtfähigkeit ist, allgemein ausgedrückt, der Antennengewinn. Der *Antennengewinnfaktor* g wird über einen *Kugelstrahler* als Bezugsantenne bestimmt, der in alle Richtungen gleichmäßig mit

$$S_{\text{ref}} = \frac{P}{4\pi r^2} \tag{95.9}$$

($4\pi r^2$: Kugeloberfläche in m^2; P: zugeführte Leistung in W) strahlt. Der Antennengewinnfaktor g ist das Verhältnis der Strahlungsdichte S_r der eigentlichen Antenne zur Strahlungsdichte S_{ref} des Kugelstrahlers als Bezugsantenne bei gleicher Entfernung r und gleicher Leistung P. Durch entsprechende Auswertung der Gleichungen (95.8) und (95.9) erhält man den maximal möglichen Antennengewinnfaktor $g_{\max}$:
Antennengewinnfaktor (allgemein):

$$g = \frac{S_r}{S_{\text{ref}}} \tag{95.10a}$$

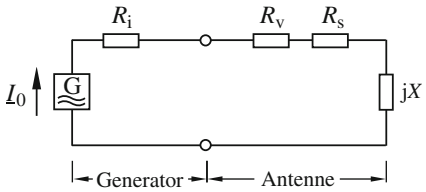


Abb. 95.6 Reihen-Ersatzschaltbild einer Antenne

maximaler Antennengewinnfaktor kurze lineare Antenne:

$$g_{max} = 1{,}5$$

maximaler Antennengewinnfaktor $\lambda/2$-Dipol:

$$g_{max} = 1{,}64 \qquad (95.10b)$$

Hinweis: Die Abhängigkeit des Strahlungswiderstandes R_s vom Verhältnis l/λ ((95.11a) und (95.11b)) führt dazu, dass sich für die Kenngrößen der kurzen linearen Antenne und des $\lambda/2$-Dipols geringfügig voneinander abweichende Werte ergeben.
Daraus leitet sich das Antennengewinnmaß G ab:

$$\textit{Gewinnmaß} \quad G = 10 \cdot \lg(g) \quad \text{in dB} \qquad (95.10c)$$

- *Eingangswiderstand*, *Fußpunktwiderstand* und *Strahlungswiderstand*: Allgemein gilt für den Antennenwiderstand $Z = U/I$. Da sich U und I aber längs der Antenne ändern (gemeint ist nicht die zeitliche, sondern die örtliche Abhängigkeit), ist auch Z ortsabhängig. Man hat deshalb den Eingangswiderstand am Fußpunkt der Antenne definiert. Dieser Fußpunktwiderstand Z_{FP} ergibt sich aus dem Verhältnis von Spannung zu Strom am Fußpunkt: $Z_{FP} = U_{FP}/I_{FP}$. Abb. 95.6 stellt das Ersatzschaltbild einer stromgespeisten Antenne dar. Es bedeuten: R_i Innenwiderstand der Quelle; R_v Verlustwiderstand der Antenne; R_s Strahlungswiderstand der Antenne; X Blindanteil der Antennenimpedanz.
  Der Eingangswiderstand (Fußpunktwiderstand) der Antenne beträgt $R_v + R_s + j\omega X$ und ist wichtig für die Anpassung Generator–Leitung–Antenne. Die Anpassung muss mit konjugiert komplexen Impedanzen erfolgen. Die Strahlungsleistung wird in R_s umgesetzt; dieser Widerstand ergibt sich zu:

Strahlungswiderstand kurze lineare Antenne

$$R_s = \frac{2}{3}\pi \cdot Z_0 \cdot \frac{l^2}{\lambda^2} \quad (l \ll \lambda) \qquad (95.11a)$$

Strahlungswiderstand $\lambda/2$-Dipol

$$R_s = 0{,}194 \cdot Z_0 = 73{,}2\,\Omega \qquad (95.11b)$$

Z_0 Wellenwiderstand des leeren Raumes, $Z_0 \approx 377\,\Omega$, l Antennenlänge in m, λ Wellenlänge in m.

- *Wirkfläche*: An einer Empfangsantenne steht die Leistung P zur Verfügung, die durch die am Ort der Antenne vorhandene Strahlungsdichte S gebildet wird. P entsteht durch Summation von S über eine Fläche A: $P = S \cdot A$. Diese Fläche wird *Wirkfläche* genannt. Es hat sich gezeigt, dass diese Fläche unabhängig von den Antennenabmessungen ist:

$$\textit{Wirkfläche} \quad A = \frac{3}{8\pi} \cdot \lambda^2 \qquad (95.12)$$

λ Wellenlänge.

Beim Empfang einer Strahlung mit $\lambda = 10\,\text{m}$ ($f = 30\,\text{MHz}$) haben eine kurze lineare Antenne und ein $\lambda/2$-Dipol die Wirkfläche $\lambda^2 \cdot \frac{3}{8\cdot\pi} = 11{,}9\,\text{m}^2$. Danach würde man die kurze lineare Antenne dem $\lambda/2$-Dipol wegen der kleineren Abmessungen vorziehen. Macht man allerdings die Abmessungen immer kleiner, nimmt der Strahlungswiderstand ab und damit wird die nutzbare Leistung verringert. Außerdem kann dann der Verlustwiderstand R_v nicht mehr vernachlässigt werden. Schließlich wird die konjugiert komplexe Anpassung nicht mehr möglich, weil der Blindwiderstand mehr als 1000mal so groß ist wie der Strahlungswiderstand. Die kapazitive Antennenkomponente wäre nur durch supraleitende Spulen (große induktive Komponente bei verschwindendem Realteil) kompensierbar.

- *Übertragungsfaktor*, *Funkfelddämpfung*: Der Übertragungsfaktor ist das Verhältnis von Empfangsantennenleistung P_e zu Sendeantennenleistung P_s und ergibt sich zu:

Übertragungsfaktor:

$$\frac{P_E}{P_S} = \left(\frac{\lambda}{4 \cdot \pi \cdot r}\right)^2 \cdot g_S \cdot g_E \qquad (95.13)$$

r Entfernung Sender–Empfänger in m; g_S, g_E Antennengewinnfaktor nach (95.10a), (95.10b) und (95.10c); λ Wellenlänge in m; Index E: Empfänger; Index S: Sender.

Eine für die Praxis nützliche Größe ist die *Funkfelddämpfung* a in dB:

Funkfelddämpfung:

$$a = 10 \cdot \lg \frac{P_S}{P_E} \quad \text{in dB} \qquad (95.14)$$

- *Freiraumdämpfung*: Dies ist die Dämpfung zwischen Sende- und Empfangsantenne, wenn beide als Kugelstrahler ausgebildet sind, und in (95.13) gilt: $g_S = g_E = 1$

Freiraumdämpfungsmaß

$$a_0 = 10 \cdot \lg\left(\frac{4 \cdot \pi \cdot r}{\lambda}\right)^2 = 20 \cdot \lg\left(\frac{4 \cdot \pi \cdot r}{\lambda}\right) \quad \text{in dB} \qquad (95.15)$$

r Entfernung Sender–Empfänger in m; λ Wellenlänge in m.

Siehe dazu auch Beispiel 1 im Kap. 101 „Richtfunktechnik“.

Beispiel 3

Bei einer Übertragung mit $\lambda/2$-Dipolen und dem maximal möglichen Antennengewinnfaktor $g_E = g_S = 1{,}64$ für Sende- und Empfangsantenne nach (95.10a), (95.10b) und (95.10c) und einer Frequenz 100 MHz beträgt die Funkfelddämpfung in einem Abstand von 10 km:

$$a = 10 \cdot \lg \frac{P_S}{P_E} = 10 \cdot \lg \left[\left(\frac{4 \cdot \pi \cdot r}{\lambda} \right)^2 \cdot \frac{1}{g_S \cdot g_E} \right]$$

$$= 10 \cdot \lg \left[\left(\frac{4 \cdot \pi \cdot 10^4 \text{m}}{3\text{m}} \right)^2 \cdot \frac{1}{1{,}64^2} \right] \quad \blacktriangleleft$$

95.3 Ausführungsformen von Antennen

Es folgt eine Zusammenstellung von Sende- und Empfangsantennen, wie sie u. a. auch in der Rundfunk- und Fernsehtechnik häufig eingesetzt werden.

95.3.1 Vertikalantenne

In Abb. 95.7 ist eine Vertikalantenne dargestellt, die als Sendeantenne für den Lang- und Mittelwellen-Rundfunk eingesetzt wird. Ihre Länge beträgt $\lambda/4$. Sie ist aus dem $\lambda/2$-Dipol entstanden, indem die untere Hälfte des Dipols an der Erdoberfläche gespiegelt wird. Damit die Leitfähigkeit der Erdoberfläche erhöht wird, bringt man ein strahlen- oder gitterförmiges Netz von Erdleitern um den Fußpunkt der Antenne herum an. Es ergibt sich ein

$$\textit{Strahlungswiderstand} \quad R_S = 36{,}6\,\Omega$$

Der Vorteil dieser Antennenbauform ist, dass ein Anschluss des Generators geerdet werden kann.

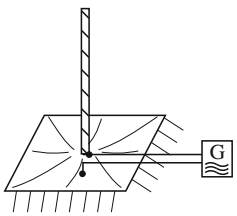


Abb. 95.7 Vertikalantenne als $\lambda/4$-Dipol

95.3.2 Rahmenantenne

Sie wird überwiegend für den Rundfunk-Empfang im Lang- und Mittelwellenbereich eingesetzt (Abb. 95.8a, b) und entspricht dem Hertz'schen Dipol bei Sendeantennen. Es gilt: $a \ll l$. Die Feldverteilung ist in Abb. 95.8c skizziert.

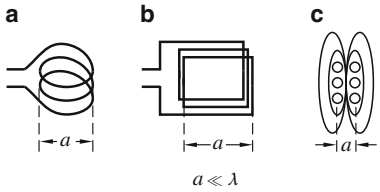


Abb. 95.8 Rahmenantenne. **a** rund, **b** quadratisch, **c** Feldverteilung

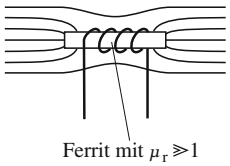


Abb. 95.9 Feldlinienverlauf bei der Ferritantenne

Die Rahmenantenne hat die gleiche Wirkfläche $A_R = \frac{3\lambda^2}{8\pi}$ wie der Hertz'sche Dipol (95.12). Weiter gilt:

Leerlaufspannung:

$$|\underline{U}_0|_R = \omega \cdot \mu_0 \cdot n \cdot A_R \cdot |\vec{\underline{H}}| \qquad (95.16)$$

n Windungszahl, A_R Fläche in m^2 ($\frac{a^2\pi}{4}$ bzw. a^2), ω Kreisfrequenz in 1/s, μ_0 magnetische Feldkonstante in V s/A m, H magnetische Feldstärke in A/m.

Demnach müsste eine Rahmenantenne die gleichen Eigenschaften wie der Hertz'sche Dipol haben. Da aber der Verlustwiderstand R_v (Abb. 95.6) groß ist im Vergleich zum Strahlungswiderstand R_S, steht nur ein geringer Teil der induzierten Spannung für den Empfänger zur Verfügung.

95.3.3 Ferritantenne

Sie besteht aus einem möglichst langgestreckten Stab aus Ferritmaterial mit $\mu_r \gg 1$. Die Feldlinien in der Umgebung des Stabes konzentrieren sich damit im Stab, siehe Abb. 95.9. Damit erhöht sich die magnetische Flussdichte $\vec{\underline{B}}$ im Stab und in der Spule auf das μ_r-fache gegenüber der Rahmenantenne ohne Eisen. Damit ist auch die Leerlaufspannung um diesen Faktor größer:

Leerlaufspannung:

$$|\underline{U}_0|_F = \omega \cdot \mu_0 \cdot \mu_r \cdot n \cdot A_R \cdot |\vec{\underline{H}}| \qquad (95.17)$$

n Windungszahl, A_R Fläche in m^2 ($A_R = \frac{a^2\pi}{4}$), ω Kreisfrequenz in 1/s, μ_0 magnetische Feldkonstante in V s/A m, H magnetische Feldstärke in A/m.

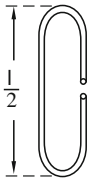


Abb. 95.10 $\lambda/2$-Faltdipol

Die maximale Empfangsleistung bei Anpassung ist zwar bei Rahmen- und Ferritantenne gleich, aber der Strahlungswiderstand R_S (Abb. 95.6) erhöht sich ebenfalls um den Faktor μ_r, sodass er nicht mehr viel kleiner als der Verlustwiderstand R_v ist und damit auch eine höhere induzierte Spannung am Empfängereingang zur Verfügung steht. Der Vorteil der Ferritantenne gegenüber der Rahmenantenne besteht darin, dass die geometrischen Abmessungen bei gleicher Spannung am Empfänger kleiner gewählt werden können. Wichtig ist die Ausrichtung der Ferritantenne in Richtung der magnetischen Feldlinien nach Abb. 95.9. In der Rundfunk-Empfangstechnik (Lang-, Mittelwelle) ist die auf dem Ferritstab aufgebrachte Spule gleichzeitig die Induktivität eines Parallelschwingkreises zur Senderselektion.

95.3.4 $\lambda/2$-Dipol, $\lambda/2$-Faltdipol

Im Bereich von Ultrakurzwellen (Meterwellen, UKW-Rundfunkempfang mit $f \geq 87\,\text{MHz}$ entsprechend $\lambda \leq 3{,}45\,\text{m}$) werden $\lambda/2$-Dipole eingesetzt, weil sie handliche Abmessungen besitzen ($\lambda/2 \leq 1{,}725\,\text{m}$). Der Strahlungswiderstand R_S ist $\approx 70\,\Omega$. Da der Dipol häufig in einiger Entfernung (Dach, Hauswand, Boden) vom Empfänger angebracht ist, sollte der Wellenwiderstand des Verbindungskabels Antenne–Empfänger an R_S angepasst werden. Mit Koaxialkabeln kann diese Bedingung erfüllt werden.

- *Sonderausführung $\lambda/2$-Faltdipol*: Es handelt sich um zwei dicht beieinander angeordnete $\lambda/2$-Dipole, die an den Enden verbunden sind, Abb. 95.10.

Beim Faltdipol sind Gesamt-Strahlungsleistung und Strahlungswiderstand viermal so groß wie beim Dipol. In der Praxis erreicht man $R_S \approx 240\,\Omega$. Auch die Blindkomponente ist beim Faltdipol kleiner und stört damit weniger bzw. kann leichter kompensiert werden.

95.3.5 Breitbanddipol

Je dicker bzw. großflächiger ein Dipol in seinen mechanischen Abmessungen ist, um so kleiner wird seine geometrische Länge. Danach erreicht man mit einer Antennenform nach Abb. 95.11

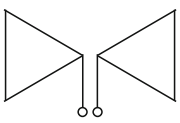


Abb. 95.11 Breitbanddipol

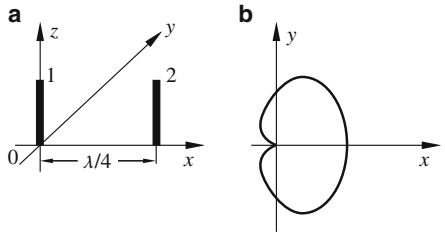


Abb. 95.12 Gruppenstrahler. **a** Antennenanordnung, **b** Richtdiagramm

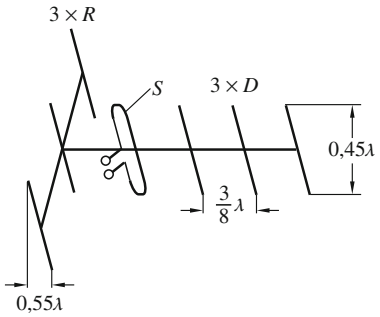


Abb. 95.13 Yagi-Antenne

eine Anpassung des Strahlungswiderstandes an ein *Frequenzband*, wie es in der Fernseh-Empfangstechnik im VHF- und UHF-Bereich auftritt.

95.3.6 Gruppenstrahler

Ordnet man nach Abb. 95.12a zwei Antennen im Abstand $\lambda/4$ an und erregt Antenne 1 mit einem Signal, das dem an Antenne 2 um 90° voreilt, ergibt sich längs der positiven x-Achse ein maximaler Feldverlauf, da sich beide Feldvektoren ohne Phasenverschiebung addieren. In Richtung der negativen x-Achse heben sich beide Felder auf (Abstand $\lambda/4$ bei der Feldberechnung vernachlässigt). Abb. 95.12b zeigt das Richtdiagramm. Der Gruppenstrahler strahlt also fast ausnahmslos in den rechten Halbraum.

95.3.7 Yagi-Antenne

Abb. 95.13 zeigt die prinzipielle Ausführungsform dieses Antennentyps. Ein Faltdipol (S) ist das eigentliche aktive Element, dem zur einen Seite bis zu 25 Direktoren (D) vorgelagert sind. Zur anderen Seite folgen meist 3 Reflektoren (R). Die in der Praxis bewährten Abmessungen und Abstände sind eingetragen.

Es ergibt sich eine verbesserte Richtcharakteristik in Richtung der Direktoren sowie ein Antennengewinn

$$\textit{Antennengewinn} \quad G = 20 \cdot \lg \sqrt{n} \quad \text{in dB} \qquad (95.18)$$

n Anzahl der Antennenelemente.

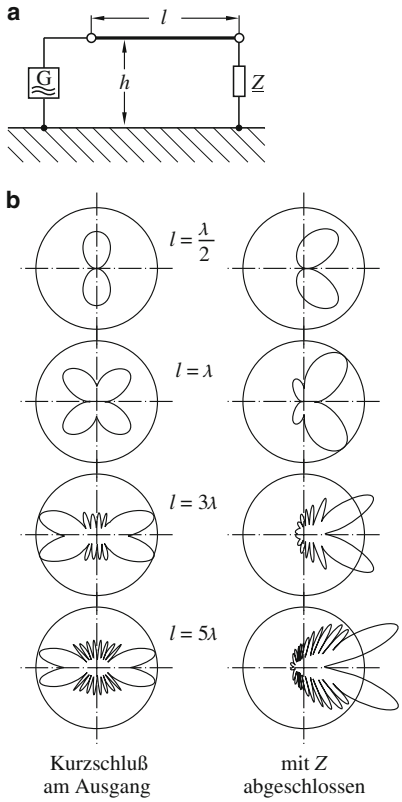


Abb. 95.14 Langdrahtantenne. **a** Anordnung, **b** Richtcharakteristik für verschiedene Verhältnisse l/λ (nach [19])

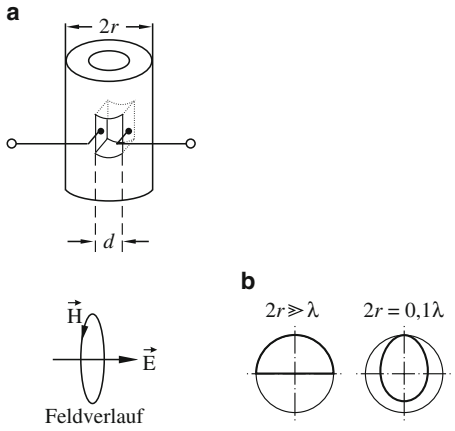


Abb. 95.15 Rohrschlitzstrahler. **a** Aufbau, **b** Richtdiagramme

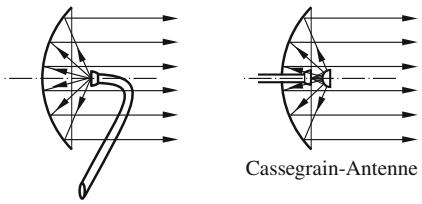


Abb. 95.16 Parabolantenne (nach [19])

95.3.8 Langdrahtantenne

Sie besteht aus einem horizontal gespannten Leiter, der mit der als ideal leitend angenommenen Erdoberfläche eine Zweidrahtleitung bildet. Sie ist am Ende reflexionsfrei mit Z abgeschlossen und wird in einer Höhe $h \approx \lambda/2 \ldots \lambda$ über dem Erdboden angebracht, Abb. 95.14a. Die Länge l wird zu $\geq \lambda/2$ gewählt. Auf der Leitung entstehen Wanderwellen. Die Langdrahtantenne hat eine um so größere Richtwirkung, je größer das Verhältnis l/λ ist, Abb. 95.14b.

95.3.9 Rohrschlitzstrahler

Eine in der Fernseh-Sendetechnik häufig verwendete Antennenbauform nach Abb. 95.15 besteht aus einem in Längsrichtung geschlitzten Rohr. In der Praxis gilt: $d \ll 2 \cdot r$. Die Speisung geschieht an den Längsseiten in Schlitzmitte. Das elektrische Feld verläuft horizontal (horizontale Polarisation), die Richtung des magnetischen Feldes ist im Bild eingetragen. Die Form des Richtdiagramms hängt vom Verhältnis $2 \cdot r/\lambda$ ab und ist in Abb. 95.15b für $2 \cdot r \gg \lambda$ und $2 \cdot r = 0{,}1 \cdot \lambda$ eingetragen.

Für den UKW-Rundfunk und den Fernsehfunk werden häufig bis zu 8 dieser Rohrschlitzstrahler übereinander angeordnet. Damit bleibt auch die horizontale Richtcharakteristik erhalten, der vertikale Öffnungswinkel wird dagegen klein, und damit steigt der Antennengewinn auf etwa 20 dB.

95.3.10 Parabolantenne

Abb. 95.16 stellt den Aufbau einer häufig verwendeten Form dar. Die Funktionsweise ist mit den Gesetzen der Optik erfassbar. Die Quelle muss im Brennpunkt angeordnet sein und wird *Primärstrahler* genannt. Sie kann, besonders bei Empfangsantennen, als Breitbanddipol ausgelegt sein.

Weitere Einzelheiten sind im Kap. 101 „Richtfunktechnik“ aufgeführt. Die *Cassegrain-Antenne* wird im Kap. 102 „Nachrichtenübertragung über Satelliten“ behandelt.

95.4 Wellenausbreitung

95.4.1 Boden- und Raumwelle

Ein Erdboden mit endlicher Leitfähigkeit entzieht dem Strahlungsfeld Energie. Damit nehmen Feldstärke und Strahlungsdichte in Bodennähe mit zunehmender Entfernung von der Sendeantenne stärker ab als in erdferneren Bereichen. Die auf diese Weise beeinflusste Welle heißt *Bodenwelle* und ist in

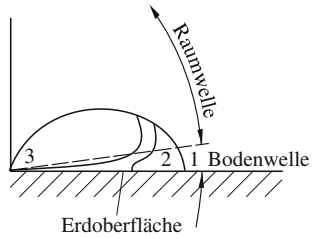


Abb. 95.17 Richtdiagramm zur Erläuterung von Boden- und Raumwelle: *1* kleine, *2* mittlere, *3* große Entfernung von der Sendeantenne (nach [18])

Abb. 95.17 für unterschiedliche Entfernungen r von der Sendeantenne eingetragen.

Ein weiteres Kennzeichen der Bodenwelle ist, dass sie auch durch die Erdkrümmung geschwächt wird. Ihre Reichweite nimmt mit wachsender Wellenlänge zu. Die Welle, die weder durch die Erdkrümmung noch durch die Absorption des Erdbodens beeinflusst wird, heißt *Raumwelle*.

95.4.2 Erdatmosphäre

In Abb. 95.18 sind Temperatur, Ozonanteil und Elektronendichte der Erdatmosphäre bis zu einer Höhe von 600 km dargestellt. Die Elektronenkonzentration nimmt ab einer Höhe von etwa 100 km durch die Sonneneinstrahlung zu, weil deren UV- und Gammastrahlung die Luftmoleküle zunehmend ionisiert. Die Schichtung wird mit D, E, F_1 und F_2 bezeichnet. Nachts verschwinden die Schichten D, E und F_1 fast vollständig.

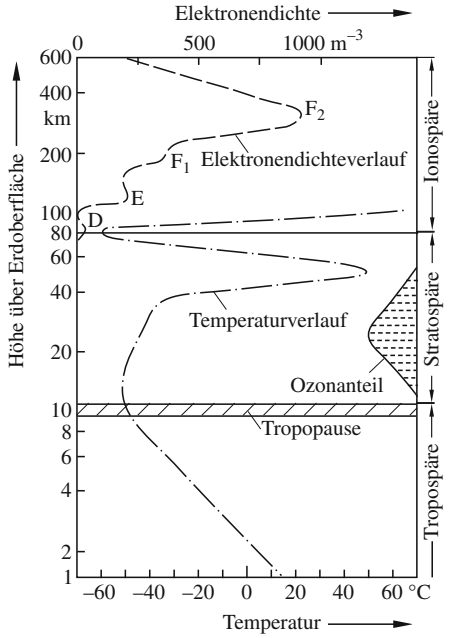


Abb. 95.18 Struktur der Erdatmosphäre mit Verlauf von Temperatur, Ozonanteil und Elektronendichte (nach [19])

95.4.3 Wellenausbreitung im Plasma

Plasma ist ein vollständig ionisiertes Gas, das in gleich viele positive und negative Ladungen aufgespalten ist und damit nach außen neutral erscheint. In der Ionosphäre ist wegen der geringen Luftdichte nahezu ein Plasma vorhanden. Trifft eine elektromagnetische Welle mit der Frequenz f auf ein Plasma, breitet sie sich dort mit der Gruppengeschwindigkeit v_g aus:

Gruppengeschwindigkeit im Plasma:

$$v_g = c \cdot \sqrt{1 - \left(\frac{f_p}{f}\right)^2} \tag{95.19}$$

$c \approx 3 \cdot 10^8$ m/s (Lichtgeschwindigkeit); f_p Plasmafrequenz in Hz mit

$$\textit{Plasmafrequenz} \quad f_p = \frac{1}{2\pi}\sqrt{\frac{n \cdot q^2}{m \cdot \varepsilon_0}} \tag{95.20}$$

n Elektronenkonzentration im Plasma in m^{-3}; q Elementarladung des Elektrons $1{,}602 \cdot 10^{-19}$ A s; m Ruhemasse des Elektrons $9{,}1083 \cdot 10^{-31}$ kg; ε_0 elektrische Feldkonstante.

Mit den in Abb. 95.18 eingetragenen Elektronenkonzentrationen errechnet sich f_p zu:

Schicht	n	f_p
E-Schicht	$\approx 200\,\text{mm}^{-3}$	≈ 4 MHz
F_1-Schicht	$\approx 400\,\text{mm}^{-3}$	$\approx 5{,}6$ MHz
F_2-Schicht	$\approx 900\,\text{mm}^{-3}$	$\approx 8{,}5$ MHz
Minimalwert	$\approx 10\,\text{mm}^{-3}$	$\approx 0{,}9$ MHz
Maximalwert	$\approx 1000\,\text{mm}^{-3}$	≈ 9 MHz

Da die Elektronenkonzentration einen kontinuierlichen Verlauf aufweist, ergibt sich für f_p ebenfalls ein Bereich, der angenähert von 1 bis 10 MHz reicht.

Die Plasmafrequenz hat Bedeutung für den Wellenwiderstand $\underline{Z}_p$ und für die Reflexionseigenschaften des Plasmas:

Wellenwiderstand des Plasmas

$$\underline{Z}_p = \frac{Z_0}{\sqrt{\varepsilon_r}} = \frac{Z_0}{\sqrt{1-\left(\frac{f_p}{f}\right)^2}} \tag{95.21}$$

Z_0 Wellenwiderstand des leeren Raumes, f_p Plasmafrequenz.

a) Für $f < f_p$ ist $\varepsilon_r < 0$ und $\underline{Z}_p$ rein imaginär (Z_0 ist reell). Die einfallende Welle wird reflektiert, Abb. 95.19a.

b) Für $f > f_p$ ist $\varepsilon_r > 0$ und $\underline{Z}_p$ reell. Die Welle breitet sich im Plasma aus. Eine Reflexion findet nur dann statt, wenn die Welle unter einem Winkel auftrifft, der kleiner ist als ein Grenzwinkel $\alpha_{\max}$, Abb. 95.19b. Die Ionisation nimmt von der Randschicht bis zum Maximum kontinuierlich zu, und damit steigt auch die Brechzahl kontinuierlich in Richtung Ionisationsmaximum. Die schräg einfallende Welle wird damit stetig in Richtung des Brechzahlgradienten gebrochen und erreicht schließlich wieder die Erdoberfläche. Wird $\alpha_{\max}$ überschritten, findet keine Reflexion statt, sodass eine *tote Zone* entsteht.

Tab. 95.1 Wellenausbreitung im Bereich 30 kHz bis 30 GHz

Bezeichnung[a]	Wellenlänge	Frequenz	Übertragung über	
			Bodenwelle	Raumwelle
			Reichweite	
Längstwellen; Langwellen LW; „Kilometerwellen“	10 … 1 km	30 … 300 kHz	Ja; wenig gedämpft; einige 1000 km	Nein
Mittelwellen MW; „Hektometerwellen“	1 km … 100 m	300 kHz … 3 MHz	Tagsüber ja; nachts ja; 100 km	Tagsüber nein; nachts ja; 1000 km
Kurzwellen KW; „Dekameterwellen“	100 … 10 m	3 … 30 MHz	Stark gedämpft	Erdballumspannend durch Mehrfachreflexionen
Ultrakurzwellen z. B. UKW-Rundfunk, VHF (Fernsehen Band I und III); „Meterwellen“	10 … 1 m	30 … 300 MHz	Nein	Optische Sicht (geringe Verluste); schwache Komponenten durch Beugung (Berge); evtl. Überreichweiten[b]; 100 km; in Ausnahmefällen 150 … 250 km
Ultra high frequency UHF (z. B. Fernsehen Band IV und V); „Dezimeterwellen“	1 m … 10 cm	300 MHz … 3 GHz	Nein	Optische Sicht (geringe Verluste); schwache Komponenten durch Beugung (Berge); evtl. Überreichweiten[b]; 50 km; in Ausnahmefällen 150 … 200 km
„Zentimeterwellen“; z. B. Rundfunk- und Fernsehübertragung über Satellit	10 … 1 cm	3 … 30 GHz		Quasioptische Sicht; Streuung und Brechung in Luft (Regen, Wolken). Reichweite: a) auf Erdoberfläche: 50 km; in Ausnahmefällen bis 150 km; b) Satellitenübertragung: ≥ 36.000 km bei direkter „Sicht“

Anmerkungen: a) Die in Anführungszeichen gesetzten Bezeichnungen entsprechen DIN 40015. b) Noch schwächere Komponenten entstehen durch Streuung bzw. Reflexion in der Ionosphäre, diese führen aber bei Überreichweiten häufig zu Signalbeeinträchtigungen

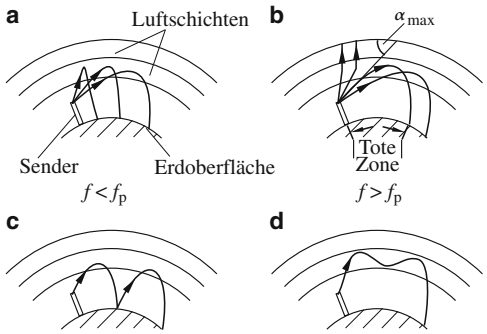


Abb. 95.19 Ausbreitung von Wellen. **a** Reflexion für $f < f_p$, **b** Brechung für $f > f_p$, **c** Mehrfachreflexion Erdoberfläche – Ionosphäre, **d** Mehrfachreflexion innerhalb der Ionosphäre

95.4.4 Wellenausbreitung im Bereich 30 kHz bis 30 GHz

Abb. 95.19c, d stellt weitere Möglichkeiten der Wellenausbreitung über größere Entfernungen dar. Entweder wird die Welle mehrfach zwischen der Ionosphäre und der Erdoberfläche reflektiert (Abb. 95.19c) oder es findet eine Mehrfachreflexion innerhalb der Ionosphäre statt, bevor die Welle endgültig wieder zum Erdboden zurückkehrt, Abb. 95.19d. Auf diese Weise können Wellen sehr große Entfernungen zurücklegen. Allerdings nimmt die Strahlungsleistung mit dem zurückgelegten Weg r aus folgenden Gründen ab:

- Elektrische und magnetische Feldstärke nehmen jeweils mit $1/r$ ab und damit die Strahlungsdichte mit $1/r^2$.
- Bei Brechung bzw. Streuung in der Ionosphäre weitet sich die Welle auf.
- Vor allem tagsüber dämpfen die unteren Schichten der Ionosphäre durch Absorption, und zwar um so mehr, je niedriger die Frequenz der Welle ist (unelastische Stöße). Diese Dämpfung ist für den Bereich Langwelle so groß, dass die Raumwelle praktisch verschwindet.
- *Überreichweiten*: Sie entstehen durch Inhomogenitäten der Atmosphäre, an denen die Welle diffus gestreut wird (engl.: scatter). Der zur Erde zurückkehrende Teil besitzt nur wenig Energie, deshalb ist einiger Aufwand für die Signalaufbereitung am Empfangsort erforderlich. Die Inhomogenitäten und damit die Überreichweiten sind zeitlich nicht konstant und erschweren die technische Nutzung dieses Effektes. Es ergeben sich folgende maximale Reichweiten:

Frequenzbereich	Streuung an	Reichweite maximal etwa
30 … 100 MHz	Troposphäre	2500 km
100 MHz … 1 GHz	Ionosphäre	2000 km

Ein Beispiel für eine Scatterverbindung ist im Kap. 101 „Richtfunktechnik“ aufgeführt.

- *Schwunderscheinungen*: Wenn zwei Wellen auf verschieden langen Wegen zum Empfänger gelangen, können sich beide so überlagern, dass sie sich je nach Phasenlage teilweise oder ganz auslöschen. In der Praxis ändern sich die Amplituden und Weglängen der einzelnen Wellen fortwährend durch den stetigen Wechsel der Bedingungen in der Ionosphäre, und damit schwankt auch die Amplitude der resultierenden Welle am Empfangsort. Diese Schwunderscheinung (engl.: fading) versucht man durch eine signalabhängige Verstärkungsregelung im Empfänger auszugleichen. (Siehe AVR im Kap. 98 „Empfängerschaltungstechnik“.)

In der Tab. 95.1 sind einige Daten über die Wellenausbreitung der technisch genutzten Frequenzbereiche zusammengestellt.

Modulation

96

Wilfried Plaßmann



96.1 Übersicht 1204
96.2 Sinusträger – mit Analogsignal moduliert 1204
96.3 Sinusträger – mit Digitalsignal moduliert 1217
96.4 Pulsträger uncodiert 1221

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_96

96.1 Übersicht

Bei der Modulation werden Signalparameter eines *Trägers* in Abhängigkeit von einem *modulierenden Signal* geändert, mit dem Ziel, den Frequenzbereich des *Basisbandes* (Frequenzbereich des Nutzsignals, modulierendes Signal) in einen anderen Frequenzbereich zu transformieren. Damit erreicht man eine Anpassung an die Kanaleigenschaften, und es wird die Übertragung mehrerer Signale gleichzeitig über einen Kanal möglich, wobei das gesamte übertragene Signal wieder eindeutig in die einzelnen Signale zerlegt werden kann. Eine Anwendung für die Modulation ist die Übertragung verschiedener Rundfunk- und Fernsehprogramme über den Kanal „freier Raum“ mit großer Reichweite, wobei der Empfänger die Auswahl eines bestimmten Programmes ermöglicht. Für den Träger wird bei der Übertragung im freien Raum ein *sinusförmiges* elektromagnetisches Feld verwendet, da andernfalls die Oberschwingungen andere Träger mit gleicher Frequenz stören würden. Bei einer leitungsgebundenen Übertragung über Koaxialkabel oder Hohlleiter werden auch Pulsträger verwendet. Wichtig ist hier, dass das elektromagnetische Feld nicht vom Leiter in den umgebenden Raum austritt. Zur Modulation wird eine nichtlineare Kennlinie eingesetzt, sodass neue Frequenzen entstehen (beabsichtigte Erzeugung von nichtlinearen Verzerrungen).

- *Unterschied Mischung – Modulation*: Bei der Mischung wird nur *eine* der entstehenden Frequenzen ausgenutzt, bei der Modulation *mehrere*. Beispiel für die Mischung: Überlagerungsempfänger, siehe Kap. 98 „Empfängerschaltungstechnik“. Vom Gesamtsignal mit den Frequenzen f_0, f_e, $f_0 \pm f_e$ ($f_0 > f_e$), gebildet aus der Eingangsfrequenz f_e und der Oszillatorfrequenz f_0, wird nur die Differenzfrequenz $f_0 - f_e = f_{ZF}$ (Zwischenfrequenz) weiterverwendet. Beispiel für die Modulation: Amplitudenmodulation (AM) mit Trägerfrequenz und einem oder zwei Seitenbändern. Es werden sowohl die Trägerfrequenz als auch mindestens die Frequenz eines Seitenbandes ausgenutzt.

Folgende Bezeichnungen werden verwendet (DIN 5488, 40110):

$s_M(t)$ Basisbandsignal, Nutzsignal, modulierendes Signal (z. B. Musik- und Sprachsignal mit dem Frequenzbereich 15 Hz bis 15 kHz bei der UKW-Rundfunkübertragung),
$s_T(t)$ Träger (Frequenz z. B. 91,2 MHz bei der UKW-Rundfunkübertragung).

Im Allgemeinen gilt: Die Frequenzen von $s_M(t)$ sind viel kleiner als die Frequenzen von $s_T(t)$.

Abb. 96.1 gibt einen Überblick über die verwendeten Modulationsverfahren.

Durch *Demodulation* gewinnt man aus dem modulierten Signal das Basisbandsignal zurück. Dies geschieht u. a. durch erneute Modulation, wie am Beispiel des Demodulators für amplitudenmodulierte Signale gezeigt wird.

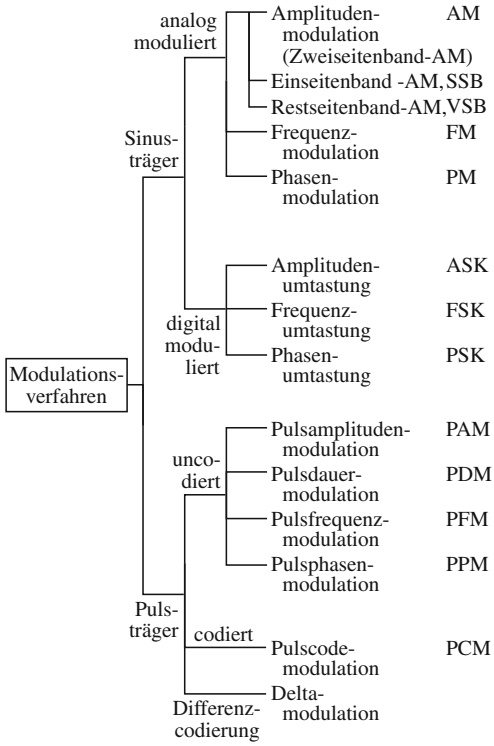


Abb. 96.1 Modulationsverfahren

96.2 Sinusträger – mit Analogsignal moduliert

Der sinusförmige Träger (Index T) hat die Darstellung

$$s_T(t) = \hat{s}_T \cdot \sin(\omega_T t + \varphi_T) \tag{96.1}$$

Die folgenden Parameter können vom modulierenden Signal geändert werden und geben dem Verfahren den Namen:

$\hat{s}_T = \hat{s}_T(t)$ Amplitudenmodulation (AM), mit den Sonderfällen EM, SSB, RM und USB (siehe Abschn. 96.2.1)
$\omega_T = \omega_T(t)$ Frequenzmodulation (FM, Abschn. 96.2.4).
$\varphi_T = \varphi_T(t)$ Phasenmodulation (PM, Abschn. 96.2.4).

Frequenz- und Phasenmodulation werden auch unter dem Oberbegriff *Winkelmodulation* zusammengefasst. Mathematisch besteht der Zusammenhang darin, dass man zur Frequenzmodulation (FM) mit dem differenzierten Phasensignal (PM) moduliert.

$$\begin{aligned} s_{PM}(t) &= \hat{s}_T \cdot \sin[\omega_T t + \varphi_T(t)] \\ &= \hat{s}_T \cdot \sin[\underbrace{\omega_T t + \varphi_T + K_\varphi \cdot u(t)}_{\text{differenzieren}}] \end{aligned} \tag{96.2}$$

mit $\varphi_T(t) = K_\varphi \cdot u(t)$. Durch Differenzieren und entsprechendes Einsetzen erhält man:

$$s_{FM}(t) = \hat{s}_T \cdot \sin\left[\underbrace{\omega_T + K_\varphi \frac{du(t)}{dt}}_{\omega_T(t)}\right] \cdot t \tag{96.3}$$

Anschaulich bedeutet dies, dass sich bei einer *Änderung* der Winkel*geschwindigkeit* auch die Frequenz ändert.

96.2.1 Amplitudenmodulation (AM)

Für das Basisband wird eine einzelne Frequenz angenommen:

$$s_M(t) = \hat{s}_M \cdot \sin(\omega_M t + \varphi_M) \tag{96.4}$$

Tritt ein nichtsinusförmiges Signal auf, wird es gemäß der Gültigkeit des Überlagerungssatzes in eine Summe aus sinusförmigen Signalen mit den entsprechenden Frequenzen und Phasenwinkeln zerlegt. Damit der Empfänger aus dem modulierten Signal das Basisbandsignal wieder zurückgewinnen kann, sind der Träger mit f_T und das obere Seitenband (OS) mit $f_T + f_M$ oder das untere Seitenband (US) mit $f_T - f_M$ oder beide erforderlich. Dabei steht f_M stellvertretend für ein *Frequenzband* (z. B. 15 Hz bis 15 kHz bei der UKW-Rundfunkübertragung).

96.2.1.1 Modulation durch Multiplikation

Es werden die zwei Zeitfunktionen nach (96.1) und (96.4) multipliziert. Diese Art der Modulation kann schaltungstechnisch durch Multiplizierer in Form integrierter Schaltkreise erfolgen. Ist allerdings die Amplitude $\hat{s}_M$ größer als $\hat{s}_T$, entsteht zusätzlich zur Amplitudenmodulation noch eine Phasenmodulation: Ein negativer Träger wird durch ein negatives modulierendes Signal positiv, und er erfährt damit einen Phasensprung von p. Um dieses Problem zu erfassen, führt man den *Modulationsgrad* m ein. Zunächst macht man für die Amplitude des Trägers nach (96.1) den Ansatz:

$$s_T^*(t) = \hat{s}_T + s_M(t) = \hat{s}_T + \hat{s}_M \cdot \sin(\omega_M t) \text{ mit } \hat{s}_T \geq \hat{s}_M \tag{96.5}$$

Der Phasenwinkel φ_M wurde der Einfachheit halber zu Null gesetzt, da hier die entstehenden Frequenzen von Interesse sind und der Phasenwinkel darauf keinen Einfluss hat. Das gilt auch für den Phasenwinkel φ_T des Trägers. Wird jetzt die Multiplikation ausgeführt, ergibt sich unter Anwendung eines Additionstheorems:

$$\begin{aligned} s_{\mathrm{AM}} &= s_T^*(t) \cdot \sin \omega_T t \\ &= [\hat{s}_T + \hat{s}_M \sin(\omega_M t)] \cdot [\sin(\omega_T t)] \\ &= \hat{s}_T \cdot \sin \omega_T t \\ &\quad + \hat{s}_M \cdot [-\cos(\omega_T + \omega_M) \cdot t + \cos(\omega_T - \omega_M) \cdot t] \cdot \frac{1}{2} \end{aligned} \tag{96.6}$$

Der Modulationsgrad wird folgendermaßen definiert (siehe Abschn. 96.2.1.2):

Modulationsgrad

$$m = \frac{\hat{s}_M}{\hat{s}_T} = \frac{\hat{s}_{M\,\mathrm{US}} + \hat{s}_{M\,\mathrm{OS}}}{\hat{s}_T} \tag{96.7}$$

$\hat{s}_{M\,\mathrm{US}}$ Trägeramplitude des unteren Seitenbandes (US) mit der Frequenz $f_T - f_M$, $\hat{s}_{M\,\mathrm{OS}}$ Trägeramplitude des oberen Seitenbandes (OS) mit der Frequenz $f_T + f_M$, $\hat{s}_{M\,\mathrm{US}}$ Trägeramplitude.

Dann folgt für (96.6):

$$s_{\mathrm{AM}} = \hat{s}_T \cdot \left\{ \sin \omega_T t + \frac{m}{2} \cdot [-\cos(\omega_T + \omega_M) \cdot t + \cos(\omega_T - \omega_M) \cdot t] \right\} \tag{96.8}$$

Bei dieser Art der Modulation entstehen nur die drei erforderlichen Kreisfrequenzen ω_T, $\omega_T + \omega_M$ und $\omega_T - \omega_M$. Damit ergibt diese Art der Amplitudenmodulation sehr geringe Verzerrungen, weil jede zusätzlich vorhandene Frequenz mit anderen Frequenzen Summen- und Differenzfrequenzen bilden kann, die im Basisband liegen.

96.2.1.2 Kenngrößen der Amplitudenmodulation

- *Modulationsgrad*: Der Modulationsgrad $m = \hat{s}_M/\hat{s}_T = (\hat{s}_{M\,\mathrm{US}} + \hat{s}_{M\,\mathrm{OS}})/\hat{s}_T$ nach (96.7) wird in der Praxis zu maximal etwa 80 % gewählt. Bei $m = 100\,\%$ verschwindet der Träger, wenn $s_M(t)$ seinen negativen Extremwert einnimmt. Unteres (US) und oberes (OS) Seitenband tragen gleichviel zu $\hat{s}_M$ bei. Modulationsgrade über 100 % sind möglich, haben aber eine zusätzliche Phasenmodulation zur Folge und werden deshalb nur in Ausnahmefällen eingesetzt. Je größer der Modulationsgrad ist, desto größer ist die Signalamplitude im Empfänger nach der Demodulation (größere Lautstärke beim Rundfunkempfang), deshalb unterschreitet man für die lautesten Töne auch nicht die angegebenen 80 % (siehe auch Abb. 96.2a).
- *Seitenbänder*: Aus (96.8) ergibt sich, dass durch die Modulation zwei neue Frequenzen entstanden sind: Neben dem bereits vorhandenen Träger mit der Frequenz f_T entstehen die Frequenzen $f_T + f_M$ und $f_T - f_M$. Da das modulierende Signal in der Praxis aus einem *Frequenzband* im Frequenzbereich von f_1 bis f_2 besteht, werden durch die Modulation zwei *Frequenzbänder* erzeugt:
  Oberes Seitenband (*OS*) mit dem Frequenzbereich
  $(f_T + f_1) \ldots (f_T + f_2)$
  unteres Seitenband (*US*) mit dem Frequenzbereich
  $(f_T - f_1) \ldots (f_T - f_2)$
  Werden beide Seitenbänder übertragen, spricht man auch von einer *Zweiseitenband-Amplitudenmodulation*.
- *Bandbreite*: Durch die Modulation hat man, bezogen auf das obere Seitenband, eine Frequenzverschiebung des Basisbandsignals um f_T erreicht. Die Übertragung erfordert eine Bandbreite von

$$\textit{Bandbreite} \quad B = 2 \cdot (f_2 - f_1) \approx 2 \cdot f_{M\,\max} \tag{96.9}$$

  für Zweiseitenband-Amplitudenmodulation.
  Für die Näherung in (96.9) wurde berücksichtigt, dass die niedrigste modulierende Frequenz f_1 meist Null ist oder aber vernachlässigt werden kann (Rundfunkübertragung: $f_1 = 15\,\mathrm{Hz} \approx 0$ verglichen mit $f_2 = 15\,\mathrm{kHz}$). Für f_2 wurde der anschauliche Wert $f_{M\,\max}$ für die maximal auftretende modulierende Frequenz eingesetzt.
- *Darstellung im Zeit- und Frequenzbereich*: Abb. 96.2 stellt einen modulierten Träger in der Zeit- und in der Frequenzebene dar.

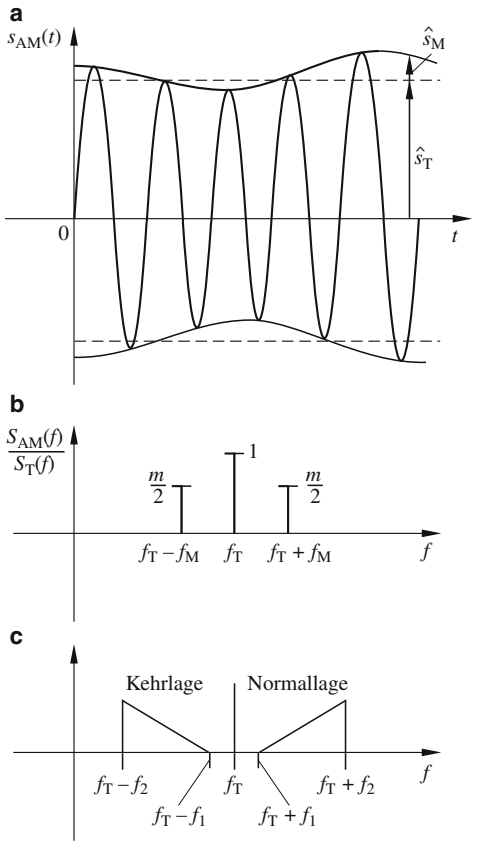


Abb. 96.2 Modulierter Träger. **a** Zeitverlauf, **b** Spektrum, **c** Normal- und Kehrlage

Im oberen Seitenband ist einer höheren Frequenz f_M auch eine höhere Frequenz in diesem Seitenband zugeordnet, man spricht deshalb von *Normallage*. Im unteren Seitenband dagegen gehört zu einer höheren Frequenz f_M eine niedrigere Frequenz im Seitenband ($f_T - f_M$), was durch den Begriff *Kehrlage* charakterisiert wird. Um Normal- und Kehrlage auch optisch erkennen zu können, verwendet man nach Abb. 96.2c die Darstellung mit Dreiecken. Die steigende Seite des Dreiecks steht für steigende Frequenzen f_M im Basisband und hat nichts mit den Amplituden der Frequenzen zu tun! Bei $f_T \pm f_1$ ist die Höhe des Dreiecks Null, bei $f_T \pm f_2$ maximal.

- *Information*: Die Information $s_M(t)$ ist in jedem der zwei Seitenbänder vollständig enthalten, aber nicht im Träger. Trotzdem wird der Träger für die Demodulation unbedingt benötigt. Zur Übertragung sind deshalb ein oder zwei Seitenbänder und der Träger erforderlich, wie sie sich nach (96.8) ergeben. Allerdings kann u. U. auf die *Übertragung* des Trägers verzichtet werden, siehe Abschn. 96.2.2.

96.2.1.3 Modulation an einer quadratischen Kennlinie

Nach Abb. 96.3 liegt in Reihe zu der Summe der Spannungen des Basisbandsignals $u_M(t)$ und des Trägersignals $u_T(t)$ eine Diode. Die Diodenkennlinie soll durch die Abhängigkeit $i = K \cdot u^2$ beschrieben werden, was durch $U_0 > 0{,}7\,\text{V}$ näherungsweise erreicht wird. In (96.4) und (96.5) werden die Amplituden s durch Spannungen ersetzt, sodass sich ergibt:

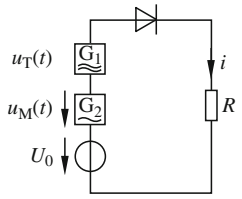


Abb. 96.3 Modulation an einer quadratischen Kennlinie

$$i(t) = K \cdot [U_0 + \hat{u}_M \sin(\omega_M t) + \hat{u}_T \sin(\omega_T t)]^2$$

Ausmultipliziert ergeben sich ein Gleichanteil sowie zeitabhängige Terme mit den Kreisfrequenzen ω_M, $2\omega_M$, $2\omega_T$, ω_T, $\omega_T + \omega_M$, $\omega_T - \omega_M$. Die ersten drei Anteile müssen durch Filter abgetrennt werden, sodass sich dann die gleichen Frequenzen wie in (96.8) ergeben.

96.2.1.4 Modulation an einer nichtlinearen nichtquadratischen Kennlinie

Die zwei oben beschriebenen Verfahren zur Modulation sind ideal, weil entweder nur die benötigten Frequenzen entstehen oder aber die nicht benötigten Frequenzen durch Filter einfach unterdrückt werden können, da sie einen genügend großen Frequenzabstand zu den erforderlichen Frequenzen haben. Im Allgemeinen liegt aber eine nichtlineare nichtquadratische Kennlinie vor, die, am Beispiel der Diode gezeigt, durch eine Reihenentwicklung dargestellt werden kann:

$$i = I_0 + K_1 u + K_2 u^2 + K_3 u^3 + \ldots \qquad (96.10)$$

Setzt man für u den Ausdruck $[U_0 + \hat{u}_M \sin(\omega_M t) + \hat{u}_T \sin(\omega_T t)]$ gemäß Abb. 96.3 ein, erhält man eine Vielzahl von neuen Frequenzen bzw. Frequenzbändern, wobei u. a. die zwei Frequenzbänder $f_T \ldots f_T + f_{M\,\max}$ und $f_T \ldots f_T + 2 \cdot f_{M\,\max}$ entstehen (Näherung wie in (96.9) verwendet). Diese zwei Bänder überlagern sich und können nicht mehr voneinander getrennt werden, sodass auch die Information, die in f_M enthalten ist, nicht mehr unverzerrt zurückgewonnen werden kann.

Allgemein entstehen an einer Kennlinie $i = K_p \cdot u^p$ die Frequenzen $f = m \cdot f_T + n \cdot f_M$ mit $m + n = p$. Für $p = 2$ ergeben sich die Frequenzen $2 \cdot f_T$, $2 \cdot f_M$, $f_T \pm f_M$. Die unerwünschten Frequenzen $2 \cdot f_T$ und $2 \cdot f_M$ lassen sich aber auf einfache Weise durch Filterschaltungen unterdrücken. Ein kubischer Anteil ($p = 3$) in der Kennlinie erzeugt die Frequenzen $3 \cdot f_T$, $3 \cdot f_M$, $2 \cdot f_T \pm f_M$, $f_T \pm 2 \cdot f_M$ und damit den oben erwähnten kritischen Anteil $f_T \pm 2 \cdot f_M$. Man versucht deshalb, eine quadratische Kennlinie anzunähern.

96.2.1.5 Zeigerdiagramm

Das Zeigerbild der AM nach (96.8) enthält drei Zeiger mit dem gleichen Drehsinn, aber unterschiedlichen Winkelgeschwindigkeiten. Man nimmt nun den Trägerzeiger als feststehend an

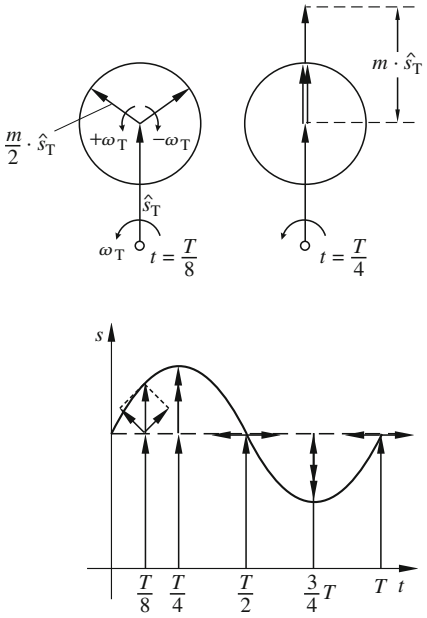


Abb. 96.4 Zeigerbild der AM zu zwei unterschiedlichen Zeitpunkten

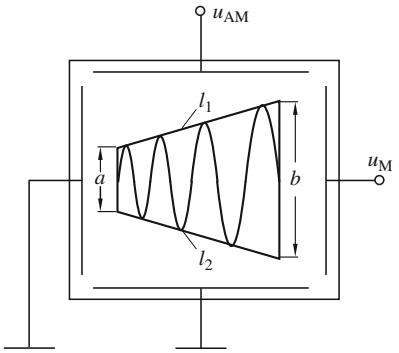


Abb. 96.5 Modulationstrapez

und lässt die Seitenbandzeiger mit den Winkelgeschwindigkeiten $+w_M$ und $-w_M$ symmetrisch zum Träger drehen, Abb. 96.4. Die Summe der Seitenbandzeiger zeigt wegen der Symmetrie stets in Richtung des Trägerzeigers und bewegt sich zwischen den Grenzen $\hat{s}_T(1+m)$ und $\hat{s}_T(1-m)$.

Der Vorteil dieser Darstellung liegt darin, dass man die Auswirkungen von Störungen oder anderen Veränderungen einfach erkennen kann. Erreichen beispielsweise die beiden Seitenbänder den Empfänger nicht mit gleicher Amplitude und ändern sich die Amplituden fortlaufend um unterschiedliche Werte, liegt der Summenzeiger nicht mehr in Richtung des Trägers, sondern schwankt um die durch den Träger gegebene mittlere Lage. Das ergibt eine zusätzliche Phasenmodulation.

96.2.1.6 Modulationstrapez

Zur Kontrolle des Modulationsgrades und der Modulationsverzerrungen verwendet man eine Messschaltung nach Abb. 96.5.

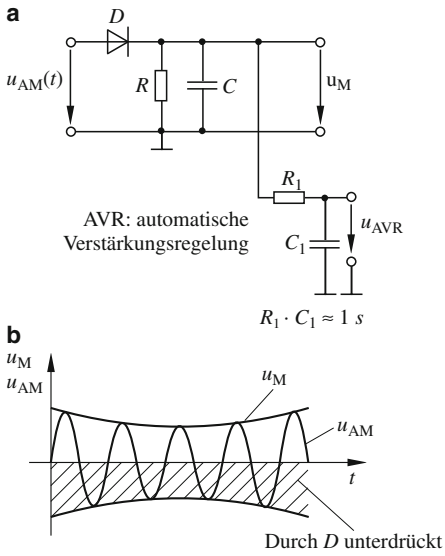


Abb. 96.6 Demodulation von AM. **a** Schaltung, **b** Betrachtung im Zeitbereich

An die Y-Ablenkplatten eines Oszilloskops wird das amplitudenmodulierte Signal u_{AM} angeschlossen, an die X-Ablenkplatten das modulierende Signal u_M. Der Schirm stellt das skizzierte Bild dar. Es gibt Auskunft über folgende Größen:

- *Modulationsgrad*: $m = \frac{b-a}{b+a}$
- *Modulationsverzerrungen*: Der Grad der Verzerrungen lässt sich an den Linien l_1 und l_2 abschätzen: Sind sie Geraden, gibt es keine Verzerrungen, andernfalls sind sie mit zunehmenden Verzerrungen stärker gekrümmt.

96.2.1.7 Demodulation von AM

Die Demodulation von amplitudenmodulierten Signalen ist sehr einfach. Abb. 96.6 zeigt die Schaltung. Sie wird auch als *Hüllkurvendemodulation* bezeichnet.

Betrachtung im Zeitbereich: Die Diode trennt die negativen Spannungswerte ab. Der Tiefpass aus R und C ist so bemessen, dass eine Spitzenwertgleichrichtung stattfindet (*Hüllkurvendetektor*, *Einhüllendendemodulation*) und die Spannung am Kondensator der Hüllkurve und damit dem Verlauf der modulierenden Spannung u_M folgt. Für die Dimensionierung gilt: $\omega_{M\max} < 1/RC < \omega_T/10$.

Betrachtung im Frequenzbereich: Die Trägerfrequenz und die Frequenzen des oberen und unteren Seitenbandes nach (96.8) gelangen als Summe auf die Diode mit der „idealen“ Diodenkennlinie (siehe Abschn. 96.2.1.3) $i = K \cdot u^2$:

$$\begin{aligned} i &= K \cdot u^2 \\ &= K \cdot \left\{ \sin \omega_T t + \frac{m}{2} \cdot [-\cos(\omega_T + \omega_M)t + \cos(\omega_T - \omega_M)t] \right\}^2 \end{aligned} \tag{96.11}$$

Nach Quadrierung des Klammerausdruckes ergibt sich zum einen die gewünschte Frequenz f_M, zum anderen Frequenzen,

die im Bereich um f_T bzw. $2 \cdot f_T$ liegen. Der Kondensator unterdrückt die letztgenannten Frequenzen, und es bleibt nur das gewünschte Signal u_M übrig. Leitet man u_M über einen Tiefpass mit einer Grenzfrequenz $f_g = 1/(2\pi R_1 C_1) \ll f_{M\,\min}$ (Abb. 96.6), kann man die Spannung u_{AVR} zur automatischen Verstärkungsregelung (AVR) heranziehen, da ihr Wert der Amplitude des Trägers proportional ist. Auf diese Weise lassen sich einigermaßen konstante Trägeramplituden und damit Lautstärken am Empfänger erreichen. Das ist besonders wichtig beim Autoradioempfang mit den ständig schwankenden Trägeramplituden.

96.2.1.8 Leistung von Träger und Seitenbändern

Der Träger benötigt die Leistung $P_T = \frac{\hat{u}_T^2}{2 \cdot Z}$, *ein* Seitenband

$$P_{\mathrm{SB}} = \frac{1}{2 \cdot Z} \left(\frac{1}{2} \hat{u}_T \cdot m \right)^2 = \frac{\hat{u}_T^2}{2 \cdot Z} \cdot \frac{m^2}{4} = P_T \cdot \frac{m^2}{4}.$$

Die Gesamtleistung ist:

$$P_{\mathrm{ges\,AM}} = P_T + 2 \cdot P_{\mathrm{SB}} = \underbrace{\frac{\hat{u}_T^2}{2 \cdot Z}}_{\text{Träger}} + \underbrace{\frac{\hat{u}_T^2}{2 \cdot Z} \cdot \frac{m^2}{2}}_{\substack{\text{beide} \\ \text{Seitenbänder}}} = \frac{\hat{u}_T^2}{2 \cdot Z} \cdot \left(1 + \frac{m^2}{2} \right) \quad (96.12)$$

Z Lastwiderstand, den die Trägerspannung speist.

Beispiel 1

a) Gilt für einen Träger der Modulationsgrad $m = 1$, entfallen nach (96.12) $\frac{1}{1+\frac{m^2}{2}} = \frac{1}{1+\frac{1}{2}} = \frac{2}{3}$ der Gesamtleistung auf den Träger und je $1/6$ auf jedes Seitenband.

b) Für $m = 0{,}8$ erhält jedes Seitenband 12,1 %, der Träger 75,8 % der Gesamtleistung. ◀

96.2.1.9 Störungen bei amplitudenmodulierten Signalen

Hier werden die zwei am häufigsten vorkommenden Störungen betrachtet, d. h. diejenigen, die durch Nachbarträger und Rauschen verursacht werden.

1. Störungen durch sinusförmigen Nachbarträger

Der Nachbarträger hat eine Trägerfrequenz f_N, die innerhalb des Spektrums des Nutzträgers $f_T \pm f_{M\,\max}$, d. h. im Bereich des unteren und oberen Seitenbandes, liegt. (Annahme: Im Empfänger werden Frequenzanteile außerhalb der Seitenbänder unterdrückt.)

a) Der Nachbarträger ist nicht moduliert. Dann entsteht nach der Demodulation ein Pfeifton mit der Differenzfrequenz $f_N - f_T$, und es ergibt sich nach dem Demodulator ein niederfrequentes Verhältnis $a_{S,N0}$ von Nutz- zu Störsignal zu:

$$a_{S,N0} = \frac{\hat{u}_T}{\hat{u}_N} \cdot m \quad (96.13)$$

$\hat{u}_T$ Amplitude des Nutzsignalträgers in V, $\hat{u}_N$ Amplitude des Nachbarsignal-(Störsignal-)Trägers in V, m Modulationsgrad des Nutzsignalträgers.
Der Nachbarsignal-(Störsignal-)Träger kann auch durch Einschaltvorgänge (Fourierzerlegung einmaliger Vorgänge), Gewitter, atmosphärische Störungen, Harmonische von nichtsinusförmigen Spannungs- und Stromverläufen in Motoren, Frequenzumrichtern und dergleichen entstehen. Deshalb werden Übertragungen im freien Raum häufig gestört. Mit wachsendem Abstand Sender–Empfänger nimmt die Häufigkeit der Störungen und deren Intensität zu, da die Amplitude der Trägerschwingung abnimmt. Deshalb ist bei Amplitudenmodulation ein störungsarmer Empfang nur dann möglich, wenn man sich im Nahbereich eines Senders befindet oder ein entfernter Sender mit entsprechend hoher Leistung und hohem Modulationsgrad arbeitet. Da oberes und unteres Seitenband zu gleichen Teilen zum Basisbandsignal beitragen (96.7), hat das demodulierte Signal bei Einseitenbandübertragung (siehe Abschn. 96.2.2.1) nur die halbe Amplitude und damit auch nur den halben Störabstand gegenüber Zweiseitenbandübertragung. Der Störabstand wird weiterhin verringert, wenn bei der Einseitenbandübertragung mit unterdrücktem Träger der im Empfänger regenerierte Träger eine Phasenverschiebung zu dem ursprünglichen Senderträger aufweist.

b) Der Nachbarträger ist mit m_N amplitudenmoduliert. Dann tritt neben dem Pfeifton nach (96.13) noch die Modulation des Nachbar-(Stör-)Trägers als *Nebensprechen* auf. Für das niederfrequente Verhältnis $a_{S,N1}$ von Nutz- zu Störsignal ergibt sich:

$$a_{S,N1} = \left(\frac{\hat{u}_T}{\hat{u}_N} \right)^2 \cdot \frac{m}{m_N} \quad (96.14)$$

$\hat{u}_T$ Amplitude des Nutzsignalträgers in V, $\hat{u}_N$ Amplitude des Nachbarsignal-(Störsignal-)Trägers in V, m Modulationsgrad des Nutzsignalträgers, m_N Modulationsgrad des Nachbarsignal-(Störsignal-)Trägers.

2. Störungen durch Rauschen

Nach dem Demodulator erhält man einen NF-Signal-Rauschabstand

$$S_{SN} = 10 \cdot \lg \frac{m^2 \cdot P_T}{2 \cdot f_{M\,\max} \cdot F \cdot k \cdot T}\,\mathrm{dB} \quad (96.15)$$

m Modulationsgrad; P_T Trägerleistung nach dem Demodulator in W; $f_{M\,\max}$ maximale Basisbandfrequenz in Hz; F Rauschzahl des Empfängers; k Boltzmann-Konstante $1{,}38 \cdot 10^{-23}$ W s/K; T absolute (thermodynamische) Temperatur in K.

96.2.1.10 Kreuzmodulation

Bei dieser Art der Modulation handelt es sich um einen *unerwünschten* Modulationseffekt. Liegen eine amplitudenmodulierte Schwingung ($\omega_T \pm \omega_M$) und ein unmodulierter Träger (ω_{T2}) an einer Kennlinie mit einem Glied 3. Ordnung (Abschn. 96.2.1.4, (96.10)), entstehen u. a. Kombinationsschwingungen mit $\omega_{T2} \pm \omega_M$. Das heißt, dass der ursprünglich unmodulierte Träger (ω_{T2}) mit ω_M *kreuzmoduliert* worden ist. Natürlich tritt dieser Effekt auch auf, wenn der Träger mit der Kreisfrequenz ω_{T2} amplitudenmoduliert ist.

96.2.2 Sonderformen der Amplitudenmodulation

Grundlage für die dargestellten Sonderformen sind (96.9) und (96.12). Da wesentlich mehr Sender als Frequenzbänder vorhanden sind, hat man versucht, den pro Sender erforderlichen Bandbreitenbedarf zu reduzieren. Außerdem hat man nach Lösungen gesucht, die im Träger enthaltene Leistung auf die Seitenbänder zu verlagern, da nur die Seitenbänder Information übertragen. Auf diese Weise kann man mit gleicher Leistung höhere Reichweiten erzielen bzw. bei gleicher Entfernung Sender–Empfänger einen störungsfreieren Empfang sicherstellen oder mit geringerer Leistung die gleiche Reichweite wie vorher erreichen.

96.2.2.1 Einseitenbandmodulation (ESB, SSB)

Da ein Seitenband die vollständige Information enthält, wird das andere Seitenband unterdrückt. (ESB-Übertragung: *E*in-*S*eiten-*B*and-Übertragung; SSB: *s*ingle-*side*band). Damit sinkt die Bandbreite nach (96.9) auf die Hälfte: $B_{\mathrm{ESB}} \approx f_{M\,\max}$. Für die Übertragung des Trägers, der nach (96.11) für die Demodulation unbedingt benötigt wird, gibt es drei Möglichkeiten:

1. Der Träger wird mit voller Amplitude übertragen. Dann wird lediglich maximal 1/6 der Gesamtleistung eingespart, die das unterdrückte Seitenband bei $m = 100\,\%$ benötigt (siehe Beispiel 1).
2. Der Träger wird mit einer Restamplitude, z. B. 10 %, übertragen: *AM mit Restträger*. Vorteile: a) Die Leistungseinsparung für den Träger kann in eine größere Leistung für das Seitenband umgesetzt werden. b) Durch den Restträger bleibt eine eventuell erforderliche Phasenbeziehung zwischen Senderträger und einem weiteren Signal erhalten (Beispiel: UKW-Stereo-Rundfunk mit Hilfsträger). Nachteile: a) Wird die sowieso schon geringe Trägeramplitude am Empfangsort durch Schwunderscheinungen zeitweise zu Null, ist wegen des fehlenden Trägers keine Demodulation und damit kein Signalempfang möglich. b) Im Empfänger muss die Amplitude des Restträgers gegenüber dem Seitenband ausreichend verstärkt werden, damit eine verzerrungsarme Demodulation möglich ist (Erklärung siehe folgender Absatz). Das erfordert sehr steilflankige und damit aufwendige Filterschaltungen.
3. Der Träger wird ganz unterdrückt. Vorteil: Die gesamte Senderleistung gelangt in ein Seitenband, und man erreicht entweder eine große Reichweite oder einen störungsärmeren Empfang. Nachteile: a) Der unterdrückte Träger muss im Empfänger wieder zugesetzt werden, da sonst keine Demodulation möglich ist. Die Frequenz dieses im Empfänger erzeugten *Hilfsträgers* muss zeitlich konstant sein und darf von der im Sender unterdrückten Trägerfrequenz nicht abweichen, andernfalls sind die vom Empfänger demodulierten Frequenzen alle um den Differenzbetrag zwischen Sender- und Empfängerträger verschoben. b) Das im Empfänger demodulierte Signal ist zusätzlich phasenmoduliert, was zu nichtlinearen Verzerrungen führt. Diese Tatsache lässt sich am Abb. 96.4 ablesen: Wenn einer der Seitenbandzeiger fehlt, zeigt der Summenzeiger aus Träger und einem Seitenband nicht mehr in Richtung des Trägers, was der genannten Phasenmodulation entspricht. Man wählt deshalb die Amplitude des Hilfsträgers ($\hat{u}_T$) groß im Verhältnis zur Amplitude des Seitenbandes ($\hat{u}_M$) und kann dadurch die Verzerrungen klein halten. Einen Anhaltswert für den die Verzerrungen kennzeichnenden Klirrfaktor gibt die folgende Formel. Da die erste Oberschwingung (oder zweite Harmonische) den Klirrfaktor wesentlich bestimmt, wurde nur ihr Einfluss berücksichtigt:

$$\textit{Klirrfaktor} \quad k \approx k_2 = \frac{\hat{u}_M}{2} \cdot \frac{1}{4 \cdot \hat{u}_T} \tag{96.16}$$

$\hat{u}_T$ Amplitude des Hilfsträgers in V, $\hat{u}_M$ Amplitude des Seitenbandes in V.
Wichtig ist, dass das Verhältnis $\hat{u}_M/\hat{u}_T$ *nicht* mehr durch den Modulationsgrad des Senders, sondern durch die empfängerseitig zugefügte Trägeramplitude $\hat{u}_T$ bestimmt wird. c) Durch die empfängerseitige Addition des Trägers geht eine eventuell erforderliche Phasenbeziehung zwischen Träger und Basisband oder einem zweiten Träger verloren (Beispiel: Stereo-Rundfunkübertragung mit zwei Trägern, deren Phasenwinkel zueinander den Stereoeffekt beeinflusst). Bei der einkanaligen Übertragung von Sprache, Musik und Bildern (Mono-Rundfunk, Schwarz-Weiß-Fernsehen) spielt ein Phasenwinkel zwischen senderseitigem Träger und dem im Empfänger zugesetzten Träger aber keine Rolle. d) Zur Verringerung des Störabstandes siehe Abschn. 96.2.1.9 „Störungen bei amplitudenmodulierten Signalen“.

Zur Einseitenbandmodulation sind spezielle Schaltungen entwickelt worden, weil es nicht ohne Weiteres möglich ist, aus einem zweiseitenbandmodulierten Signal mit Träger den Träger und ein Seitenband mit einer Filterschaltung herauszufiltern. Das ist genau dann schwierig, wenn die Frequenz $f_{M\,\min}$ sehr klein ist. Bei der Rundfunkübertragung ist $f_{M\,\min} = 15\,\mathrm{Hz}$. Verwendet man $f_T = 1\,\mathrm{MHz}$, wäre ein Filter erforderlich, dass bis einschließlich 1,000 MHz alle Frequenzen unterdrückt, ab 1,000015 MHz die entsprechenden Frequenzen durchlässt. Erreicht $f_{M\,\min}$ den Wert Null (Fernsehen: Bild schwarz), können die erforderlichen Filterschaltungen prinzipiell nicht realisiert werden.

Abb. 96.7 zeigt einen Ringmodulator, der als Ausgangssignal nur die zwei Seitenbänder enthält und den Träger unterdrückt.

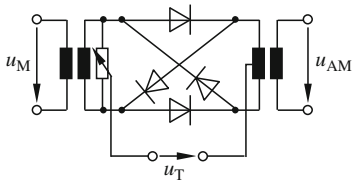


Abb. 96.7 Ringmodulator

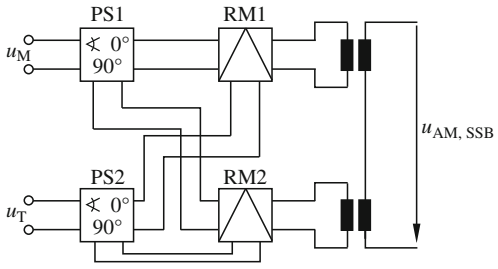
Ein Seitenband muss noch, sofern das technisch möglich ist, durch geeignete Filterschaltungen abgetrennt werden. Eine Weiterentwicklung ist in Abb. 96.8 dargestellt. Die beiden Ringmodulatoren liefern die zwei Seitenbänder ohne Träger. Durch die Phasenschieber sind die beiden oberen Seitenbänder der Modulatoren um 180° verschoben und heben sich über die Summenbildung in dem Transformator auf. Die unteren Seitenbänder dagegen sind in Phase und addieren sich.

96.2.2.2 Restseitenbandmodulation (RM, VSB)

Die Restseitenbandübertragung (*R*estseitenband*m*odulation, *v*estigial *s*ide*b*and) wird bei der Fernsehübertragung angewendet. Dort ist $f_{M\,\max} \approx 5{,}5\,\text{MHz}$, sodass bei Zweiseitenbandübertragung eine Bandbreite von etwa 11 MHz erforderlich gewesen wäre. Damit hätte man zu wenig Sender unterbringen können. Bei der Bildübertragung ist außerdem die Übermittlung eines Gleichwertes erforderlich (z. B. Dunkelsteuern der Bildröhre beim Umschalten im Studio). Senderseitig wären damit bei der Trägerfrequenz f_T „unendlich steile" Filterkurven erforderlich gewesen, um genau ab der Trägerfrequenz das zweite Seitenband vollständig zu unterdrücken. Da dies technisch nicht möglich ist, setzt man die Restseitenbandübertragung nach Abb. 96.9 ein.

Ein Seitenband wird vollständig, das andere nur teilweise übertragen. Das dargestellte Frequenzspektrum lässt sich senderseitig durch geeignete Filterschaltungen realisieren. Wird dieses Signal demoduliert, erhält man die Anteile mit Frequenzen bis $f_M = 0{,}75\,\text{MHz}$ mit doppelter Amplitude, da beide Seitenbänder gleichgewichtig zum Gesamtsignal beitragen.

Bis 1,25 MHz sinkt die Amplitude kontinuierlich auf den einfachen Wert, der auch für noch höhere Frequenzanteile entsteht. Um diese Signalverzerrungen empfängerseitig auszugleichen,



PS1, 2: Phasenschieber mit 0° und 90°
RM1, 2: Ringmodulatoren

Abb. 96.8 Modulator für Einseitenband-AM

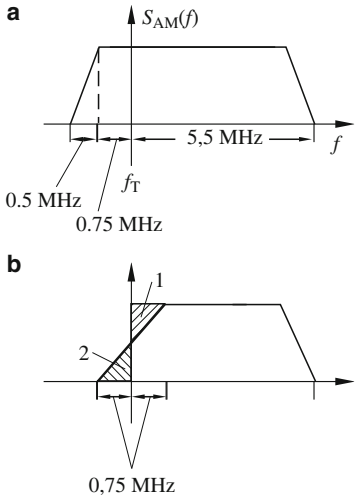


Abb. 96.9 Restseitenbandübertragung (Beispiel Fernsehen). **a** Frequenzspektrum des Senders, **b** Durchlasskurve des Empfängers

wird eine Durchlasskurve nach Abb. 96.9b eingesetzt. Der Verlauf der *Nyquistflanke* sollte möglichst gerade sein und muss durch geeignete Filterschaltungen realisiert werden. Wichtig ist weiterhin, dass die Flächen 1 und 2 gleich groß sind, denn dann erhält man alle Frequenzanteile mit gleicher Amplitude.

Der Vorteil der Restseitenbandübertragung liegt darin, dass, am Beispiel Fernsehen gezeigt, senderseitig nicht mehr 11 MHz, sondern nur noch etwas weniger als 7 MHz Bandbreite benötigt werden. Damit können mehr Sender in Betrieb genommen werden, was wegen der Forderung nach flächendeckender Versorgung der Bevölkerung mit mehreren Fernsehprogrammen erforderlich ist. Die technische Realisierung ist sender- und empfängerseitig mit vertretbarem Aufwand möglich.

96.2.2.3 Quadraturmodulation

Diese auch in der Farbfernsehtechnik angewendete Modulationsart beruht darauf, dass man zwei Träger mit gleicher Frequenz f_T, aber mit einem Phasenwinkel von 90° zueinander erzeugt. Dem einen Träger wird ein Signal mit ω_{M1} und dem anderen ein Signal mit ω_{M2} aufmoduliert. Die phasenverschobenen Träger werden unterdrückt, damit sie im Empfänger keine Störungen verursachen. Die Amplituden- und Phasenbeziehung der zwei Signale mit ω_{M1} und ω_{M2} zueinander muss nach der Demodulation erhalten bleiben, weil daraus über eine Summenbildung das Farbsignal gebildet wird. Setzt man den Träger empfängerseitig frequenzstabil mit 0° bzw. 90° zu (Quarzoszillator mit 90°-Phasenschieber), ist diese Bedingung erfüllt. Weitere Einzelheiten zur Anwendung der Quadraturmodulation in der Farbfernsehtechnik zeigt Abschn. 96.2.2.

96.2.3 Technische Ausführung der Amplitudenmodulation

Abb. 96.10 stellt Schaltungen mit Transistoren zur Zweiseitenband-Amplitudenmodulation dar, und zwar mit Modulation an

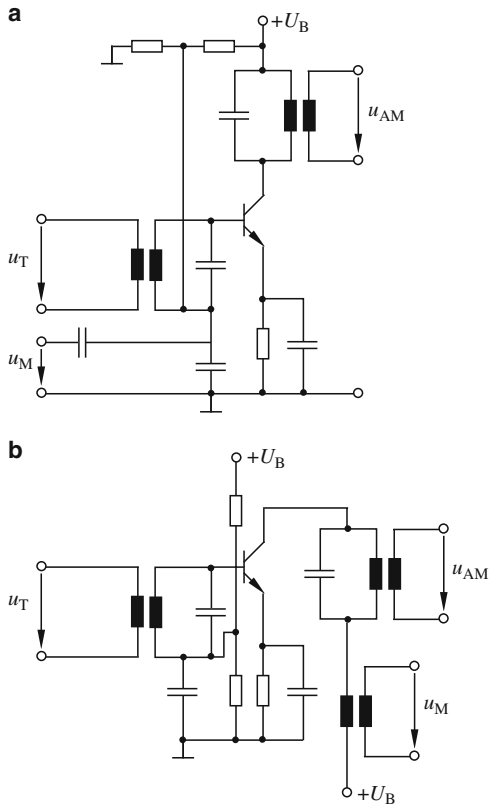


Abb. 96.10 Schaltungen zur Amplitudenmodulation. **a** Modulation an der Basis (additiv), **b** Modulation am Kollektor (multiplikativ)

der Basis (*additives Verfahren*) oder am Kollektor (*multiplikatives Verfahren*).

Bei der Modulation an der Basis wird die nichtlineare nichtquadratische Kennlinie $i_B = f(u_{BE})$ des Transistors ausgenutzt, was nach Abschn. 96.2.1.4 zu Verzerrungen führt. Deshalb sind Modulationsgrade von maximal 70 % zulässig. Es werden Wirkungsgrade von höchstens 50 % erreicht. Vorteil dieser Modulationsschaltung ist, dass nur eine geringe Leistung für das modulierende Signal erforderlich ist. Wesentlich verzerrungsärmer ist die Schaltung mit Modulation im Kollektorkreis (Modulation durch Multiplikation nach Abschn. 96.2.1.1). Allerdings muss die Leistung des modulierenden Signals in der Größenordnung der an die Antenne abgegebenen Leistung liegen. Es sind Modulationsgrade bis 100 % möglich. Der Wirkungsgrad kann bis zu 90 % betragen. In Rundfunk- und Fernsehsendern werden anstelle der Transistoren Senderöhren mit Ausgangsleistungen im kW-Bereich eingesetzt.

In der Rundfunkübertragungstechnik hat man zunächst die einfach zu handhabende Zweiseitenband-AM eingesetzt. Der Übergang zur Einseitenbandmodulation mit unterdrücktem Träger wurde noch nicht vollzogen, weil der Anwender dann die Trägerfrequenz genau kennen und auch frequenzstabil zur Verfügung haben müsste. Bei der Fernseh-Bildübertragung hat man die Restseitenbandübertragung gewählt, weil die Einseitenbandübertragung technisch nicht realisierbar ist (siehe Abschn. 96.2.2.2). Für den Kurzwellenbereich gibt es aber Überlegungen, die besonders dort eingesetzten hohen Senderleistungen bei gleicher Reichweite durch Einseitenbandmodulation zu reduzieren. Um auch die bisher benutzten Empfänger weiterhin verwenden zu können, soll in einem ersten Schritt die Trägeramplitude auf die Hälfte gesenkt werden (Einseitenbandmodulation mit Restträger).

Für die Rundfunk- und Fernsehübertragung gelten folgende Daten (siehe auch Kap. 99, „Ton- und Bildübertragung“):

1. Lang-, Mittel- und Kurzwellenrundfunk: Zweiseitenband-AM mit einer höchsten Basisbandfrequenz $f_{M\,max} = 4{,}5$ kHz. Die Senderleistungen betragen maximal etwa 1 MW. Wegen der Vielzahl der unterzubringenden Sender wurde die Bandbreite des Basisbandes auf maximal 4,5 kHz beschränkt.
2. Fernsehfunk: Bildübertragung in Restseitenbandmodulation, Daten siehe Abb. 96.9. Die Tonübertragung geschieht frequenzmoduliert.

96.2.4 Winkelmodulation

Unter dem Begriff *Winkelmodulation* werden *Frequenz-* (FM) und *Phasenmodulation* (PM) zusammengefasst.

96.2.4.1 Grundlagen

Es wird wieder ein sinusförmiger Träger nach (96.1) verwendet:

$$s_T(t) = \hat{s}_T \cdot \sin \underbrace{(\omega_T t + \varphi_T)}_{\Phi(t)} \tag{96.17}$$

Der Winkel $\Phi(t)$ kann auf zwei Arten in Abhängigkeit vom modulierenden Signal geändert werden.

1. *Phasenmodulation* (PM):

$$\Phi(t) = \omega_T t + \underbrace{\alpha_P \cdot s_M(t)}_{\varphi(t)}; \tag{96.18}$$

2. *Frequenzmodulation* (FM):

$$\Phi(t) = \underbrace{[\omega_T + \alpha_F \cdot s_M(t)]}_{\Omega(t)} \cdot t + \varphi_T; \tag{96.19}$$

α_P und α_F sind Modulatorkonstanten. Der Zusammenhang zwischen beiden Modulationsarten ist gegeben durch

$$\Omega(t) = \frac{d\Phi(t)}{dt} \approx \frac{\Delta\Phi(t)}{\Delta t} \tag{96.20}$$

Die Frequenz ist gleich der Änderungsgeschwindigkeit des Phasenwinkels, geteilt durch 2π (Zusammenhang zwischen Frequenz- und Phasenmodulation siehe auch Abschn. 96.4.2 „Pulsfrequenz- und Pulsphasenmodulation“). Größter Frequenz- und größter Phasenhub treten nicht gleichzeitig auf, da sie über einen Differenzialquotienten miteinander

verknüpft sind. Eine cosinusförmige modulierende Spannung bei FM hat einen Phasenhub, der sinusförmig ist, also um 90° nacheilt. Setzt man wieder ein sinusförmiges modulierendes Signal nach (96.4) voraus und setzt den Winkel $\varphi_M = 0$, so folgt: $s_M(t) = \hat{s}_M \cdot \sin(\omega_M t)$. Eingesetzt in (96.18) und (96.19) ergibt sich für die beiden Modulationsarten:

PM:

$$s_{\mathrm{PM}}(t) = \hat{s}_T \cdot \sin(\omega_T t + \underbrace{\alpha_P \cdot \hat{s}_M}_{\text{Phasenhub } \Delta\Phi} \cdot \sin \omega_M t) \qquad (96.21)$$

FM:

$$s_{\mathrm{FM}}(t) = \hat{s}_T \cdot \sin[(\omega_T t + \underbrace{\alpha_F \cdot \hat{s}_M}_{\text{Frequenzhub } \Delta\Omega} \cdot \sin \omega_M t) + \varphi_T] \qquad (96.22)$$

Im Phasenhub $\alpha_P \cdot \hat{s}_M$ ist damit die Information über die *Amplitude* des modulierenden Signals (Lautstärke bei Tonübertragung) enthalten. Senderseitig wird für die größte zu übertragende Amplitude (größte Lautstärke) ein Wert $\alpha_P \cdot \hat{s}_M\big|_{\max}$ vorgegeben. Das Gleiche gilt für den (Kreis-)Frequenzhub $\alpha_F \cdot \hat{s}_M$. Für die größte zu übertragende Amplitude (lautester zu übertragender Ton) ist beim UKW-Rundfunk $\Delta\Omega\big|_{\max} = 2\pi \cdot f_{\max} = 2\pi \cdot 75\,\mathrm{kHz}$ festgelegt. Die Information über die *Frequenz* des modulierenden Signals ist in der Anzahl der Wechsel zwischen maximaler und minimaler Frequenz pro Sekunde enthalten.

Für den UKW-Rundfunk gilt: Bei einer Trägerfrequenz $f_T = 100\,\mathrm{MHz}$ und einer Frequenz des modulierenden Signals von $f_M = 4\,\mathrm{kHz}$ und maximaler Lautstärke ändert sich die Frequenz des Senders 4000mal pro Sekunde zwischen den Werten $100\,\mathrm{MHz} - 75\,\mathrm{kHz}$ und $100\,\mathrm{MHz} + 75\,\mathrm{kHz}$. Bei einem leiseren Ton gleicher Frequenz und halber Lautstärke ändert sich die Frequenz des Senders wiederum 4000mal pro Sekunde, aber jetzt beispielsweise zwischen den Werten $100\,\mathrm{MHz} - 30\,\mathrm{kHz}$ und $100\,\mathrm{MHz} + 30\,\mathrm{kHz}$ (Hinweis: Kein linearer Zusammenhang zwischen Frequenzhub und f_M, siehe dazu *Preemphasis* und *Deemphasis*).

96.2.4.2 Kenngrößen

Mit (96.18) bis (96.22) ergeben sich folgende Kenngrößen:

- *Phasenhub PM*: $\Delta\Phi = \alpha_P \cdot s_M$
- *(Kreis-)Frequenzhub PM*: $\alpha_P \cdot \hat{s}_M \cdot \omega_M$
  (Folgt durch Differenzieren von (96.21) gemäß (96.20) von $\alpha_P \cdot \hat{s}_M \cdot \sin \omega_M \cdot t$.)
- *(Kreis-)Frequenzhub FM*: $\Delta\Omega = \alpha_F \cdot \hat{s}_M$
- *Phasenhub FM*: $\Delta\Phi = \frac{\alpha_F \cdot \hat{s}_M}{\omega_M}$
  (Folgt durch Integrieren von $\alpha_F \cdot \hat{s}_M \cdot \sin \omega_M \cdot t$.)

Auch bei FM ist der *Phasenhub* wichtig, weil sich der Einfluss von Störungen und die Bandbreite auf den Phasenhub beziehen. Daher wurde der Modulationsindex als Kenngröße eingeführt:

- *Modulationsindex η bei FM*:

$$\eta = \Delta\Phi = \frac{\Delta\Omega}{\omega_M} = \frac{\Delta f_{T\max}}{f_M} = \frac{\Delta F}{f_M} \qquad (96.23)$$

ΔF Spitzenhub.

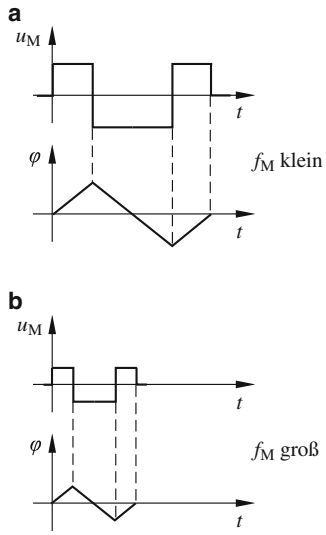


Abb. 96.11 Zusammenhang zwischen Phasenhub φ und Frequenz f_M. **a** f_M klein, **b** f_M groß

Für den UKW-Rundfunk gilt:

$f_{M\max} = 15\,\mathrm{kHz}$; $\Delta F = 75\,\mathrm{kHz}$; $\Rightarrow \eta = \eta_{\min} = 75/15 = 5$. Der Modulationsindex ist abhängig von der Frequenz des modulierenden Signals und wird häufig für $f_{M\max}$ angegeben. Er steigt mit sinkender modulierender Frequenz f_M und kann sehr groß werden. Damit nimmt auch der Dynamikbereich (Lautstärkebereich) große Werte an. Das gilt ebenso für den Phasenhub, wie aus Abb. 96.11 zu ersehen ist.

Hinweis: Der Modulationsgrad $m = \frac{\Delta F}{f_T}$, der bei der Amplitudenmodulation eine wichtige Kenngröße ist, hat hier keine Bedeutung.

96.2.4.3 Zeigerdarstellung

Der Zeiger des Trägers dreht sich ohne Modulation mit konstanter Winkelgeschwindigkeit proportional zur Frequenz f_T. Er dient als Bezugszeiger. Wird mit einem sinusförmigen Signal nach Abb. 96.12 moduliert, erhöht sich für positives u_M der Phasenwinkel j, und der Trägerzeiger eilt dem Bezugszeiger vor, maximal bis $+\Delta F$. Das ist gleichbedeutend mit einer Frequenzänderung Δf, die der Winkeländerung nach (96.20) um 90° voreilt. Bei negativem u_M eilt der Trägerzeiger dem Bezugszeiger nach, maximal bis $-\Delta F$. Wie bei der Zeigerdarstellung der Amplitudenmodulation stellt man sich den Bezugszeiger ruhend vor, sodass der Trägerzeiger im Rhythmus von u_M zwischen $+\Delta F$ und $-\Delta F$ hin und her pendelt (*Pendelzeiger*).

96.2.4.4 Spektrum und Bandbreitenbedarf

Wird ein Träger mit der Frequenz f_T mit einem Signal der Frequenz f_M frequenzmoduliert, so entstehen außer f_T folgende Frequenzen:

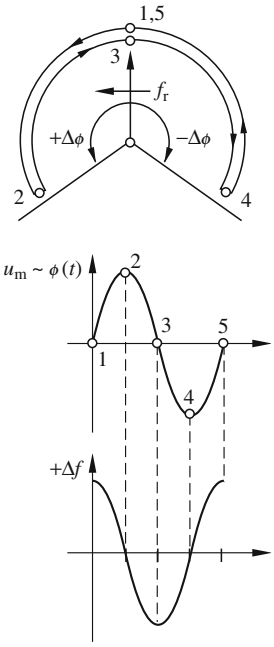


Abb. 96.12 Zeigerdarstellung bei der Winkelmodulation

Frequenzspektrum bei FM:

$$f_T \pm f_M; \quad f_T \pm 2f_M; \quad f_T \pm 3f_M;$$
$$f_T \pm 4f_M; \quad \ldots; \quad f_T \pm nf_M; \quad \ldots$$

Es entstehen unendlich viele Frequenzen im Abstand von $\pm f_M$ zueinander. Die Amplituden dieser einzelnen Schwingungen lassen sich Abb. 96.13 entnehmen. Mathematisch handelt es sich um *Besselfunktionen 1. Art n-ter Ordnung*. Es bedeuten:

$J_0(\eta)$ Amplitude der Trägerschwingung mit der Frequenz f_T
$J_1(\eta)$ Amplitude der Schwingungen mit den Frequenzen $f_T + f_M$ und $f_T - f_M$
$J_2(\eta)$ Amplitude der Schwingungen mit den Frequenzen $f_T + 2f_M$ und $f_T - 2f_M$
$J_3(\eta)$ Amplitude der Schwingungen mit den Frequenzen $f_T + 3f_M$ und $f_T - 3f_M$

usw.

Die Amplituden hängen ab vom Modulationsindex η. Für $\eta = 0$, d. h. ohne Modulation, existiert nur die Trägerschwingung, was auch anschaulich einsichtig ist. Für $\eta \approx 2{,}4$ beispielsweise verschwindet die Trägerschwingung ($J_0 = 0$, gilt auch für $\eta \approx 5{,}5$ usw.), für $\eta \approx 3{,}8$ gibt es keine Schwingungen mit den Frequenzen $f_T + f_M$ und $f_T - f_M$ ($J_1 = 0$, gilt auch für $\eta \approx 7$ usw.). Negative Werte von J_n bedeuten eine Phasenverschiebung von 180° gegenüber positiven Werten.

Da die Amplituden der Besselfunktionen erst für $\eta \to \infty$ gegen Null gehen, ist theoretisch eine unendlich große Bandbreite erforderlich. In der Praxis kann und muss man die Bandbreite begrenzen. Die dadurch entstehenden nichtlinearen Verzerrungen lassen sich berechnen und dienen als Grundlage für die folgenden Dimensionierungen. Man unterteilt, abhängig von η, in zwei Modulationsarten:

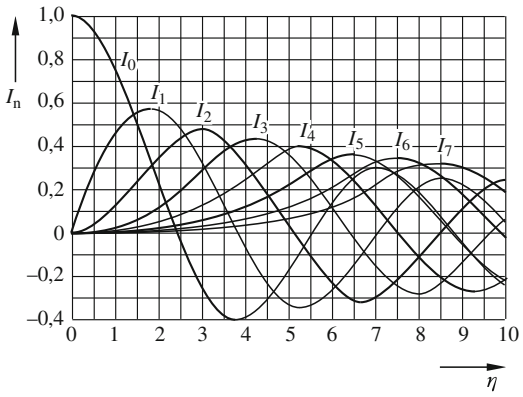


Abb. 96.13 Besselfunktionen $J_n(\eta)$ für $n = 1 \ldots 7$

1. *Schmalband-Frequenzmodulation*: Es gilt: $\eta < 1$. Dann ergeben nur J_0 und J_1, also Trägerfrequenz und die Frequenzen $f_T + f_M$ und $f_T - f_M$, nennenswerte Amplituden. Damit ist das Spektrum aber gleich dem der Amplitudenmodulation, wenn man die übrigen Frequenzanteile vernachlässigt.
Bandbreite bei Schmalband-Frequenzmodulation:

$$B_{\mathrm{FM}} \approx 2B_M \approx 2f_{M\,\max}$$

2. *Breitband-Frequenzmodulation*: Es gilt: $\eta > 1$. Dann begrenzt man die Bandbreite bei einem Klirrfaktor $k \approx k_3 \leq 1\,\%$ bei der Rundfunkübertragung auf:
Bandbreite bei Breitband-Frequenzmodulation:

$$B_{\mathrm{FM}} \approx 2(\Delta F + B_M) \approx 2(\Delta F + f_{M\,\max})$$

(siehe (96.23)). In dieser sog. *Carson-Formel* wurde berücksichtigt, dass erst die dritte Harmonische der Signalschwingung zum Klirrfaktor beiträgt. Beim UKW-Rundfunk ergibt sich $B_{\mathrm{UKW}} \approx 2(\Delta F + f_{M\,\max}) \approx 2(75\,\mathrm{kHz} + 15\,\mathrm{kHz}) = 180\,\mathrm{kHz}$. Bei der Übertragung von Trägerfrequenz-Vielkanal-Fernsprechsignalen durch Frequenzmodulation in der Richtfunktechnik werden teilweise Klirrfaktoren von 0,01 % gefordert. Dann ist die Übertragungsbandbreite, verglichen mit der Bandbreite bei 1 % Klirrfaktor, um den Faktor 1,5 höher zu wählen.

96.2.4.5 Störungen bei winkelmodulierten Signalen

Die hier behandelten Störungen werden verursacht durch benachbarte Träger und Rauschen. Ihr Einfluss auf frequenzmodulierte Signale wird im Ergebnis angegeben, weitere Einzelheiten sind beispielsweise [20] zu entnehmen.

1. Störungen durch sinusförmige Nachbarträger

Nachbarträger sind sinusförmige Stör-Träger mit einer Frequenz, die innerhalb der Bandbreite des Nutzträgers von $f_T \pm B_T/2$ liegt. Für das niederfrequente Störverhältnis $a_{\mathrm{FM,NF}}$ nach dem Demodulator bei Frequenzmodulation ergibt sich:

Niederfrequentes Störverhältnis

$$a_{\mathrm{FM,NF}} = \frac{\hat{u}_N}{\hat{u}_{\mathrm{NF}}} \approx \frac{\hat{u}_N}{\hat{u}_S} \cdot \frac{|f_T - f_N|}{\Delta f_T} \tag{96.24a}$$

$\hat{u}_N$ Amplitude des Störsignalträgers in V, $\hat{u}_{\mathrm{NF}}$ Amplitude des Niederfrequenzsignals in V, $\hat{u}_S$ Amplitude des Nutzsignalträgers in V, $|f_T - f_N|$ Frequenzabstand zwischen Nutz- und Störsignalträger in Hz, Δf_T Frequenzhub in Hz

Das niederfrequente Störverhältnis nimmt zu durch die folgenden drei Ursachen:

a) Steigende Amplitude des Störsignalträgers
b) Zunehmender Abstand $|f_T - f_N|$ zwischen Nutz- und Störträger. Dieser Wert ist maximal an der oberen Frequenzgrenze des modulierenden Signals:

$$a_{\mathrm{FM,NF}}\big|_{\max} \approx \frac{\hat{u}_N}{\hat{u}_S} \cdot \frac{f_{M\max}}{\Delta \mathrm{F}} = \frac{\hat{u}_N}{\hat{u}_S} \cdot \frac{1}{\eta_{\min}} \tag{96.24b}$$

Index S: (Nutz-)Signal; Index N: Noise (Störsignal).
Durch Frequenzmodulation erhält man gegenüber Amplitudenmodulation eine Verbesserung des Störabstandes S:

$$S_{\mathrm{AM}} = \frac{\hat{u}_{S(\mathrm{HF})}}{\hat{u}_{N(\mathrm{HF})}}; \quad S_{\mathrm{FM}} = \frac{\hat{u}_{S(\mathrm{HF})}}{\hat{u}_{N(\mathrm{HF})}} \cdot \eta$$

nach (96.24b) folgt daraus:

$$\frac{S_{\mathrm{FM}}}{S_{\mathrm{AM}}} = \eta \quad \text{bzw.} \quad \frac{S_{\mathrm{FM}}}{S_{\mathrm{AM}}}\bigg|_{\min} = \eta_{\min} = \eta\big|_{f_{T\max}} \tag{96.25}$$

Das Verhältnis ist minimal bei maximaler Frequenz des modulierenden Signals $\hat{s}_M$.
c) Sinkender Frequenzhub, also bei kleineren Amplituden des modulierenden Signals $\hat{s}_M$. Dieses Ergebnis ist auch anschaulich zu erwarten.

2. Störungen durch Rauschen

Die im folgenden dargestellten Ergebnisse sind [21] entnommen und ermöglichen den Vergleich des Leistungsverhältnisses von Nutz- (P_S) zu Störsignal (P_N) auf der Hochfrequenzseite (Index HF) und auf der Niederfrequenzseite (Index NF):

a) Schmalband-FM ($\eta < 1$):

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}} \cdot 3\eta_{\min}^2$$

b) Breitband-FM ($\eta > 1$):

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}} \cdot 3\eta_{\min}^3 \tag{96.26}$$

c) Zum Vergleich: Amplitudenmodulation mit $m = 1$:

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}}$$

(siehe (96.13)).

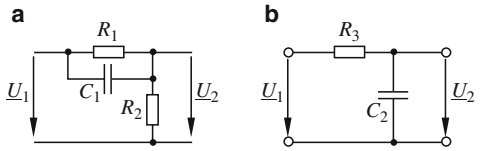


Abb. 96.14 Schaltungen zur Preemphase (**a**) und Deemphase (**b**)

Damit ermöglicht die Breitband-Frequenzmodulation ($\eta_{\min} > 1$) eine störungsfreiere Übertragung als die Amplitudenmodulation (Faktor $3\eta_{\min}^3$). Breitband-FM ist auch weniger störanfällig als Schmalband-FM ($\eta_{\min} < 1$).

Beispiel 2

Die Leistung eines Nutzsignalträgers (P_S) und die eines Störsignalträgers (P_N) seien gleich groß. Das Nutzsignal wird zum Vergleich amplituden- und frequenzmoduliert übertragen.

a) Bei Amplitudenmodulation ergibt sich:

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}} = 1 \quad (m = 1)$$

b) Schmalband-Frequenzmodulation mit $0{,}5 \le \eta \le 0{,}9$.
Für $\eta = \eta_{\min} = 0{,}5$ folgt:

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}} \cdot 3 \cdot \eta_{\min}^2 = 1 \cdot 3 \cdot 0{,}5^2 = 0{,}75$$

c) Breitband-Frequenzmodulation mit $1{,}2 \le \eta \le 5$.
Für $\eta = \eta_{\min} = 1{,}2$ folgt:

$$\frac{P_S}{P_N}\bigg|_{\mathrm{NF}} \approx \frac{P_S}{P_N}\bigg|_{\mathrm{HF}} \cdot 3 \cdot \eta_{\min}^3 = 1 \cdot 3 \cdot 1{,}2^3 = 5{,}2$$

Das Verhältnis von Nutz- zu Störsignal ist bei der Breitband-Frequenzmodulation am größten. Schmalband-Frequenzmodulation ergibt gegenüber Amplitudenmodulation u. U. sogar ein ungünstigeres Verhältnis. ◀

96.2.4.6 Preemphase, Deemphase

Aus (96.26) geht hervor, dass der Störabstand um so größer ist, je größer η und damit je kleiner die Basisbandfrequenz f_M ist. Deshalb erhöht man senderseitig die Amplituden der höheren Frequenzen, bevor moduliert wird. Dieser als *Preemphase* bezeichnete Vorgang setzt beispielsweise beim UKW-Rundfunk ab etwa 3,2 kHz ein und wird schaltungstechnisch über einen Hochpass realisiert. Er wird im Empfänger durch einen RC-Tiefpass mit gleicher Grenzfrequenz rückgängig gemacht (*Deemphase*), um einen linearen Amplitudenfrequenzgang zu erhalten. Abb. 96.14 zeigt Schaltungen, mit denen die Preemphase (a) bzw. Deemphase (b) erreicht wird.

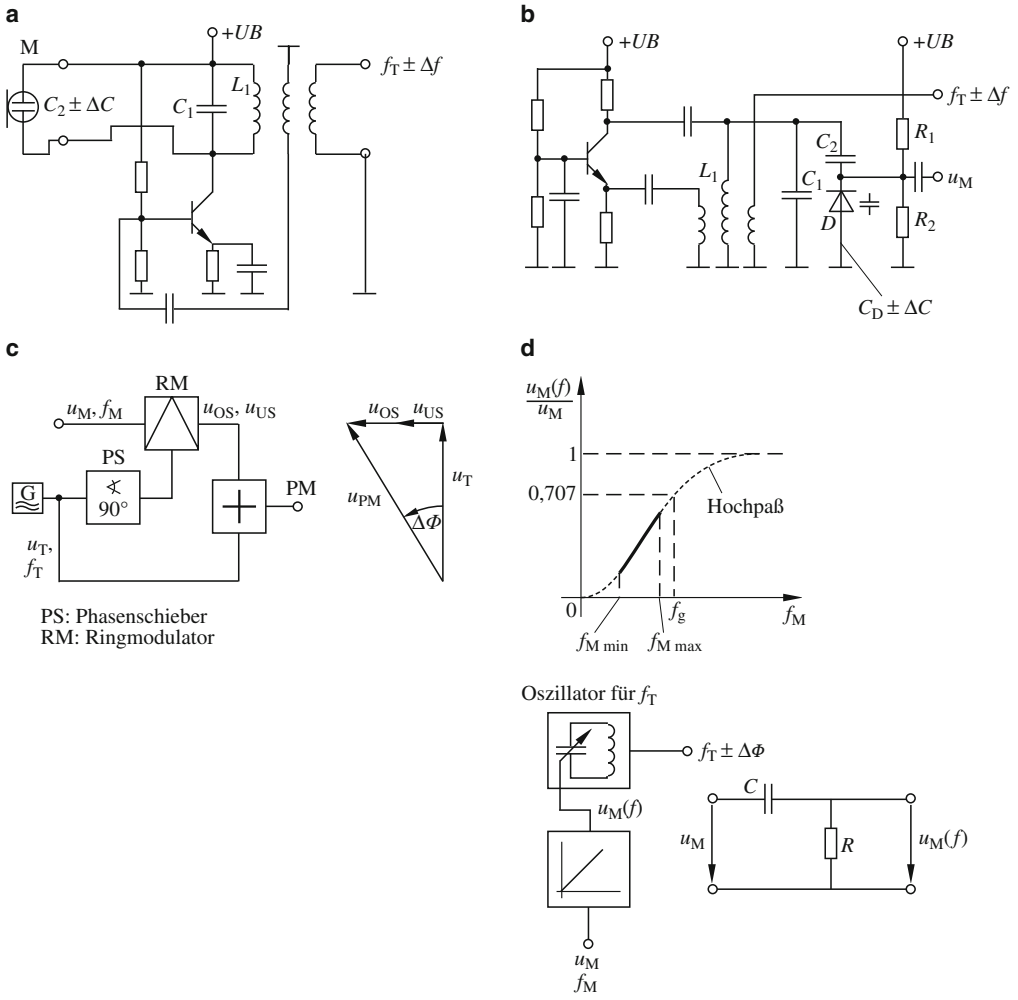


Abb. 96.15 Erzeugung von Winkelmodulation. **a** FM mit Kondensatormikrofon, **b** FM mit Kapazitätsdiode, **c** PM durch Überlagerung von zwei um 90° verschobenen Trägern, **d** PM durch FM und Vorverzerrung des Basisbandes

96.2.4.7 Erzeugung von Frequenz- und Phasenmodulation

Abb. 96.15a zeigt eine Oszillatorschaltung, in der eines der frequenzbestimmenden Elemente ein Kondensatormikrofon M ist, dessen Kapazität $C_2 \pm \Delta C$ sich linear mit dem Schalldruck ändert. Damit ändert sich auch die Frequenz $f_T = 1/[2\pi\sqrt{L_1(C_1 + C_2 \pm \Delta C)}]$ für $C_1 + C_2 \gg \Delta C$ nahezu linear mit der Amplitude des modulierenden Signals (Sprache, Musik). Dabei wurde von der Beziehung $1/(1 \pm \Delta x) \approx 1 \mp \Delta x$ für $|\Delta x \ll 1$ Gebrauch gemacht. Nach Abb. 96.15b liegt parallel zum Schwingkreis aus L_1 und C_1 eine Kapazitätsvariationsdiode D, deren Kapazität sich in erster Näherung linear zur modulierenden Spannung u_M ändert. Der Spannungsteiler aus R_1, R_2 legt den Arbeitspunkt der Diode D in Sperrrichtung und damit eine Kapazität C_D fest. Mit dem Kondensator C_2 wird diese Gleichspannung vom Schwingkreis abgetrennt, und außerdem kann man damit auch den Frequenzhub einstellen. In der Praxis wird man, sofern f_T konstant ist (Rundfunksender, Funkfernsteuersen- der), den frequenzbestimmenden Schwingkreis durch einen Schwingquarz ersetzen, der eine wesentlich höhere Frequenzkonstanz aufweist als der LC-Schwingkreis.

Zur Erzeugung von Phasenmodulation zeigt Abb. 96.15c ein Verfahren, das seltener angewendet wird, dafür aber das Prinzip deutlich zeigt. Der um 90° verschobene Träger (f_T) wird im Ringmodulator mit dem modulierenden Signal (f_M) amplitudenmoduliert. Der Träger ist unterdrückt. Dieses Signal wird zum nicht verschobenen nicht modulierten Träger hinzuaddiert. Trägerspannung u_T und die Spannungen der Seitenbänder, u_{OS}, u_{US}, stehen senkrecht aufeinander. Steigt die Amplitude von u_M, steigen proportional die Spannungen u_{OS}, u_{US}. Damit dreht sich der Zeiger von u_{PM} um den Winkel $\Delta\Phi$ aus der Lage von u_T, was der gewünschten Phasenmodulation entspricht:

$$\textit{Phasenhub} \quad \Delta\Phi = \arctan \frac{\hat{u}_{OS} + \hat{u}_{US}}{\hat{u}_T} \tag{96.27}$$

Ausnutzbar ist ein Bereich bis $\Delta\Phi_{max} \approx 0{,}5$. Das ergibt einen Klirrfaktor von 2,5 %. Allerdings ist die Amplitude von u_{PM} größer geworden als die von u_T, was einer zusätzlichen unerwünschten Amplitudenmodulation entspricht. Diese muss durch

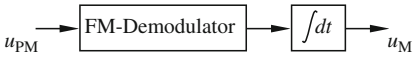


Abb. 96.16 Demodulation von PM

eine Begrenzerschaltung unwirksam gemacht werden. Ein häufig angewendetes Schaltungsprinzip zur Phasenmodulation zeigt Abb. 96.15d. Die modulierende Spannung u_M passiert zunächst ein Hochpassfilter aus R und C, für dessen Grenzfrequenz gilt: $f_g = 1/(2\pi RC) \gg f_{M\,\max}$. Damit ist die modulierende Spannung $u_M(f)$ frequenzabhängig. Nimmt man eine niedrige Frequenz f_M für u_M an, ergibt sich für $u_M(f)$ eine niedrige Amplitude, sodass die Oszillatorfrequenz nur geringfügig im Rhythmus von u_M geändert wird. Hat dagegen u_M bei gleicher Frequenz eine große Amplitude, ist auch $u_M(f)$ entsprechend groß, sodass die Frequenz des Oszillators in einem größeren Bereich geändert wird. Hält man die Amplitude von u_M konstant und vergrößert die Frequenz f_M, nimmt die Amplitude von $u_M(f)$ ebenfalls zu. ΔF darf sich aber wegen der als konstant angenommenen Amplitude von u_M nicht ändern. Da andererseits der Phasenhub nach Abb. 96.11 mit wachsender Frequenz f_M abnimmt, wird diese Abnahme genau durch die über den Hochpass bewirkte Zunahme der Amplitude von $u_M(f)$ kompensiert. Man erkennt hier den engen Zusammenhang zwischen Frequenz- und Phasenmodulation.

96.2.4.8 Demodulation von Frequenz- und Phasenmodulation

Die nachfolgend dargestellten Demodulatorschaltungen dienen zur Demodulation von frequenzmodulierten Signalen. Sollen phasenmodulierte Signale demoduliert werden, ist nach Abb. 96.16 dem FM-Demodulator eine Integrierstufe nachzuschalten (siehe dazu (96.20)).

Da in der Amplitude des frequenzmodulierten Trägers keine Information enthalten ist, andererseits Störungen häufig als Änderungen der Amplitude auftreten, schaltet man dem Demodulator eine Amplitudenbegrenzerschaltung vor. Die Oberschwingungen des jetzt nicht mehr sinusförmigen Trägers unterdrückt man durch Selektionsstufen in Form von Schwingkreisen. Prinzipiell unterscheidet man vier Arten der Demodulation, Abb. 96.17.

Beim ersten Verfahren (a) verwendet man die frequenzabhängigen Blindwiderstände von Spulen und Kondensatoren oder die frequenzabhängige Impedanz von Schwingkreisen bzw. deren Phasendrehung. Man überführt die Frequenzmodulation in eine Amplitudenmodulation. Zwei Beispiele dazu sind in Abb. 96.18 dargestellt.

Nach Verfahren zwei (b) gelangt der frequenzmodulierte Träger einmal direkt auf den einen Eingang eines Multiplizierers, zum anderen über ein frequenzabhängiges Phasendrehglied mit $\varphi(f) = 90° \pm \Delta\varphi(f)$ auf den anderen Eingang. Am Ausgang des Multiplizierers steht u. a. auch das modulierende Signal zur Verfügung (Umsetzung der Frequenzmodulation in *P*uls*d*auer*m*odulation PDM). Der Tiefpass unterdrückt die auch entstehenden höherfrequenten Anteile. Für diese Demodulationsart, auch mit *Koinzidenzdemodulator* bezeichnet, gibt

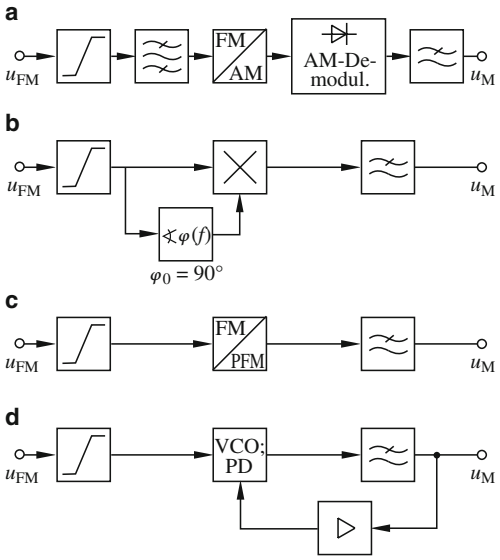


Abb. 96.17 Verschiedene Prinzipien der FM-Demodulation. **a** FM-AM-Umsetzung, **b** FM-PDM-Umsetzung, **c** FM-PFM-Umsetzung, **d** Demodulation mit PLL, Erläuterungen siehe Text

es Integrierte Schaltungen, die die erforderlichen Baugruppen einschließlich Amplitudenbegrenzerstufe enthalten (z. B. TBA 120).

Im dritten Verfahren (c) gibt man sich ein Zeitintervall vor und zählt die darin enthaltenen Nulldurchgänge der frequenzmodulierten Schwingung. Das geschieht auf einfache Weise dadurch, dass man den sinusförmigen Träger in Rechteckimpulse umwandelt und diese dann differenziert. Über eine Kondensator-Umladung mit anschließender Mittelwertbildung lässt sich das modulierende Signal zurückgewinnen (Umsetzung Frequenzmodulation in *P*uls*f*requenz*m*odulation).

Verfahren vier (d) verwendet einen Phasenregelkreis (PLL, *p*hase-*l*ocked *l*oop). Eine sich durch Frequenzmodulation ändernde Trägerfrequenz wird durch den Phasendiskriminator erkannt, der eine entsprechende Spannung zur Nachführung des spannungsgesteuerten Oszillators ausgibt. Diese Spannung ist proportional zur Abweichung der Trägerfrequenz vom Sollwert und damit proportional zur modulierenden Spannung.

Abb. 96.18a, b zeigt zwei Demodulatorschaltungen, die zwar nicht mehr eingesetzt werden, die aber das Prinzip des ersten Verfahrens nach Abb. 96.17a auf einfache Weise darstellen. Für a) muss gelten: $\omega L \ll R$. Daraus folgt für sinusförmiges u_{FM}: $u_L \approx u_{FM} \cdot \omega \cdot L/R$. u_{FM} ändert nicht die Amplitude, aber die Frequenz: $\hat{u}_L \approx \hat{u}_{FM} \cdot (\omega_T + \Delta\omega_T) \cdot L/R$. Die Spannung an L ist proportional der modulierenden Spannung. Der Nachteil dieser Schaltung liegt darin, dass u_L für $\omega L \ll R$ sehr klein und damit schwierig auswertbar ist. Deshalb ist der im Bild eingetragene Verstärker unbedingt erforderlich. In Abb. 96.18b hat der Schwingkreis aus L_1, C_1 und R_1 die dargestellte Durchlasskurve. Die Trägerfrequenz f_T liegt auf einer Flanke dieser Durchlasskurve in der Mitte eines möglichst linearen Bereiches. Anschließend wird mit der Diode D demoduliert. Einzelheiten

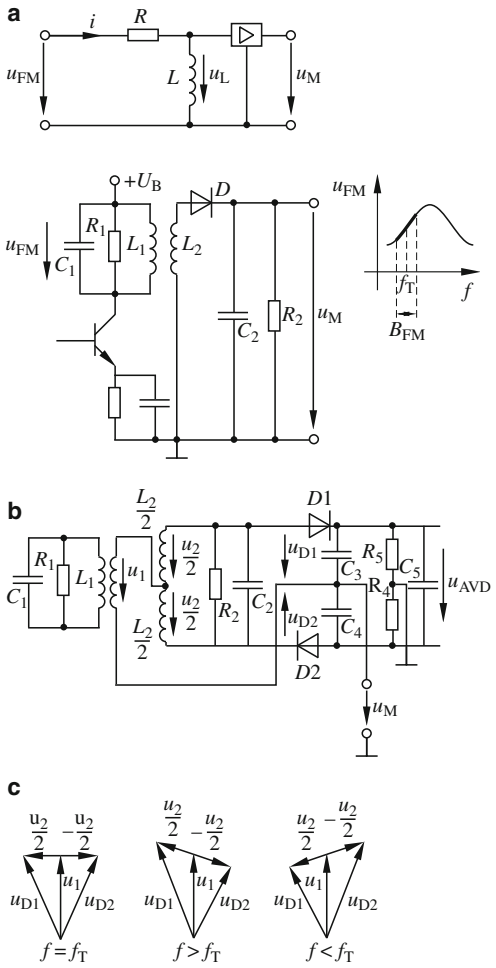


Abb. 96.18 Frequenzdemodulation. **a** mit frequenzabhängigem Blindwiderstand, **b** mit Flankendemodulator, **c** mit Ratiodetektor

hierzu können dem Kapitel über die Demodulation von AM entnommen werden. Steigt die Frequenz des Trägers, steigt auch die Spannung am Schwingkreis und damit die Amplitude von u_M.

Abb. 96.18c stellt die heute in der Rundfunkübertragungstechnik üblicherweise verwendete Schaltung des *Ratiodetektors* (Prinzip nach Abb. 96.17a) dar. Die beiden Schwingkreise aus L_1, C_1 und L_2, C_2 sind auf die Trägerfrequenz f_T abgestimmt. Die Spannungen u_1 und u_2 stehen für $f = f_T$ senkrecht aufeinander. Weicht f von der Resonanzfrequenz ab, ändert sich der Phasenwinkel von u_2 gegenüber u_1. Es gilt:

$$u_{D1} = u_2/2 + u_1; \quad u_{D2} = -u_2/2 + u_1; \quad u_M = u_{D1} - u_{D2}$$

Nach Spitzengleichrichtung steht die modulierende Spannung wieder zur Verfügung. Der Ratiodetektor hat die besondere Eigenschaft, dass er kurzzeitige Amplitudenänderungen von u_1 unterdrückt: Der Kondensator C_5 muss auf die neue Amplitude umgeladen werden, was aber wegen seiner Größe (einige µF) nur langsam (0,1 bis 1 s) erfolgen kann. Damit werden Amplitudenstörungen nahezu unwirksam.

96.3 Sinusträger – mit Digitalsignal moduliert

Die im Abschn. 96.2 behandelten Modulationsverfahren und die dort dargestellten Eigenschaften sind unabhängig vom zeitlichen Verlauf des modulierenden Signals und deshalb auch hier gültig. Dass trotzdem speziell für Digitalsignale eigene Bezeichnungen eingeführt worden sind, liegt an der großen Bedeutung dieser Signalform vor allem für die Datenfernübertragung.

Abb. 96.19 zeigt die jeweilige Trägerschwingung für die drei Modulationsarten:

- *Amplitudenumtastung* (ASK: *a*mplitude *s*hift *k*eying, entspricht der Amplitudenmodulation),
- *Frequenzumtastung* (FSK: *f*requency *s*hift *k*eying, entspricht der Frequenzmodulation),
- *Phasenumtastung* (PSK: *p*hase *s*hift *k*eying, entspricht der Phasenmodulation).

96.3.1 Amplitudenumtastung (ASK)

Nach Abb. 96.19a wird der positiveren Amplitude von u_M eine im Prinzip beliebige Trägeramplitude zugeordnet, der negativeren Amplitude von u_M eine weitere Trägeramplitude. Zur Demodulation kann eine Schaltung nach Abb. 96.6 verwendet werden. Man muss aber sicherstellen, dass für beide Amplitudenwerte von u_M eine ausreichende Mindest-Trägeramplitude vorhanden ist. Sofern möglich, nimmt man die Amplitudenumschaltung in den Nulldurchgängen der Trägerschwingung vor, weil dann das Einschwingen der Spannung am Ausgang des Demodulators auf den neuen Wert optimal geschieht, d. h. in kürzester Zeit und mit minimalem Überschwingen.

Die erforderliche Bandbreite ist theoretisch zwar unendlich groß, wenn die Signalspannung u_M periodisch angenommen und in eine Fourierreihe entwickelt wird, aber in der Praxis kann der rechteckförmige Verlauf durch einen sinusförmigen Verlauf angenähert werden, weil sich daraus die Rechteckform wieder eindeutig zurückgewinnen lässt. Das modulierende Signal ist damit die Grundschwingung der Fourierreihe mit der Frequenz $f = 1/T_S$. Für die erforderliche Bandbreite bei Zweiseitenbandmodulation folgen für den Minimalwert $B_{\text{min-ASK}}$ und den in der Praxis verwendeten Wert B_{ASK}:

$$\begin{aligned} &\textit{Bandbreite} && B_{\text{min-ASK}} = 2 \cdot \frac{1}{T_S}; \\ &\textit{Praxis} && B_{\text{ASK}} \approx 1{,}5 \cdot B_{\text{min-ASK}} = 3 \cdot \frac{1}{T_S} \end{aligned} \qquad (96.28)$$

Als Beispiel für einen ASK-modulierten Sender sei der Normalfrequenz- und Zeitzeichensender DCF 77 der Physikalisch-Technischen Bundesanstalt in Braunschweig angeführt. Seine Trägerfrequenz beträgt 77,5 kHz, und die Informationen über Zeit, Datum, Sommer-/Winterzeit und einige andere Größen werden durch Absenken der Trägeramplitude zu Beginn jeder Sekunde (Ausnahme: 59. Sekunde) auf 25 % des Nenn-

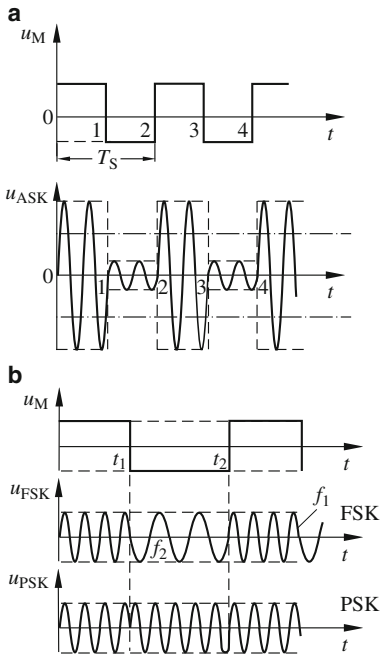


Abb. 96.19 Modulation eines Sinusträgers mit Digitalsignal. **a** Amplitudenumtastung (ASK), **b** Frequenzumtastung (FSK) und Phasenumtastung (PSK)

wertes für 0,1 s (entspricht logisch 0) oder 0,2 s (entspricht logisch 1) übertragen.

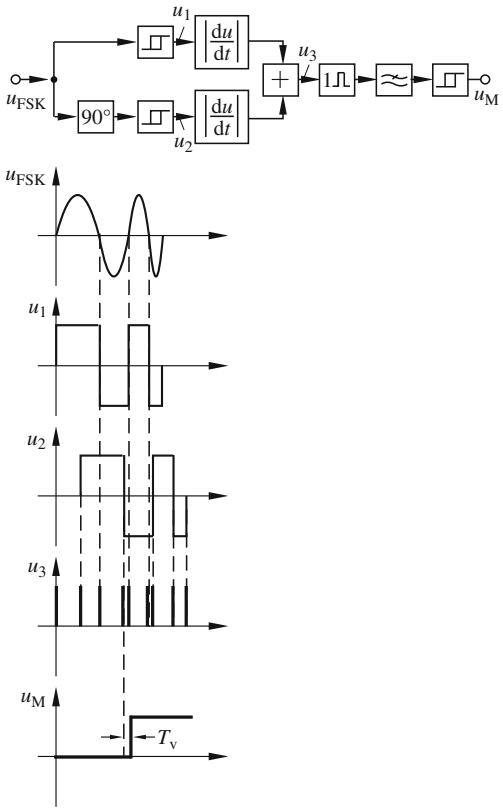


Abb. 96.20 Demodulation bei Frequenzumtastung (FSK)

96.3.2 Frequenzumtastung (FSK)

Nach Abb. 96.19b wird ein Oszillator in Abhängigkeit vom modulierenden Signal u_M zwischen zwei Frequenzen (f_1, f_2) umgeschaltet. Diese Umschaltung sollte ohne Amplitudensprünge geschehen, damit bei der Demodulation ein optimales Einschwingen auf den neuen Amplitudenwert möglich wird. Deshalb ist die Umschaltung zwischen zwei getrennten Oszillatoren problematisch. In der Praxis geschieht sie durch Ändern von L bzw. C des frequenzbestimmenden Oszillatorschwingkreises. Auch hier lassen sich geringe Amplitudensprünge nicht vermeiden, weil es sich bei L und C um Energiespeicher handelt. Quarzoszillatoren mit Kapazitätsvariationsdioden als frequenzänderndem Element ergeben relativ geringe und über die Ansteuerelektronik kalkulierbare Amplitudensprünge.

Für die Abschätzung der erforderlichen Bandbreite nimmt man an, dass die zwei Frequenzen f_1 und f_2 durch Frequenzmodulation einer Mittenfrequenz $f_M = (f_1 + f_2)/2$ entstehen. Dann lassen sich die Formeln des Abschn. 96.2.4 verwenden. Auch hier gilt, dass der rechteckförmige Verlauf von u_M durch einen sinusförmigen Verlauf mit gleicher Periodendauer ersetzt werden kann (siehe Abschn. 96.3.1).

Zur Demodulation können zwei FM-Demodulatoren nach Abschn. 96.2.4.8, abgestimmt auf f_1 und f_2, verwendet werden. Allerdings kann es Probleme mit dem Einschwingverhalten des jeweiligen Demodulators beim Eintreffen des neuen Frequenzwertes geben. Deshalb hat sich u. a. eine Schaltung nach Abb. 96.20 bewährt. Die modulierte sinusförmige Trägerspannung u_{FSK} wird einmal im Original und einmal um 90° verschoben in ein Rechtecksignal umgewandelt (u_1, u_2) und differenziert. Von den entstehenden Impulsen wird der Betrag gebildet, sodass sich u_3 ergibt. Durch die Addition des um 90° verschobenen Signalanteiles erhält man für u_3 die doppelte Anzahl von Impulsen, was zu einem schnelleren Einschwingen des demodulierten Signals auf den neuen Wert führt. Über eine monostabile Kippstufe werden die Nadelimpulse in solche mit konstanter Zeitdauer umgesetzt, sodass sich über einen Tiefpass mit nachfolgendem Schmitt-Trigger die modulierende Spannung u_M zurückgewinnen lässt. Die Verzögerungszeit T_v ist ohne Wirkung auf die enthaltene Information.

96.3.3 Phasenumtastung (PSK)

Die Modulation durch Phasenumtastung der Trägerschwingung ist sehr unempfindlich gegenüber Störungen. Kennzeichen ist die Zahl der unterschiedlichen Phasenzustände, zwischen denen umgeschaltet werden kann. Abb. 96.19b zeigt die Verwendung von zwei Phasenzuständen (2-PSK). Daneben werden häufig vier Phasenzustände eingesetzt (4-PSK, QPSK: *q*uadrature *p*hase *s*hift *k*eying). Als Quadraturmodulation, allerdings für ein analoges modulierendes Signal, findet dieses Verfahren in der Farbfernsehtechnik Anwendung (Abschn. 96.2.2.3).

Abhängig von der Art der Informationskodierung gibt es zur Demodulation von PSK zwei Verfahren: 1. Jedem der übertragenen Zustände entspricht eine festgelegte Phasenlage. Dann muss entweder der Träger mit der definierten Nullphase mit übertragen werden, oder aber es müssen periodisch wiederkehrende Muster zur Synchronisation des im Empfänger erzeugten Trägers gesendet werden (siehe Abschn. 96.2.2.3). 2. Die Information steckt in der *Differenz* zweier aufeinanderfolgender Phasenlagen. Hier bietet sich zur Demodulation an, den Träger aus dem übertragenen Signal zurückzugewinnen. Auch wenn dies stets nur fehlerhaft möglich ist, spielt das wegen der Phasen*differenzbildung* keine Rolle (*Phasendifferenzkodierung*, s. u.).

96.3.3.1 Zweiphasenumtastung (2-PSK)

Abb. 96.21a stellt die modulierende Spannung u_M und die daraus gebildete Trägerspannung u_T in der Zeitebene dar.

- *Modulation, Einfluss von Störsignalen*: Die Modulation des Trägers geschieht in einem Ringmodulator nach Abb. 96.21b bzw. nach Abb. 96.7, an dessen Ausgang nur die zwei Seitenbänder vorhanden sind. Die Zuordnung der zwei Phasenzustände ist so gewählt, dass sie sich maximal voneinander unterscheiden, also zu 0° und 180°. Das hat den Vorteil, dass bei einer auftretenden Störspannung u_N nach Abb. 96.21c die „1“ ($\hat{=}\,0°$) theoretisch solange sicher erkannt wird, wie der Summenzeiger u_r, gebildet aus u_T und u_N, rechts von der Ordinate liegt. In der Praxis darf eine Minimalamplitude $u_{T\,\min}$ nicht unterschritten werden, sodass u_r rechts von der im Bild dargestellten gestrichelten Senkrechten liegen muss. Das Gleiche gilt für die kodierte „0“, sodass jede der beiden Codierungen am unempfindlichsten gegenüber Störungen ist. Die Amplitude von u_N kann theoretisch zwar sehr groß werden, wenn sie in einem Winkelbereich $-90° \ldots +90°$ zu u_T liegt, praktisch aber sollte sie den Wert $u_T - u_{T\,\min}$ in keiner der möglichen Phasenlagen überschreiten. In Abb. 96.21a findet der Phasenwechsel in den Nulldurchgängen der Trägerschwingung statt. Das ist vorteilhaft, weil es keine Amplitudensprünge gibt und sich die Zeit für die zur Dekodierung notwendige Trägersynchronisierung bzw. Trägerrückgewinnung verringert. Damit können dann u. U. höhere Datenraten übertragen werden.
  Zur Abschätzung der erforderlichen Bandbreite wird wiederum ein periodisches Rechtecksignal für u_M angenommen, das durch die Grundschwingung des nach Fourier dargestellten Signals angenähert wird. Damit kann das Ergebnis von (96.28) übernommen werden.
  minimal erforderliche Bandbreite:

$$B_{\text{min-2PSK}} = 2 \cdot \frac{1}{T_S};$$

$$\textit{Praxis} \quad B_{\text{2PSK}} \approx 1{,}5 \cdot B_{\text{min-2PSK}} = 3 \cdot \frac{1}{T_S} \qquad (96.29)$$

Für den Rauschabstand folgt mit (92.26):

$$\textit{Rauschabstand} \quad a_r = 10 \cdot \lg \frac{P_S}{P_N} \quad \text{in dB} \qquad (96.30)$$

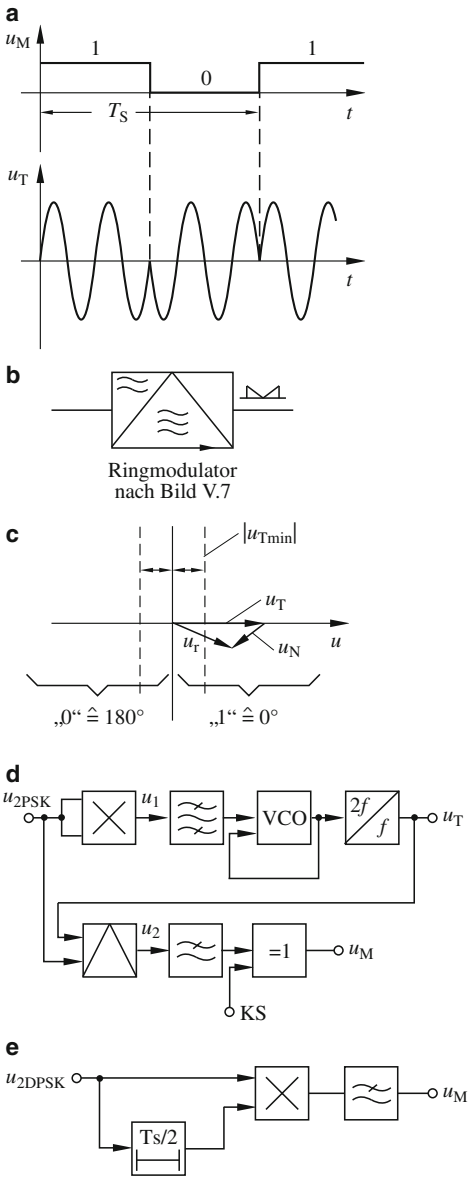


Abb. 96.21 Zweiphasenumtastung (2-PSK), Abkürzungen siehe Text. **a** Signalverläufe, **b** Erzeugung, **c** Einfluss von Störsignalen, **d** Demodulation, **e** Demodulation von Phasendifferenzcodierung (2-DPSK)

- *Demodulation, Trägerrückgewinnung*: Die Demodulation eines 2-PSK-Signals geschieht z. B. mit einer Schaltung nach Abb. 96.21d. Die Spannung $u_{\text{2PSK}} = \pm\hat{u}_T \cdot \cos(\omega_T \cdot t + k \cdot \pi)$, $k = 0$ oder 1, wird im Multiplizierer quadriert und ergibt

$$u_1 = \frac{\hat{u}_T^2}{2} \cdot (1 + \cos(2 \cdot \omega_T \cdot t + 2 \cdot k \cdot \pi)) \qquad (96.31)$$

Nach einem Bandfilter folgt ein spannungsgesteuerter Oszillator (VCO) mit eingebautem Phasenvergleicher und Tiefpassfilter. Die nach Gleichung (96.31) vorhandene doppelte Trägerfrequenz wird in einem Frequenzteiler auf die Trägerfrequenz f_T heruntergeteilt (u_T). Durch diese Teilung ergibt

Tab. 96.1 Zu Beispiel 3

Zu übertragende Bitfolge	(1) Startbit	1	1	0	0	1	0	1	1	0
gewählter Phasenwinkel senderseitig	0°	0°	0°	180°	0°	0°	180°	180°	180°	0°
Phasenwinkel am Empfänger, 1. Möglichkeit	0°	0°	0°	180°	0°	0°	180°	180°	180°	0°
Phasenwinkel am Empfänger, 2. Möglichkeit	180°	180°	180°	0°	180°	180°	0°	0°	0°	180°
Decodierung am Empfänger		1	1	0	0	1	0	1	1	0

sich eine Phasenunsicherheit von $k \cdot (2 \cdot \pi)/2 = k \cdot \pi$. In einem Demodulator wird aus den Seitenbändern (u_{2PSK}) und dem rückgewonnenen Träger (u_T) das modulierende Signal u_M erzeugt. Da die absolute Phasenlage des Trägers nicht mit übertragen wird, erkennt man nicht, wie den zwei empfangenen Werten senderseitig die „1“ bzw. „0“ zugeordnet wurde (Phasenunsicherheit $k \cdot \pi$). Man erhält für u_M also entweder die richtige oder die invertierte Bitfolge. Durch die Übertragung geeigneter Codewörter lässt sich die korrekte Bitfolge durch Anlegen eines Korrektursignals KS an ein Exklusiv-Oder-Glied einstellen.

- *Phasendifferenzcodierung (2-DPSK)*: Man kann auf die Übertragung des Absolutwertes der Trägerphase verzichten, wenn man die Phasendifferenzcodierung anwendet: Für $u_M =$ „1“ bleibt die momentan vorhandene Phase bestehen, für $u_M =$ „0“ ändert sich die Phase. Daraus kann u_M eindeutig gewonnen werden, siehe Beispiel 3. Die Decodierung geschieht nach Abb. 96.21e: In einer Verzögerungsschaltung wird das Signal u_{2DPSK} um $T_S/2$ verzögert und mit dem unverzögerten Signal multipliziert. Bei gleicher Phase ist das Produkt positiv und wird der „1“ zugeordnet; bei unterschiedlicher Phase ist es negativ und bedeutet „0“.

Beispiel 3

Übertragen werden soll die Bitfolge

(1) 1 1 0 0 1 0 1 1 0

in Phasendifferenzcodierung (Tab. 96.1). Der erste Wert kann nicht wiederhergestellt werden, weil er keinen Vorgänger zur Differenzbildung hat (Startbit). ◄

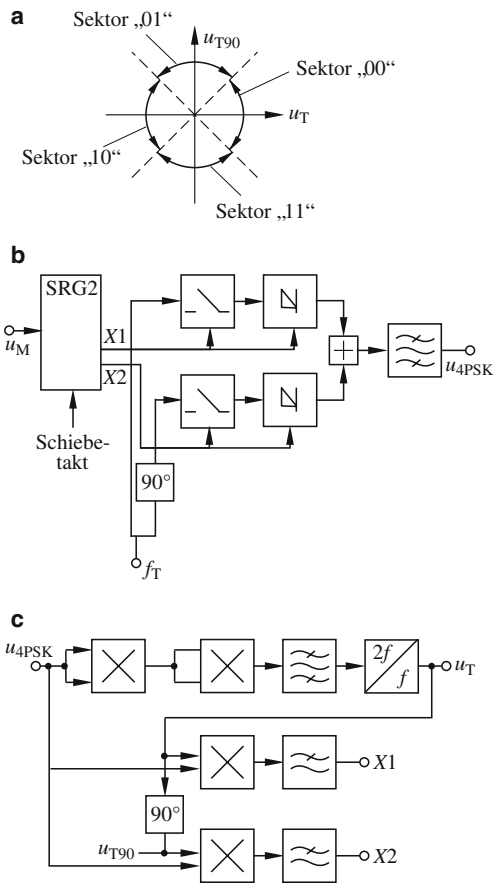


Abb. 96.22 Vierphasenumtastung (4-PSK). **a** Prinzip, **b** Erzeugung, **c** Trägerrückgewinnung und Demodulation

96.3.3.2 Vierphasenumtastung (4-PSK)

- *Modulation*: Man setzt zwei Träger gleicher Frequenz, aber mit 90° Phasenverschiebung ein, siehe u_T, u_{T90} in Abb. 96.22a. Außerdem kann die Phase beider Träger zwischen 0° und 180° umgeschaltet werden. Bei der Modulation wird der Träger unterdrückt. Der Winkelbereich −45° bis +45°, bezogen auf u_T, entspricht digital „00“ und wird dadurch erreicht, dass die Amplitude von u_T maximal, die von u_{T90} null ist. Der Bereich +45° … +135° entspricht digital „01“ und erfordert eine maximale Amplitude von u_{T90} und die Amplitude null für u_T. Für die restlichen zwei Kodierungen muss der mit maximaler Amplitude gesendete Träger „negativ“ sein, d. h. eine Phasenverschiebung von 180° aufweisen. Für die zu übertragende Bitfolge werden jeweils zwei aufeinanderfolgende Bits zusammengefasst, und daraus wird nach Abb. 96.22a die Phasenlage zu u_T bestimmt (Tab. 96.2).

Die Erzeugung von 4-PSK geschieht nach Abb. 96.22b. Nachdem je 2 seriell eintreffende Bits in einem 2-Bit-Seriell-Parallel-Umsetzer (Schieberegister SRG2) parallel als X1 und X2 anstehen, werden die Trägeramplituden entsprechend Tab. 96.2 eingestellt.

- *Demodulation, Trägerrückgewinnung*: Die Trägerrückgewinnung geschieht nach Abb. 96.22c durch zweimalige Multiplikation der Spannung u_{4PSK} mit sich selbst:

$$u_1 = \left[\left(\hat{u}_T \cdot \cos\left(\omega \cdot t + k \cdot \frac{\pi}{2}\right)\right)^2\right]^2$$

Tab. 96.2 Beispiel zur Ansteuerung bei Vierphasenumtastung (4-PSK)

Zu übertragende Bitfolge	0	0	1	1	0	1	1	1	1	0
Phase gegenüber u_T	0°		270°		90°		270°		180°	
X1 (siehe Abb. 96.22b), steuert u_T:		0		1		0		1		1
a) Amplitude von u_T	Maximal		Null		Null		Null		Maximal	
b) Phase von u_T	0°								180°	
X2 (siehe Abb. 96.22b), steuert u_{T90}:		0		1		1		1		0
a) Amplitude von u_{T90}	Null		Maximal		Maximal		Maximal		Null	
b) Phase von u_{T90}			180°		0°		180°			

mit $k = 0, 1, 2$ oder 3

$$u_1 = \frac{\hat{u}_T^4}{8}[3 + 4 \cdot \cos(2 \cdot \omega_T \cdot t + k \cdot \pi) + \cos(4 \cdot \omega_T \cdot t + 2 \cdot k \cdot \pi)] \qquad (96.32)$$

Die Komponente mit der doppelten Trägerfrequenz wird über einen Bandpass ausgefiltert, die Komponente mit der vierfachen Trägerfrequenz durch einen Frequenzteiler 4 : 1 auf die Trägerfrequenz geteilt (u_T). Aus dem rückgewonnenen Träger (u_T) wird ein zweiter mit 90° Phasenverschiebung erzeugt (u_{T90}). In zwei Multiplizierschaltungen wird das übertragene Signal u_{4PSK} mit u_T und u_{T90} multipliziert. Es ergeben sich insgesamt 4 Möglichkeiten für die Vorzeichenkombinationen, sodass das modulierende Signal zurückgewonnen werden kann. Wegen der Frequenzteilung bleibt eine Phasenunsicherheit von $(2 \cdot k \cdot \pi)/4 = k \cdot (\pi/2)$.

Da zwei Bits mit nur einem Phasenwinkel codiert werden, ist die erforderliche Bandbreite nur noch halb so groß wie bei 2-PSK:

minimal erforderliche Bandbreite

$$B_{\text{min-4PSK}} = 2 \cdot \frac{1}{2 \cdot T_S} = \frac{1}{T_S};$$

$$\textit{Praxis} \quad B_{\text{PSK}} \approx 1{,}5 \cdot B_{\text{min-PSK}} = 1{,}5 \cdot \frac{1}{T_S} \qquad (96.33)$$

Für den Rauschabstand lässt sich allgemein sagen, dass er in der Praxis bei 4-PSK gegenüber 2-PSK um etwa 3 dB abnimmt, weil bei 4-PSK pro Amplitudenwert zwei Bit übertragen werden gegenüber einem Bit bei 2-PSK.

Auch hier kann die Phasenunsicherheit durch Differenzphasenmodulation beseitigt werden.

96.3.3.3 *n*-Phasen-Umtastung

Das Prinzip der Phasenumtastung lässt sich auf n Phasen erweitern, wobei n in der Praxis Werte von 8, 16, 32 oder 64 annimmt. Bei der Auswahl eines geeigneten Verfahrens sind die zwei folgenden Ergebnisse von Bedeutung:

1. Die erforderliche Bandbreite nimmt mit steigendem n gemäß $1/n$ ab.
2. Der Rauschabstand verringert sich bei gleicher Sender- und Rauschleistung mit zunehmendem n um etwa $n \cdot (3 \ldots 5)$ dB.
3. Der Einfluss von Störspannungen nimmt mit wachsendem n zu, weil immer kleinere Segmente pro Codierung zur Verfügung stehen. In Abb. 96.21c würde die Spitze des Zeigers u_r bei einer 16-Phasen-Umtastung bereits in das benachbarte Segment hineinragen und damit einen Fehler verursachen.

96.3.3.4 QAPSK kombinierte Amplituden-Phasenumtastung

Diese Modulationsart wird in der Ausführung 16-QAPSK beim terrestrischen Digitalfernsehen eingesetzt. Siehe dazu Kap. 99 „Ton- und Bildübertragung“, Abschn. 99.2 „Digitale Farbfernsehtechnik“, Trägermodulation: 16-QAPSK.

96.4 Pulsträger uncodiert

In Analogie zum Sinusträger gibt es hier die in Abb. 96.23 dargestellten Verfahren:

- *Pulsamplitudenmodulation* (PAM: *p*ulse *a*mplitude *m*odulation),
- *Pulsfrequenzmodulation* (PFM: *p*ulse *f*requency *m*odulation),
- *Pulsphasenmodulation* (PPM: *p*ulse *p*hase *m*odulation),
- *Pulsdauermodulation* (PDM: *p*ulse *d*uration *m*odulation).

Lediglich das zuletzt genannte Verfahren hat kein direktes Analogon zur Sinusträgermodulation.

96.4.1 Pulsamplitudenmodulation (PAM)

Sie wird relativ selten eingesetzt, hat aber eine gewisse Bedeutung beim *Zeitmultiplexverfahren* (siehe Abschn. 100.1). Die Zeit T_a in Abb. 96.23 ist durch das Abtasttheorem von Shannon vorgegeben, die Zeit ΔT kann so klein wie möglich gemacht werden. Dann kann man in den Zeitraum $T_a - \Delta T$ weitere Impulse von anderen Signalen einfügen. Abb. 96.24 zeigt eine von mehreren Möglichkeiten zur Modulation und Demodulation bei PAM. Die Spannung U_0 sorgt dafür, dass auch bei der größten negativen Amplitude des modulierenden Signals die Diode noch in Durchlassrichtung betrieben wird. Der Schalter S1 schließt für die Zeitdauer ΔT. Der Demodulator besteht aus einer Abtast-Halte-Schaltung, deren Schalter S2 genau dann schließt, wenn das vom geschlossenen Schalter S1 ausgesandte Signal eintrifft. Deshalb ist eine geeignete Synchronisation zwischen beiden Schaltern erforderlich. Diese Art der Modulation heißt unipolar, weil die Amplituden neben dem Wert Null nur eine Polarität, hier die positive, aufweisen.

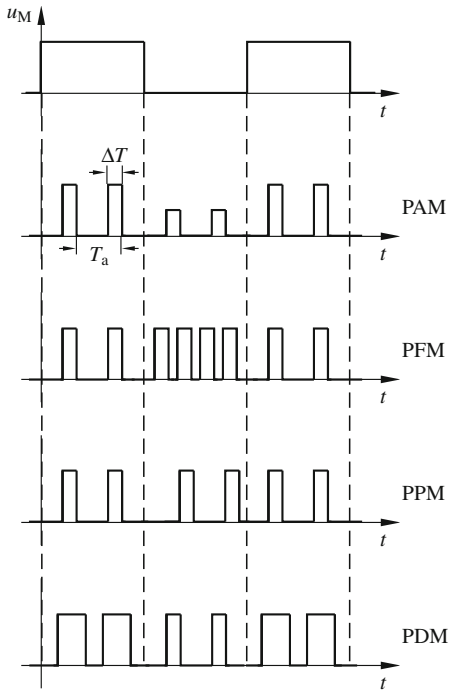


Abb. 96.23 Verfahren zur Pulsträgermodulation (Pulse unipolar)

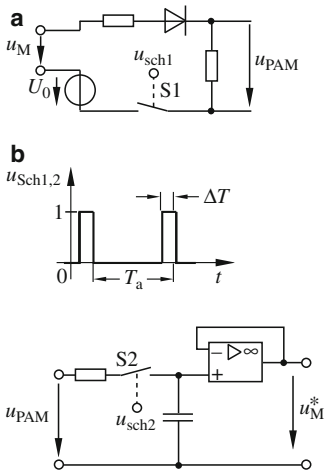


Abb. 96.24 Unipolare Pulsamplitudenmodulation. **a** Modulation, **b** Demodulation

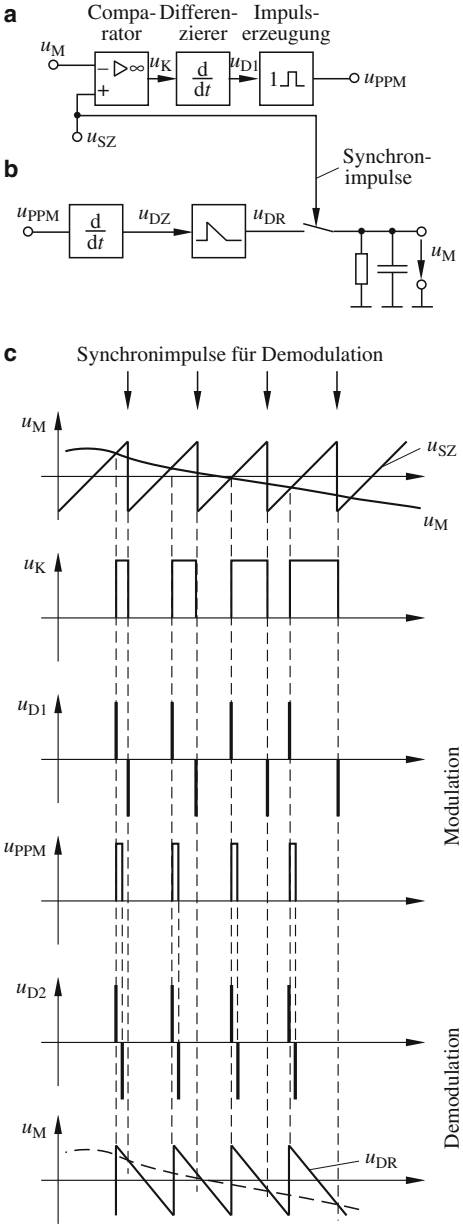


Abb. 96.25 Pulsphasenmodulation (PPM). **a** Modulator, **b** Demodulator, **c** Wirkungsweise

96.4.2 Pulsfrequenz- und Pulsphasenmodulation (PFM, PPM)

Wie beim Sinusträger stehen auch hier Pulsfrequenz- und Pulsphasenmodulation in einem engen Zusammenhang. An Abb. 96.25 kann man erkennen, dass eine Phasenmodulation (Verschiebung der Impulse) zunächst wie eine Frequenzmodulation (wiederum Verschiebung der Impulse) erscheint. Erst verschieben sich die Impulse, verglichen mit der regelmäßigen Aufeinanderfolge vorher, dann resultiert daraus eine Frequenzänderung. Die größere Bedeutung für die Praxis hat die Pulsphasenmodulation.

Abb. 96.25 stellt eine von mehreren Möglichkeiten zur Erzeugung von Pulsphasenmodulation sowie eine Schaltung zur Demodulation dar. Zur Demodulation müssen Synchronimpulse entweder getrennt übertragen oder aus dem PPM-Signal z. B. durch PLL-Schaltungen (PLL: phase locked loop) empfängerseitig erzeugt werden. Die Spannung u_{DR} hat stets den gleichen zeitlichen Verlauf. Die Information liegt in der zeitlichen Verschiebung der Impulse gegenüber ihrem Auftreten

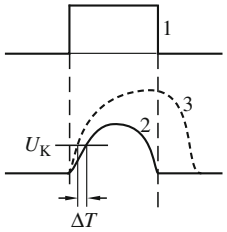


Abb. 96.26 Überlagerte Störspannung bei PPM. *1* idealer Rechteckimpuls, *2* verformter Rechteckimpuls bei begrenzter Bandbreite, *3* wie *2*, mit überlagerter Störspannung

in gleichen Zeitabständen. Demnach ist das Verfahren gegenüber Störungen, die als Amplitudenänderungen auftreten, unempfindlich. Da allerdings nur eine begrenzte Bandbreite zur Signalübertragung zur Verfügung steht, ergibt sich ein verformter Rechteckimpuls nach Abb. 96.26. Überlagert sich diesem verformten Impuls eine Störspannung, ergeben sich für den Komparator mit der Vergleichsspannung U_K unterschiedliche Schaltzeitpunkte mit der Zeitdifferenz ΔT, die als zusätzliche Phasenverschiebung ein fehlerhaftes Signal ergeben. Pulsphasenmodulation wird dann vorteilhaft eingesetzt, wenn breitbandige Übertragungskanäle zur Verfügung stehen. Der Gewinn an Signal-Rauschabstand beträgt nach [22]:

$$a_{SN} = 10 \cdot \lg \left[\frac{1}{8} \cdot \left(\frac{B_N}{B_S} \right)^2 \right] \text{ dB} \qquad (96.34)$$

B_N Bandbreite des durch Rauschen gestörten Übertragungskanals in Hz, B_S Signalbandbreite in Hz.

Bereits für $B_N > 2{,}83 \cdot B_S$ ergibt sich ein Gewinn an Signal-Rauschabstand gegenüber der Pulsamplitudenmodulation, was auch anschaulich nach Abb. 96.26 zu erwarten ist: Der übertragene Impuls nähert sich der Rechteckform, sodass ΔT kaum noch wirksam werden kann.

96.4.3 Pulsdauermodulation (PDM)

Ein Verfahren zur Pulsdauermodulation ist in Abb. 96.27 dargestellt. Die Demodulation geschieht mit einem Tiefpass, dessen Grenzfrequenz geringfügig über der höchsten modulierenden Frequenz liegt. Für die Codierung von Signalen in der Nachrichtentechnik hat diese Modulationsart geringe Bedeutung.

Allerdings wird die Pulsdauermodulation auch außerhalb der Signalkodierung angewendet, wie die zwei folgenden Beispiele zeigen:

a) Das An- und Abschalten von Gleichspannungsquellen an einen Verbraucher mit elektronischen Schaltern (Transistoren) kann sehr verlustleistungsarm geschehen. Soll ein Verbraucher mit einer wenigstens näherungsweise sinusförmigen Spannung aus einer Gleichspannungsquelle versorgt

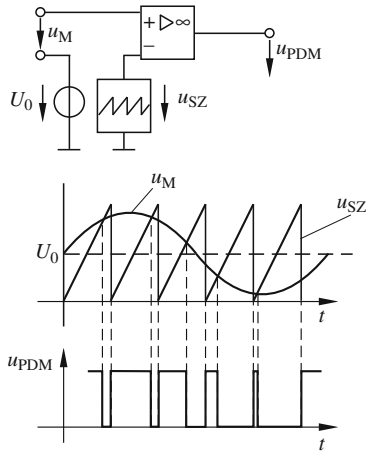


Abb. 96.27 Erzeugung von Pulsdauermodulation

werden, könnte man in Reihe zum Verbraucher einen Transistor schalten, der so angesteuert wird, dass die Spannung am Verbraucher sinusförmigen Verlauf hat. Der Nachteil dieses Konzeptes liegt darin, dass der Transistor mit einer Verlustleistung in der Größenordnung der Verbraucherleistung beaufschlagt wird.
Deshalb schaltet man nach Abb. 96.28 die Gleichspannungsquelle mit der Spannung U_0 für unterschiedliche Zeitdauern auf den Verbraucher. Die Zeit ΔT ist viel kleiner als die Periodendauer des anzunähernden sinusförmigen Spannungsverlaufes, sodass der Verbraucher in der Regel wegen seines Tiefpassverhaltens (z. B. Motor mit Induktivität) den im Bild eingezeichneten realen Mittelwert $\bar{u}(t)$ bildet, der den idealen Verlauf oft ausreichend genau annähert. Angewendet wird dieses Verfahren beispielsweise bei unterbrechungsfreien Stromversorgungen für Rechner, die das Versorgungsnetz mit sinusförmigem Verlauf und der Spannung 230 V für die Dauer der Unterbrechung mit einem eingebauten Gleichspannungsakkumulator nachbilden. Hier kommt es auf möglichst geringe Verlustleistung an, um mit kleinen handlichen Geräten einen möglichst lang andauernden Notstrombetrieb zu erreichen und um die Probleme mit der Abfuhr der Wärme, verursacht durch die Verlustleistung, gering zu halten.

b) Dieses Verfahren wird auch angewendet bei der Amplitudenmodulation im Anodenkreis von Senderöhren (multiplikatives Verfahren mit niedrigem Klirrfaktor nach Abb. 96.10), weil die Leistung des modulierenden Signals in der Größenordnung der Trägerleistung liegen muss und damit die gleichen Probleme mit der Verlustleistung auftreten wie unter a) beschrieben.

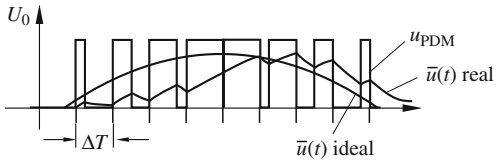


Abb. 96.28 Pulsdauermodulation in der Energietechnik

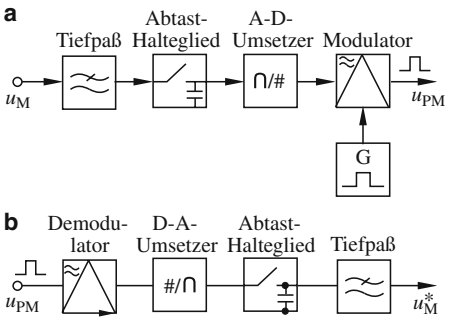


Abb. 96.29 Prinzip der Signalübertragung mit Pulscodemodulation. **a** Sender, **b** Empfänger

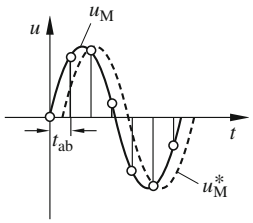


Abb. 96.30 Pulscodemodulation: ausgesendetes und empfangenes Signal

96.4.4 Pulscodemodulation (PCM)

Dies ist die heute überwiegend eingesetzte Art der Pulsmodulation, sodass prinzipiell Pulsamplituden-, Pulsfrequenz-, Pulsphasen- und Pulsdauermodulation möglich sind. Das zu übertragende Signal wird geeignet codiert.

96.4.4.1 Prinzip

Das Prinzip zeigt Abb. 96.29.

Das zu übertragende Signal mit der Spannung u_M wird zunächst mit einem Tiefpass auf die maximale Frequenz $f_{M\,max}$ begrenzt und dann mit schmalen Impulsen der Dauer t_s im Abstand von t_{ab} ($1/t_{ab} > 2f_{M\,max}$, Abtasttheorem von Shannon) abgetastet. Ein Halteglied „friert“ den Abtastwert bis zur Aktualisierung durch den neuen Wert ein, damit der Analog-Digital-Umsetzer einen konstanten und damit eindeutigen Wert zur Umsetzung zur Verfügung hat. An seinem Ausgang steht ein treppenförmiges wertkontinuierliches Signal, mit dem ein Pulsträger moduliert wird. Zur Rückgewinnung des Signals im Empfänger ist der Bittakt mit der Frequenz $f_{ab} = 1/t_{ab}$ notwendig, der aber nicht getrennt mit übertragen wird, sondern durch entsprechende senderseitige Codierung regeneriert werden kann (siehe Abschn. 96.4.4.7). Der Ausgangswert des Digital-Analog-Umsetzers wird in einem Abtast-Halte-Glied solange gespeichert, bis er durch einen neuen Wert ersetzt wird. Dieses zeit- und wertdiskrete Signal ist nach dem Tiefpass in ein zeit- und amplitudenkontinuierliches Signal u_M^* umgewandelt worden. Abb. 96.30 stellt das ursprüngliche Signal u_M und das empfangene Signal u_M^* einander gegenüber.

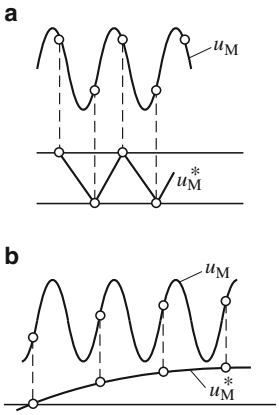


Abb. 96.31 Aliasing-Effekt: u_M abgetastetes Basisbandsignal, u_M^* demoduliertes Signal. **a** $f_{ab} = 2 \cdot f_M$, **b** $f_{ab} \approx 1{,}1 \cdot f_M$

Die Kurvenform ist nahezu unverändert, die Verzögerungszeit wird wesentlich durch die Abtast-Halte-Glieder, die Umsetzzeiten der Umsetzer und die Laufzeit im Kanal bestimmt. Da dieses Verfahren der Codierung auch im Prinzip bei der Speicherung von Musik auf Compact Disc (CD) verwendet wird, soll dieses Verfahren etwas näher betrachtet werden.

96.4.4.2 Aliasing-Effekt

Zunächst stellt man sich das abzutastende Signal so vor, dass es gemäß Fourieranalyse aus Grund- und Oberschwingungen besteht, die alle sinusförmigen Verlauf haben. Obwohl die Signale der Nachrichtentechnik grundsätzlich nicht periodisch sind, kann man auch für sie die Fourierzerlegung durchführen (Abschn. 92.4.9.3). Die höchste noch berücksichtigte sinusförmige Oberschwingung hat im Folgenden die Bezeichnung $f_{M\,max}$.

Wird das abzutastende Signal nicht so frequenzbandbegrenzt, dass die Abtastfrequenz f_{ab} mindestens doppelt so groß ist wie die höchste Signalfrequenz $f_{M\,max}$, entsteht bei der Demodulation eine Signalverfälschung, wie in Abb. 96.31b dargestellt. Das nach der Abtastung wiederhergestellte Signal mit der Spannung u_M^* ist für den Fall eingezeichnet, dass die Abtastfrequenz f_{ab} nur geringfügig über der höchsten Signalfrequenz $f_{M\,max}$ liegt: $f_{ab} \approx 1{,}1 \cdot f_{M\,max}$. Abb. 96.31a zeigt das demodulierte Signal u_M^* für den Grenzfall der doppelten Abtastfrequenz gegenüber der höchsten Signalfrequenz: $f_{ab} = 2 \cdot f_{M\,max}$. Auch hier kann es zu Signalverfälschungen kommen, wenn nämlich die Abtastzeitpunkte genau im Nulldurchgang des Signals liegen oder wenn, wie eingezeichnet, das wiederhergestellte Signal nicht ebenfalls auf $f_{M\,max}$ in seiner Frequenz begrenzt wird. Nur dann bleibt von dem eingezeichneten dreieckförmigen Verlauf die Grundschwingung mit sinusförmigem Verlauf übrig, die mit dem abgetasteten Signal identisch ist. Wichtig ist, dass der theoretische Wert nach Shannon, $f_{ab} = 2 \cdot f_{M\,max}$, nicht ausreicht, um die höchste Signalfrequenz (stets sinusförmig wegen der Bandbegrenzung) fehlerfrei wiederherzustellen. Man wählt deshalb in der Praxis:

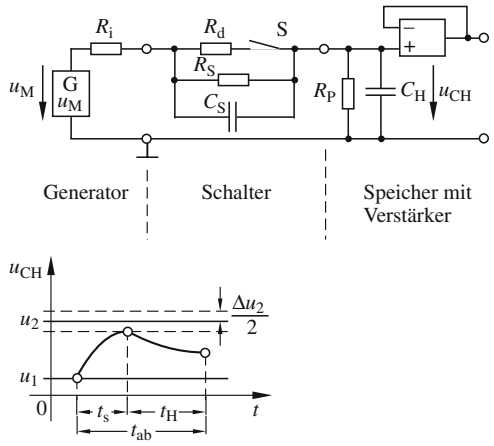


Abb. 96.32 Abtast-Halte-Schaltung. **a** Ersatzschaltung, **b** Spannungsverlauf am Haltekondensator C_H

Abtastfrequenz in der Praxis:

$$f_{ab} = (2{,}2 \ldots 4) \cdot f_{M\,max} \tag{96.35}$$

Anti-Aliasing-Filter: Um die genannten Signalverfälschungen zu vermeiden, wird vor der Abtastung das Basisbandsignal mit einem Tiefpassfilter frequenzbandbegrenzt, um die Bedingung nach (96.35) sicher zu erfüllen.

96.4.4.3 Abtast-Halte-Glied

Abb. 96.32a stellt das Ersatzschaltbild eines Abtast-Halte-Gliedes dar. Es bedeuten:

R_i	Innenwiderstand der Signalquelle mit der modulierenden Spannung u_M,
R_d	Widerstand des elektronischen Schalters S im durchgeschalteten Zustand,
R_s	Widerstand des elektronischen Schalters S im Sperrzustand,
C_s	Kapazität des elektronischen Schalters S im Sperrzustand,
R_p	erfasst den Verlustwiderstand des Kondensators C_H und den Eingangswiderstand des Operationsverstärkers,
C_H	Halte-(Speicher-)Kondensator mit der Spannung u_{CH},
u_1	Spannungswert von u_M zu einem angenommenen Zeitpunkt, auf den sich der Kondensator C_H aufgeladen hat,
u_2	neuer Spannungswert von u_M, auf den sich C_H durch Schließen von S aufladen soll,
$\Delta u_2/2 = \lvert\pm\Delta u_2/2\rvert$	Betrag der maximal möglichen Abweichung zwischen Soll- und Istwert bei der Analog-Digital-Umsetzung, ergibt sich bei einem N-Bit-Umsetzer zu

$$(u_{2\,max} - u_{2\,min})\frac{1}{2} \cdot \frac{1}{2^N - 1} \tag{96.36}$$

t_s	Zeit, während der der elektronische Schalter S geschlossen ist,
$t_{ab} = 1/f_{ab}$	Abstand der Abtastimpulse.

1. Abweichung am Ende der Abtastphase

Es wird der ungünstigste Fall angenommen: $u_1 = u_{M\,min}$ und $u_2 = u_{M\,max}$. Dann lädt sich C_H gemäß

$$(u_{M\,max} - u_{M\,min}) \cdot \left(1 - \exp\left(\frac{-t_s}{\tau_{auf}}\right)\right) \quad \tau_{auf} = C_H(R_i + R_d)$$

auf. Soll am Ende der Abtastung die Kondensatorspannung höchstens um $\Delta u_2/2$ vom Sollwert abweichen (sinnvolle Annahme), muss gelten:

$$\exp\left(-\frac{t_s}{\tau_{auf}}\right) = \frac{1}{2} \cdot \frac{1}{2^N - 1} \approx \frac{1}{2^{N+1}}, \quad \text{mit } N \gg 1$$

Nach einigen Zwischenrechnungen folgt daraus:

$$(R_i + R_d)C_H \approx \frac{t_s}{\ln 2^{N+1}} \tag{96.37}$$

2. Abweichung am Ende der Haltephase

Bei geöffnetem Schalter S entlädt sich C_H über R_p und R_s. Der ungünstigste Fall tritt auf, wenn u_{CH} maximal ist und u_M minimal. Dann kann man für den Entladestrom die zwei Widerstände näherungsweise als parallel geschaltet ansehen, sodass sich die Kondensatorspannung ergibt:

$$u_{CH}(t) \approx u_{CH}(t_0) \cdot \exp\left(-\frac{t_H}{t_{ent}}\right)$$
$$\approx u_{CH}(t_0) \cdot \exp\left(-\frac{t_{ab}}{t_{ent}}\right)$$

mit $t_H \approx t_{ab}$ (t_0 Beginn des Haltezyklus). Die hierdurch verursachte Spannungsänderung liegt wieder in den oben genannten Grenzen:

$$\exp\left(-\frac{t_H}{\tau_{ent}}\right) \approx 1 - \frac{1}{2} \cdot \frac{1}{2^N - 1} \approx 1 - \frac{1}{2^{N+1}} \quad \text{mit } N \gg 1,$$
$$\tau_{ent} \approx C_H(R_p \parallel R_d)$$

Daraus folgt:

$$C_H(R_p \parallel R_d) \approx t_{ab}\frac{1}{\lvert\ln(1 - \frac{1}{2^{N+1}})\rvert} \quad \text{mit } t_H \approx t_{ab} \tag{96.38}$$

3. Einfluss des Kondensators C_s

Bei geöffnetem Schalter bildet C_s mit C_H einen Spannungsteiler, wenn sich die Spannung u_M ändert. Der Einfluss von C_s ist um so geringer, je kleiner sein Wert im Vergleich zu C_H ist. In Analogie zu den Abschätzungen oben kann man ansetzen:

$$C_s < \frac{1}{2^{N+1}} \cdot C_H \tag{96.39}$$

4. Weitere Einflussgrößen

Die Zeit t_s muss so klein sein, dass sich u_M während dieser Zeit nicht wesentlich ändert. Auch hier kann ein Anhaltswert analog zu den oben genannten Größen ohne Ableitung angeführt werden:

$$\Delta u_M(t_s) < (u_{M\max} - u_{M\min}) \cdot \frac{1}{2} \cdot \frac{1}{2^N - 1} \tag{96.40}$$

Dabei wurde (96.36) verwendet.

Beispiel 4

Die Spannung u_M hat die maximale Frequenz $f_{M\max} = 15\,\text{kHz}$ und wird mit einer Frequenz $f_{ab} = 1/t_{ab} = 45\,\text{kHz}$ ($t_{ab} = 22{,}2\,\mu\text{s}$) abgetastet. Der nachgeschaltete Umsetzer hat 12 Bit. C_s liegt bei Feldeffekttransistoren in der Größenordnung 10 pF. Dann muss C_H nach (96.39) zu mindestens $(10\,\text{pF}) \cdot 2^{12+1} \approx 8200\,\text{pF}$ gewählt werden. Die Zeit t_s ist mit (96.40) zu maximal $(22{,}2\,\mu\text{s}) \cdot 1/2^{13} \approx 2{,}7\,\text{ns}$ festgelegt (Sinusschwingung mit der höchsten Frequenz 15 kHz hat im Nulldurchgang die maximale Steigung mit dem Zahlenwert $+1$ bzw. -1). Mit $C_H = 10\,\text{nF}$ folgt mit (96.37):

$$(R_i + R_d)_{\max} \approx (2{,}7\,\text{ns})/((10\,\text{nF}) \cdot \ln 2^{13}) \approx 0{,}03\,\Omega\,.$$

Für $R_p \parallel R_d$ schließlich ergibt sich mit (96.37):

$$(R_p \parallel R_d)_{\min} \approx (22{,}2\mu\text{s})/\left((10\,\text{nF}) \cdot \left|\ln\left(1 - \frac{1}{2^{13}}\right)\right|\right) \approx 18{,}5\,\text{M}\Omega$$

◀

Zu diesem Beispiel sind zwei Hinweise erforderlich:

1. In der Praxis sind die Anforderungen an das Abtast-Halte-Glied nicht so hoch wie hier dargestellt, weil der angenommene Wechsel von u_M zwischen Maximal- und Minimalwert innerhalb der Zeit zwischen zwei Abtastungen so gut wie nie auftritt.
2. Zur Vermeidung von Problemen wird man die Abtastfrequenz (hier zu 45 kHz angenommen) in der Praxis nicht unbedingt zu einem ganzzahligen Vielfachen der höchsten Basisbandfrequenz (hier 15 kHz) wählen.

96.4.4.4 Quantisierung

Zur Quantisierung wird der gesamte Wertebereich des Basisbandsignals in eine endliche Anzahl von Quantisierungsintervallen eingeteilt. Jedem Wert (hier am Beispiel einer Spannung), der innerhalb eines bestimmten Intervalles liegt, wird ein Codewort mit N Elementen (Bits je Codewort) zugeordnet. Die Codierung geschieht im Dualsystem, und damit ist die Zahl der Quantisierungsintervalle zu $2^N - 1$; $N = 1, 2, 3, \ldots$ gegeben. Macht man alle Intervalle gleich groß, hat ein Intervall die Breite $\Delta u = |u_{\max} - u_{\min}| \cdot \frac{1}{2^N - 1}$. Es wurde vorausgesetzt, dass der Wertebereich der Spannung alle möglichen Codierungen erfordert. Damit ergibt sich ein

maximaler Amplituden- oder Quantisierungsfehler

$$F = \pm(u_{\max} - u_{\min}) \cdot \frac{1}{2} \cdot \frac{1}{2^N - 1} \tag{96.41}$$

Dieser Fehler lässt sich nach [23] durch den Klirrfaktor ausdrücken:

$$\textit{Klirrfaktor} \quad k \approx \frac{1}{2^N} \tag{96.42}$$

96.4.4.5 Quantisierungsgeräusch

Abb. 96.33 stellt zu einem sinusförmigen Signalverlauf das zugehörige quantisierte Signal dar. Da beide Verläufe nicht mehr übereinstimmen, am Empfangsort aber das quantisierte Signal zur Verfügung steht, interpretiert der Empfänger das Signal als Sinussignal mit überlagerter Stör- oder Geräuschspannung u_{NQ}, auch als *Quantisierungsrauschen* oder *Quantisierungsgeräusch* bezeichnet. Man kann zeigen, dass sich bei *Gleichwahrscheinlichkeit aller Quantisierungsintervalle* ein Signal-Quantisierungs-Geräuschabstand ergibt von:

Signal-Quantisierungs-Geräuschabstand

$$a_{\text{SQ}} = 20 \cdot \lg 2^N\,\text{dB} \tag{96.43}$$

Die spektrale Verteilung des Quantisierungsgeräusches ist nahezu konstant im Bereich von der kleinsten Basisbandfrequenz bis zur Abtastfrequenz. Die Geräuschleistung, summiert über den gesamten Bereich, ist ein konstanter Wert. Erhöht man demnach die Abtastfrequenz, verringert sich die auf den Basisbandfrequenzbereich entfallende Geräuschleistung. Diese Maßnahme wird als *Überabtastung* oder *oversampling* bezeichnet. Einer Verdopplung der Abtastfrequenz entspricht eine Erhöhung des Signal-Quantisierungs-Geräuschabstandes von $10 \cdot \lg 2 = 3\,\text{dB}$. Ebenso verdoppelt sich die Bitrate (Zahl der zu übertragenden Bits pro Zeiteinheit). Erhöht man andererseits die Zahl N der Bits pro Codewort, erhält man eine Erhöhung des Signal-Quantisierungs-Geräuschabstandes um $20 \cdot \lg 2 = 6\,\text{dB}$. Die Bitrate *erhöht* sich nur um den Faktor $\frac{N+1}{N}$, ist also kleiner als 2. Danach ist das Prinzip der Überabtastung wenig sinnvoll, trotzdem wird es angewendet, weil es bei der Rückgewinnung des Basisbandsignales keine steilflankigen Tiefpassfilter erfordert (Abb. 96.29b). Die störende Abtastfrequenz ist weiter von der Grenzfrequenz des Filters entfernt.

Die Maximal- und Minimalwerte der in Abb. 96.33 eingetragenen Geräuschspannung u_{NQ} nach (96.41) sind unabhängig von den Maximal- und Minimalwerten der Spannung u_M. Mit sinkendem Scheitelwert von u_M nimmt lediglich die Zahl der Maxima und Minima von u_{NQ} ab.

Beispiel 5

Für Sprache ist ein *Dynamikbereich* (Verhältnis $\hat{u}_{M\max}/\hat{u}_{M\min}$) von 40 dB erforderlich ($\hat{u}_{M\max}/\hat{u}_{M\min} = 100$). Damit das Quantisierungsgeräusch auch bei den kleinsten Amplituden nicht stört, muss gelten: $\hat{u}_{M\min} \geq 20 \cdot \hat{u}_{\text{NQ}}$. Daraus ergibt sich: $\hat{u}_M = 100 \cdot 20 \cdot \hat{u}_{\text{NQ}} =$

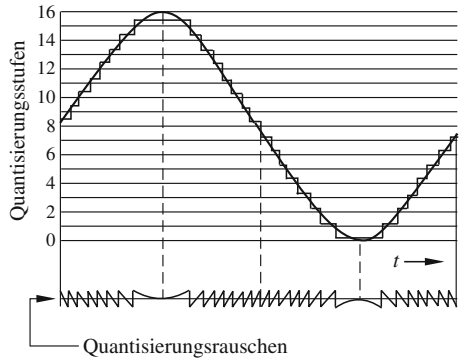


Abb. 96.33 Entstehung des Quantisierungsrauschens (nach [19])

$2000 \cdot \hat{u}_{NQ}$. Da der Wertebereich von u_M von $-\hat{u}_M$ bis $+\hat{u}_M$ reicht, folgt: $2 \cdot \hat{u}_M = 2 \cdot 100 \cdot 20 \cdot \hat{u}_{NQ} = 4000 \cdot \hat{u}_{NQ}$. Die Zahl der Quantisierungsintervalle ist demnach zu ≥ 4000 zu wählen, was mit einer Codierung mit 12 Bit je Codewort erreicht wird. ◀

96.4.4.6 Kompandierung

Die Deutsche Post AG hat einen Signal-Quantisierungs-Geräuschabstand $a_{SQ} = 20 \cdot \lg 2^N \,\text{dB} = 65\,\text{dB}$ (96.43) festgelegt, was $N = 11$ erfordert. Damit besteht jedes Codewort aus 11 Bit, und es sind entsprechend breitbandige Übertragungskanäle erforderlich. Durch Kompandierung kann N auf 7 verringert werden. Dazu werden die *kleinen* Signalamplituden, die besonders durch das Quantisierungsgeräusch gestört werden, zunächst verstärkt und dann erst in einen Digitalwert umgesetzt. Die großen Amplitudenwerte werden komprimiert. Abb. 96.34 zeigt die Übertragungsstrecke und den prinzipiellen Verlauf der Kompressions- und Expanderkennlinien. Über weitere Einzelheiten zu den Kennlinien informieren die Veröffentlichungen des CCITT. Dieses Verfahren wird wegen der Probleme mit der exakten Abstimmung von Kompressor- und Expanderkennlinie zueinander nicht mehr angewendet, denn je mehr beide voneinander abweichen, desto größer ist der Klirrfaktor. Man setzt das Verfahren der digitalen Kompandierung nach Abb. 96.34b ein. Das Basisbandsignal wird in einem Analog-Digital-Umsetzer hoher Stufenzahl (z. B. 12 Bit) gleichmäßig codiert, sodass alle Quantisierungsstufen gleich groß sind. In einem digitalen Kompressor werden den 12-Bit-kodierten Werten neue Werte mit z. B. nur 9 Bit zugeordnet, indem man Teilbereiche zusammenfasst. In Bild 34b ist dieser Vorgang vereinfacht anhand eines 3-Bit-Umsetzers und eines Kompressors auf 2 Bit gezeigt.

96.4.4.7 Codierung

Die Codierung kann nach einer beliebigen Zuordnung zwischen Signalamplitude und Codewort geschehen. In der Praxis werden häufig Codierungen nach dem Binärcode, dem BCD-Code (binär codierte Dezimalzahlen) und dem Gray-Code (benachbarte Signalwerte unterscheiden sich im Codewort nur durch Änderung an *einer* Stelle) verwendet. Hierfür gibt es industriell gefertigte Analog-Digital-Umsetzer, bei denen teilweise das Abtast-Halte-Glied (Abb. 96.29 und 96.32) integriert ist. Lediglich C_H ist extern anzuschließen. Entsprechend stehen die *zugehörigen* Digital-Analog-Umsetzer zur Demodulation zur Verfügung.

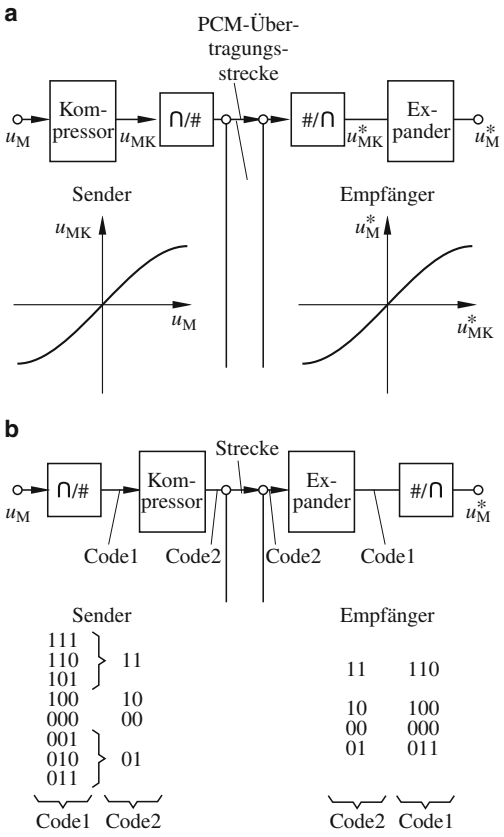


Abb. 96.34 Kompandierung. **a** Prinzip, **b** digitale Kompandierung

Bevor der Pulsträger mit dem Digitalsignal moduliert wird, muss es häufig noch aufbereitet werden. Dazu sind folgende Zielsetzungen von Bedeutung:

1. Das Signal soll keinen Gleichanteil enthalten (wichtig für transformatorische Kopplung zur galvanischen Trennung von Signalkreisen),
2. die Übertragungsfrequenz soll so klein wie möglich sein,
3. aus der übertragenen Impulsfolge soll sich die beim Sender benutzte Taktfrequenz zurückgewinnen lassen.

Abb. 96.35 listet entsprechende Verfahren auf.

Die Abkürzungen haben folgende Bedeutung:

- *RZ-Verfahren* (return to zero): Binär-Null entspricht Signal-Null; Binär-Eins entspricht Signal-Eins für eine halbe Taktdauer, dann Rückkehr zu Signal-Null. Nachteile: Folgen mehrere Binär-Einsen aufeinander, ist die zu übertragende Frequenz maximal gleich der Taktfrequenz. Folgen mehrere Binär-Nullen aufeinander, kann der Takt nicht zurückgewonnen werden. Außerdem ist Signal-Null, über längere Zeit übertragen, relativ störanfällig.

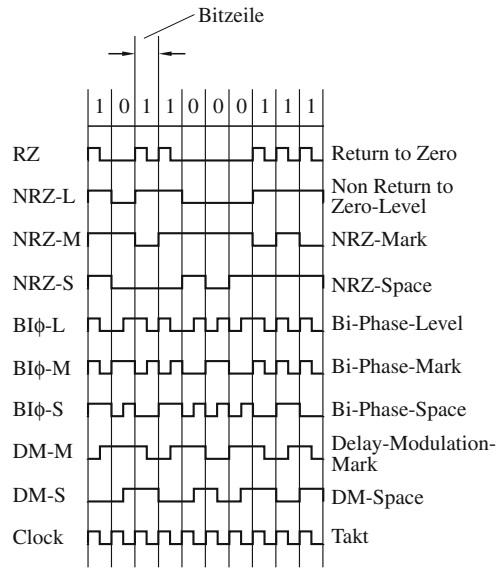


Abb. 96.35 Verwendete Formate bei der Pulscodemodulation (nach [19])

- *NRZ-Verfahren* (not return to zero): Es gibt kein Signal-Null, sondern Signal-Plus-Eins und Signal-Minus-Eins, z. B. +5 und −5 V.
  NRZ-L (L von level): Binär-Eins als Signal-Plus-Eins oder Signal-Minus-Eins, Binär-Null entgegengesetzt.
  NRZ-M (M von mark): Binär-Eins als Signaländerung, Binär-Null als keine Signaländerung.
  NRZ-S (S von space): Binär-Null als Signaländerung, Binär-Eins als keine Änderung.
  Nachteile: Keine Taktinformation bei Aufeinanderfolge mehrerer gleicher Bits. Es müssen Gleichspannungen übertragen werden.
  Vorteil: Die Übertragungsfrequenz ist höchstens halb so groß wie die Taktfrequenz.
- *Bi-Phase-Verfahren*: Die Binärzeichen werden mit zwei unterschiedlichen Perioden (Frequenzen) dargestellt, deshalb wird es auch mit *Zweifrequenzverfahren* oder *double frequency method* bezeichnet.
  Bi-Phase-L (BiΦ-L): Binär-Eins als Wechsel von Signal-Plus-Eins nach Signal-Minus-Eins in Taktmitte. Binär-Null als Wechsel von Signal-Minus-Eins nach Signal-Plus-Eins in Taktmitte.
  Bi-Phase-M (BiΦ-M): Binär-Null durch Wechsel der Signalpolarität zum Taktende bzw. Taktanfang, Binär-Eins durch zusätzlichen Wechsel in Taktmitte.
  Bi-Phase-S (BiΦ-S): Darstellung invers zu Bi-Phase-M.
  Nachteil: Die Übertragungsfrequenz ist maximal gleich der Taktfrequenz.
  Vorteile: Es ist keine Gleichspannung zu übertragen. Der Takt kann einfach zurückgewonnen werden.
- DM-Verfahren (D von delay, auch Miller-Code):
  DM-M: Binär-Eins durch Wechsel der Signalpolarität in Taktmitte, Binär-Null nach Binär-Eins kein Polaritätswechsel, Binär-Null nach Binär-Null Polaritätswechsel zum Taktbeginn.
  DM-S-Verfahren: Binär-Null durch Wechsel der Signalpolarität in Taktmitte, Binär-Eins nach Binär-Null kein Polaritätswechsel, Binär-Eins nach Binär-Eins Polaritätswechsel zum Taktbeginn.
  Vorteile: Die Übertragungsfrequenz ist maximal gleich der halben Taktfrequenz. Es ist keine Gleichspannung zu übertragen. Der Takt kann einfach zurückgewonnen werden.

Das DM-Verfahren stellt somit eine optimale Lösung dar. Es wird häufig zur Ton- und Messwertaufzeichnung auf Magnetbandspeichern eingesetzt, weil auch die Speicherung von Gleichgrößen möglich ist. Weiterhin findet es Anwendung bei der Telefon-Sprachübertragung und der PCM-Telemetrie.

96.4.4.8 Deltamodulation (DM) und Differenz-Pulscodemodulation (DPCM)

Das übertragene Signal weist einen beträchtlichen Teil an Redundanz auf, weil sich aufeinanderfolgende Abtastwerte im Allgemeinen nur geringfügig voneinander unterscheiden. Es liegt deshalb nahe, nur jeweils deren Differenz, codiert mit wenigen Bit, zu übertragen. Damit können die Bitrate und die erforderliche Bandbreite des Übertragungskanals verringert werden.

- *Deltamodulation (DM)*: Es wird jeweils nur ein Bit übertragen, das das Vorzeichen der Änderung zum vorhergehenden Wert enthält. Unterscheiden sich allerdings zwei benachbarte Werte um mehr als ein Bit, kommt es zu einer *Steigungsüberlastung* (engl. slope overload). Das codierte Signal kann dem Basisbandsignal nur verzögert folgen, da pro Abtastung immer nur die kleinste Stufenhöhe zur Verfügung steht. Daraus resultiert ein Quantisierungsgeräusch besonderer Art, das *Granulargeräusch* (engl. granular noise). Man kann diesen Effekt dadurch verringern, dass man die Abtastfrequenz erhöht.
- *Delta-Sigma-Modulation*: Das modulierende Signal wird zunächst einem Integrierglied zugeführt, an dessen Ausgang die steilen Signalanstiege flacher sind. Damit kann der Delta-Modulator den Signalverlauf ohne Steigungsüberlastung darstellen. Im Empfänger wird das demodulierte Signal über ein Differenzierglied geführt, um die im Sender erfolgte lineare Signalverzerrung rückgängig zu machen.
- *Adaptive Deltamodulationsverfahren*: Hier gibt es eine Menge von Verfahren zur Codierung. So kann man beispielsweise bei einer bestimmten Anzahl aufeinanderfolgender positiver Impulse (z. B. drei) die Quantisierungsstufenhöhe verdoppeln, beim Auftreten einer entsprechenden Anzahl negativer Impulse wieder halbieren. Damit lässt sich bei gleicher Abtastfrequenz die Steigungsüberlastung verringern.
- *Differenz-Pulscodemodulation (DPCM)*: Bei diesen vorwiegend zur Fernsehbildübertragung eingesetzten Verfahren werden zwei aufeinanderfolgende Abtastwerte je nach deren Differenz auch mit mehr als einem Bit codiert. Treten allerdings große Differenzen in der Bildhelligkeit auf, werden ihnen nur wenige Bit zugeordnet, da das Auge für diese Änderungen wenig empfindlich ist.

Filter

Wilfried Plaßmann

97

Teil XIII



97.1 Einfache passive R-C-Filter 1230
97.2 Weitere Ausführungsformen 1235
97.3 Allgemeiner Filterentwurf 1238

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_97

97.1 Einfache passive R-C-Filter

Die Ableitungen setzen sinusförmige Signale und den eingeschwungenen Zustand im Netzwerk voraus.

97.1.1 Allgemeines

Filter sind Vierpole und werden häufig auch innerhalb dieses Themenkreises behandelt. Sie erhalten hier aber im Hinblick auf die speziellen Anwendungen in der Nachrichtentechnik ein eigenes Kapitel.

Filter grenzen ein gegebenes Frequenzspektrum ein. Damit soll eines der folgenden Ziele erreicht werden:

- Aus einem Frequenzgemisch mit vielen Signalen wird ein einzelnes Signal ausgefiltert. Beispiel: Senderauswahl beim Rundfunk- und Fernsehempfang.
- Ein vorhandenes Signal wird so verändert, dass es weniger Bandbreite im Übertragungskanal benötigt; die damit verbundene Signalverfälschung muss vertretbar sein. Beispiel: Farbfernsehübertragung; hier können nicht alle darstellbaren Bildpunkte in aufeinander folgenden Bildern unterschiedliche Farben und unterschiedliche Helligkeiten aufweisen.
- Eine oder mehrere Frequenzanteile sollen gezielt aus dem Spektrum entfernt werden. Beispiel: Störsignale mit Netzfrequenz 50 Hz.

Um die Eigenschaften eines Filters zu charakterisieren, werden in der Regel das Verhältnis Ausgangs- zu Eingangsspannung und die Phasenverschiebung dieses Verhältnisses in Abhängigkeit von der Frequenz angegeben. Damit die genannten Größen über mehrere Dekaden dargestellt werden können, verwendet man logarithmische Skalenteilungen.

- *Frequenzgang*: Darstellung des Amplitudenverhältnisses von Ausgangs- zu Eingangsspannung in Abhängigkeit von der Frequenz (siehe (97.16a) und Hinweis dazu).
- *Amplitudenfrequenzgang*: Häufig doppelt-logarithmische Darstellung des Verhältnisses von Ausgangs- zu Eingangsspannung. Für die Ordinatenbeschriftung gilt:

$$20 \cdot \lg(|\underline{U}_{\text{aus}}/\underline{U}_{\text{ein}}|) \quad \text{in dB}$$

  (siehe (97.16b) und Hinweis dazu).
- *Phasengang*: Darstellung des Phasenwinkels von Ausgangs- zu Eingangsspannung in Abhängigkeit von der Frequenz. Die Frequenzachse ist logarithmisch, die Phasenachse ist linear geteilt.

  Je nachdem, wie die Frequenzanteile des Originalsignals das Filter passieren, spricht man von:
- *Tiefpassfilter*, wenn die im Signal enthaltenen Frequenzanteile von Gleichspannung bis zur Grenzfrequenz durchgelassen werden;
- *Hochpassfilter*, wenn ab der Grenzfrequenz die im Signal enthaltenen Frequenzanteile durchgelassen werden;
- *Bandpassfilter*, wenn nur die im Signal enthaltenen Frequenzanteile eines meist eng begrenzten Frequenzbereiches durchgelassen werden;
- *Bandsperre*, wenn alle im Signal enthaltenen Frequenzanteile außer denen in einem meist eng begrenzten Frequenzbereich durchgelassen werden.

Der Übergang vom Sperr- in den Durchlassbereich und umgekehrt wird in der Praxis durch die Angabe der Grenzfrequenz bzw. der Resonanzfrequenz als scharf begrenzt angenommen, obwohl es sich um einen allmählichen Übergang handelt. Kenngrößen für die praktische Anwendung sind:

- *Grenzfrequenz* f_g: Frequenz, bis zu der (Tiefpass) bzw. ab der (Hochpass) die im Signal enthaltenen Frequenzanteile durchgelassen werden. Bei den Filtern erster Ordnung gilt: 1. Der Betrag des Verhältnisses Ausgangs- zu Eingangsspannung ist vom Maximalwert auf das 0,707fache des Maximalwertes abgesunken. 2. Der Phasenwinkel zwischen Ein- und Ausgangsspannung beträgt $-45°$ (Tiefpass) bzw. $+45°$ (Hochpass).
- *Durchlassbereich*: Frequenzbereich bis zu f_g (Tiefpass) bzw. ab f_g (Hochpass) bzw. innerhalb der Bandbreite B (Bandpass).
- *Sperrbereich*: Frequenzbereich ab f_g (Tiefpass) bzw. bis zu f_g (Hochpass) bzw. innerhalb der Bandbreite B (Bandsperre).
- *Phasenwinkel* φ_g: Phasenwinkel bei der Grenzfrequenz f_g.
- *Grad (Ordnung) n eines Filters*: Höchster im Nenner vorkommender Exponent der Kreisfrequenz ω für das Verhältnis von Ausgangs- zu Eingangsspannung, wenn der Zähler eine Konstante ist. Er wird vorwiegend für Hoch- und Tiefpassfilter angegeben. Beispiel: Tiefpassfilter 3. Grades $U_{\text{aus}}/U_{\text{ein}} = A_0/(-j\omega^3 - B\omega^2 + Dj\omega + E)$; A_0, B, D, E, Konstante, dimensionslos oder Kombinationen aus R, L und C.
- *Steigung des Amplitudenganges*: Sie wird im Sperrbereich angegeben und ist ein Maß für die Fähigkeit, Frequenzanteile außerhalb des Durchlassbereiches zu unterdrücken. Sie wird bestimmt durch den Grad n des Filters und in dB/Dekade angegeben. Allgemein gilt: Steigung eines Filters n-ten Grades: $-n \cdot 20$ db/Dekade für einen Tiefpass bzw. $+n \cdot 20$ db/Dekade für einen Hochpass.
- *Bandbreite B*: Frequenzbereich, der durchgelassen (Bandpass) bzw. gesperrt (Bandsperre) wird. Sie ist bestimmt durch den Bereich, in dem die Ausgangsgröße beim Bandpass zwischen Maximalwert und 0,707 · Maximalwert liegt bzw. bei der Bandsperre zwischen Minimalwert und 1,41 · Minimalwert liegt.
- *Güte Q*: Mass für die Eigenschaften von Bandpass bzw. Bandsperre im Durchlass- bzw. Sperrbereich. Siehe dazu auch Abb. 97.4.
- *Dämpfungsmaß*: $d = 1/Q$
- *Resonanzfrequenz* f_0: Frequenz, die ein Bandpassfilter optimal durchlässt bzw. eine Bandsperre maximal unterdrückt.

Für die zuletzt genannten Kenngrößen besteht folgende Beziehung:

$$f_0 = B \cdot Q \tag{97.1}$$

Hinweis: In einer verallgemeinerten Filtertheorie spricht man z. B. auch beim Hoch- und Tiefpass erster Ordnung von Güte und Resonanzfrequenz. Hier werden dagegen die üblicherweise in der praktischen Nachrichtentechnik verwendeten Bezeichnungen eingesetzt.

97.1.2 Passive *R*-*C*-Filter

Die hier behandelten *R*-*C*-Filter lassen sich prinzipiell auch als *R*-*L*-Filter aufbauen. Der Einsatz von Induktivitäten erfordert besondere Aufmerksamkeit, weil sie stets nichtideal sind (Reihenschaltung aus der Induktivität L und dem Wicklungswiderstand R_L) und bei der Verwendung von Eisen ein nichtlineares Bauteil vorliegt (die Induktivität L ist über $\mu_r = f(i)$ stromabhängig). Außerdem haben Induktivitäten größere Abmessungen und lassen sich in Integrierten Schaltkreisen weniger gut fertigen. Aktive RC-Filter werden hier nicht behandelt.

Im folgenden sind die Eigenschaften einiger Filterschaltungen ohne Ableitung aufgelistet. Die Berechnung geschieht entweder mit der komplexen Rechnung oder unter Anwendung der Vierpoltheorie. Es werden sinusförmige Spannungs- und Stromverläufe im gesamten Filter vorausgesetzt.

97.1.2.1 Passive *R*-*C*-Tiefpassfilter

Passive R-C-Tiefpassfilter erster, zweiter und dritter Ordnung: Abb. 97.1 zeigt die Realisierung sowie das Verhältnis von Ausgangs- zu Eingangsspannung mit linearer Skalenteilung.

Beispiel 1

Ein Tiefpassfilter dritter Ordnung soll in einer Sinusoszillatorschaltung eingesetzt werden. Dazu muss die Ausgangsspannung gegenüber der Eingangsspannung bei *einer* Frequenz um 180° (Vorzeichen ohne Bedeutung) phasenverschoben sein. Das ist genau dann der Fall, wenn der Imaginärteil des Spannungsverhältnisses $\underline{U}_2/\underline{U}_1$ Null ist. Aus der Tabelle folgt damit: $j(6\omega_0 RC - \omega_0^3 R^3 C^3) = j\omega_0 RC(6 - \omega_0^2 R^2 C^2) = 0$ und daraus $\omega_0 = \sqrt{6}/RC$. Die negative Lösung der quadratischen Gleichung scheidet aus, da es keine negativen Frequenzen gibt. In den verbleibenden Realteil eingesetzt ergibt sich: $\underline{U}_2/\underline{U}_1(\omega_0) = -1/29$. Die Amplitude der Ausgangsspannung ist bei ω_0 auf $1/29$ der Amplitude der Eingangsspannung abgesunken, der Phasenwinkel beträgt $-180°$. Damit ist ein Verstärker mit der Verstärkung -29 erforderlich.

Allgemein gilt: Der Übergang vom Durchlass- in den Sperrbereich ist um so schmaler, je höher die Ordnung n des Filters ist. ◀

97.1.2.2 Passive *R*-*C*-Hochpassfilter

Passive R-C-Hochpassfilter erster, zweiter und dritter Ordnung: Abb. 97.2 zeigt die Realisierung und den Amplitudenfrequenzgang.

Das Hochpassfilter dritter Ordnung wird ebenfalls für Sinusoszillatorschaltungen eingesetzt und hat gegenüber dem Tiefpassfilter dritter Ordnung den Vorteil, dass Gleichspannung nicht durchgelassen wird (wichtig in Transistorschaltungen, bei denen man die Eingangsspannung des Filters am Kollektor abgreift und vom Ausgang zur Basis zurückführt).

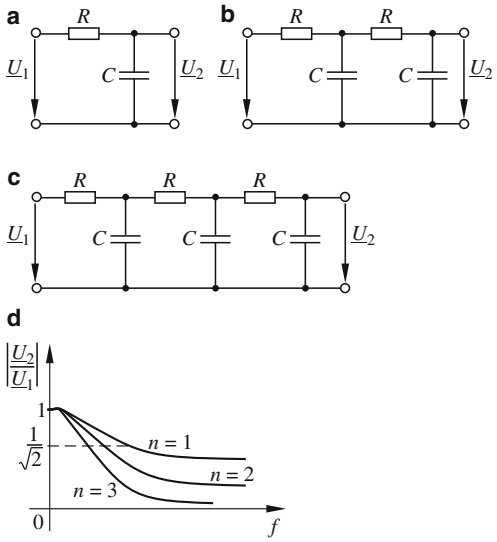


Abb. 97.1 *R*-*C*-Tiefpassfilter. **a** erster ($n = 1$), **b** zweiter ($n = 2$), **c** dritter Ordnung ($n = 3$), **d** Amplitudenfrequenzgang

Tab. 97.1 Übertragungsverhalten von Tiefpässen erster bis dritter Ordnung

Tiefpass erster Ordnung ($n = 1$), Abb. 97.1a				
$\frac{\underline{U}_2}{\underline{U}_1}$	φ	f_g	φ_g	Steigung
$\frac{1}{1+j\omega RC}$	$-\arctan \omega RC$	$\frac{1}{2\pi RC}$	$-45°$	
Näherung für $\omega \ll 1/RC$				
≈ 1	$\approx 0°$			≈ 0
Näherung für $\omega \gg 1/RC$				
$\approx \frac{1}{j\omega RC}$	$= -\arctan \omega RC \approx -90°$			$\approx \frac{-20\,\text{dB}}{\text{Dekade}}$

Tiefpass zweiter Ordnung ($n = 2$), Abb. 97.1b		
$\frac{\underline{U}_2}{\underline{U}_1}$	φ	Steigung
$\frac{1}{1+3j\omega RC-\omega^2R^2C^2}$	$-\arctan \frac{3\omega RC}{1-\omega^2R^2C^2}$ [a]	
Näherung für $\omega \ll 1/RC$		
≈ 1	$\approx 0°$	≈ 0
Näherung für $\omega \gg 1/RC$		
$\approx \frac{-1}{\omega^2R^2C^2}$	$\approx -180° + \arctan \frac{3}{\omega RC}$ [a, b] $\approx -180°$ [a, c]	$\approx \frac{-40\,\text{dB}}{\text{Dekade}}$

Tiefpass dritter Ordnung ($n = 3$), Abb. 97.1c		
$\frac{\underline{U}_2}{\underline{U}_1}$	φ	Steigung
$\frac{1}{1+6j\omega RC-5\omega^2R^2C^2-j\omega^3R^3C^3}$	$-\arctan \frac{6\omega RC-\omega^3R^3C^3}{1-5\omega^2R^2C^2}$ [a]	
Näherung für $\omega \ll 1/RC$		
≈ 1	≈ 0	≈ 0
Näherung für $\omega \gg 1/RC$		
$\approx \frac{-1}{j\omega^3R^3C^3}$	$\approx -180° - \arctan \frac{\omega RC}{5}$ [a, b] $\approx -270°$ [a, c]	$\approx \frac{-60\,\text{dB}}{\text{Dekade}}$

[a] $\tan(x) = \tan(x \pm n \cdot 180°)$; $n = 1, 2, 3, \ldots$;
[b] erste Näherung;
[c] zweite, grobe Näherung, häufig zulässig

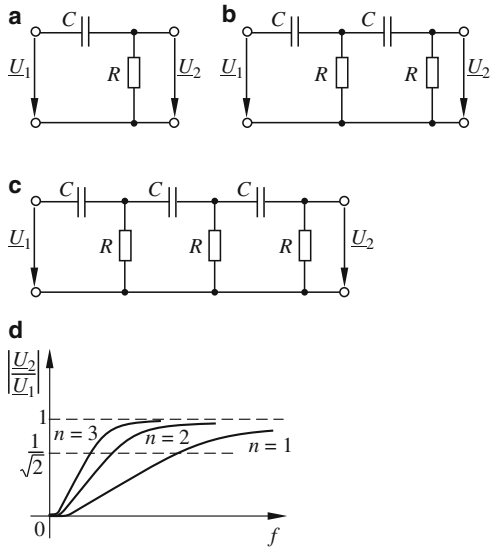


Abb. 97.2 RC-Hochpassfilter. **a** erster ($n = 1$), **b** zweiter ($n = 2$), **c** dritter Ordnung ($n = 3$), **d** Amplitudenfrequenzgang

Tab. 97.2 Übertragungsverhalten von Hochpässen erster bis dritter Ordnung

Hochpass erster Ordnung ($n = 1$), Abb. 97.2a				
$\frac{U_2}{U_1}$	φ	f_g	φ_g	Steigung
$\frac{j\omega RC}{1+j\omega RC}$	$90° - \arctan \omega RC$	$\frac{1}{2\pi RC}$	45°	
Näherung für $\omega \ll 1/RC$				
$\approx j\omega RC$	$\approx 90°$			$\approx \frac{20\,\text{dB}}{\text{Dekade}}$
Näherung für $\omega \gg 1/RC$				
≈ 1	$= 90° - \arctan \omega RC \approx 0$			≈ 0

Hochpass zweiter Ordnung ($n = 2$), Abb. 97.2b		
$\frac{U_2}{U_1}$	φ	Steigung
$\frac{-\omega^2R^2C^2}{1+3j\omega RC-\omega^2R^2C^2}$	$180° - \arctan \frac{3\omega RC}{1-\omega^2R^2C^2}$ [a]	
Näherung für $\omega \ll 1/RC$		
$\approx -\omega^2R^2C^2$	$\approx 180° + \arctan 3\omega RC$ [a, b] $\approx 180°$ [a, c]	$\approx \frac{40\,\text{dB}}{\text{Dekade}}$
Näherung für $\omega \gg 1/RC$		
≈ 1	$\approx \arctan \frac{3}{\omega RC}$ [a, b] $\approx 0°$ [a, c]	≈ 0

Hochpass dritter Ordnung ($n = 3$), Abb. 97.2c		
$\frac{U_2}{U_1}$	φ	Steigung
$\frac{-j\omega^3R^3C^3}{1+5j\omega RC-6\omega^2R^2C^2-j\omega^3R^3C^3}$	$270° - \arctan \frac{5\omega RC-\omega^3R^3C^3}{1-6\omega^2R^2C^2}$ [a]	
Näherung für $\omega \ll 1/RC$		
$\approx -j\omega^3R^3C^3$	$\approx 270° - \arctan 5\omega RC$ [a, b] $\approx 270°$ [a, c]	$\approx \frac{60\,\text{dB}}{\text{Dekade}}$
Näherung für $\omega \gg 1/RC$		
≈ 1	$\approx 90° - \arctan \frac{\omega RC}{6}$ [a, b] $\approx 0°$ [a, c]	≈ 0

[a] $\tan(x) = \tan(x \pm n \cdot 180°)$; $n = 1, 2, 3, \ldots$;
[b] erste Näherung;
[c] zweite, grobe Näherung, häufig zulässig

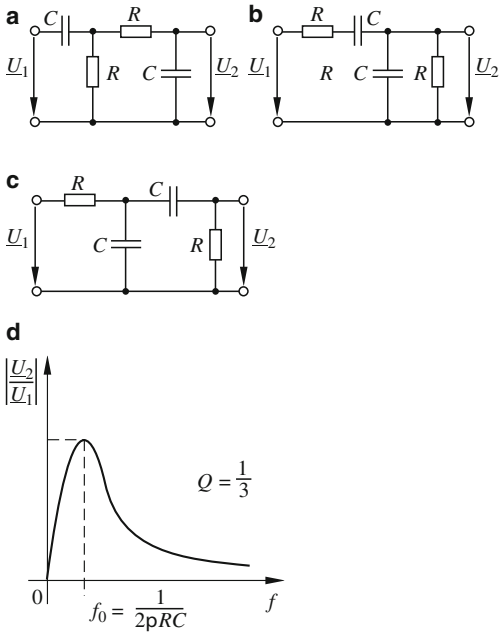


Abb. 97.3 Bandpass aus Hoch- und Tiefpass. **a–c** 3 Schaltungen mit identischem Verhalten, **d** Amplitudenfrequenzgang

97.1.2.3 Bandpass aus *R*-*C*-Hoch- und Tiefpassfilter

Nach Abb. 97.3 kann man Hoch- und Tiefpassfilter zusammenschalten und erhält drei Filterschaltungen mit identischem Verhalten.

$$\frac{\underline{U}_2}{\underline{U}_1} = \frac{j\omega RC}{1 + 3j\omega RC - \omega^2R^2C^2} \tag{97.2}$$

Die maximale Ausgangsspannung ergibt sich für $\omega_0 = 1/RC$. Man erhält folgende Näherungen:

$$\begin{aligned} \omega \ll 1/RC \quad &\Rightarrow \quad \frac{\underline{U}_2}{\underline{U}_1} \approx j\omega RC, \quad \text{Steigung } \frac{20\,\text{dB}}{\text{Dekade}} \\ \omega \gg 1/RC \quad &\Rightarrow \quad \frac{\underline{U}_2}{\underline{U}_1} \approx \frac{1}{j\omega RC}, \quad \text{Steigung } \frac{-20\,\text{dB}}{\text{Dekade}} \end{aligned} \tag{97.3}$$

Wegen der geringen Steigung außerhalb des Durchlassbereiches wird diese Filterschaltung nur selten eingesetzt. Die Steigung −20 dB/Dekade bedeutet, dass bei einer Frequenzverzehnfachung die Ausgangsspannung auf 1/10 sinkt. Die Forderungen in der Nachrichtentechnik zur Senderselektion dagegen lauten, dass bei einer Frequenzänderung von ±1 % gegenüber der Resonanzfrequenz $f_0 = \omega_0/2\pi$ die Amplitude der Ausgangsspannung um den Faktor $10^{-1} \ldots 10^{-4}$ kleiner wird.

Beispiel 2

Ein Sender mit der Frequenz $f_0 = 1{,}000\,\text{MHz}$ soll empfangen werden, ein Nachbarsender mit der Frequenz $f_N = 1{,}010\,\text{MHz}$ stört den Empfang und soll mit einem Filter unterdrückt werden. Beide Sender haben die gleiche

Amplitude. Am Ausgang des Filters soll die Amplitude des Nachbarsenders auf 10^{-1} derjenigen des zu empfangenen Senders abgesunken sein. Verwendet werden soll ein Bandpassfilter aus *R-C*-Hoch- und Tiefpassfilter. Mit (97.2) ergibt sich für das Verhältnis von Nachbarsignalamplitude zu Empfangssignalamplitude am Ausgang des Filters mit $f_N = 1{,}010\,\text{MHz} = 1{,}000\,\text{MHz} \cdot 1{,}01 = f_0 \cdot 1{,}01$:

$$\left|\frac{\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}\Big|_{f_N}}{\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}\Big|_{f_0}}\right| = \left|\frac{\frac{j\cdot\omega_N\cdot R\cdot C}{1+3\cdot j\cdot\omega_N\cdot R\cdot C-\omega_N^2\cdot R^2\cdot C^2}}{\frac{j\cdot\omega_0\cdot R\cdot C}{1+3\cdot j\cdot\omega_0\cdot R\cdot C-\omega_0^2\cdot R^2\cdot C^2}}\right|$$

$$= \left|\frac{\frac{j\cdot\omega_0\cdot R\cdot C\cdot 1{,}01}{1+3\cdot j\cdot\omega_0\cdot R\cdot C\cdot 1{,}01-(\omega_0 R\cdot C)^2\cdot 1{,}01^2}}{\frac{j}{1+3j-1}}\right|_{\omega_0=1/RC}$$

$$= \left|\frac{\frac{j\cdot 1{,}01}{1+3\cdot j\cdot 1{,}01-1{,}01^2}}{\frac{j}{3\cdot j}}\right| = \frac{\frac{1{,}01}{\sqrt{(-0{,}0201)^2+(3\cdot 1{,}01)^2}}}{\frac{1}{3}}$$

$$\approx \frac{\frac{1{,}01}{\sqrt{9{,}1813}}}{\frac{1}{3}} \approx 3\cdot\frac{1{,}01}{3{,}0301} \approx 0{,}99997$$

Waren beide Amplituden am Filtereingang gleich groß, hat sich die Amplitude des Nachbarsenders am Filterausgang auf das 0,99997fache der Amplitude des empfangenen Senders „verringert". Die gestellte Forderung ist mit dieser Filterschaltung nicht zu erfüllen. Theoretisch könnte man eine Vielzahl solcher Filterschaltungen hintereinanderschalten (Kettenschaltung, siehe Kap. 93). Das ergäbe dann $n = \lg(10^{-1})/\lg(0{,}99997) \approx 76.752$ Bandpässe. Dieses Konzept ist technisch nicht ausführbar, selbst wenn man von einer konstanten Resonanzfrequenz ausgeht. ◀

Die im Beispiel 2 gestellte Forderung kann aber mit vergleichsweise geringem Aufwand mit LC-Schwingkreisen erfüllt werden.

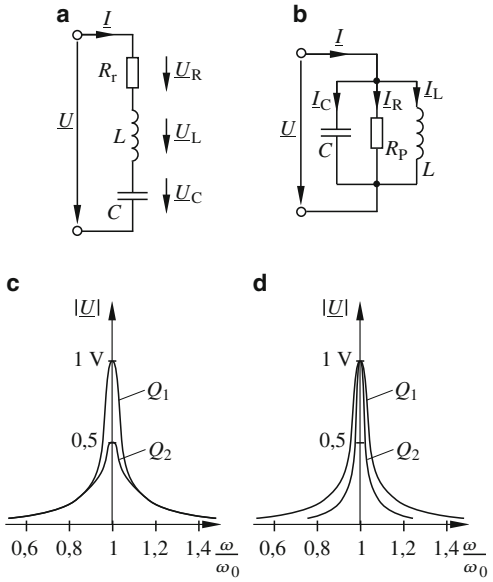


Abb. 97.4 Schwingkreise als Bandfilter. **a** Reihenschwingkreis, **b** Parallelschwingkreis, **c** Spannung am Parallelschwingkreis bei Stromeinprägung bei $Q_2 = 2Q_1$, **d** wie **c**, aber auf den Maximalwert normiert

Tab. 97.3 Formeln zum Parallel- und Reihenschwingkreis

	Reihenschwingkreis	Parallelschwingkreis
Allgemeine Beziehung	$\underline{I} = \frac{\underline{U}_0}{R_r+j(\omega L-\frac{1}{\omega C})}$ Speisung aus einer Spannungsquelle mit der Spannung $\underline{U}_0$	$\underline{U} = \underline{I}_0\frac{1}{\frac{1}{R_p}+j(\omega C-\frac{1}{\omega L})}$ Speisung aus einer Stromquelle mit dem Strom $\underline{I}_0$
Resonanz für Imaginärteil von $\underline{Z}=0$: $\Rightarrow \omega_0 = \frac{1}{\sqrt{LC}}$	$\underline{I}(\omega_0) = \frac{\underline{U}_0}{R_r}$ $\underline{U}_R = \underline{U}_0$; $\underline{U}_C = \frac{\underline{U}_0}{R_r j\omega_0 C}$; $\underline{U}_L = \frac{\underline{U}_0}{R_r}\cdot j\omega_0 C$ $\lvert\underline{U}_C\rvert = \lvert\underline{U}_L\rvert$	$\underline{U} = \underline{I}_0\cdot R_p$ $\underline{I}_R = \underline{I}_0$; $\underline{I}_C = \underline{I}_0\cdot R_p\cdot j\omega_0 C$; $\underline{I}_L = \underline{I}_0\cdot R_p\frac{1}{j\omega_0 L}$ $\lvert\underline{I}_C\rvert = \lvert\underline{I}_L\rvert$
Güte Q	$Q_r = \frac{\omega_0 L}{R_r} = \frac{1}{R_r\omega_0 C} = \frac{1}{R_r}\sqrt{\frac{L}{C}}$	$Q_p = \frac{R_p}{\omega_0 L} = R_p\omega_0 C = R_p\sqrt{\frac{C}{L}}$
Dämpfungsmaß d	$d = 1/Q$	$d = 1/Q$
Bandbreite B	$B = \frac{f_0}{Q}$	$B = \frac{f_0}{Q}$

97.1.2.4 *R-L-C*-Bandpass und -Bandsperre

Bandpass und Bandsperre werden als Reihen- bzw. Parallelschwingkreis realisiert. Setzt man mehrere derartige Filter ein, spricht man von *Bandfiltern* (siehe unten). Beim Reihenschwingkreis erfasst der ohmsche Widerstand R_r nach Abb. 97.4a den Wicklungswiderstand der Spule und eventuell den zum Kondensator parallel liegenden Isolationswiderstand, der in einen Reihenwiderstand umgerechnet worden ist. Beim Parallelschwingkreis rechnet man den zur Spule in Reihe liegenden Wicklungswiderstand in einen Parallelwiderstand um und fasst ihn mit dem Isolationswiderstand des Kondensators zu R_p zusammen (Abb. 97.4b). Formeln siehe Tab. 97.3.

In der Nachrichtentechnik werden häufig Parallelschwingkreise eingesetzt, die im Kollektorkreis eines Transistors liegen und deshalb von einer Stromquelle gespeist werden. Es handelt sich damit um einen Bandpass, der im weiteren betrachtet werden soll.

Die Güte beeinflusst sowohl die Höhe der Spannung im Resonanzfall als auch die Bandbreite B. Die Güte wird bestimmt durch den ohmschen Widerstand und das Verhältnis C/L: Je größer R_p und je größer C im Verhältnis zu L ist, um so größer ist die Güte Q. In der Empfängertechnik (z. B. bei Rundfunk- und Fernsehempfängern) werden Schwingkreise als Bandpass eingesetzt, um aus einer Vielzahl von zur Verfügung stehenden Sendern *einen* auszuwählen. Je größer die Güte ist, um so wirkungsvoller werden störende benachbarte Sender unterdrückt. Allerdings sinkt auch die Bandbreite, die aber einen bestimmten Wert nicht unterschreiten darf (siehe Modulation). Damit beide Eigenschaften gleichzeitig genutzt werden können, wird ein Weg nach Abb. 97.5 gewählt. Man verwendet mehrere Schwingkreise, bei denen die Güte möglichst groß

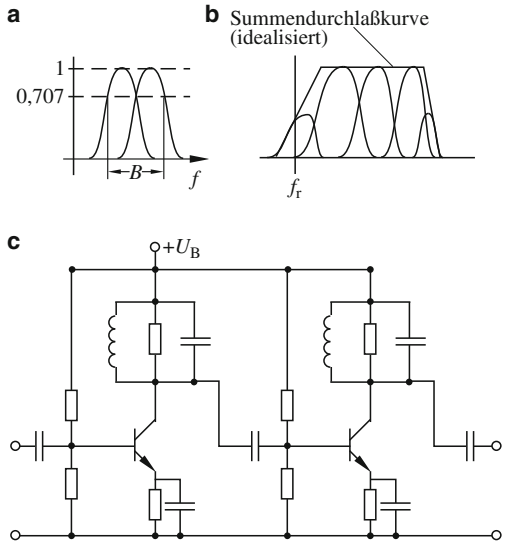


Abb. 97.5 Durchlasskurve mit mehreren Bandfiltern. **a** zwei Bandfilter zur Erhöhung der Bandbreite, **b** fünf Bandfilter für die Restseitenbandübertragung (Fernsehen), **c** Schaltungstechnische Realisierung

und die Bandbreite demzufolge zu klein ist (siehe auch der folgende Abschn. 97.1.2.5 „Bandfilter“). Erreicht wird damit eine gute Unterdrückung von Nachbarsendern. Um die geforderte Bandbreite zu realisieren, verwendet man mehrere Filter, deren Resonanzfrequenzen geringfügig gegeneinander verschoben sind. Als resultierende Durchlasskurve ergeben sich die in Abb. 97.5 dargestellten Formen. Auf diese Weise wird auch die in der Fernsehempfangstechnik verwendete Filterkurve für die Restseitenbandübertragung erreicht (Kap. 96 „Modulation“, Abb. 96.9).

Beispiel 3

Es gelten die gleichen Vorgaben wie im Beispiel 2. Als Filter soll ein Parallelschwingkreis verwendet werden, der im Kollektorkreis eines Transistors angeordnet ist und damit für beide Frequenzen von einer Stromquelle angesteuert wird. Mit $\omega_0 = 1/\sqrt{LC}$ folgt für das Spannungsverhältnis:

$$\left|\frac{\underline{U}_{f_N}}{\underline{U}_{f_0}}\right| = \left| \frac{\underline{I}\cdot \frac{1}{\frac{1}{R_p}+j(\omega_0\cdot 1{,}01\cdot C-\frac{1}{\omega_0\cdot 1{,}01\cdot L})}}{\underline{I}\cdot \frac{1}{\frac{1}{R_p}+j\underbrace{(\omega_0\cdot C-\frac{1}{\omega_0\cdot L})}_{=0}}} \right|_{\omega_0=1/\sqrt{LC}}$$

$$= \left| \frac{1}{\frac{1}{R_p}+j(\omega_0\cdot 1{,}01\cdot C-\frac{1}{\omega_0\cdot 1{,}01\cdot L})} \cdot \frac{1}{\frac{1}{R_p}} \right|$$

Nach einigen Umformungen und der Bedingung, dass für dieses Verhältnis der Wert 1/10 gelten muss, folgt:

$$\left|\frac{\underline{U}_{f_N}}{\underline{U}_{f_0}}\right| = \frac{1}{\sqrt{1+\underbrace{R_p^2\left(\frac{1{,}01^2\cdot\omega_0^2\cdot L\cdot C-1}{1{,}01\cdot\omega_0\cdot L}\right)^2}_{=99}}}$$

$$= \frac{1}{10} \quad \Rightarrow \quad R_p^2\left(\frac{1{,}01^2-1}{1{,}01\cdot\omega_0\cdot L}\right)^2 = 99$$

Daraus ergibt sich: $R_p(\frac{1{,}01^2-1}{1{,}01\cdot\omega_0\cdot L})^2 = \sqrt{99}$; und schließlich: $\frac{R_p}{\omega_0\cdot L} \approx 500$. Das ist aber nach Tab. 97.3 die Güte Q des Schwingkreises. Ein Wert in dieser Größenordnung ist in der Praxis kaum erreichbar, kann aber mit zwei hintereinandergeschalteten Schwingkreisen nach Abb. 97.5c realisiert werden. Allerdings muss überprüft werden, ob die aus $Q = 500$ folgende Bandbreite nach Tab. 97.3 für eine Signalübertragung ausreicht: $B = \frac{f_0}{Q} = \frac{1{,}000\,\text{MHz}}{500} = 2\,\text{kHz}$. Ist eine größere Bandbreite gefordert, lässt sich ein Bandfilter nach Abschn. 97.1.2.5 einsetzen. ◀

97.1.2.5 Bandfilter

Setzt man zwei Filter ein, die magnetisch (gekennzeichnet durch die *Gegeninduktivität M* in Abb. 97.6a) oder kapazitiv gekoppelt sind, spricht man von einem *Bandfilter*. Je nach Kopplungsgrad erhält man als resultierende Filterkurve die in Abb. 97.6b dargestellten Verläufe. Vom Transformator werden folgende Begriffe übernommen:

$$\textit{Kopplungsfaktor} \quad k = \frac{M}{\sqrt{L_1 L_2}} \tag{97.4}$$

M Gegeninduktivität in H; L_1, L_2 Induktivität der Primär- bzw. Sekundärwicklung in H.

Weiterhin gilt auch:

$$\textit{Kopplungsfaktor} \quad k = \sqrt{1-\sigma} \tag{97.5}$$

σ Streufaktor.

Es ergeben sich die unterschiedlichen Resonanzkurven nach Abb. 97.6b.

Man unterscheidet: a) $k < d$: *unterkritische Kopplung*; b) $k = d$: *kritische Kopplung*; c) $k > d$: *überkritische Kopplung* (d siehe Tab. 97.3). Ist das Dämpfungsmaß $d = 1/Q$ (manchmal auch als Dämpfungsfaktor bezeichnet) beider Schwingkreise unterschiedlich, spricht man anstatt der kritischen Kopplung ($k = d$) auch von *transitionaler Kopplung* und definiert den Kopplungsfaktor k_{tr} wie folgt:

transitionaler Kopplungsfaktor

$$k_{tr} = \sqrt{\frac{1}{2}(d_1^2 + d_2^2)} \quad \text{für kritische Kopplung} \tag{97.6}$$

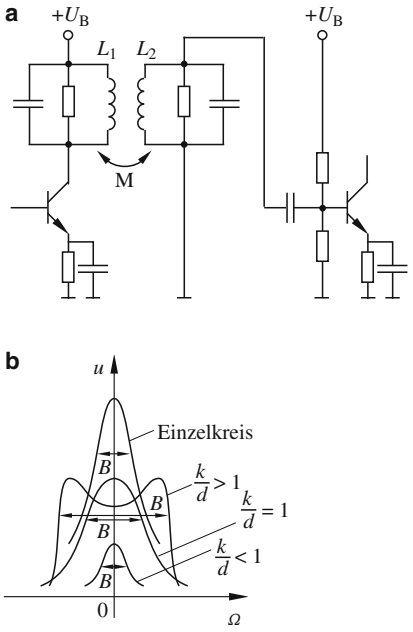


Abb. 97.6 Bandfilter. **a** Realisierung mit Transistorschaltung, **b** Resonanzkurven bei unterschiedlicher Kopplung

In allen anderen Fällen gilt:

Dämpfungsmaß bei Bandfiltern:

$$d = \sqrt{d_1 d_2} \tag{97.7}$$

Normierte Verstimmung Ω: Bei der Betrachtung von Schwingkreisen wird vorteilhaft die *normierte Verstimmung* Ω eingesetzt:

$$\begin{aligned}\Omega &= Q \cdot v = Q \cdot \left(\frac{f}{f_0} - \frac{f_0}{f}\right) = Q \cdot \left(\frac{f^2 - f_0^2}{f \cdot f_0}\right) \\ &\approx Q \cdot \left(\frac{(f + f_0)(f - f_0)}{f_0^2}\right)_{f \approx f_0} \approx Q \cdot \frac{2 \cdot f_0 \cdot \Delta f}{f_0^2} \\ &\approx Q \cdot \frac{2 \cdot \Delta f}{f_0} \approx Q \cdot \frac{B}{f_0} \Rightarrow \Omega = 1 \quad \text{mit } Q \cdot B = f_0 \end{aligned} \tag{97.8}$$

v Verstimmung, f_0 Resonanzfrequenz in Hz.

Für die Spannung am zweiten Schwingkreis errechnet sich:

$$\underline{U} = \frac{\underline{I} \cdot R_p \cdot 2 \cdot \frac{k}{d}}{\sqrt{\left(1 + \frac{k^2}{d^2}\right)^2 - 2\Omega^2\left(\frac{k^2}{d^2} - 1\right) + \Omega^4}}; \quad d = \frac{1}{Q} \tag{97.9}$$

Abb. 97.6b stellt die unterschiedlichen Filterkurven einander gegenüber. Ein Einzelkreis hat zwar die höhere Spannung gegenüber einem Bandfilter, aber dessen Bandbreite ist bei unterkritischer Kopplung kleiner. Die Bandbreite nimmt bei überkritischer Kopplung gegenüber dem Einzelkreis zu. Die normierte Verstimmung berechnet sich jeweils dadurch, dass die Spannung in (97.9) auf den Wert 0,707 gegenüber dem Maximalwert abgesunken ist.

Bandbreite bei einem einzelnen Schwingkreis:

$$\Omega_{\text{ein}} = 1 \quad \Rightarrow \quad B = f_0/Q \tag{97.10}$$

Bandbreite bei unterkritischer Kopplung ($k/d < 1$): Mit (97.9) und einigen Umformungen ergibt sich:

$$\Omega_{\text{un}} = \sqrt{\frac{k^2}{d^2} - 1 + \sqrt{2\left(1 + \frac{k^4}{d^4}\right)}} \approx 0{,}64$$

$$\text{mit } \frac{k^2}{d^2} \ll 1, \ \frac{k^4}{d^4} \ll 1 \quad \Rightarrow \quad B_{\text{un}} \approx 0{,}64 \cdot f_0/Q \tag{97.11}$$

Die Bandbreite bei unterkritischer Kopplung (Näherungen nach (97.11) müssen erfüllt sein) ist also nur 64 % derjenigen eines einzelnen Schwingkreises.

Bandbreite bei kritischer Kopplung ($k/d = 1$):

$$\Omega_{\text{kri}} = \sqrt{\sqrt{2 \cdot 2}} = 1{,}41 \quad \Rightarrow \quad B_{\text{kri}} = 1{,}41 \cdot f_0/Q \tag{97.12}$$

Die Bandbreite ist um den Faktor 1,41 größer als bei einem Einzelkreis.

Bandbreite bei überkritischer Kopplung ($k/d > 1$):

$$\Omega_{\text{üb}} = \pm\sqrt{\frac{k^2}{d^2} + 2\frac{k}{d} - 1} \tag{97.13}$$

Allerdings darf der Sattelpunkt bei $\Omega = 0$ nicht unter dem Wert 0,707 · Maximalwert liegen, weil man sonst nicht mehr von Bandbreite sprechen kann. Das wird erreicht für $k/d \approx 2{,}4$, sodass folgt:

$$\Omega_{\text{üb max}} \approx 3 \quad \Rightarrow \quad B_{\text{üb max}} \approx 3 \cdot f_0/Q \tag{97.14}$$

Wenn der Sattelpunkt also −3 dB unter dem Maximalwert liegt, ist die Bandbreite etwa dreimal so groß wie bei einem Einzelkreis (Abb. 97.6b).

97.2 Weitere Ausführungsformen

Der technische Fortschritt hat durch die Ergebnisse der Digitaltechnik, durch die Fertigung Integrierter Schaltkreise und durch die Umsetzung theoretischer Ergebnisse in die Praxis zu neuen Filterprinzipien geführt. Einige davon sollen in ihrer grundsätzlichen Wirkungsweise vorgestellt werden. Für die Entwicklung von Filterschaltungen ist eine weitergehende Einarbeitung erforderlich.

97.2.1 Quarzfilter, keramische Filter

Quarzfilter und keramische Filter nutzen die Eigenschaften eines Quarzes als Reihen- oder Parallelschwingkreis. Bei keramischen Filtern sind neben den Quarzen noch *L*-*C*-Schwingkreise vorhanden. Quarze haben zwei Resonanzfrequenzen, eine Reihenresonanzfrequenz $f_r = 1/2\pi\sqrt{LC_r}$ (Abb. 97.7a) und eine

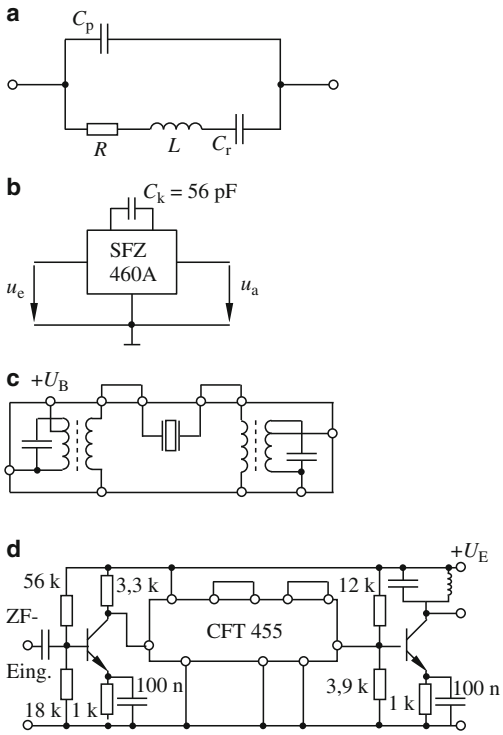


Abb. 97.7 Quarzfilter, keramische Filter. **a** Ersatzschaltbild eines Quarzes, **b** Quarzfilter (Fa. Stettner), **c** Keramikfilter, Innenaufbau (Fa. Murata), **d** Schaltungsauszug mit Filter nach **c**

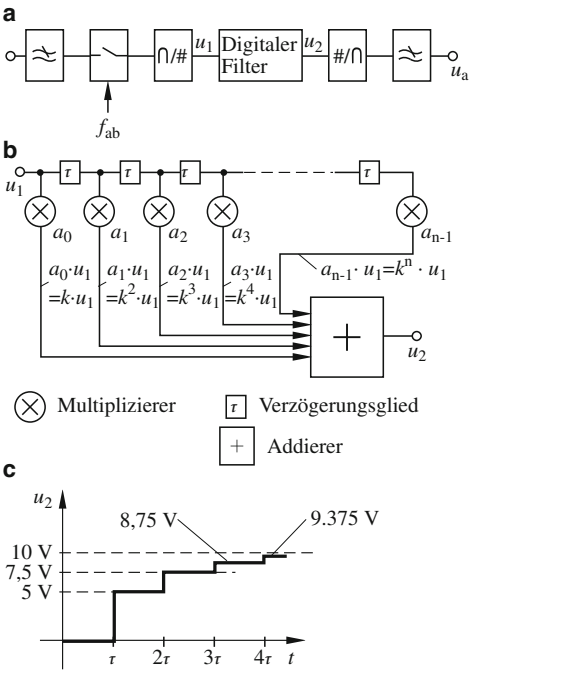


Abb. 97.8 Prinzip digitaler Filter am Beispiel eines Tiefpasses. **a** Filter mit Peripherie, **b** Digitales Tiefpassfilter, nichtrekursiv, **c** Verlauf der Ausgangsspannung

Parallelresonanzfrequenz $f_p = 1/2\pi\sqrt{L \cdot C_r \cdot C_p/(C_r + C_p)}$, die sehr dicht beieinander liegen. Für den Quarz SQ 4815 (Fa. Valvo) gilt für die in Abb. 97.7a eingetragenen Größen: $L = 100\,\text{mH}$, $C_r = 0{,}015\,\text{pF}$, $C_p = 5\,\text{pF}$, $R = 100\,\Omega$, $f_r = 4\,\text{MHz}$. Das ergibt nach Tab. 97.3 für den Quarz in Reihenresonanz eine Güte $Q = (1/R)\sqrt{L/C_r} = 26.000$ und eine Bandbreite $B = 154\,\text{Hz}$. Die hohe Güte von Quarzen bewirkt eine relativ geringe Bandbreite, die häufig zur Signalübertragung nicht ausreicht. In einem Quarzfilter werden in der Regel mehrere Quarz-Schwingkreise nach der Art eines Bandfilters so zusammengeschaltet, dass für eine vorgegebene Frequenz die Impedanz besonders groß ist. Die Resonanzfrequenzen der Einzelquarze werden geringfügig gegeneinander verschoben, um nach Abb. 97.5a eine ausreichende Bandbreite zu erhalten. So ergeben sich für das Bandfilter F 480 (Fa. Quarzkeramik) folgende Werte: Resonanzfrequenz 142,5 kHz; Bandbreite 1440 Hz; bei einer Frequenz $(142{,}5 \pm 2{,}5)\,\text{kHz}$ ist die Ausgangsspannung auf 1/10.000 derjenigen bei der Resonanzfrequenz abgesunken.

Für das in Abb. 97.7b dargestellte Quarzfilter SFZ 460 A (Fa. Stettner) werden bei einem Koppelkondensator $C_k = 56\,\text{pF}$ folgende Daten angegeben: Mitten-(Resonanz-)Frequenz 460 kHz, Bandbreite 4,5 kHz, bei einer Eingangsfrequenz von $(460 \pm 10)\,\text{kHz}$ sinkt die Ausgangsspannung auf typisch 0,056 des Wertes bei der Mittenfrequenz 460 kHz. Über den Wert des Koppelkondensators kann nach Abschn. 97.1.2.5 und Abb. 97.6 die Bandbreite in gewissen Grenzen eingestellt werden. Dieses Filter ist für den Einsatz in Kurzwellenempfängern bei Amplitudenmodulation (AM) und mit einer Zwischenfrequenz von 460 kHz entwickelt worden.

Die Bandbreite von 4,5 kHz eignet sich für einseitenbandmodulierte Trägerschwingungen mit einer Nutzsignalbandbreite von 4,5 kHz oder für zweiseitenbandmodulierte Trägerschwingungen mit einer dann für Sprachübertragung ausreichenden Signalbandbreite von 2,25 kHz. Abb. 97.7c, d zeigt die Innenschaltung des Keramikfilters vom Typ CFT 455 (Fa. Murata) und dessen Einsatz in einer elektronischen Schaltung. Die Anordnung der Anschlüsse ist in beiden Bildern gleich. Diese Filter sind z. B. mit 15 Resonatoren erhältlich (Typ CFS 455). Die Bandbreite kann ausgewählt werden, sie wird durch einen angehängten Buchstaben gekennzeichnet, z. B. CFT 455 *F*, *F* bedeutet 8,4 kHz. Die Zahl der Quarzresonatoren legt die Steilheit der Filterkurve außerhalb der Mittenfrequenz fest, während über die Koppelkondensatoren (Abb. 97.7b, c) bzw. die Schwingkreise die Bandbreite eingestellt wird.

97.2.2 Digitale Filter

Die Theorie digitaler Filter würde den Rahmen dieses Buches sprengen, deshalb soll die prinzipielle Wirkungsweise nach Abb. 97.8 am Beispiel eines Tiefpassfilters erläutert werden. Abb. 97.8a stellt das Filter mit zugehöriger Peripherie dar. Das zu filternde Signal mit der Spannung u_e wird nach dem Durchgang durch ein Antialiasing-Filter zur Frequenzbandbegrenzung mit der Frequenz f_{ab} abgetastet und in eine digital

kodierte Spannung u_1 umgesetzt. Diese gelangt auf das digitale Filter, dessen Prinzip im Abb. 97.8b gezeigt ist. Die Ausgangsspannung u_2 des Filters wird wieder in einen Analogwert umgesetzt und bei Bedarf von eventuell vorhandenen, durch die Digital-Analog-Umsetzung hervorgerufenen höherfrequenten Signalanteilen, befreit und steht als u_a zur Verfügung.

Zur Erläuterung der Wirkungsweise des digitalen Filters nach Abb. 97.8b soll die Spannung u_1 für $t < 0$ den Wert 0 haben und zum Zeitpunkt $t = 0$ auf den Wert U_0 springen. Sie gelangt auf eine Laufzeitkette mit den Einzelverzögerungszeiten τ. Die Ausgangsspannung u_2 wird durch Summation der Ausgangsspannungen der Verzögerungsglieder gebildet, nachdem diese noch mit den Faktoren a_n multipliziert worden sind. Für das Beispiel gilt: $k = 0{,}5$; $a_0 = k = 0{,}5$; $a_1 = k^2 = 0{,}5^2 = 0{,}25$; $a_2 = k^3 = 0{,}5^3 = 0{,}125$ usw.; $u_1 = U_0 = 10\,\text{V}$. Dann gilt für die Ausgangsspannung u_2:

$$
\begin{aligned}
t = 0 \ldots \tau: \quad & u_{2,0} = a_0 \cdot u_1 = k \cdot u_1 = 0{,}5 \cdot u_1 = 5\,\text{V} \\
t = \tau \ldots 2\tau: \quad & u_{2,\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 \\
& = k \cdot u_1 + k^2 \cdot u_1 = 5\,\text{V} + 2{,}5\,\text{V} = 7{,}5\,\text{V} \\
t = 2\tau \ldots 3\tau: \quad & u_{2,2\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 + a_2 \cdot u_1 \\
& = k \cdot u_1 + k^2 \cdot u_1 + k^3 \cdot u_1 = 8{,}75\,\text{V} \\
t = 3\tau \ldots 4\tau: \quad & u_{2,3\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 + a_2 \cdot u_1 + a_3 \cdot u_1 \\
& = k \cdot u_1 + k^2 \cdot u_1 + k^3 \cdot u_1 + k^4 \cdot u_1 \\
& = 9{,}375\,\text{V}
\end{aligned}
$$

In Abb. 97.8c ist der Verlauf der Spannung u_2 eingetragen. Sind theoretisch unendlich viele Verzögerungsglieder vorhanden, geht die Spannung u_2 gegen den Wert 10 V, wenn man die Formel $\sum_{k=1}^{\infty}(\frac{1}{2})^k = 1$ verwendet. Der in Abb. 97.8c gezeigte Verlauf der Ausgangsspannung ist der digital codierte Verlauf der Funktion $u_2 = 10\,\text{V}(1 - \exp(t/T))$. Er entspricht damit der Sprungantwort eines *Tiefpassfilters*. Über die Verzögerungszeit τ wird die Grenzfrequenz des Filters eingestellt. In der Praxis wird man den idealen Verlauf durch eine endliche Anzahl von Verzögerungsgliedern annähern.

Durch entsprechende Wahl der Koeffizienten $a_0, a_1, a_2, \ldots$ lassen sich verschiedene Filtercharakteristiken einstellen. Hierin liegt auch die große Bedeutung der digitalen Filter. Soll beispielsweise das soeben dargestellte Tiefpassfilter in ein Hochpassfilter umprogrammiert werden, sind die Koeffizienten $a_0, a_1, a_2, \ldots$ folgendermaßen einzustellen, wenn $k = 0{,}5$ unverändert übernommen wird: $a_0 = +k^0 = 1$; $a_1 = -k^1 = -0{,}5$; $a_2 = -k^2 = -0{,}5^2 = -0{,}25$; $a_3 = -k^3 = -0{,}5^3 = -0{,}125$; usw. Damit ergibt sich für u_2:

$$
\begin{aligned}
t = 0 \ldots \tau: \quad & u_{2,0} = a_0 \cdot u_1 = k \cdot u_1 = 1 \cdot u_1 = 10\,\text{V} \\
t = \tau \ldots 2\tau: \quad & u_{2,\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 \\
& = 1 \cdot u_1 - k \cdot u_1 = 10\,\text{V} - 5\,\text{V} = 5\,\text{V} \\
t = 2\tau \ldots 3\tau: \quad & u_{2,2\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 + a_2 \cdot u_1 \\
& = 1 \cdot u_1 - k \cdot u_1 - k^2 \cdot u_1 \\
& = 10\,\text{V} - 5\,\text{V} - 2{,}5\,\text{V} = 2{,}5\,\text{V} \\
t = 3\tau \ldots 4\tau: \quad & u_{2,3\tau} = a_0 \cdot u_1 + a_1 \cdot u_1 + a_2 \cdot u_1 + a_3 \cdot u_1 \\
& = 1 \cdot u_1 - k \cdot u_1 - k^2 \cdot u_1 - k^3 \cdot u_1 \\
& = 1{,}25\,\text{V}
\end{aligned}
$$

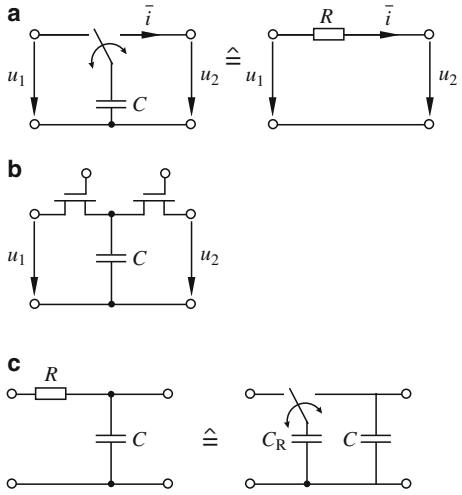


Abb. 97.9 Filter mit geschalteten Kapazitäten (SC-Filter). **a** Darstellung eines Widerstandes R, **b** Realisierung in CMOS-Technologie, **c** Tiefpass erster Ordnung

Die Spannung u_2 verläuft gemäß der Sprungantwort eines Hochpassfilters.

Die Operationen „Multiplizieren“ und „Summieren“ der Digitalwerte nach Abb. 97.8b geschehen in einem Rechner, bei dem man durch entsprechende Programmierung die Filtercharakteristik wählen bzw. ändern kann. Außerdem sind die Filtereigenschaften nicht von Bauteiletoleranzen bzw. -änderungen abhängig. Für die Nachrichtentechnik sind diese Filter nur begrenzt verwendbar, weil bei hohen Signalfrequenzen die erforderliche Anzahl von Rechenoperationen zeitlich nicht mehr durchführbar ist.

Folgende zwei Arten von Filtern werden unterschieden:

- *Nichtrekursive Filter*: Die Ausgangsspannung u_2 wird *nicht* auf das Filter zurückgeführt.
- *Rekursive Filter*: Die Ausgangsspannung u_2 gelangt auf eine zweite Laufzeitkette, deren Ausgangsspannungen auf unterschiedliche Punkte innerhalb des Filters zurückgeführt werden. Auf diese Weise lassen sich sehr komplexe Filterfunktionen erreichen.

97.2.3 Filter mit geschalteten Kondensatoren, SC-Filter

Das Prinzip dieser Filter (engl. *s*witched *c*apacitor filter) ist in Abb. 97.9a dargestellt.

Der Schalter schaltet periodisch zwischen den beiden Stellungen hin- und her mit einer Frequenz f ($f = 1/T$), die wesentlich größer ist als die höchste Signalfrequenz der Spannung u_1. Auf dem Kondensator entstehen Ladungsänderungen $\Delta Q = C(u_1 - u_2)$, die einen mittleren Strom $\bar{i} = \Delta Q/T = C(u_1 - u_2)/T$ vom Eingang zum Ausgang zur Folge haben. Dieser Strom entspricht einem Widerstand R der Größe $R = (u_1 - u_2)/\bar{i} = T/C$.

Tab. 97.4 Auswahl industriell gefertigter SC-Filter

Typ	Hersteller	Filtertyp	Ordnung	Taktfrequenz
MAX 262	Maxim	Universalfilter	2 · 2ter Ordnung	40 Hz … 4 MHz
MF 10	National	Universalfilter	2 · 2ter Ordnung	10 Hz … 1 MHz
CS 7008	Crystal	Universalfilter	8ter Ordnung	Max. 1 MHz
XR 1003	Exar	Bessel-Tiefpass	4ter Ordnung	100 Hz … 1 MHz
MF 6	National	Butterworth-Tiefpass	6ter Ordnung	Max. 250 kHz
TSG 8532	Thomson	Tschebyscheff-Hochpass	6ter Ordnung	5 kHz … 1,8 MHz
MF 8	National	Tschebyscheff/Butterworth-Bandpass	2 · 4ter Ordnung	100 Hz … 20 kHz
RM 5614	Reticon	Tschebyscheff-Bandpass	6ter Ordnung	60 Hz … 1 MHz
LMF 90	National	Cauer-Bandsperre	4ter Ordnung	10 Hz … 3 MHz

Bei einem vorgegebenen Kondensator C kann der Wert von R also über die Periodendauer T eingestellt werden. Schaltungstechnisch realisiert werden diese Filter in Form von Integrierten CMOS-Schaltungen, die aus Schaltertransistoren, Verstärkern und Kapazitäten bestehen (Abb. 97.9b). Abb. 97.9c zeigt die Realisierung eines Tiefpassfilters erster Ordnung. Der Widerstand R wird durch den geschalteten Kondensator C_R dargestellt. Die Zeitkonstante ergibt sich damit zu: $\tau = C \cdot R = C \cdot T/C_R$. Daraus folgen zwei Vorteile dieser Filter:

1. Die Zeitkonstante hängt vom Kapazitäts*verhältnis* ab, nicht von den Einzelwerten.
2. Die Zeitkonstante kann über die Periodendauer T des Taktes eingestellt werden. Tab. 97.4 gibt eine kleine Auswahl von industriell gefertigten Filtertypen. Aus der Taktfrequenz erhält man für Tiefpassfilter eine Größenordnung für die Grenzfrequenz, wenn man sie durch 100 teilt.

97.3 Allgemeiner Filterentwurf

Die Anforderungen an die Eigenschaften von Filterschaltungen werden wesentlich von der Anwendung bestimmt, sodass die im Abschn. 97.1 beschriebenen passiven *R*-*C*-Filter, bei denen die Widerstandswerte und die Kapazitätswerte jeweils gleich sind, nicht immer die optimale Lösung darstellen. Die Betrachtungen werden auch auf nichtsinusförmige periodische und einmalige Signale ausgedehnt. Deshalb folgt ein erster Überblick über die Grundlagen der Filterentwurfstheorie.

97.3.1 Allgemeine Eigenschaften und Entwurfskriterien

Für den Einsatz von Filterschaltungen gibt es vielfältige Anwendungen. Die Auswahl einer geeigneten Filterschaltung und deren Dimensionierung sind aber relativ aufwändig, weil sie eingehende Kenntnisse der Filtertheorie und praktische Erfahrung erfordern. Deshalb sollen hier nur einige prinzipielle Überlegungen bei der Auswahl einer Filterschaltung dargestellt werden, und zwar bevorzugt am Beispiel von Tiefpassfiltern. Daneben gibt es noch Hochpassfilter, Bandpassfilter und Bandsperren (Kurzform: Tiefpass, Hochpass, Bandpass, Bandsperre). Auch wenn es für den Anwender u. U. ungewohnt ist, sollte er den Einsatz von Filter-Entwurfsschaltungen und Simulationsschaltungen in Erwägung ziehen. DIE optimale Filterschaltung gibt es in der Regel nicht, vielmehr muss aus dem Spektrum der möglichen Filterrealisierungen der für den speziellen Anwendungsfall geeignete Kompromiss ausgesucht werden.

Es gibt folgende Unterteilung:

1. Passive Filter: Sie bestehen aus Kondensatoren, Spulen und ohmschen Widerständen.
2. Aktive Filter: Sie setzen Operationsverstärker, Kondensatoren und ohmsche Widerstände ein. Spulen finden Verwendung, wenn die reale Spule (R und L in Reihe) ohne Eisenkern eingesetzt wird. Aktive Filter haben den Vorteil, dass sie mit weniger Bauteileaufwand die gleichen Ergebnisse erreichen wie passive Filter, allerdings müssen die Bauteilewerte je nach Schaltung evtl. mit (deutlich) weniger als $\pm 1\,\%$ Abweichung vom errechneten Wert eingesetzt werden. Das liegt zum Teil daran, dass eine bestimmte Filtereigenschaft um so optimaler wird, je mehr man sich durch die Dimensionierung der Bauteile einer Nullstelle im Nennerpolynom des Frequenzganges oder einem kritischen Phasenwinkel des Phasenganges nähert. Ein Erreichen der kritischen Werte führt häufig zur Instabilität des Filters und zum Schwingen.

Die Betrachtungen erfolgen sowohl in der Zeit- als auch in der Frequenzebene. Das gegebene Signal, das periodisch oder einmalig sein kann, ist nach Fourier zusammengesetzt aus sinusförmigen Signalen, deren Frequenzen, Amplituden und Phasenverschiebungen berechnet werden können. Es treten folgende Probleme auf:

1. Häufig überstreichen die im gegebenen Signal enthaltenen Signalfrequenzen einen Bereich, der nicht zur Verfügung gestellt werden kann oder muss. Beispiele: Bei der Fernsehbild-Übertragung steht jedem Sender nur ein begrenzter Frequenzbereich zur Verfügung; bei der Rundfunkübertragung können evtl. Frequenzanteile oberhalb des Hörbereiches des menschlichen Ohres unterdrückt werden. Anwendung: Tiefpassfilter.
2. Im gegebenen Signal sind unerwünschte Frequenzanteile enthalten. Beispiele: a) 50-Hz-Netzstörungen in Niederfrequenzverstärkern; b) Frequenzanteile oberhalb der halben Abtastfrequenz bei der Digitalisierung von Analogsignalen (Aliasing-Effekt). Anwendung: Tiefpassfilter, Bandsperre.

3. Aus einem gegebenen Signal sollen bestimmte charakteristische Frequenzanteile herausgefiltert und weiterverarbeitet werden. Beispiel: Schadenfrüherkennung in Kugellagern durch Änderung des Lagergeräusches bei der Beschädigung von Kugeln. Anwendung: Bandpassfilter.
4. In der Unterhaltungselektronik gibt es Niederfrequenzverstärker, bei denen die tiefen Töne oder die hohen Töne oder aber bestimmte Tonbereiche innerhalb des Hörbereiches abgeschwächt oder bevorzugt werden können. Anwendung: Tiefpass-, Hochpass-, Bandpassfilter und Bandsperren.
5. In der Regelungstechnik sollen häufig *Änderungen* einer Größe möglichst unverzögert und signalgetreu erkannt werden, während ein stationärer Zustand weniger interessant ist. Anwendung: Hochpassfilter (D-Regler, siehe Regelungstechnik).

In der Praxis hat sich gezeigt, dass die Begriffe Tiefpass, Hochpass, Bandpass und Bandsperre für die Anwendung zu wenig aussagekräftig sind. Vielmehr sind die folgenden Filtereigenschaften, die für alle Filtertypen gleichsam gelten, für den Anwender häufig von wesentlich größerer Aussagekraft:

1. Übertragungsfunktion, Amplitudenfrequenzgang

Angegeben wird das Verhältnis von Ausgangs- zu Eingangsamplitude in Abhängigkeit von der Frequenz. In der Regel bezieht man sich bei den Amplituden auf Spannungen. Zwei Ausdrücke sind von Bedeutung (In der Literatur werden die Begriffe „Übertragungsfunktion“, „Frequenzgang“ und „Amplitudenfrequenzgang“ evtl. abweichend verwendet.):

Übertragungsfunktion:

$$\frac{U_a(s)}{U_e(s)}; \tag{97.15}$$

Frequenzgang:

$$\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}} \quad \text{oder} \quad \frac{\underline{U}_a}{\underline{U}_e} \tag{97.16a}$$

(beides wird verwendet);

Amplitudenfrequenzgang:

$$\left|\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}\right| \quad \text{oder} \quad \left|\frac{\underline{U}_a}{\underline{U}_e}\right| \tag{97.16b}$$

(beides wird verwendet).

$\underline{U}_{\text{ein}}, \underline{U}_e$ Eingangsspannung, $\underline{U}_{\text{aus}}, \underline{U}_a$ Ausgangsspannung → komplexe Ebene; $U_e(s)$ Eingangsspannung, $U_a(s)$ Ausgangsspannung → Laplace-Ebene (s-Ebene).

Der Amplitudenfrequenzgang wird häufig im Bode-Diagramm dargestellt, in dem beide Achsen logarithmisch geteilt sind: Die Frequenzachse dekadisch, die Achse für das Spannungsverhältnis ebenfalls dekadisch, aber mit einer dB-Skalierung, was dann eine lineare Skalenteilung ergibt.

Beim Tiefpass wird z. B. häufig angestrebt, dass alle Frequenzanteile unterhalb der Grenzfrequenz mit dem gleichen Faktor multipliziert am Ausgang erscheinen. Oberhalb der Grenzfrequenz soll dieser Wert möglichst klein sein. Zu beachten ist, dass in jedem Fall eine Frequenzabhängigkeit des Frequenzganges eine Signalverfälschung zwischen Eingangs- und Ausgangssignal zur Folge hat, erklärbar mit der Fouriersynthese von Eingangs- und Ausgangssignal.

2. Signallaufzeit

Das Ausgangssignal ist gegenüber dem Eingangssignal zeitlich verzögert, hervorgerufen durch das Nacheilen der Spannung gegenüber dem Strom beim Kondensator bzw. umgekehrt bei der Spule. Dies kann zu Problemen führen, wenn z. B. bei Fernsehaufnahmen Bild- und Tonsignal diverse unterschiedliche Filter durchlaufen, deren Laufzeiten unterschiedlich sind. Dann sind Bild- und Tonsignal nicht mehr lippensynchron. In Regelsystemen erhöht die Signallaufzeit die Ausregelzeit von Störgrößen. Außerdem können rückgekoppelte Systeme dadurch zum Schwingen angeregt werden.

3. Gruppenlaufzeit

Innerhalb des Durchlassbereiches kann es zu einer weiteren Signalverfälschung kommen, nämlich dann, wenn die Laufzeit der frequenzabhängigen Signalanteile vom Eingang zum Ausgang von der Frequenz abhängt. Ziel ist es, dass alle Frequenzanteile die gleiche Zeit zum Durchlaufen des Filters benötigen, denn nur dann setzen sie sich am Ausgang wieder zum Originalsignal zusammen (Fouriersynthese). Diese „Durchlaufzeit“ wird als Gruppenlaufzeit v_{gr} bezeichnet und lässt sich berechnen als Ableitung des Phasenwinkels φ des Frequenzganges nach der Kreisfrequenz:

$$v_{\text{gr}} = \frac{d(\phi(\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}))}{d\omega} \tag{97.17}$$

Wünschenswert ist:

$$v_{\text{gr}} = \text{const} \tag{97.18}$$

Hinweis: Die Ergebnisse einer Berechnung nach (97.17) sollten grafisch dargestellt und dann interpretiert werden, siehe dazu die Erläuterungen zu Abb. 97.13.

4. Abfall im Sperrbereich

Der Filtergrad bestimmt im Wesentlichen diesen Abfall, siehe Erläuterungen zu (97.20) unten. Die Dimensionierung der Bauelemente hat Einfluss auf den Verlauf im Übergangsbereich vom Durchlass- in den Sperrbereich. Er nähert sich einem „Knick“, wenn man in Kauf nimmt, dass der Amplitudenfrequenzgang innerhalb des Durchlassbereiches oder des Sperrbereiches eine Welligkeit aufweist und nahe der Grenzfrequenz Überschwingen auftritt, siehe dazu Cauer- und Tschebyscheff-Filter weiter unten.

5. Optimierte Reaktion auf vorgegebene Eingangssignale

Ein Filter kann z. B. daraufhin optimiert werden, dass Spannungssprünge am Eingang möglichst schnell den (nahezu) stationären Endwert am Ausgang erreichen, je nach Anwendung mit oder ohne Überschwingen.

6. Wellenwiderstand

Für Anwendungen, vor allem bei höheren Frequenzen, kann es erforderlich sein, dass das Filter einen vorgegebenen Wellenwiderstand besitzt, in Vorwärts- und Rückwärtsrichtung gleich oder unterschiedlich, siehe Kap. 93 „Zweitore".

Für den Anwender werden einige der genannten Kriterien von besonderer Bedeutung sein, während andere in den Hintergrund treten. Im Abschn. 97.3.2 „Tiefpassfilter" weiter unten wird anschaulich gezeigt, dass es zur Realisierung eines Tiefpasses bei gegebener Grenzfrequenz „unendlich viele" Möglichkeiten der Realisierung durch unterschiedliche Arten der Bauteiledimensionierung gibt. Allerdings hängen die oben beschriebenen 6 Filtereigenschaften von dieser Dimensionierung ab. Theoretiker und Praktiker haben sich dieser Problematik angenommen und Bauteiledimensionierungen für bestimmte Anwendungen optimiert: Butterworth, Bessel, Cauer und Tschebyscheff (in der Literatur auch Chebyshev), um nur einige zu nennen.

97.3.2 Tiefpassfilter

Der Frequenzgang eines passiven Tiefpassfilters als Kettenschaltung aus T- und π-Zweitoren mit den Elementen R, L und C kann durch geeignete Umformungen auf die folgende Form gebracht werden:

$$\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}} = \frac{a_0}{1 + a_1 \cdot j \cdot \omega - a_2 \cdot \omega^2 - a_3 \cdot j \cdot \omega^3 + a_4 \cdot \omega^4 + a_5 \cdot j \cdot \omega^5 \mp \ldots} \tag{97.19}$$

Darin ist a_0 die sog. „Gleichspannungsverstärkung". Sie ist bei passiven R-L-C-Filtern ≤ 1, bei aktiven R-L-C-Filtern liegt sie häufig im Bereich von $1/10$ bis 100. Der Grad des Filters ist gleich dem höchsten Exponenten von ω. Für steigende Frequenzen oberhalb der Grenzkreisfrequenz ω_g bestimmt dieser Ausdruck zunehmend das Verhalten des Filters und damit auch die Unterdrückung von Frequenzen oberhalb von ω_g:

$$\left.\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}\right|_{\omega > \omega_g} \approx \frac{a_0}{a_k \cdot j^k \cdot \omega^k} \quad \text{bzw.} \quad \left.\left|\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}}\right|\right|_{\omega > \omega_g} \approx \frac{a_0}{a_k \cdot \omega^k} \tag{97.20}$$

k Grad des Filters, a_0 Gleichspannungsverstärkung.

Bei Verzehnfachung der Frequenz sinkt der Betrag des Ausdruckes auf $1/(10^k)$, und die Steigung der Geraden im Bode-Diagramm beträgt, unabhängig von der Frequenz, $-k \cdot 20\,\text{dB/Dekade}$. Je höher der Grad des Filters ist, desto wirkungsvoller werden die Frequenzen oberhalb ω_g unterdrückt. Die Konstanten $a_1, a_2, a_3, \ldots$ in (97.19) sind dimensionsbehaftete Kombinationen der verwendeten Bauelemente $R_1, R_2, \ldots, L_1, L_2, \ldots, C_1, C_2, \ldots$. Sie sind positiv bei passiven R-L-C-Filtern. Bei aktiven R-L-C-Filtern können sie auch negativ sein, was bei entsprechender Dimensionierung Nullstellen in der positiven komplexen Ebene zur Folge hat und damit eine Schwingneigung des Filters gegeben ist. Es lässt sich zeigen, dass eine gegebene Filterschaltung genau *eine* Übertragungsfunktion (*einen* Frequenzgang) hat, dass aber eine gegebene Übertragungsfunktion (Frequenzgang) durch „unendlich viele" Filterschaltungen realisiert werden kann. Daraus folgt, dass es für die Lösung von Filterschaltungen allgemein und damit auch von den hier betrachteten Tiefpassfiltern „unendlich viele" Möglichkeiten der Realisierung gibt. Anschaulich lässt sich das dadurch erklären, dass sich z. B. bei einem passiven R-C-Tiefpass zweiter Ordnung, gebildet aus 2 Widerständen R_1, R_2 und 2 Kondensatoren C_1, C_2, der Frequenzgang

$$\frac{\underline{U}_{\text{aus}}}{\underline{U}_{\text{ein}}} = \frac{1}{1 + j \cdot \omega \cdot \underbrace{(R_1 \cdot C_1 + R_1 \cdot C_2 + R_2 \cdot C_2)}_{a_1} - \omega^2 \cdot \underbrace{(R_1 \cdot C_1 \cdot R_2 \cdot C_2)}_{a_2}} \tag{97.21}$$

ergibt. Die Grenzkreisfrequenz ω_g wird durch den Ausdruck $\frac{1}{a_2} = \frac{1}{R_1 \cdot C_1 \cdot R_2 \cdot C_2} = \omega_g^2$, also das *Produkt* aller 4 Bauelemente, bestimmt; das Verhalten innerhalb des Durchlassbereiches wesentlich durch den Ausdruck a_1. Damit stehen zur Berechnung der 2 Größen a_1 und a_2 mehr als 2 Variablen zur Verfügung. Dies gilt, verallgemeinert, für Filterschaltungen beliebiger Ordnung. Beim Entwurf eines Filters wählt man die Bauelemente zahlenmäßig so, dass die gewünschten Kriterien mehr oder weniger gut erfüllt werden. Häufig gilt, dass mit steigendem Filtergrad auch gewisse Eigenschaften optimaler werden. Allerdings muss beachtet werden, dass die Bauteiledimensionierung in immer engeren Grenzen verläuft, weil man sich evtl. in der Nähe von kritischen Punkten der Übertragungsfunktion (des Frequenzganges) befindet (z. B. Nullstelle des Nennerpolynoms, kritischer Phasenwinkel).

Beim Filterentwurf haben sich ausgewählte Dimensionierungen für bestimmte Anwendungen als besonders vorteilhaft erwiesen. Sie sind meist nach der Person benannt, die sie erstmalig eingesetzt hat. Tab. 97.5 zeigt einige Beispiele.

In den Abb. 97.10 bis 97.14 sind die in Tab. 97.5 aufgeführten Filtertypen in ihrem Verhalten in Abhängigkeit von der Frequenz gezeigt. Es handelt sich um passive R-L-C-Tiefpass-Filter fünften Grades mit einer sinusförmigen Eingangsspannung $|\underline{U}_e| = 10\,\text{V}$. Der Wellenwiderstand beträgt eingangsseitig (Vorwärtsrichtung) und ausgangsseitig (Rückwärtsrichtung) jeweils $50\,\Omega$.

Abb. 97.10 zeigt das Schaltbild mit der Dimensionierung der Bauteile.

Abb. 97.11 stellt die Ausgangsspannungen $|\underline{U}_a|$ der 4 Filter dar. Die in der Tab. 97.5 aufgeführten Eigenschaften sind anschaulich nachvollziehbar. Die Gruppenlaufzeiten v_{gr} der Filter nach Abb. 97.10 zeigen Abb. 97.12 und 97.13.

Im Abb. 97.13 wurde die Skalierung gegenüber der im Abb. 97.12 geändert. Bei der Darstellung der Gruppenlaufzeit des Cauer-Filters in Abb. 97.13 zeigt sich eine gewisse Problematik bei der Auswertung (97.17). Das Ergebnis ist die (Lauf-) Zeit, die sich aus dem Phasenwinkel zwischen Eingangs- und Ausgangsspannung bei bekannter Periodendauer und sinusförmigem Signal bestimmen lässt. Aber es wird nicht die Periodizität des Phasenwinkels von $2 \cdot \pi$ berücksichtigt. Gemäß Formel

Tab. 97.5 Eigenschaften häufig eingesetzter Filtertypen (Auswahl)

Name des Filters	Eigenschaften Tiefpassfilter Diese Eigenschaften sind sinngemäß auf Hochpass, Bandpass und Bandsperre übertragbar
Bessel-Filter	Konstante Gruppenlaufzeit. Die Signalform wird nicht geändert, sofern man nur die Frequenzanteile im Durchlassbereich berücksichtigt. Übergang in den Sperrbereich gleitend (nicht scharf abgeknickt). Nur geringes Überschwingen bei der Sprungantwort
Butterworth-Filter	Amplitudenfrequenzgang nahezu konstant im Durchlassbereich. Relativ scharfes Abknicken des Amplitudenfrequenzganges bei der Grenzfrequenz (relativ abrupter Übergang vom waagerechten Verlauf in die Gerade mit der Steigung $-n \cdot (20\,\mathrm{dB/Dek})$), nimmt zu mit wachsendem Filtergrad. Starkes Überschwingen bei der Sprungantwort, nimmt zu mit wachsendem Filtergrad
Cauer-Filter oder elliptisches Filter	Optimiert auf sehr scharfes Abknicken des Amplitudenfrequenzganges bei der Grenzfrequenz (sehr abrupter Übergang vom waagerechten Verlauf in die Gerade mit der Steigung $-n \cdot (20\,\mathrm{dB/Dek})$). Der Amplitudenfrequenzgang besitzt sowohl im Durchlass- als auch im Sperrbereich eine ausgeprägte Welligkeit (oszillierendes Verhalten). Durch ausgeprägte Phasenverzerrungen und damit verbunden große Gruppenlaufzeitänderungen treten starke Signalverzerrungen auf
Tschebyscheff-Filter Typ 1 Typ 2 (auch inverses Filter genannt)	Optimiert auf relativ scharfes Abknicken des Amplitudenfrequenzganges bei der Grenzfrequenz. Abknicken umso abrupter, je größer der Filtergrad. Dadurch bedingt tritt Welligkeit (oszillierendes Verhalten) entweder im Durchlassbereich (Typ 1) oder im Sperrbereich (Typ 2) auf. Geht die Welligkeit durch entsprechende Dimensionierung gegen Null, geht dieses Filter in das Butterworth-Filter über. Starkes Überschwingen bei der Sprungantwort, nimmt zu mit zunehmendem Filtergrad. Keine konstante Gruppenlaufzeit und damit Signalverzerrungen

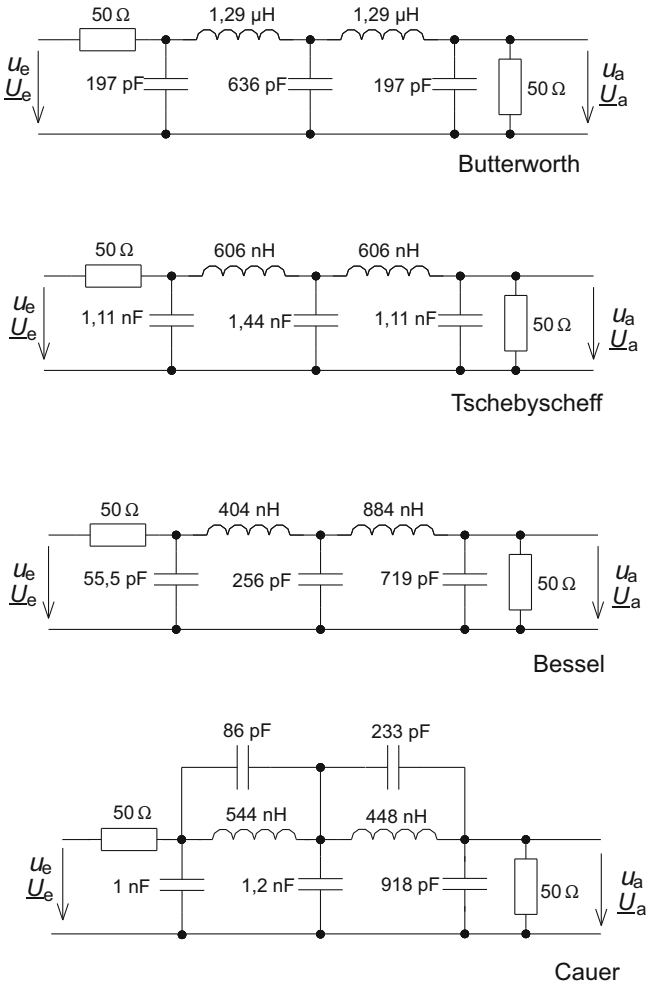


Abb. 97.10 Schaltbild Bessel-, Butterworth-, Cauer- und Tschebyscheff-Filter, weitere Einzelheiten siehe Text

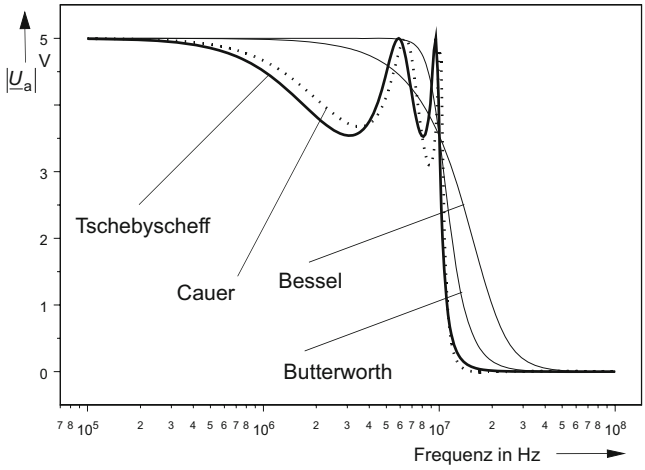


Abb. 97.11 Ausgangsspannungen der 4 Filter nach Abb. 97.10

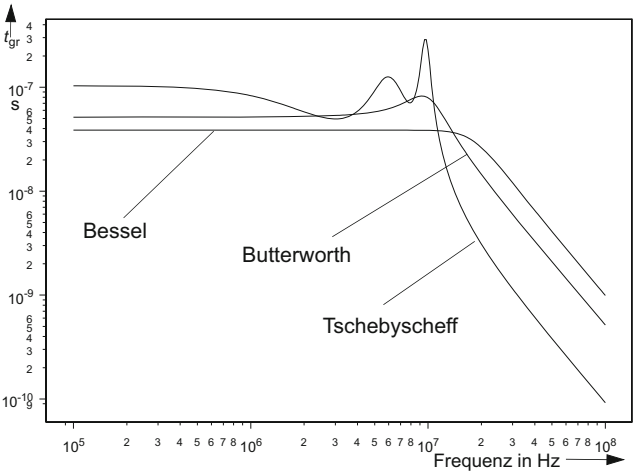


Abb. 97.12 Gruppenlaufzeit von Bessel-, Butterworth- und Tschebyscheff-Filter nach Abb. 97.10

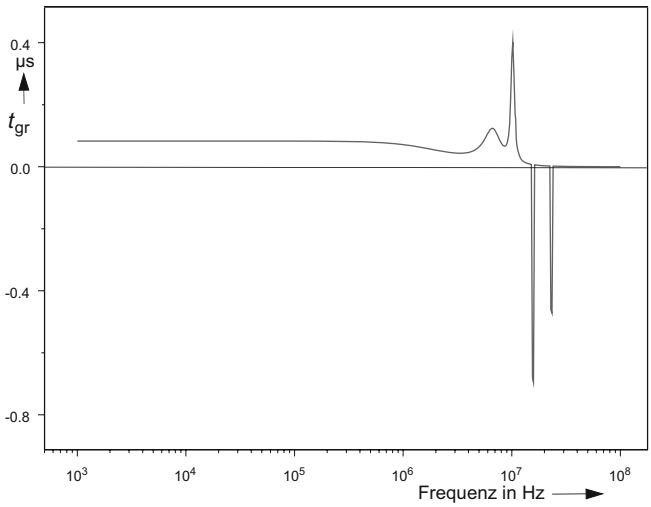


Abb. 97.13 Gruppenlaufzeit des Cauer-Filters nach Abb. 97.10; Hinweise siehe Text

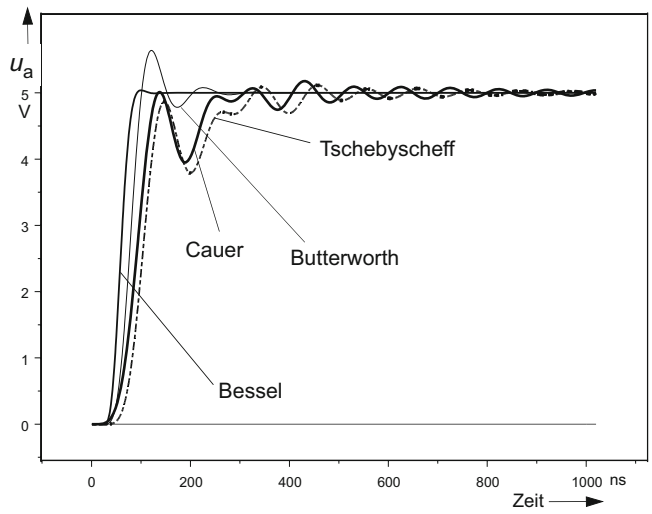


Abb. 97.14 Sprungantwort der 4 Filter nach Abb. 97.10

werden nur die *Änderungen* erfasst, was zu positiven (Zunahme des Phasenwinkels) und negativen (Abnahme des Phasenwinkels) Laufzeitwerten führt. Man kann die Laufzeit je nach Sichtweise durchaus positiv oder negativ interpretieren, aber beide Vorzeichen zusammen in einem Bild sind physikalisch schwierig zu interpretieren, denn dann eilt die Ausgangsspannung in einigen Bereichen der Eingangsspannung voraus. Die weiter unten genannten Simulationsprogramme stellen deshalb z. B. folgendes dar:

$$v_{gr} = -\frac{d(\phi(\frac{\underline{U}_{aus}}{\underline{U}_{ein}}))}{d\omega} \quad \text{bei „PSpice“ oder}$$

$$v_{gr} = \left|\frac{d(\phi(\frac{\underline{U}_{aus}}{\underline{U}_{ein}}))}{d\omega}\right| \quad \text{bei „AADE Filter Design“}$$

Es ist gegebenenfalls sinnvoll, sich zusätzlich zur Gruppenlaufzeit den Verlauf des Phasenwinkels, $\varphi(\underline{U}_{aus}/\underline{U}_{ein})$, anzeigen zu lassen.

Auch hier sind die in Tab. 97.5 aufgeführten Eigenschaften anschaulich nachvollziehen.

Wenn Filter entworfen werden sollen, sind die eingangs angesprochenen Programme sehr hilfreich und führen schnell zum Ziel. Hier folgen dazu einige nähere Angaben. Über eine Internet-Recherche kann man weitere Programme finden, von denen es häufig Probe- oder Testversionen gibt, die man testen und dann entscheiden kann, welches Programm den persönlichen Ansprüchen und Vorstellungen entspricht. Ohne Anspruch auf Vollständigkeit und Rechteverwertung seien hier 4 Beispiele genannt (alle Programme in englischer Sprache):

1. Simulationsprogramme für elektronische Schaltungen auf der Grundlage von Spice/PSpice:
   1.1 „OrCAD-Capture“ über Fa. Cadence. Internetadresse: www.flowcad.de. Eine Studentenversion mit eingeschränktem Funktionsumfang ist kostenlos erhältlich, die uneingeschränkte Vollversion ist kostenpflichtig.
   1.2 „LTspice“ über Fa. Linear Technology. Internetadresse: www.linear.com. Die uneingeschränkte Vollversion ist kostenlos.
   Hinweis: Beide Programme ähneln sich in der Bedienung und im Funktionsumfang, weil sie den gleichen Programmkern verwenden.
2. Filter-Entwurfsprogramme:
   Beide Programme sind so angelegt, dass u. a. vorgewählt werden können:
   Filterart (Tiefpass, Hochpass, Bandpass, Bandsperre), Filtergrad, Filtertyp (nach Tab. 97.5 sowie noch weitere Filtertypen), Grenzfrequenz, Welligkeit und Wellenwiderstand. Gemäß diesen Vorgaben wird das Filter dimensioniert, es erscheint das Schaltbild mit den Bauteilewerten.
   In Diagrammen lassen sich u. a. darstellen: Amplitudenfrequenzgang, Phasengang, Gruppenlaufzeit und Sprungantwort. Jedes Programm hat darüber hinaus noch spezielle Funktionen eingebaut.
   2.1 „Filter Free“ über Fa. Nuhertz Technologies. Internetadresse: www.nuhertz.com. Die eingeschränkte Testversion für Filter bis dritten Grades ist kostenlos, die uneingeschränkte Vollversion ist kostenpflichtig.
   2.2 „AADE Filter Design“. Internetadresse: www.aade.com. Das Programm ist kostenlos.

97.3.3 Filtertransformationen

Sollen Hochpässe, Bandpässe oder Bandsperren entworfen werden, können die Ergebnisse der Tiefpässe auf diese Filter durch eine Frequenztransformation übertragen werden.

Im allgemeinen bestehen passive Filterschaltungen je nach Filtergrad aus mehreren Elementen L und C und nur zwei ohmschen Widerständen. Einer der ohmschen Widerstände ist ein Reihenwiderstand am Eingang des Filters, der andere ist ein Parallelwiderstand zu den beiden Ausgangsanschlüssen. Sie bilden

den eingangs- und ausgangsseitigen Wellenwiderstand (Wellenwiderstand in Vorwärts- und Rückwärtsrichtung, siehe Kap. 93 „Zweitore“) des Filters und sind meist gleich groß, z. B. 50 Ω, siehe dazu Abb. 97.10. Wegen des erhöhten Aufwandes bei der Filtertransformation sind hier nur einige Grundgedanken angegeben. Es sei auf die Anwendung von Filterentwurfsschaltungen hingewiesen.

97.3.3.1 Hochpassfilter

Der Ausdruck $j\omega$ wird durch $1/(j\omega)$ ersetzt. Das heißt in der Praxis:

1. Einen Kondensator durch eine Spule ersetzen.
2. Eine Spule durch einen Kondensator ersetzen.
3. Ein Widerstand bleibt als Widerstand erhalten. Sein Wert kann geändert werden.
   Die zwei Widerstände bleiben aber meist unverändert, weil sie den Wellenwiderstand vorgeben (siehe letzter Abschnitt oben).
4. Die Werte und Dimensionen bei den Bauelementen müssen angepasst werden.

In der Praxis heißt das:

Bei den Ersetzungen $C \to L$ und $L \to C$ bleibt der Zahlenwert erhalten, nur die Einheit wird angepasst:

$$\begin{aligned} C \to L: &\quad \underbrace{330\,\mathrm{n}}_{u}\,\underbrace{\mathrm{F}}_{a} \to \underbrace{330\,\mathrm{n}}_{u}\,\underbrace{\mathrm{H}}_{a}; \\ L \to C: &\quad \underbrace{4{,}7\,\mu}_{u}\,\underbrace{\mathrm{H}}_{a} \to \underbrace{4{,}7\,\mu}_{u}\,\underbrace{\mathrm{F}}_{a} \end{aligned} \qquad (97.22)$$

u: unverändert, a: angepasst.

Damit ist die Transformation für den Fall der unveränderten Wellenwiderstände und gleichbleibender Grenzfrequenz abgeschlossen. Sollen zusätzlich die zwei oben erwähnten ohmschen Widerstände um den gleichen Faktor geändert werden, ist so vorzugehen:

Um den neuen Widerstandswert zu erhalten, werden die alten Widerstandswerte mit dem Faktor A (positive beliebige Zahl) multipliziert. Anschließend müssen alle Kapazitätswerte mit dem Faktor $1/A$ und alle Induktivitätswerte mit A multipliziert werden.

Beispiel

Ein Tiefpass soll durch einen Hochpass ersetzt werden, und der Wellenwiderstand des Tiefpasses von 600 Ω in Vorwärts- und Rückwärtsrichtung zu 60 Ω im Hochpass geändert werden.

Daraus ergibt sich: $(60\,\Omega)\big|_{\mathrm{neu}} = A \cdot ((600\,\Omega)\big|_{\mathrm{alt}}) \Rightarrow A = 1/10$. Die Spulen werden durch Kondensatoren ersetzt und die Kondensatoren durch Spulen. Die Widerstände bleiben als Widerstände erhalten, ihr Wert ändert sich von 600 Ω in den Wert $A \cdot (600\,\Omega) = (1/10) \cdot (600\,\Omega) = 60\,\Omega$.

Ein Kondensator mit dem Wert 500 nF ginge dann in eine Spule mit dem Wert $(1/A) \cdot (500\,\mathrm{nF}) = (1/(1/10)) \cdot (500\,\mathrm{nF}) = 10 \cdot (500\,\mathrm{nF}) = 5\,\mu\mathrm{F} \to 5\,\mu\mathrm{H}$ und eine Spule mit dem Wert 500 mH in den Wert $A \cdot (500\,\mathrm{mH}) = (1/10) \cdot (500\,\mathrm{mH}) = 50\,\mathrm{mH} \to 50\,\mathrm{mF}$ über. ◀

97.3.3.2 Bandpassfilter

Ein Bandpassfilter entsteht, anschaulich und vereinfacht dargestellt, dadurch, dass die Übertragungsfunktion (der Frequenzgang) eines Tiefpassfilters mit linearer Achsenskalierung an der senkrechten Achse bei der Frequenz Null gespiegelt wird und der sich ergebende Verlauf um den Wert der Mittenfrequenz des Bandpassfilters nach rechts verschoben wird.

Damit sind folgende Transformationen vorzunehmen:

Ein Kondensator geht in einen Parallelschwingkreis über.

Eine Spule geht in einen Reihenschwingkreis über.

Enthält das Tiefpassfilter bereits Schwingkreise, sind die Einzelelemente getrennt zu transformieren:

Ein Parallelschwingkreis geht in einen Parallelschwingkreis (für C) und dazu parallel einen Reihenschwingkreis (für L) über.

Ein Reihenschwingkreis geht in einen Reihenschwingkreis (für L) in Reihe zu einem Parallelschwingkreis (für C) über.

Die Transformation gestaltet sich etwas aufwändig, so dass der Einsatz von Filterentwurfsprogrammen zu erwägen ist. Außerdem lassen sich dann die tatsächlichen Eigenschaften des Filters sehr schnell erfassen und gegebenenfalls korrigieren.

97.3.3.3 Bandsperre

Anschaulich erhält man diesen Filtertyp, indem die Übertragungsfunktion (der Frequenzgang) eines Hochpassfilters mit linearer Achsenskalierung an der senkrechten Achse bei der Frequenz Null gespiegelt wird und der sich ergebende Verlauf um den Wert der Mittenfrequenz der Bandsperre nach rechts verschoben wird.

Ein Kondensator geht in einen Reihenschwingkreis über.

Eine Spule geht in einen Parallelschwingkreis über.

Enthält das Tiefpassfilter bereits Schwingkreise, sind die Einzelelemente getrennt zu transformieren:

Ein Parallelschwingkreis geht in einen Parallelschwingkreis (für L) und dazu parallel einen Reihenschwingkreis (für C) über.

Ein Reihenschwingkreis geht in einen Reihenschwingkreis (für C) in Reihe zu einem Parallelschwingkreis (für L) über.

Auch hier gestaltet sich die Transformation etwas aufwändig, so dass der Einsatz von Filterentwurfsprogrammen zu empfehlen ist. Außerdem lassen sich die tatsächlichen Eigenschaften des Filters sehr schnell erfassen und gegebenenfalls korrigieren.

97.3.4 Besondere Filterschaltungen

97.3.4.1 Allpassfilter

Dieses Filter ist ein Sondertyp mit folgenden Eigenschaften:

1. Der Amplitudenfrequenzgang ist unabhängig von der Frequenz.
2. Der Phasenwinkel ist – bewusst – eine Funktion der Frequenz. Damit kann ein Allpass die Gruppenlaufzeit eines vorhandenen Filters wenigstens teilweise den Erfordernissen anpassen, z. B. nahezu frequenzunabhängig machen.
3. Dieses Filter kann zur Impedanztransformation eingesetzt werden. Durch entsprechende Dimensionierung lässt sich neben einer evtl. erforderlichen Gruppenlaufzeitkorrektur z. B. auch noch der Wellenwiderstand in Vorwärts- und Rückwärtsrichtung der gegebenen Schaltung anpassen.

Auch hier führt der Einsatz von Filterentwurfsprogrammen einschließlich der Simulation der Ergebnisse schnell zu dem gewünschten Ergebnis.

97.3.4.2 Sallen-Key-Filter

Es handelt sich um aktive Filterschaltungen, bei denen Operationsverstärker, Widerstände und Kondensatoren eingesetzt werden. Verglichen mit den passiven Filterschaltungen ist bei den Sallen-Key-Filtern ein geringerer Bauteileaufwand erforderlich, weil die Verstärker auch mit positiver Rückkopplung betrieben werden. Sie sind i. Allg. gegenüber Bauteiletoleranzen relativ unempfindlich. Allerdings können keine Bandsperren realisiert werden, und hohe Gütefaktoren sind auch nicht erreichbar. Wegen der relativ konstanten Gruppenlaufzeit und dem relativ abrupten Übergang vom Durchlass- in den Sperrbereich werden sie häufig als Antialiasingfilter vor einem Analog-Digital-Umsetzer eingebaut.

Empfängerschaltungstechnik

98

Wilfried Plaßmann



98.1 Geradeausempfänger 1246
98.2 Überlagerungsempfänger 1246
98.3 Automatische Verstärkungsregelung (AVR) 1246
98.4 Weitere Schaltungskonzepte 1247

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_98

98.1 Geradeausempfänger

Beim Empfang eines Sendersignals und damit einer bestimmten Information müssen alle anderen Sendersignale wirkungsvoll unterdrückt werden, wobei die Möglichkeit bestehen muss, aus der Vielzahl der angebotenen Informationen (Sender) eine bestimmte auszuwählen. Damit muss aus dem angebotenen Frequenzspektrum ein bestimmtes Frequenzband ausgewählt und der Rest wirkungsvoll unterdrückt werden. Zur Selektion werden mehrere Hochfrequenzstufen (HF-Stufen) eingesetzt, von denen jede einen Verstärker und einen Schwingkreis verwendet (siehe dazu auch Abschn. 97.1.2.4 und Beispiel 3). Damit entsteht das in Abb. 98.1a gezeigte Prinzip eines *Geradeausempfängers*. Die Resonanzfrequenz aller Schwingkreise ist gleich der gewünschten Senderfrequenz f_e. Um verschiedene Sender empfangen zu können, müssen die Resonanzfrequenzen einstellbar und stets genau gleich sein. Diese Schaltungstechnik hat folgende Nachteile:

1. Bei der heutigen Senderdichte sind 5 bis 10 Schwingkreise für einen einigermaßen störungsfreien Empfang erforderlich. Das exakte Einstellen dieser Schwingkreise auf genau die gleiche Resonanzfrequenz, wenn sich diese außerdem noch in einem weiten Bereich ändert (z. B. UKW-Rundfunk: 87 bis 108 MHz), ist nicht mehr möglich.
2. Die resultierende Bandbreite ändert sich mit der Senderfrequenz ($B = f_e/Q_p = f_e/(R_p \cdot 2 \cdot p \cdot f_e \cdot C) = f(1/C)$, Tab. 97.3). Sie kann deshalb nicht optimal über den gesamten Frequenzbereich eingestellt werden.
3. Beim Fernsehempfang müssten die Schwingkreise in ihrer Resonanzfrequenz außerdem noch so gegeneinander versetzt werden, dass eine Bandbreite von einigen MHz mit konstantem Betrag der Übertragungsfunktion und mit steil abfallenden Flanken an den Rändern erreicht wird. Das ist technisch nicht zu verwirklichen.

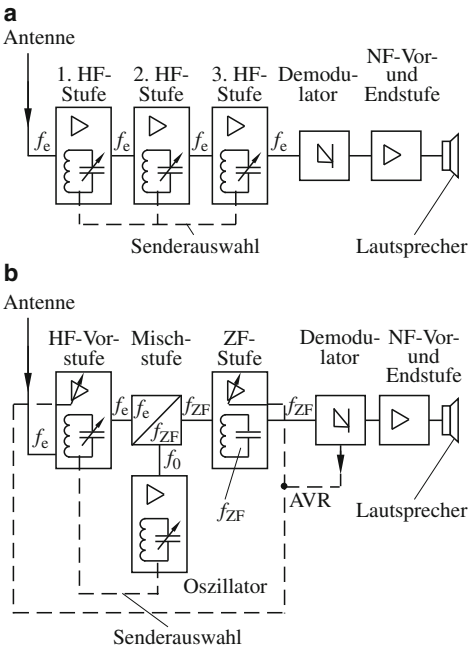


Abb. 98.1 Schaltungsprinzipien von Empfängern. **a** Geradeausempfänger, **b** Überlagerungsempfänger

98.2 Überlagerungsempfänger

Man hat deshalb das Konzept des *Überlagerungsempfängers* nach Abb. 98.1b entwickelt. Empfangen werden soll ein Sender mit der Frequenz f_e. Das von der Antenne kommende Signal wird in einer Hochfrequenz-(HF-)Vorstufe verstärkt, deren Schwingkreis auf die Resonanzfrequenz f_e eingestellt wird. Zusammen mit dieser Einstellung wird die Frequenz f_o eines Oszillators so verändert, dass die in der Mischstufe gebildete Differenzfrequenz $f_o - f_e$ für alle möglichen Frequenzen f_e konstant ist: $f_o - f_e = \text{const} = f_{ZF}$. Diese konstante Frequenz heißt *Zwischenfrequenz (ZF)*. Ihr Wert ist ca. 460 kHz beim Empfang auf Lang-, Mittel- und Kurzwelle, 10,7 MHz beim UKW-Rundfunkempfang und u. a. etwa 30 MHz beim Fernsehempfang. Die Zwischenfrequenz wird in einem ZF-Verstärker, der aus mehreren Transistor-Verstärkerstufen und aus mehr als 10 Schwingkreisen bestehen kann, verstärkt. Da es sich um eine konstante Frequenz handelt, werden die Schwingkreise bei der Fertigung auf die erforderliche Zwischenfrequenz abgeglichen. Auch spezielle Filterkurven lassen sich einstellen. Beim Überlagerungsempfänger müssen also nur zwei Schwingkreise aufeinander abgestimmt werden, was technisch möglich und praktisch nur näherungsweise erforderlich ist. Es lässt sich nämlich zeigen, dass die Eigenschaften des Empfängers nicht wesentlich schlechter werden, wenn diese Abstimmung nur angenähert erreicht wird. Beim Empfang von frequenzmodulierten Sendern wird die Verstärkung der ZF-Stufe so groß gewählt, dass eine Amplitudenbegrenzung eintritt und damit eine zusätzliche Störunterdrückung erreicht wird (siehe auch Abschn. 96.2.4.5 „Winkelmodulation“).

Das Prinzip des Überlagerungsempfängers ist noch erweitert worden zum Mehrfach-Überlagerungsempfänger. Dabei wird die erste Zwischenfrequenz f_{ZF1} mit einem zweiten Oszillator auf eine zweite niedrigere Zwischenfrequenz f_{ZF2} umgesetzt. Die niedrigere Zwischenfrequenz erlaubt schmalbandige Filter hoher Güte, was mit einem direkten Übergang $f_e - f_{o2} = f_{ZF2}$ nicht möglich ist ($B = f_0/Q$, siehe Kap. 97 „Filter“, Abschn. 97.1.2.4 „*R-L-C*-Bandpass und -Bandsperre“).

98.3 Automatische Verstärkungsregelung (AVR)

Wie in Abb. 96.6 im Abschn. 96.2.1.7 „Demodulation von AM“ eingetragen, wird am Demodulatorausgang eine Spannung u_{AVR} gewonnen, deren Wert dem Mittelwert der Amplitude des Zwischenfrequenzsignals und damit des Sendersignals proportional ist. Diese Spannung wird auf die ZF- und HF-Vorstufe zurückgeführt, um die Verstärkung dieser Stufen einzustellen: Steigendes u_{AVR} bewirkt sinkende Verstärkung beider Stufen. Damit

ist die Amplitude am Eingang des Demodulators und daraus folgend auch die am Lautsprecher trotz schwankender Amplitude des Sendersignals nahezu konstant. Diese Schaltungstechnik wird als *a*utomatische *V*erstärkungs*r*egelung (AVR) bezeichnet. Auch der Demodulator für frequenzmodulierte Signale stellt eine derartige Spannung zur Verfügung, siehe dazu Abb. 96.18c, u_{AVD}, Abschn. 96.2.4.8.

98.4 Weitere Schaltungskonzepte

Moderne Empfänger weisen noch einige der folgenden Schaltungskonzepte auf, deren Realisierung z. B. so aussieht:

- Die Abstimmung der Schwingkreise von Vor- und Oszillatorstufe geschieht über Kapazitätsvariationsdioden, deren Kapazität über eine Gleichspannung eingestellt wird. Diese gewinnt man über einen Spannungsteiler von einer temperaturkompensierten Z-Diode als Konstantspannungsquelle.
- Für die Speicherung von Sendern zum einfachen Abruf über Tastendruck wird ein zugeordnetes Digitalwort gespeichert, das nach Senderauswahl als Eingangsgröße eines Digital-Analog-Umsetzers dient, dessen Ausgangsspannung über die Kapazitätsdioden den gewünschten Sender einstellt. Zur Anzeige der Frequenz des empfangenen Senders misst man die Oszillatorfrequenz und zieht vom erhaltenen Wert die Zwischenfrequenz ab. Die Eingangsfrequenz direkt zu messen ist wegen der kleinen schwankenden Amplituden nicht möglich, und außerdem erhielte man prinzipiell keine Frequenzanzeige, wenn kein Sendersignal vorhanden wäre.
- Überwiegend beim UKW- und Fernsehempfang sorgt eine automatische Senderabstimmung (AFC, *a*utomatic *f*requency *c*ontrol) dafür, dass eine ungenaue Abstimmung auf die Senderfrequenz korrigiert wird.
- Die Rundfunksender übertragen häufig neben dem eigentlichen Programm noch Zusatzinformationen, wie z. B. Sendername, Art des gerade übertragenen Programms (Nachrichten, E-Musik, U-Musik), Uhrzeit und Signale zur Steuerung des Empfängers für Verkehrsdurchsagen. Besonders die Autoradioempfänger nutzen den Sendernamen und die Kennung für Verkehrsdurchsagen: Der Empfänger sucht selbsttätig aus dem Angebot an Sendern gleichen Namens den am besten zu empfangenden aus, und der ausgeschaltete Empfänger wird speziell für die Verkehrsdurchsagen eingeschaltet oder der eingeschaltete Empfänger auf eine Mindestlautstärke eingestellt.
- Bei der Fernsehübertragung werden von etlichen Fernsehanstalten zusätzlich zu Bild und Ton noch weitere Informationen als *Videotext* übertragen. Es handelt sich dabei um Zusatzinformationen zu den Sendungen, Nachrichten, Berichte usw. Man nutzt dazu vorhandene Lücken bei der Bildübertragung (Kap. 99).

Ton- und Bildübertragung

99

Wilfried Plaßmann



99.1 Rundfunk-Stereoübertragung 1250
99.2 Digitale Farbfernsehtechnik 1251

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_99

99.1 Rundfunk-Stereoübertragung

Aus dem Laufzeitunterschied und der Lautstärke, mit dem ein Schallsignal das linke und das rechte Ohr erreicht, lässt sich die Richtung der Schallquelle erkennen. Bei der *Mono-Übertragung* wird das gesamte Schallsignal über *einen* Kanal übertragen, sodass der Mensch als Empfänger zwar die Richtung des Gesamtsignals, d. h. den Ort des Lautsprechers, erkennt, aber nicht erfährt, wo bei der Aufnahme z. B. der Sänger und wo die Musikinstrumente gestanden haben. Da dies der täglichen Praxis des Hörens widerspricht, hat man den räumlichen Eindruck durch Einführung der *Stereo-Übertragung* wiederhergestellt. Bei der Rundfunkübertragung gilt der Grundsatz der *Kompatibilität*, d. h. der Höreindruck eines Rundfunkempfängers ohne Stereoempfangseinrichtung darf durch die Aussendung des Stereosignals nicht beeinträchtigt werden. Deshalb war es auch nicht vertretbar, den naheliegenden Weg zu gehen, nämlich für Rundfunkaufnahmen zwei Mikrofone räumlich voneinander getrennt links und rechts im Raum anzuordnen und für die Monowiedergabe beide Signale zu addieren. Es hat sich gezeigt, dass dies zu einer „Verflachung" des Höreindruckes führt.

Für die Aufnahme und Wiedergabe von Musikcassetten und Compact-Disks (CDs) ordnet man die Mikrofone häufig nicht links und rechts im Raum an, sondern direkt neben- oder übereinander, Abb. 99.1a, wobei deren Richtcharakteristik nach links bzw. rechts weist. Diese Aufnahmetechnik wird mit *Links-Rechts-* oder *X-Y-Stereofonie* bezeichnet.

Bei Rundfunkaufnahmen dagegen setzt man zwei Mikrofone nach Abb. 99.1b ein. Das eine steht in der Mitte und nimmt die Schallsignale aus allen Richtungen gleich gut auf (kreisförmige Richtcharakteristik, siehe Abschn. 105.2 „Schallempfänger"), das andere steht ebenfalls in der Mitte und bevorzugt diejenigen Signale, die genau von rechts und genau von links kommen (Richtcharakteristik in Form einer Acht). Bei dieser mit *M-S-(Mitte-Seiten-)Stereofonie* bezeichneten Aufnahmetechnik wird das Mittensignal für die Monowiedergabe verwendet. Es hat sich gezeigt, dass die *X-Y-* und die *M-S*-Signale ineinander überführt werden können:

$$X + Y \mathrel{\widehat{=}} M \qquad M + S \mathrel{\widehat{=}} X$$
$$\text{bzw.} \qquad (99.1a)$$
$$X - Y \mathrel{\widehat{=}} S \qquad M - S \mathrel{\widehat{=}} Y$$

Damit genügt *eine* der beiden Aufnahmetechniken für alle Anwendungsfälle.

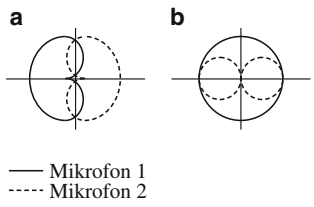


Abb. 99.1 Mikrofonanordnungen mit Richtcharakteristik für Stereo-Aufnahmen. **a** *X-Y*-Stereofonie (Links-Rechts), **b** *M-S*-Stereofonie (Mitte-Seiten)

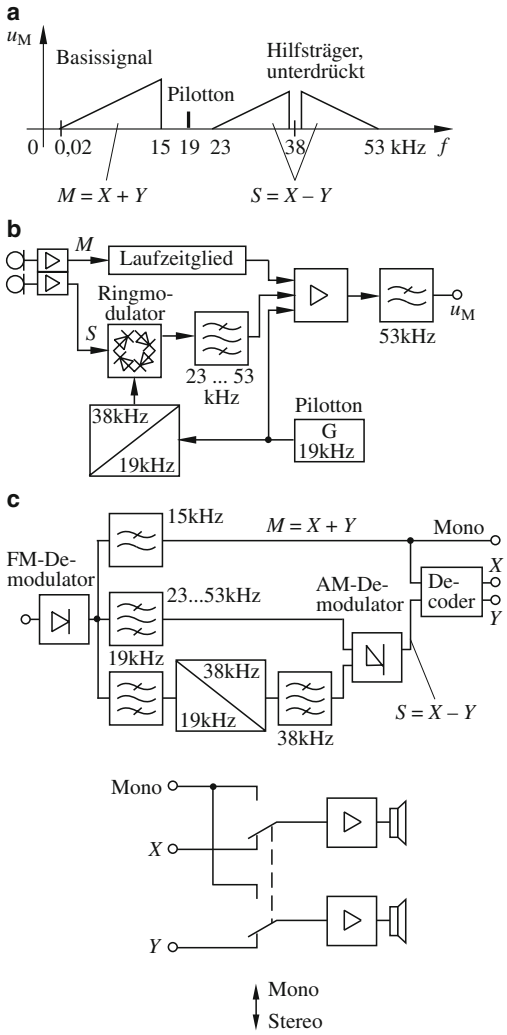


Abb. 99.2 Stereo-Rundfunkübertragung. **a** Spektrum des Stereosignals (senderseitig), **b** Erzeugung des Stereo-Modulationssignals (senderseitig), **c** Stereo-Empfänger (ab Demodulator dargestellt)

Die Stereo-Rundfunkübertragung wird im UKW-Bereich eingesetzt, weil hier durch die verwendete Frequenzmodulation eine hohe Übertragungsqualität erreicht wird. Abb. 99.2a zeigt das Spektrum eines Stereosignals vor der Frequenzmodulation des Trägers. Für Monoempfang steht das Basissignal $M = X+Y$ zur Verfügung. Ein *Hilfsträger* mit der Frequenz 38 kHz wird mit dem Signal $S = X - Y$ zweiseitenband-amplitudenmoduliert. Da der Träger bei der anschließenden Frequenzmodulation und Demodulation wegen seiner großen Amplitude unerwünschte Frequenzanteile erzeugen könnte, wird er unterdrückt, was mit geeigneten Modulatorschaltungen möglich ist (siehe Abschn. 96.2.2.1 „Einseitenbandmodulation", und Abb. 96.7). Auch eines der beiden Seitenbänder hätte theoretisch noch unterdrückt werden können, ist aber praktisch wegen der erforderlichen steilflankigen Filter nicht möglich: Bei Anstieg der Frequenz von 37,980 auf 38,020 kHz wäre ein Abfall der Ausgangsspannung auf weniger als 1/1000 notwendig gewesen. Da

für die Demodulation im Empfänger der Träger frequenzgenau wieder zugesetzt werden muss und außerdem seine Phasenlage zu den Seitenbändern zur Erhaltung des Stereo-Effektes wichtig ist, wird ein *Pilotton* übertragen. Verglichen mit dem (unterdrückten) Hilfsträger hat er die halbe Frequenz (19 kHz), die gleiche Phasenlage und 1/10 der Amplitude. Abb. 99.2b stellt das Prinzip der Signalaufbereitung bis zur Modulationsstufe dar. Durch das Laufzeitglied gelangen das durch Modulation und Filterung verzögerte Signal $S = X - Y$ und das Signal $M = X + Y$ ohne Phasenverschiebung zueinander auf den Tiefpass und anschließend zur Modulationsstufe. Ohne dieses Laufzeitglied wäre der Stereoeffekt verfälscht, weil eine Phasenverschiebung zwischen S und M vom Ohr als Laufzeitunterschied zwischen den Schallsignalen aufträte und als falsche Rauminformation interpretiert würde.

Der Empfänger nach Abb. 99.2c führt in der Schalterstellung „Mono" das Signal $M = X + Y$ den beiden Verstärkern parallel zu. Für den Stereoempfang wird die Pilottonfrequenz 19 kHz in einer Frequenzverdopplerschaltung in die Frequenz 38 kHz umgewandelt. Dazu wird häufig der sinusförmige Pilotton brückengleichgerichtet. Es entsteht gemäß Fourierzerlegung u. a. die Frequenz 38 kHz, phasenrichtig zu 19 kHz. Die höherfrequenten Anteile sowie der Anteil mit 19 kHz werden durch das Bandpassfilter unterdrückt. Der Anteil mit 19 kHz tritt nicht mehr auf. Nach der Demodulation bildet der Dekoder die Signale für den rechten und linken Kanal durch Anwendung von (99.1a):

$$\begin{aligned} (X+Y)+(X-Y) &= 2X \quad \text{linker Kanal} \\ (X+Y)-(X-Y) &= 2Y \quad \text{rechter Kanal} \end{aligned} \tag{99.1b}$$

99.2 Digitale Farbfernsehtechnik

Zielsetzung, Vorgaben

Zielsetzung bei der Entwicklung des sogenannten Digitalfernsehens war die Übertragung von Bild und Ton in qualitativ hochwertiger Form (gute Bild- und Ton-Qualität bei minimalen Störungen) bei geringer Senderleistung und Empfang der Sender mit Stabantenne. Mit dieser Antennenform können Mehrfachreflexionen aber nicht mehr ausgeblendet werden. Es galt außerdem die Nebenbedingung, dass nur eine begrenzte und damit eine *zunächst* nicht ausreichende Bandbreite zur Verfügung stand. Es musste somit versucht werden, durch Redundanz- und Irrelevanzreduktion sowie durch den Einsatz geeigneter nachrichtentechnischer Verfahren eine vertretbare Bandbreitenverringerung zu erreichen. Grundsätzlich galt, dass das Ziel nur mit der Übertragung der Signale in digital codierter Form erreicht werden konnte. Wie noch gezeigt wird, erfordern Sender und auch Empfänger Bausteine, die kleinen Rechnern vergleichbar sind.

Die folgenden Teilgebiete der Nachrichtentechnik boten die erforderlichen Lösungsansätze: Modulationstechnik, Codierverfahren, Fehler-Erkennung und -Korrektur, Redundanzreduktion, Irrelevanzreduktion und Kanalcodierung. Weiterhin waren Untersuchungsergebnisse über das Verhalten von Auge und Ohr bei der Signalverarbeitung von großem Nutzen. Das Ergebnis ist sehr komplex, und deshalb kann hier nur ein erster grober und vereinfachender Überblick gegeben werden, für Einzelheiten wird auf die spezielle Fachliteratur verwiesen. Trotzdem kann es interessant sein zu sehen, wie die Lösung der Probleme aussieht.

DVB-Projekt

Seit Beginn der 1990er Jahre besteht das DVB-Projekt (*D*igital *V*ideo *B*roadcasting), in dem einheitliche und verbindliche Verarbeitungs- und Übertragungssysteme für das Digitale Farbfernsehen erarbeitet worden sind. Dabei wurde dreigleisig verfahren, abhängig davon, ob das Fernseh- und das begleitende Tonsignal über Kabel (Entwicklung *DVB-C*), Satellit (Entwicklung *DVB-S*) oder terrestrisch (Entwicklung *DVB-T*; Funkübertragung von Sender- zu Empfänger-Antenne) übertragen werden soll. Am komplexesten ist das Verfahren für die terrestrische Übertragung, weil hier der Einfluss von Störungen am größten ist und Lösungen eingesetzt worden sind, die die Auswirkungen dieser Störeinflüsse minimieren. Im Folgenden soll das DVB-T-Verfahren wenigstens in einigen prinzipiellen Eigenschaften dargestellt werden.

Ein Fernsehbild wird zerlegt in waagerechte schmale Streifen, sog. Zeilen. Die Zeilen wiederum bestehen aus einzelnen Punkten. Insgesamt besteht das Fernsehbild aus etwa 415.000 einzelnen Punkten. Für jeden Punkt müssen Helligkeit und Farbe übertragen werden. Außerdem sind weitere zusätzliche Signale erforderlich. Die Fernsehkameras tasten das zu übertragende Bild punktweise so ab, dass die Informationen über Helligkeit und Farbe von jedem der etwa 415.000 Bildpunkte zur Verfügung stehen (415.000 Abtastwerte). Der Sender bereitet die Signale noch entsprechend auf, bevor sie gesendet werden.

Bei den Abtastwerten pro Zeile muss unterschieden werden zwischen den Abtastwerten, die direkt den Bildpunkten auf dem Empfängerschirm zugeordnet sind, und den Daten, die zwischen dem Ende der auf dem Bildschirm dargestellten Zeile n und dem Beginn der folgenden auf dem Bildschirm dargestellten Zeile $(n + 2)$ liegen (siehe Zeilensprungverfahren weiter unten). Das Bereitstellen eines Bereiches für Daten zwischen 2 Zeileninhalten wird mit *Horizontalaustastung* bezeichnet, und es entsteht eine *horizontale Austastlücke*. Darin enthalten sind u. a. die Ton- und die Synchronisations-Informationen. Ebenso gibt es eine *vertikale Austastlücke*, die entsteht, wenn nach Erreichen des Endes der letzten Zeile des momentanen Bildes bis zu Beginn der ersten Zeile des folgenden neuen Bildes ebenfalls ein Zeitbereich für weitere Informationen zur Verfügung gestellt wird. In dieser *vertikalen Austastlücke* werden neben allgemein nutzbaren Informationen auch Informationen über den Bildaufbau übertragen.

Der *Studiostandard* ITU-R BT.601 enthält die Festlegungen für das Digitalfernsehen *im Studio*. Hier einige Auszüge, Erläuterungen folgen weiter unten:

- 720 Abtastwerte je aktive Zeile für das Y-Signal.
- 864 Abtastwerte pro Zeile insgesamt (720 aktiv plus 144 für Zusatzinformationen).
- 576 aktive Zeilen (d. h. für das Bild) bei insgesamt 625 Zeilen pro Vollbild.

Tab. 99.1 Wichtung der Farben rot, grün und blau

Farbe	Wellenlänge in nm	Helligkeitseindruck, relativ zueinander
Rot (R)	615	0,30
Grün (G)	532	0,59
Blau (B)	470	0,11
		$\sum = 1{,}00$

- Codierung des Bildsignals mit 10 Bit.
- Abtastraster 4 : 2 : 2.
- Bitrate des seriellen Studiosignals: 270 MBit/s.
- Bitrate für das eigentliche Videosignal: 207 MBit/s.

Die Praxis zeigt, dass diese Bitrate nicht für die Übertragung des Farbfernsehens zur Verfügung steht!

Damit ergibt sich die Forderung, dass die Bitrate wesentlich verringert werden muss, mit der Nebenbedingung, die Fernsehbild- und Ton-Übertragungsqualität gegenüber der Studioqualität nur minimal zu verringern.

MPEG-2 Standard; ETSI 300 744

Eine von der **I***nternational* **E***lectrotechnical* **C***om*mission (IEC), der **I***nternational* **S***tandards* **O***rganisation* (ISO) und dem *Joint Technical Committee* (JTC) eingesetzte Arbeitsgruppe erarbeitete weltweit verbindliche Richtlinien zur Codierung von Audio- und Video-Signalen, zunächst für nicht bewegte Bilder, dann erweitert auf bewegte Bilder: **M***oving* **P***ictures* **E***xpert* **G***roup* (MPEG). ISO und IEC haben 1994 den MPEG-2 Standard, ISO/IEC 13818, geschaffen.

Einzelheiten zur DVB-T-Übertragung sind in der ETSI 300 744 (European Telecommunications Standards Institute) enthalten.

Kenngrößen des Bildsignals

Zunächst wird ein Farbbild analog aufgezeichnet, indem senderseitig 3 Bildaufnahmegeräte mit vorgeschalteten Farbfiltern, durchlässig für die Farben Rot, Grün und Blau, verwendet werden. Durch entsprechende Mischung dieser 3 Farben lassen sich nahezu alle in der Natur vorkommenden Farben darstellen.

Folgende Signale müssen für die Übertragung des Farbbildes erzeugt werden:

- *Helligkeitssignal* oder *Leuchtdichtesignal* oder *Luminanzsignal* oder *Y-Signal*:

Dieses Signal entspricht den „Grauwerten“. Entsprechend der spektralen Empfindlichkeit des Auges werden die 3 Farbsignale nach Tab. 99.1 gewichtet:

Hierfür wird die folgende vereinfachte Schreibweise gewählt:

$$Y = 0{,}30 \cdot R + 0{,}59 \cdot G + 0{,}11 \cdot B \qquad (99.2)$$

- *Farbartsignal oder Chrominanzsignal*:

Es setzt sich zusammen aus:

a) *Farbton*: Das ist die Farbe, z. B. orange.
b) *Farbsättigung*: Sie gibt an, ob die Farbe kräftig oder blass ist. 100-%-Amplitude heißt „kein weiß enthalten“, 50-%-Amplitude bedeutet: 50 % Farbe, 50 % weiß.

Die Farbsättigung ist bereits im *Y-Signal* enthalten. Es werden deshalb übertragen:

$$(R-Y)' = (R-Y) \cdot 0{,}88$$
$$(B-Y)' = (B-Y) \cdot 0{,}49 \qquad (99.3)$$

Die Bezeichnungen $(R-Y)'$ bzw. C_R werden beide verwendet, ebenso $(B-Y)'$ bzw. C_B. Die Faktoren 0,88 und 0,49 enthalten die Farbwichtung.

Redundanz- und Irrelevanzreduktion

Bei der Redundanzreduktion erleidet das Signal keinen Informationsverlust, es wird lediglich *überflüssige*, d. h. bereits vorhandene, Information entfernt. Bei der Irrelevanzreduktion dagegen wird diejenige Information entfernt, auf die der Empfänger erfahrungsgemäß ohne allzu großen Nachteil verzichten kann oder aber die der Empfänger im Allgemeinen nicht auswertet bzw. auswerten kann. Als Empfänger dienen in diesem Fall Auge und Ohr des Menschen.

- *Zeilensprungverfahren*: Für die Zahl der Zeilen hat sich durch Tests ein Wert von mindestens 500 als sinnvoll erwiesen, gewählt wurden 625 Zeilen. Davon dienen 576 Zeilen (sog. aktive Zeilen) zur Darstellung des Bildes, der Rest wird zur Übertragung von Zusatzinformationen verwendet (Synchronisationsimpulse, Trägerrückgewinnung, Fehler-Erkennung und -Korrektur, Zeiteinblendung, Informationen über Codierung, Videotext usw.; Einzelheiten folgen weiter unten). Die Zahl der Bildpunkte pro Zeile ergibt sich damit durch das gewählte Bildformat, dass von bisher 4 : 3 (Verhältnis Breite : Höhe) zunehmend auf 16 : 9 geändert wird. Für ein flimmerfreies Bild müssen mindestens 16 Einzelbilder pro Sekunde (16/s) dargestellt werden. Die Kinotechnik verwendet 24 Bilder pro Sekunde, beim Fernsehbild reicht dieser Wert nicht aus. Die Übertragung von 50 Bildern pro Sekunde ist für das Auge ausreichend, scheitert aber an der zu hohen erforderlichen Bandbreite. Werte unterhalb 50/s (z. B. wären 40/s durchaus für das Auge vertretbar) sind wegen möglicher Interferenzen mit der Netzfrequenz 50 Hz nicht sinnvoll. Eine Lösung ergibt sich durch das Zeilensprungverfahren, das eine Bandbreite gemäß 25 Bildern pro Sekunde erfordert, physiologisch aber vom Auge als flimmerfrei mit etwa 40 Bildern pro Sekunde empfunden wird. In einem ersten Durchlauf werden auf dem Bildschirm die ungeraden Zeilen 1, 3, 5, … dargestellt, im anschließenden zweiten Durchlauf die geraden Zeilen 2, 4, 6, … Das ergibt 25 ganze Bilder oder 50 Halbbilder pro Sekunde. Das Verfahren wird ein weiteres Mal eingesetzt, indem ein Vollbild elektronisch gespeichert wird und das ursprüngliche Bild nach der halben Zeit kurz unterbrochen und durch das gespeicherte Bild (auch mit gleichem Bildinhalt) ersetzt wird. Diese Technik hat man schon bei der Kinotechnik angewendet durch eine „Unterbrecherblende“. Damit stellt man – fiktiv – 100 Bilder pro Sekunde dar, bei einem Bandbreitenbedarf entsprechend 25/s. Das Auge empfindet dadurch eine Bildfolgefrequenz, die zwischen 50/s und 100/s liegt und ein flimmerfreies Bild ergibt.

Das analoge Bildsignal durchläuft Analog-Digital-Umsetzer. Das Auge kann etwa 250 Helligkeitsstufen (Graustufen) unter-

scheiden, so dass eine 8-bit-Codierung ausreicht. Allerdings erkennt das Auge bei monochromen Bildern ein nicht zu vernachlässigendes Quantisierungsgeräusch (Quantisierungsrauschen), so dass im Studio mit 10-bit-Umsetzern gearbeitet wird.

- *Abtastraster* 4 : 2 : 2 und 4 : 2 : 0: Das menschliche Auge ist für das Farbsehen weniger empfindlich als für das Hell-Dunkel-Sehen (Stäbchen für das Hell-Dunkel-Sehen, Zäpfchen für das Farbsehen). Deshalb wird in den Zeilen zwingend jeder einzelne Bildpunkt für das Luminanzsignal (Y-Signal) abgetastet, während für das C_R- und das C_B-Signal nur jeder zweite Bildpunkt abgetastet wird. Diese Abtastung wird bezeichnet mit dem Begriff *Abtastraster* 4 : 2 : 2. Die Abtastfrequenz für das Y-Signal ist damit doppelt so groß (4) wie die für die beiden Chrominanzsignale C_R und C_B (2 : 2). Das ergibt eine erste Datenreduzierung für die Fernsehbildübertragung.

Diese Datenreduzierung lässt sich weiter fortführen, indem auch in vertikaler Richtung nur jeder zweite Bildpunkt für das Chrominanzsignal abgetastet wird. Das Problem dabei ist, dass dann die Chrominanzsignale wegen des Zeilensprungverfahrens *nur* aus den Werten des ersten (oder zweiten) Halbbildes gebildet würden, was das Auge als Flimmern empfände. Deshalb bildet man aus 2 nebeneinanderliegenden Zeilenwerten des ersten Halbbildes und den zwei genau darunterliegenden Zeilenwerten des zweiten Halbbildes einen Mittelwert, der dann für die 4 Bildpunkte gemeinsam gilt (Abtastraster 4 : 2 : 0).

- *Differenz-Pulscodemodulation*

Die Erfahrung zeigt, dass sich die Bildinformation benachbarter Bildpunkte im Mittel nur wenig voneinander unterscheidet. Bildet man die Differenz, ergibt sich, dass der Wert 0 am häufigsten vorkommt und dass die Häufigkeit mit wachsendem Differenzwert stark abnimmt. Deshalb werden kleine Differenzwerte mit wenigen Bit codiert, große Werte mit größerer Bit-Zahl. Dieses Verfahren ist vom Morse-Alphabet bekannt. Es handelt sich hierbei um eine Redundanzreduktion, weil z. B. in 2 Codewörtern mit 8 Bit, bei denen nur das niederwertigste Bit unterschiedlich ist, die übrigen 7 Bit gleich und damit redundant sind. Es wird also nur die Differenz einschließlich Vorzeichen zum vorhergehenden Wert übertragen. Allerdings liegt eine Gefahr darin, dass bei einem falschen Wert, wie auch immer er entstanden ist, alle folgenden durch Addition des vorzeichenrichtigen Differenzwertes falsch sind. Es muss deshalb in geeigneten zeitlichen Abständen, deren Wert durch Erfahrung und Statistik zu ermitteln ist, ein kompletter ungekürzter Wert übertragen werden, von dem aus dann wieder neue Differenzadditionen gestartet werden können. Der Sender führt die gleiche Differenzaddition durch wie der Empfänger, um Rundungs- bzw. Quantisierungs-Fehler des Empfängers nicht zu groß werden zu lassen: *Differenz-Pulscodemodulation mit prädiktiver Codierung*. Die Differenzbildung kann erfolgen für benachbarte Bildpunkte in einer Zeile oder für nebeneinander liegende Bildpunkte in zwei aufeinander folgenden Zeilen eines Halb- oder Vollbildes. Insgesamt ergibt sich durch diese Art der Codierung im Mittel eine geringere Datenrate. Allerdings bringt diese Maßnahme nicht automatisch eine Reduktion der Bandbreite mit sich, sofern man als größte Differenz den Unterschied zwischen dem Wert 0 und dem Maximalwert zulässt. Der Übertragungskanal muss für diesen zwar seltenen aber möglichen Fall die Kanalbandbreite bereithalten. Da derart große Differenzen aber nur selten vorkommen und das Auge diesen Änderungen zeitlich nur verzögert folgen kann, werden solche Differenzen in ihrem Wert begrenzt. Eine Änderung um 8 Bit wird senderseitig z. B. auf 4 Bit begrenzt, und dieser Wert wird übertragen. Anschließend wird in einem oder mehreren Schritten senderseitig auf den neuen Wert abgeglichen. Das Auge merkt häufig nichts von dieser Vorgehensweise, und da dieser Fall auch noch selten vorkommt, ergeben sich kaum negative Eindrücke für das Auge.

- *Diskrete Kosinus-Transformation*

Es handelt sich bei diesem Ansatz um eine Irrelevanzreduktion, bei der ausgenutzt wird, dass das menschliche Auge feiner werdende Bildstrukturen immer unvollkommener auflöst, so dass diese mit weniger Bit quantisiert werden können.

Die Methode, das Fernsehbild über Bandpässe in mehrere Frequenzbereiche aufzuteilen und die höherfrequenten Anteile mit weniger Bit zu quantisieren, ist wegen der schwierig zu realisierenden steilflankigen Filter mit ihren Laufzeitverzerrungen kaum anwendbar. Das Problem wird auf die Ebene der Digitaltechnik in Form der schnellen diskreten Fouriertransformation verlagert. Sender und Empfänger haben einen *Block* gespeichert, der aus 8×8 Basismustern besteht. Das Basismuster in der linken oberen Ecke ist ohne Änderung, nach rechts in waagerechter Richtung nimmt die Horizontalauflösung zu bis zu maximal 4 Wechseln ganz rechts, in senkrechter Richtung nimmt die Vertikalauflösung zu, wieder bis zu maximal 4 Wechseln ganz unten. Das Basismuster in der rechten unteren Ecke hat damit je 4 Wechsel in horizontaler wie auch in vertikaler Richtung (Schachbrettmuster). Das auf dem Fernsehschirm dargestellte Bild, bestehend aus 720×576 Pixeln, wird unterteilt in 6480 Blöcke zu je 8×8 Pixeln. ($720 \cdot 576 = 6480 \cdot 8 \cdot 8$). Die gespeicherten Blöcke werden mit den Blöcken des aktuellen Fernsehbildes verglichen. Über die hier eingesetzte Diskrete Kosinus-Transformation, die nur die Kosinusanteile erfasst, werden die Frequenzanteile und deren Amplituden des Blocks erfasst. Es zeigt sich, dass durch geschicktes Umsortieren der Positionen der Pixel der Bereich der Blocks in Richtung rechts-unten häufig den Wert Null annehmen. Damit muss man nicht die einzelnen Nullwerte angeben, sondern nur die Zahl, wie oft die Null noch auftritt. Außerdem werden die Werte, die nicht Null sind, noch so umcodiert, dass häufig auftretende Werte mit weniger Bit codiert werden als weniger häufig auftretende, so wie beim Morse-Alphabet.

- *Fehler-Erkennung und -Korrektur*

Mit den bisher geschilderten Maßnahmen lässt sich in der Praxis eine Datenreduktion um den Faktor 20 bis 50 ohne gravierende Qualitätseinbuße erreichen. Das heißt aber auch, dass kaum noch Redundanz zur Verfügung steht, um eventuell auftretende Fehler zu erkennen und zu korrigieren. Damit wird das Übertragungssystem anfällig gegenüber Störungen. Bereits *ein* falsches Bit kann dann gravierende negative Folgen haben, wenn es z. B. die Synchronisation, die Trägerrückgewinnung, die Differenz-Pulscodemodulation oder die Verarbeitungsbits für das Signal betrifft. Deshalb wurde dem Datenstrom wieder Redundanz

zugefügt, um eine Fehler-Erkennung und -Korrektur zu ermöglichen. Dabei gilt, dass kein Zugriff bzw. Rückgriff auf den Sender möglich ist, also eine Fehlererkennung stets eine Fehlerkorrektur beinhalten muss. Damit wird der bisher erreichte Faktor der Datenreduktion wieder verringert.

Reed-Solomon-Verfahren: Die Fehlerkorrektur beinhaltet zunächst die Fehlererkennung und daran anschließend die Fehlerkorrektur. Beides ist nur dadurch erreichbar, dass zusätzlich Redundanz in Form von zusätzlichen Bits bzw. Bytes eingefügt wird. Ein geeignetes Verfahren ist das nach den Entwicklern benannte Reed-Solomon-Verfahren (siehe dazu Spezialliteratur). Es ist besonders geeignet beim Auftreten von Burstfehlern, das sind Fehler, bei denen fehlerhafte Bits in einer Kette hintereinander liegen. Als zweckmäßig hat sich erwiesen, dem Transportpaket mit **188** Byte ein weiteres Paket (Block) mit 16 Bytes anzuhängen, so dass ein Transportpaket jetzt aus **204** Bytes besteht. Damit können maximal **8** fehlerhafte Bytes erkannt und korrigiert werden. Der Code wird auch mit RS (204, 188, 8) bezeichnet (die Zahlen 204, 188 und 8: siehe Unterstreichungen im Text). Treten in einem Transportpaket Burstfehler auf, und dieser Fall ist bei der Fernsehbildübertragung am Häufigsten, ist die folgende Vorgehensweise vorteilhaft: Die 204 Bytes eines Transportpaketes in der Reihenfolge 0 … 203 werden in 17 Teilpakete zu je 12 Bytes umsortiert und dann gesendet:

$$\left.\begin{array}{l} \underbrace{0\text{-}17\text{-}34\text{-}51\text{-}68\text{-}85\text{-}102\text{-}119\text{-}136\text{-}153\text{-}170\text{-}187}_{12\ \text{Bytes}} \\ -\underbrace{1\text{-}18\text{-}35\text{-}\ldots\text{-}188}_{12\ \text{Bytes}}\text{-}\ldots\text{-}\underbrace{16\text{-}33\text{-}\ldots\text{-}203}_{12\ \text{Bytes}} \end{array}\right\} \begin{array}{l} \text{Transportpaket mit} \\ 17 \cdot 12 = 204\ \text{Bytes,} \\ \text{umsortiert} \end{array}$$

Der Empfänger stellt die Originalreihenfolge 0, 1, 2, …, 202, 203 wieder her. Wenn jetzt ein Burstfehler auftritt, in dem z. B. die Bytes 17, 34, 51, 68 als *aufeinander folgende* Bytes fehlerhaft sind, z. B. durch eine Störung während der Funkübertragung, liegen die fehlerhaften Bytes im Empfänger nach Umsortierung in die Originalreihenfolge nicht mehr direkt nebeneinander. Man kann zeigen, dass die Fehlerkorrektur des Reed-Solomon-Verfahrens durch diese Maßnahme effektiver wird. Ganz allgemein hat das Reed-Solomon-Verfahren den Vorteil, dass der Decodieralgorithmus relativ einfach ist. Das Verhältnis (Zahl der Bits für das eigentliche Nutzsignal)/(Zahl der Bits für das eigentliche Nutzsignal + Fehlerkorrektur-Bits) wird mit *Coderate* bezeichnet.

Faltungscode mit Viterbi-Algorithmus: Treten mehr als 8 fehlerhafte Bytes auf, ist das Reed-Solomon-Verfahren überfordert. Dann greift ein zweiter Fehlerkorrektur-Mechanismus ein, der Faltungscode mit Viterbi-Algorithmus. Dieser Algorithmus ist prinzipiell im Kap. 108 „Optimierte Nachrichten- und Datenübertragung“ näher beschrieben.

- *Energieverwischung*

Die vom Sender übertragenen Informationen sind in Blöcken, sog. Transportpaketen, zu je 188 Byte (204 Byte, siehe Fehler-Erkennung und -Korrektur) organisiert. Dabei kann es vorkommen, dass längere Bit-Folgen mit nur Nullen oder nur Einsen auftreten. Für diesen Fall erfolgt, wenn auch nur kurzzeitig, die Aussendung des Trägers. Der Träger wurde

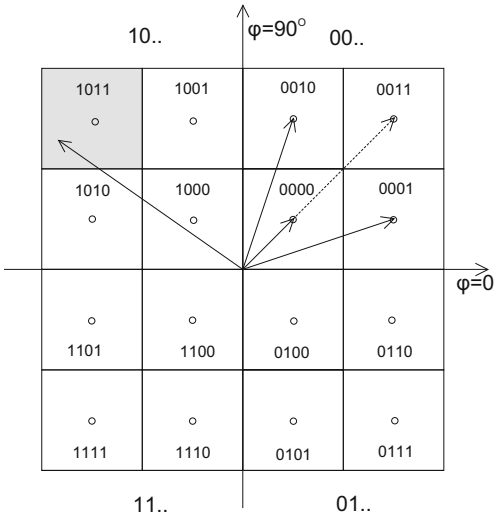


Abb. 99.3 Prinzip einer 16-QAPSK-Modulation

bei der Modulation aber absichtlich unterdrückt, sodass jetzt Probleme mit der im Fernsehempfänger durchgeführten Trägerrückgewinnung und der Demodulation auftreten. Deshalb wird der Datenstrom, aber ohne das Synchronisations-Byte, in eine Pseudo-Zufallsfolge umsortiert. Das geschieht mit Hilfe eines rückgekoppelten Schieberegisters. Nach jeweils 8 Transportpaketen wird das Schieberegister in den Ausgangszustand zurückgesetzt. Der Empfänger wandelt diese Zufallsfolge wieder in die Originalreihenfolge zurück. Der Rücksetzpunkt wird durch ein invertiertes Synchronisations-Byte gekennzeichnet.

- *Trägermodulation: 16-QAPSK, unterdrückter Träger, COFDM*

Zur Übertragung der Fernsehsignale einschließlich Tonsignal wird das 16-QAPSK-Verfahren eingesetzt. Es werden 2 Träger mit gleicher Frequenz und einer Phasenverschiebung von 90° zueinander verwendet (Sinus- und Kosinus-Schwingung). Codiert wird mit einem 4-Bit-Code entsprechend 16 unterscheidbaren Zuständen. Es ist eine kombinierte Amplituden-Phasenumtastung. Abb. 99.3 zeigt die prinzipielle Zuordnung zwischen den zwei Träger-Amplituden und -Phasen für die 16 Zustände. Die ideale Zuordnung zwischen Codewort und Phasen-Amplitudenkombination ist im ersten Quadranten durch die Pfeilspitzen dargestellt. Das jeweilige 4-Bit-Wort wird in der Praxis aber auch dann eindeutig erkannt, wenn sich die Zeigerspitze in dem der Codierung zugeordneten Quadrat befindet, siehe Codewort 1011 links oben. Sofern möglich unterscheiden sich benachbarte Codewörter in nur einem Bit (Gray-Code). Tritt eine Störung auf, wird laut Statistik sehr häufig nicht das zugeordnete Codewort, sondern ein Nachbarwort erkannt. Da sich aber beide nur in einem Bit unterscheiden, wird für die Fehler-Erkennung und -Korrektur nur eine Minimalzahl an Bits benötigt. Bei der Modulation wird der Träger unterdrückt, und damit steht mehr Leistung für die Seitenbänder zur Verfügung. Außerdem werden Probleme bei der Demodulation durch schwankende Trägeramplituden und Laufzeiten vermieden. Im Fernsehempfänger muss der Träger wieder phasenrichtig er-

zeugt werden. Dazu werden Synchronisationssignale für den im Empfänger vorhandenen Oszillator übertragen. Als weitere Besonderheit wird das Signal nicht einem einzelnen Träger, sondern 8192 Unterträgern mit einem Abstand von 1,116 kHz aufmoduliert, sog. 8k-Modus. Dieses Verfahren wird mit COFDM, *C*oded *O*rthogonal *F*requency *D*ivision *M*ultiplexing bezeichnet. Es werden aber nur 6817 Unterträger benutzt, von denen wiederum nur 6048 für die eigentlichen Nutzdaten verwendet werden. Jeder der Träger wird mit 4 Bit moduliert. Man erreicht dadurch eine höhere Störsicherheit bei schwankenden Signalamplituden, wie sie bei terrestrischer Übertragung häufiger auftreten.

- *Gleichwellennetz*, **SFN**

Ausgangspunkt ist der Betrieb von räumlich benachbarten Sendern auf der gleichen Frequenz mit synchroner Signalabstrahlung (Gleichwellennetz, *s*ingle *f*requency *n*etwork, SFN). Dann können sich an der Empfängerantenne die Signalamplituden beider Sender durchaus addieren. Diese Addition ist aber praktisch nur dann vorhanden, wenn die Senderamplituden am Empfangsort etwa gleich groß sind. Kritisch wird es dann, wenn die Laufzeit Sender-Empfänger für *beide* Signale so unterschiedlich ist, dass sich aufeinander folgende Signalamplituden überlagern und damit den Empfang des korrekten Signals verfälschen. Dieser Effekt tritt auch auf, wenn von *einem* Sender das direkte Signal und ein reflektiertes Signal empfangen werden und das reflektierte und damit zeitlich verzögerte Signal am Empfänger mit dem bereits neuen Signal zusammentrifft. Um diesem Problem zu begegnen, wurde ein *Schutzintervall* (guard interval) eingeführt. Zwischen der Aussendung zweier benachbarter Signalamplituden wird ein Zeitintervall ohne Signalaussendung eingeschoben, während dessen die verzögerten Signale eintreffen können. Sie verursachen damit aber keine Signalverfälschung. Die Länge des Schutzintervalls legt gleichzeitig fest, wie weit die Sender maximal voneinander entfernt sein dürfen. Dieses Schutzintervall wird häufig zu 1/4 gewählt, d. h., seine Dauer betragt 1/4 der Dauer der Symbole, wobei die Symboldauer diejenige Zeit ist, während der die 6817 Träger ihre Informationen unverändert anliegen haben und aussenden.

- *Statistischer Multiplex*

Der Vorteil der geschilderten Bandbreitenreduzierungen, bei denen die Statistik eine Rolle spielt (mittlerer Bandbreitenbedarf), wird auch dadurch unterstützt, dass sich 4 Programme einen Träger teilen. Die für alle 4 Sender gemeinsam zur Verfügung stehende Bandbreite ist so bemessen, dass sie für den weitaus größten Teil der Übertragung ausreicht (Erfahrungswert, Statistik). Hat ein Sender momentan besonders ausgeprägte Bildinhalts*änderungen*, bekommt er temporär einen größeren Anteil an Bandbreite zugeteilt, die die anderen Programmen in der Regel gerade nicht benötigen. Im Mittel funktioniert diese Aufteilung problemlos. In ganz wenigen Ausnahmefällen kann es vorkommen, dass 3 oder 4 Programme zum gleichen Zeitpunkt über einen etwas längeren Zeitraum viele Bildinhaltswechsel haben. Das führt dann zur sogenannten Klötzchenbildung (Artefakte), im Extremfall zum kurzzeitigen Einfrieren von einem oder mehreren Bildern. Die Programme (Fernsehanstalten) versuchen aber häufig, statt der Klötzchenbildung lieber unscharfe Bilder zu übertragen, weil das Auge und damit die Zuschauer das als weniger störend empfinden.

Tab. 99.2 Daten zum DVB-T (Nach Unterlagen des Norddeutschen Rundfunks, NDR)

Kanalbandbreite (steht den Sendern zur Verfügung)	7 MHz (VHF) bzw. 8 MHz (UHF)
Erforderlich für Bild und Stereoton	(2,5 . . . 4) MBit/s (Bild) und ca. 0,2 MBit/s (Stereoton)
Schutzintervall	1/4 (der Symboldauer)
Coderate	2/3
Modus	8k
Symboldauer	898 µs
Datenrate (8-MHz-Kanal; Coderate 2/3; Schutzintervall 1/4, 8k-Modus)	13,27 Mbit/s
Ergibt bei 4 Fernsehprogrammen	3,31 Mbit/s im Mittel für jedes Programm
Belegte Bandbreite	7,61 MHz
Maximaler Senderabstand	67 km
Gesamt-Symboldauer (Summe aus Schutzintervalldauer und Symboldauer)	1,120 ms
Trägeranzahl (theoretisch), COFDM)	8192
Trägeranzahl (praktisch), COFDM)	6817
Trägerabstand, COFDM	1,116 kHz
Schutzintervalllänge	224 µs
Senderleistung (Fernsehsender Hemmingen bei Hannover des NDR)	5 kW

- *Daten zum DVB-T*

In Tab. 99.2 sind allgemeine Daten von aktuell eingesetzten DVB-T-Systemen aufgeführt.

- *Tonübertragung*

Auch bei der Tonübertragung wurde eine Datenreduktion durchgeführt. Aufgrund von umfangreichen Untersuchungen zum Hörverhalten des menschlichen Ohres haben sich Möglichkeiten zur Irrelevanzreduktion ergeben. Dazu folgen einige Anmerkungen:

Tritt ein lautes Tonsignal auf, nimmt das Ohr frequenzmäßig benachbarte leisere Tonsignale nicht mehr wahr. Also müssen sie auch nicht übertragen werden. Dazu wird der gesamte Frequenzbereich des Tonsignals über Bandpässe in Teilfrequenzblöcke von z. B. 2 kHz zerlegt. Ein lauter Ton bei 3 kHz erlaubt es damit, den Frequenzbereich von 4 bis 6 kHz auszublenden, sofern die darin enthaltenen Amplituden wesentlich kleiner sind.

Die Empfindlichkeit des menschlichen Ohres ist frequenzabhängig und im Frequenzbereich von etwa 200 Hz bis 6 kHz am größten. Damit können tiefere und höhere Frequenzanteile mit weniger Bit codiert werden.

Für die Tonübertragung stehen 4 Kanäle zur Verfügung: Linker Kanal, rechter Kanal, Zusatz-Tonkanal (Kommentar, Fremdsprache) und getrennter Musikkanal. Es besteht die Möglichkeit, weitere Tonsignale im sog. Auxiliary-Data-Bereich (AUX; in den Austastlücken der Fernsehbildübertragung verfügbar) zu senden. Für die Übertragung von „Raumklang“ mit 5 Kanälen

entsprechend 5 zugeordneten Lautsprechern (vor dem Zuhörer: Stereo-Links, Stereo-Rechts, Tiefton-Mitte und hinter dem Zuhörer: Surround-Links, Surround-Rechts) ergibt sich eine Datenreduktion dadurch, dass in den hohen Frequenzen kaum Richtungsinformation steckt und dass sowohl in den Kanälen Stereo-Links-Rechts als auch in Surround-Links-Rechts Redundanz enthalten ist.

Die Codierung des Tonsignals muss wegen des Dynamikumfanges des Ohres mit mindestens 12 bit erfolgen, es werden üblicherweise 16 bit verwendet.

Mehrfachübertragung – Multiplexverfahren

100

Teil XIII

Wilfried Plaßmann



100.1 Zeitmultiplexverfahren 1258
100.2 Frequenzmultiplexverfahren 1259
100.3 Raummultiplexverfahren 1260

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_100

Ziel der Mehrfachübertragung ist es, einen vorhandenen Übertragungskanal optimal zu nutzen. Optimal im strengen Sinn heißt, dass die Bandbreite des Kanals voll ausgenutzt wird und dass es keine Zeiträume gibt, in denen keine Information übertragen wird. In der Praxis versucht man, diesem Ziel so nahe wie möglich zu kommen, weil der Übertragungskanal einen erheblichen Anteil an den Gesamtkosten der Nachrichtenübertragungsstrecke hat. Der Kanal kann drahtlos oder drahtgebunden sein, und der Begriff „gleichzeitig" muss beim Zeitmultiplexverfahren durch „quasi gleichzeitig" ersetzt werden. Für das Verfahren der Mehrfachübertragung müssen die drei folgenden Voraussetzungen erfüllt sein:

1. Das empfangene Gesamtsignal muss sich wieder eindeutig in die Einzelsignale (Basissignale) zerlegen lassen.
2. Die Bandbreite des Übertragungskanals muss größer sein als die Bandbreite *eines* Basissignals.
3. Die zu übertragenden Signale müssen jedes für sich frequenzbandbegrenzt sein, damit das *Shannon'sche Abtasttheorem* nach Abschn. 92.4.10 erfüllt wird.

100.1 Zeitmultiplexverfahren

Das Frequenzspektrum der in Abb. 100.1 dargestellten Signalverläufe u_{S1}, u_{S2} und u_{S3} wurde zunächst über ein Tiefpassfilter auf die höchste sinusförmige Signalfrequenz f_g begrenzt und dann mit einer Frequenz $f_{ab} = 1/\Delta T$ abgetastet, die von f_g und der Anzahl n der Signale (hier 3) abhängt:

$$\textit{Abtastfrequenz} \quad f_{ab} \geq n \cdot 2{,}2 \cdot f_g \qquad (100.1)$$

n Zahl der zu übertragenden Basissignale, f_g höchste in den Signalen enthaltene Frequenz.

Dabei wurde das Abtasttheorem nach Abschn. 92.4.10, (92.18), verwendet. In (100.1) ist berücksichtigt worden, dass, wie in der Praxis üblich, alle Signale mit der gleichen Frequenz abgetastet werden. Diese wiederum wird durch das Signal mit der höchsten zu berücksichtigenden Signalfrequenz vorgegeben. Die zu den diskreten Zeitpunkten abgetasteten Signalamplituden (1, 2, 3, ...) werden über den Schalter S1 nacheinander über den Übertragungskanal übertragen und im Empfänger mit dem Schalter S2 dem entsprechenden Kanal wieder zugeordnet. Man spricht deshalb von „quasi gleichzeitiger" Übertragung. Um die Zuordnung der Schalterstellungen eindeutig zu machen, ist ein Synchronisationssignal u_{syn} erforderlich, das man entweder über eine zusätzliche Schalterstellung überträgt oder das über Zusatzinformationen zusammen mit den Signalamplituden gewonnen wird. Je nach Anforderung an die Störsicherheit der Übertragung muss entweder jede einzelne Schalterstellung eindeutig codiert dem Empfänger übermittelt werden, oder aber es genügt eine Kennung zur Weiterschaltung auf die nächste Schalterstellung bei gelegentlicher eindeutiger Synchronisation der Schalterstellungen. Die Übertragung kann in analoger (Abb. 100.1) oder digital codierter Form erfolgen. Bei der digitalen Codierung bewährt sich ein Verfahren nach Abschn. 96.4.4.7 und Abb. 96.35, das die Gewinnung des im Sender verwendeten Taktes ermöglicht.

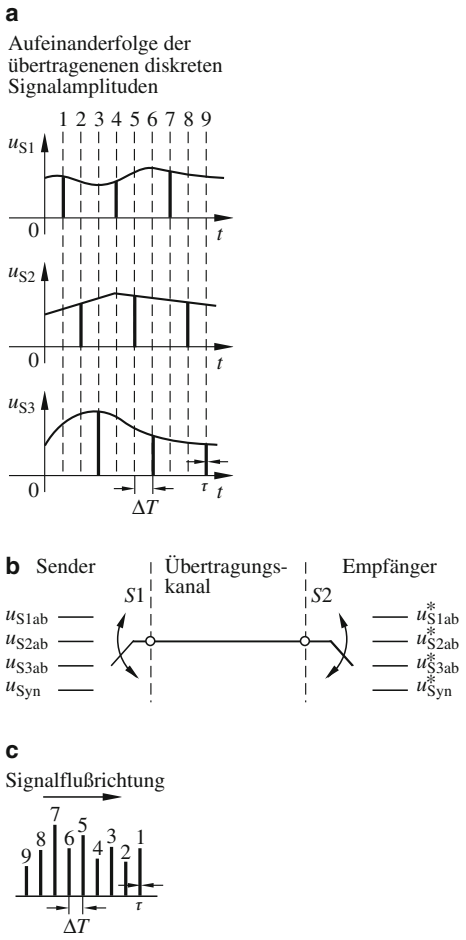


Abb. 100.1 Zeitmultiplexverfahren, analoge Übertragung. **a** Drei Signalverläufe, **b** Prinzip der Übertragung, **c** Signalverlauf auf dem Übertragungskanal

100.1.1 Analoge Signalübertragung

Die für den Übertragungskanal erforderliche minimale Bandbreite bei analoger Signalübertragung berechnet sich zu:

Minimale erforderliche Bandbreite des Kanals:

$$B_{an,\,min} = n \cdot f_g + f_{syn} \qquad (100.2)$$

n Zahl der zu übertragenden Basissignale, f_g höchste in den Basissignalen enthaltene Frequenz (Sinusform) in Hz, f_{syn} Frequenz des Synchronisationssignals in Hz.

Die Frequenz f_{syn} des Synchronisationssignals ergibt sich aus dem verwendeten Übertragungsverfahren. Ein vollständiger Zyklus (Schalterumlauf) ist periodisch mit der Zeit t_{um}:

Zykluszeit eines Schalterumlaufes:

$$t_{um} = n \cdot \tau + n \cdot t_{sch} + t_{syn} \qquad (100.3)$$

n Zahl der zu übertragenden Basissignale, τ Dauer des abgetasteten Signals in s, t_{sch} Zeit zum Weiterschalten auf den nächsten Kanal in s, t_{syn} Zeit für die Schaltersynchronisation in s.

In der Zeit t_{sch} ist neben der reinen Umschaltzeit meist noch ein Sicherheitsabstand enthalten, damit ein ungestörter Übergang auf das nächste Signal sichergestellt ist. Um die Zeit t_{sch} klein zu halten, werden Feldeffekt-Transistor-Schalter mit Schaltzeiten im ns-Bereich eingesetzt.

Die Zahl der zu übertragenden Signale kann bei gegebener Kanalbandbreite auch durch die endliche Dauer t der Abtastimpulse (Abb. 100.1a, c) begrenzt werden, denn es muss gelten:

$$\tau < \Delta T = 1/f_{ab} \tag{100.4}$$

100.1.2 Digital codierte Signalübertragung

Die in Abb. 100.1a dargestellten Amplitudenwerte werden in einem Analog-Digital-Umsetzer in digital codierte Signale umgesetzt und dann über den Kanal übertragen. Die erforderliche minimale Bandbreite lässt sich bestimmen:

Minimale erforderliche Bandbreite des Kanals:

$$B_{di,\,min} = n \cdot k \cdot f_g + f_{syn} \tag{100.5}$$

n Zahl der zu übertragenden Basissignale, k Zahl der Bits pro Amplitudenwert (Abtastwert), f_g höchste in den Basissignalen enthaltene Frequenz (Sinusform) in Hz, f_{syn} Zahl der Synchronisationsbits in $1/s$.

Die Übertragung digital codierter Amplitudenwerte erfordert nach Gleichung (100.5) eine höhere Bandbreite als die analoge Übertragung nach Gleichung (100.2), dafür ist sie aber auch weniger störanfällig. Deshalb findet auch der *Digitale Satellitenrundfunk* (DSR) im unteren GHz-Bereich statt, weil dort die erforderliche große Bandbreite zur Verfügung steht. Auch bei ausreichender Kanalbandbreite muss die Bedingung

$$\Delta T > k \cdot T_{bit} \tag{100.6}$$

(k Zahl der Bits pro Amplitudenwert; T_{bit} Zeitdauer für ein Bit in s) eingehalten werden.

Die Deutsche Telekom verwendet für den Fernsprechverkehr u. a. auch das Zeitmultiplexverfahren. Aus der Vielzahl der verwendeten Systeme seien zwei aufgeführt: 1. *System PCM 30*: Es werden 30 Gespräche gleichzeitig übertragen. Ein *Pulsrahmen* (entspricht einem Schalterumlauf) erzeugt 30 Codewörter zu je 8 Bit, die verschlüsselt die Amplitudeninformationen der 30 Kanäle enthalten. Es werden noch zwei weitere 8-Bit-Codewörter hinzugefügt, die u. a. Informationen über die Kennzeichen der Kanäle und eine Kennung des Rahmens enthalten. Jeder Kanal wird 8000mal pro Sekunde abgetastet, sodass sich ein Informationsfluss von $8000 \cdot 8 \cdot 32\,\text{bit/s} = 2048\,\text{kbit/s}$ ergibt. 2. *System PCM 30720*: Es werden 30.720 Gespräche gleichzeitig über eine Glasfaser von ca. 125 mm Dicke über Hunderte von Kilometern übertragen.

100.2 Frequenzmultiplexverfahren

Das Frequenzmultiplexverfahren wurde und wird von der Deutschen Telekom u. a. in großem Umfang im Fernsprechnetz eingesetzt und soll hier anhand der TF-Technik (*Trägerfrequenz-Technik*) erläutert werden. Dieses Verfahren wurde für die analoge Telefontechnik verwendet. Zwar ist der Fernsprechverkehr in der Zwischenzeit fast vollständig durch die Digitaltechnik mit ihren erweiterten Möglichkeiten ersetzt worden, aber an diesem System lässt sich das Verfahren anschaulich und nachvollziehbar darstellen. Die Zielsetzung war, so viel Gespräche wie möglich über einen Kanal gleichzeitig zu übertragen und dabei eine ausreichende *Sprachverständlichkeit* zu gewährleisten. Die Übertragung der Sprache im Original wurde nicht angestrebt. Der Frequenzbereich für die analoge Gesprächsübermittlung (Basisband) wurde deshalb von 300 Hz bis 3,4 kHz festgelegt, weil Untersuchungen ergeben hatten, dass Frequenzen unterhalb 300 Hz die Sprachverständlichkeit erschweren, Frequenzen oberhalb 3,4 kHz sie *nicht weiter* erhöhen. Man kann den Frequenzbereich oberhalb 3,4 kHz deshalb abschneiden und Kanalbandbreite einsparen. Lediglich die Frequenz 3,85 kHz muss noch übertragen werden, weil sie die Rufnummer enthält. Der Frequenzbereich unterhalb 300 Hz ließ sich für z. B. für den Telefaxverkehr einsetzen.

Die zunächst naheliegende Vorgehensweise, die 3,85 kHz breiten Kanäle den Trägerfrequenzen im Abstand von 4 kHz in Einseitenbandtechnik aufzumodulieren, und zwar mit 12 kHz beginnend bis in den 100-kHz-Bereich, scheitert an den nicht realisierbaren Bandpassfiltern, die z. B. die Frequenz 119,85 kHz (Rufnummer-Frequenz des Sprechkanals 116 kHz bis 119,4 kHz) noch vollständig unterdrücken müssen, 120,00 kHz bis 123,85 kHz durchlassen und 124,00 kHz wieder vollständig unterdrücken müssen. Das Problem lässt sich aber durch die im Folgenden dargestellte Technik lösen.

In Abb. 100.2a sind drei Telefonkanäle dargestellt, wobei auch die Verbindung zum hausinternen Telefon einen Kanal darstellt. Zur Übermittlung des Kennzeichens (Rufnummer) wird die Frequenz 3,85 kHz verwendet. Im folgenden wird das mit BW 52 und BW 7 bezeichnete Verfahren erläutert. Kanal 1 wird einer Trägerschwingung mit 12 kHz, Kanal 2 einer mit 16 kHz und Kanal 3 einer mit 20 kHz in Einseitenband-Amplitudenmodulation (ESB-AM) ohne Träger in Normallage aufmoduliert, sodass das in Abb. 100.2b eingezeichnete Frequenzspektrum dieser *Vorgruppe* von 12,3 bis 23,85 kHz entsteht. In einer zweiten Modulationsstufe werden vier Vorgruppen der Frequenzlage von je 12 bis 24 kHz einer Trägerfrequenz 120 bzw. 108 bzw. 96 bzw. 84 kHz in ESB-AM und Kehrlage aufmoduliert und bilden die sogenannte *Grundprimärgruppe* (Abb. 100.2c). Damit können bereits 12 Gespräche gleichzeitig über einen Übertragungskanal mit dem Frequenzbereich 60,15 bis 107,7 kHz übertragen werden. Bei der Deutschen Telekom werden bzw. wurden noch weitere Gruppen gebildet (Sekundär-,

Tab. 100.1 Bei der Deutschen Telekom eingesetzte Frequenzmultiplexverfahren (Auswahl)

System	Übertragungs-Frequenzband	Zahl der Gespräche (gleichzeitig)	Kabeltyp, Kabeldurchmesser innen/außen, Abstand der Zwischenverstärker
V300	60 bis 1296 kHz	300	Koaxialleitung, 1,2/4,4 mm, 4 km
V960	60 bis 4028 kHz	960	Koaxialleitung, 1,2/4,4 mm, 4 km
V2700	312 bis 12.388 kHz	2700	Koaxialleitung, 2,6/9,5 mm, 4,65 km
V10800	4,332 bis 59,684 MHz	10800	Koaxialleitung, 2,6/9,5 mm, 1,55 km

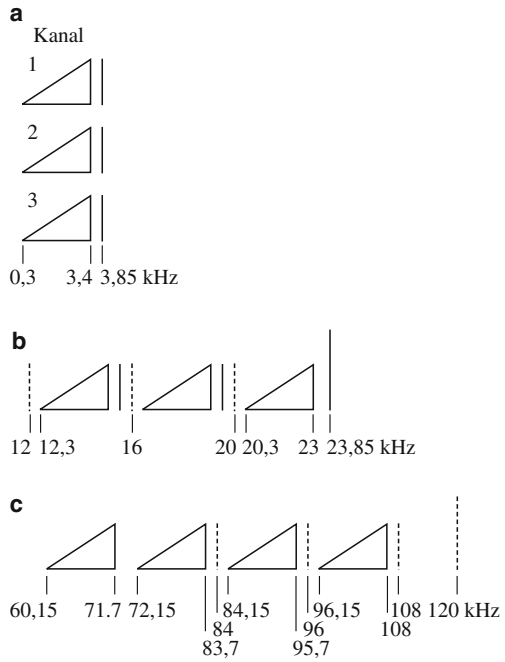


Abb. 100.2 Frequenzmultiplexverfahren am Beispiel des Fernsprechnetzes der Deutschen Telekom, unterdrückte Träger *gestrichelt* dargestellt. **a** Drei Fernsprechkanäle, **b** Bildung der Vorgruppe, **c** Grundprimärgruppe

Tertiär-, Quartärgruppen), sodass für die Übertragung u. a. folgende Systeme zur Verfügung stehen (Tab. 100.1).

Dabei ist es möglich, den beim Verfahren V10800 nicht benutzten Frequenzbereich bis 4,332 MHz durch ein System V960 zu belegen. Der dann noch verbleibende freie Bereich von 0 bis 60 kHz wird von 6 bis 54 kHz durch eine der Trägerfrequenz 114 kHz aufmodulierte Grundprimärgruppe belegt. Die Bildung von Gruppen und damit die Mehrfachmodulationen (Träger 12, 16, 20 kHz, dann 120, 108, 96, 84 kHz) ist erforderlich, damit einerseits die Fernsprechkanäle dicht beieinander liegen (Abstand 150 Hz bei der Vorgruppe, 450 Hz bei der Grundprimärgruppe) und so möglichst wenig Kanalbandbreite ungenutzt bleibt, andererseits die einzelnen Kanäle wieder eindeutig voneinander getrennt werden können. Außerdem ist die Einseitenbandfilterung in der Grundprimärgruppe erst möglich, weil die Seitenbänder 24 kHz auseinanderliegen (deshalb auch die niedrigste Trägerfrequenz 12 kHz in der Vorgruppe und nicht 8 oder 4 kHz). Verallgemeinert gilt für das Frequenzmultiplexverfahren, dass es dann einsetzbar ist, wenn der vorhandene Übertragungskanal (z. B. Koaxialkabel) eine viel größere Bandbreite besitzt als *ein* Basissignal.

Die erforderliche Kanalbandbreite B_K ergibt sich aus den hier als gleich angenommenen Bandbreiten B_S der Basissignale:

$$B_K = n \cdot B_S + n \cdot \Delta f$$

B_S Signalbandbreite eines Basissignals in Hz; n Zahl der Basissignale; Δf Frequenzabstand zwischen zwei Basissignalen nach der Modulation zur Kanaltrennung in Hz.

100.3 Raummultiplexverfahren

Beim Raummultiplexverfahren (SDM, *S*pace *D*ivision *M*ultiplex) werden *unterschiedliche* Signale über *unterschiedliche* Übertragungswege übertragen, und zwar *gleichzeitig*. Dieses Verfahren wird vielfach eingesetzt, auch wenn es zunächst nicht einsichtig erscheint. Es wird unterschieden zwischen drahtgebundenem und drahtlosem (kabellosem) Raummultiplexverfahren.

- *Drahtgebundenes Raummultiplexverfahren*: Eingesetzt wird dieses Verfahren in großem Rahmen in den Koppelnetzen der Telefon-Vermittlungstechnik, siehe dazu auch Abb. 106.3 im Kap. 106 „Vermittlungstechnik“. Es können m ankommende Signale auf n abgehende Signale gleichzeitig weitergeleitet werden, wobei Überschneidungen unzulässig sind. Im ganz kleinen Rahmen wird es verwendet in Rundfunk- und Fernsehstudios, wo die Signale von Mikrofonen, Fernsehkameras und Bild- und Tonträgern auf laufende Programme und andere Bild- und Tonträger übertragen werden müssen.
- *Drahtloses Raummultiplexverfahren*: Bei diesem Verfahren wird eine Funkverbindung in mehreren räumlich weit voneinander entfernten Regionen auf gleicher Frequenz betrieben. Beim UKW-Rundfunk und beim terrestrischen Fernsehen werden gleiche oder auch unterschiedliche Programme auf der gleichen Frequenz übertragen. Das ist auch auf nationalem Gebiet möglich, weil die Reichweite der verwendeten Frequenzen (ab etwa 85 MHz) nur bedingt über die optische Sicht hinausgeht und damit gegenseitige Beeinflussungen (Störungen) minimal gehalten werden können. Das Raummultiplexverfahren wird weiterhin angewendet im Mobilfunk. Die Leistung der Handys ist vergleichsweise niedrig und die Leistung der Sender wird bewusst niedrig gehalten. Damit können die Senderabstände gering gehalten und die gleichen Frequenzen vielfach vergeben werden. Auch Richtfunkstrecken verwenden das Multiplexverfahren. Allerdings ist die gegenseitige Beeinflussung durch die ausgeprägte Richtwirkung vom Prinzip her schon gering.

Diese Technik wird bewusst eingesetzt, weil andernfalls nicht genügend Frequenzbereiche zur Verfügung stehen.

Richtfunktechnik

Wilfried Plaßmann

101

Teil XIII



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_101

Die Nachrichtenübertragung über Richtfunkstrecken wird neben der Übertragung über Kabel eingesetzt. Gegenüber der Rundfunk- und Fernsehübertragung weist sie folgende Unterschiede auf:

1. Sender und Empfänger sind ortsfest. Da es nur einen Empfänger gibt, kann die Übertragung mit einem gerichteten elektromagnetischen Feld erfolgen.
2. Der verwendete Frequenzbereich liegt zurzeit zwischen 300 MHz und 30 GHz. Damit werden Parabolantennen als Richtantennen eingesetzt, weil sie gute Richteigenschaften und handhabbare Abmessungen (Durchmesser $< 4{,}5$ m) aufweisen.
3. Die Übertragung geschieht quasioptisch, sodass Entfernungen bis maximal etwa 100 km überbrückbar sind. Bei größeren Entfernungen werden Zwischenstationen eingesetzt.

Durch diese Art der Übertragung ergeben sich folgende Vorteile:

1. Die verwendete Frequenz kann mehrfach benutzt werden, sofern die Richtfunkstrecken räumlich voneinander getrennt sind.
2. Die Richtfunkstrecke kann parallel zur Kabelübertragung eingesetzt werden und vermindert damit die Ausfallrate der Übertragungsstrecke.
3. Sie wird dort eingesetzt, wo eine Verlegung von Kabeln aus geografischen oder politischen Gründen nicht möglich ist. So wurde der Telefonverkehr mit Berlin während der Existenz der DDR über Richtfunkstrecken abgewickelt (Torfhaus–Berlin).
4. Richtfunkstrecken erfordern relativ wenig Aufwand und sind deshalb auch für mobile Anlagen geeignet. Häufig kann ein vorhandener hoher Turm für die Antennenmontage mitverwendet werden, sodass dadurch die Reichweite erhöht wird.

Der verwendete Frequenzbereich von 300 MHz bis 30 GHz ist aus folgenden Gründen besonders gut geeignet:

1. Durch die quasioptische Ausbreitung und die Bündelung treten wenig Überlagerungen mit anderen Signalen auf.
2. Das kosmische Rauschen ist genau in diesem Bereich besonders gering („Mikrowellenfenster“).
3. Die Dämpfung der Atmosphäre ist gering, und an den Schichten der Ionosphäre finden kaum Reflexionen statt.
4. Die Beugung und Reflexion an Regentropfen ist noch gering ($f = 10\,\text{GHz} \mathrel{\widehat{=}} \lambda = 3\,\text{cm}$).
5. Die Antennenabmessungen sind handlich.

Ein wesentlicher Nachteil ist die *Zusatzdämpfung* bei starken Niederschlägen, die ab etwa 10 GHz einsetzt.

Zur Erhöhung der Verfügbarkeit einer Übertragungsstrecke, die durch Schwunderscheinungen (Überlagerung von zwei Feldern mit 180° Phasenverschiebung) verringert wird, verwendet man:

- *Frequenzdiversity*: Es wird das gleiche Signal über zwei parallele Übertragungskanäle mit unterschiedlichen Frequenzen übertragen.
- *Raumdiversity*: Die Übertragung geschieht räumlich auf zwei unterschiedlichen Wegen, wobei auch die Empfänger örtlich voneinander getrennt sind.

Einige weitere für die Übertragung wichtige Formeln und Begriffe sind hier aufgelistet.

- *Freiraumdämpfung*: Dies ist die Dämpfung zwischen Sende- und Empfangsantenne, wenn beide als Kugelstrahler ausgebildet sind:
  Freiraumdämpfungsmaß

$$a_0 = 20 \cdot \lg \frac{4\pi l}{\lambda} \quad \text{in dB} \tag{101.1}$$

  l Entfernung in m, λ Wellenlänge in m.
- *Schwund*: Störungen der Ausbreitungsbedingungen. *Interferenzschwund*: Überlagerung von direkt empfangener und reflektierter Welle und damit teilweise oder totale Auslöschung. *Absorptions- und Streuverluste*: Streuung und Dämpfung durch Wassertropfen und Wasserdampf, besonders für Frequenzen ab 10 GHz.
- *Antennen*: Es werden hauptsächlich Parabolantennen mit Durchmessern $< 4{,}5$ m eingesetzt. Unter *Halbwertsbreite* versteht man den Winkel gegenüber der Hauptstrahlrichtung, bei der der Antennengewinn auf die Hälfte (−3 dB) abgesunken ist.
  Bei Parabolantennen liegt der Winkel zwischen 0,5° und 4°. Die Halbwertsbreite sinkt mit steigendem Antennendurchmesser, sodass große Antennen eine größere Richtwirkung besitzen als kleine. (Siehe auch Abschn. 95.3.10 „Parabolantennen“).
- *Antennengewinnmaß*: Angegeben wird das logarithmische Maß (siehe auch Kap. 95 „Antennen“, (95.10a)–(95.10c)):
  Antennengewinnmaß

$$G = 10 \cdot \lg \frac{4r^2\pi^2 q}{\lambda^2}\,\text{dB} \tag{101.2}$$

  r Radius der Antenne in m, q Flächenausnutzungsfaktor, $q \approx 0{,}6$, λ Wellenlänge in m.
  Dieses Antennengewinnmaß gilt sowohl für die Sende- als auch für die Empfangsantenne.
- *Gesamt-Systemdämpfung*: Das Gesamtdämpfungsmaß a_{ges} zwischen Sender und Empfänger lässt sich wie folgt berechnen:
  Gesamt-Systemdämpfungsmaß:

$$a_{\text{ges}} = a_0 + a_L - (G_S + G_E) \tag{101.3}$$

  a_0 Freiraumdämpfungsmaß in dB nach (101.1), a_L Zuleitungsdämpfungsmaß in dB, G_S, G_E Antennengewinnmaß der Sende- bzw. Empfangsantenne nach (101.2).
  Weitere Einzelheiten sind im Beispiel 1 aufgeführt.
- *Scatterverbindung*: Bei dieser auch als *Überreichweitenanlage* bezeichneten Richtfunkstrecke besteht zwischen den zwei Richtfunkstellen keine optische Sicht. Der Einsatz von

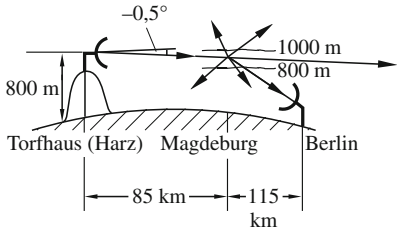


Abb. 101.1 Scatterverbindung im Telefonverkehr Bundesrepublik–Westberlin

Relaisstationen ist aus geografischen oder politischen Gründen nicht möglich. Dann richtet man beide Antennen nach Abb. 101.1 auf einen gemeinsamen Punkt in der Troposphäre und nutzt die dort auftretenden Brechwertinhomogenitäten aus, die einen geringen Teil der Energie zur jeweils anderen Antenne reflektieren. Da dieser Anteil sehr klein ist, sind relativ hohe Sendeleistungen erforderlich. Dieses Verfahren wurde im Telefonverkehr zwischen der Bundesrepublik und Westberlin während der Existenz der DDR eingesetzt. Die Entfernung beträgt etwa 200 km. Auf dem Torfhaus (Harz) wurde in 800 m Höhe über NN unter einem Winkel von –0,5° in Richtung Berlin gesendet. Etwa über Magdeburg traten Beugungen in 800 ... 1000 m Höhe auf, sodass ein kleiner Teil des Hauptstrahles auf die Empfangsantenne in Berlin umgelenkt wurde. Das Gleiche fand in umgekehrter Richtung statt. Die Sendeleistung, die bei optischer Sicht mit etwa 5 W ausgereicht hätte, musste auf 1 kW erhöht werden. Sende- und Empfangsantennen hatten bei einer Frequenz im Bereich von 2,1 ... 2,3 GHz einen Durchmesser von 10 m bzw. im Frequenzbereich von 1,7 ... 2,1 GHz einen Durchmesser von 18 m. Es wurde mit Raum-Diversity (s. o.) gearbeitet.

Beispiel 1

Ein Sender hat am Ausgang eine Leistung von 1 W bei einer Frequenz $f = 12\,\text{GHz}$ ($\lambda = 2{,}5\,\text{cm}$) und speist über eine Zuleitung der Länge $l_L = 10\,\text{m}$ und dem Dämpfungsmaß $a_L^* = 12\,\text{dB}/100\,\text{m}$ eine Parabolantenne mit einem Durchmesser von 2 m. In einer Entfernung $l = 15\,\text{km}$ steht die Empfangsantenne mit dem gleichen Durchmesser. Mit $q = 0{,}6$ und (101.3) folgt:

$$\begin{aligned} a_{\text{ges}} &= 20 \cdot \lg \frac{4\pi l}{\lambda} + \frac{12\,\text{dB}}{100\,\text{m}} \cdot 10\,\text{m} - 2 \cdot 10 \cdot \lg \frac{4r^2\pi^2 q}{\lambda^2} \\ &= 20 \cdot \lg \frac{4 \cdot \pi \cdot 15 \cdot 10^3\,\text{m}}{2{,}5 \cdot 10^{-2}\,\text{m}}\,\text{dB} + 1{,}2\,\text{dB} \\ &\quad - 20 \cdot \lg \frac{4 \cdot 1\,\text{m}^2 \cdot \pi^2 \cdot 0{,}6}{2{,}5^2 \cdot 10^{-4}\,\text{m}^2} \\ &= 137{,}5\,\text{dB} + 1{,}2\,\text{dB} - 91{,}6\,\text{dB} = 47{,}1\,\text{dB} \end{aligned}$$

Die Leistung P_E an der Empfängerantenne ergibt sich zu:

$$47{,}1\,\text{dB} = 10 \cdot \lg \frac{1\,\text{W}}{P_E} \quad \Rightarrow \quad P_E = \frac{1\,\text{W}}{10^{4{,}71}} = 19{,}5\,\mu\text{W}.$$

◀

Heute werden von der Telekom überwiegend **D***igital*-**R***icht*-*funk***s***ysteme* (*DRS*) eingesetzt. Ein solches System ist z. B. DRS 34/15000. Es ist ein PCM-System mit **P**uls-**C**ode-**M**odulation und einer Übertragungsrate von 34 Mbit/s und überträgt im 15-GHz-Bereich. Damit lassen sich etwa 500 Fernsprechkanäle gleichzeitig übertragen.

Nachrichtenübertragung über Satellit

102

Teil XIII

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_102

Satelliten werden in ca. 36.000 km Höhe in eine geostationäre Umlaufbahn um die Erde gebracht, sodass sie die gleiche Umlaufzeit wie die Erde haben. Allerdings ist nur eine der unendlich vielen möglichen Umlaufbahnen optimal, nämlich die, bei der sich der Satellit wie die Erde dreht, von West nach Ost, über dem Äquator positioniert (*erd-* oder *geosynchrone Umlaufbahn*). Der Satellit ändert seine Position gegenüber der Erdoberfläche nicht. Die Erde erscheint dem Satelliten unter einem Winkel von 17,15 Grad, und er erfasst etwa 42 % der Erdoberfläche. Allerdings ist seine Nutzung bei flachen Einfallswinkeln kaum möglich. Es werden heute ausschließlich *aktive* Satelliten eingesetzt, die über eine eigene Empfangs- und Sendeanlage verfügen. Die in der Anfangszeit verwendeten *passiven* Satelliten hatten nur Reflektoren. Eine Nachrichtenübertragung über Satellit besteht aus einer oder mehreren Erdefunkstellen und dem Satelliten. Die Erdefunkstelle unterscheidet sich von einer Richtfunkstelle (siehe Kap. 101) durch die folgenden Punkte:

- eine nachführbare Antenne mit hohem Gewinn, da der Satellit nicht exakt ortsfest ist und die Entfernung ca. 36.000 km beträgt;
- eine hohe Sendeleistung wegen der großen Entfernung;
- aufwendige Empfängerschaltung wegen der geringen empfangenen Leistung.

Bei der Übertragung gibt es zwei Möglichkeiten:

- *Punkt-zu-Punkt-Verbindung*: Der Satellit wird als Relaisstation benutzt. Beispiel: Verbindung Europa–Amerika.
- *Vielfachzugang*: Mehrere Bodenstationen benutzen die gleiche Trägerfrequenz, aber das Basisband von z. B. 0 ... 500 MHz wird bereichsweise auf die Stationen aufgeteilt. Es wird das Frequenzmultiplexverfahren angewendet und einseitenbandmoduliert mit unterdrücktem Träger.

Für die Übertragung über Satellit ist der Frequenzbereich 300 MHz bis 30 GHz besonders geeignet, siehe Kap. 101 „Richtfunktechnik“. Dadurch ist eine Signalübertragung trotz hoher Freiraumdämpfung durch die große Entfernung mit vertretbarem Empfängeraufwand möglich.

Beispiel 1

Das Freiraumdämpfungsmaß ergibt nach (101.1) bei einer Frequenz $f = 10\,\text{GHz}$ ($\hat{=}\ \lambda = 3\,\text{cm}$) und einer Entfernung $l = 36.000\,\text{km}$:

$$a_0 = 20 \cdot \lg \frac{4\pi l}{\lambda} = 20 \cdot \lg \frac{4\pi \cdot 3{,}6 \cdot 10^9\,\text{cm}}{3\,\text{cm}} \approx 203\,\text{dB}.$$

Benutzt die Erdefunkstelle eine Antenne mit dem Radius $r_E = 12{,}5\,\text{m}$ (u. a. in Raisting bei München verwendet) und der Satellit eine mit $r_S = 1\,\text{m}$, ergeben sich die Antennengewinnmaße nach (101.2) mit $q = 0{,}6$ zu:

$$\textit{Antennengewinnmaß} \quad G = 10 \cdot \lg \frac{4r^2\pi^2 q}{\lambda^2}\,\text{dB}$$

r Radius der Antenne in m, q Flächenausnutzungsfaktor, $q \approx 0{,}6$, λ Wellenlänge in m.

$$G_E = 10 \cdot \lg \frac{4 \cdot 12{,}5^2\,\text{m}^2 \cdot \pi^2 \cdot 0{,}6}{3^2\,\text{cm}^2} = 66{,}1\,\text{dB}$$

$$G_S = 10 \cdot \lg \frac{4 \cdot 1\,\text{m}^2 \cdot \pi^2 \cdot 0{,}6}{3^2\,\text{cm}^2} = 44{,}2\,\text{dB}$$

$$G_E + G_S = 110{,}3\,\text{dB}$$

Resultierend erhält man ein Gesamtdämpfungsmaß von $203\,\text{dB} - 110{,}3\,\text{dB} = 92{,}7\,\text{dB}$. Wird von einer Antenne die Leistung 1 kW abgegeben, empfängt die andere Antenne bei Vernachlässigung weiterer Dämpfungen die Leistung P_e

$$92{,}7\,\text{dB} = 10 \cdot \lg \frac{1\,\text{kW}}{P_e} \quad \Rightarrow \quad P_e = 540\,\text{nW}. \quad \blacktriangleleft$$

Beispiel 1 zeigt, dass in der Erdefunkstelle Antennen mit möglichst großen Durchmessern verwendet werden müssen, und dass trotz hoher Senderleistung des Satelliten an die Empfänger auf der Erde hohe Anforderungen gestellt werden.

Intelsat betreibt derzeit mehr als 15 Satelliten, die mit über 800 über die Erdoberfläche verteilten Antennen erreicht werden können. Etwa 2/3 des interkontinentalen Fernsprechverkehrs und alle interkontinentalen Live-Sendungen des Fernsehens benutzen die Satellitenverbindung. Auch regionale und nationale Nachrichtenübertragungen werden zunehmend über Satelliten abgewickelt. Für Fernseh- und Rundfunkprogramme gibt es preiswerte Satellitenempfangsanlagen, die eine Alternative zum Programmempfang über Kabel darstellen und eine Fülle an unterschiedlichen Programmen bieten. Da die Satellitensender für Rundfunk- und Fernsehprogramme mit einer mehr als 10fachen Strahlungsleistung gegenüber den übrigen Nachrichtensatelliten senden, ist auf der Erde auch mit geringem Aufwand ein störungsfreier Empfang möglich.

Die Deutsche Bundespost hat 1989 einen nationalen Fernmeldesatelliten *DFS-Kopernikus* mit folgenden Zielen in Betrieb genommen:

- Fernsprech- und Datenübertragung zwischen der Bundesrepublik und Westberlin: 2 Kanäle mit je 90 MHz Bandbreite auf 14/12 GHz und ein Kanal mit ebenfalls 90 MHz Bandbreite auf 30/20 GHz.
- Fernsehübertragung zwischen der Bundesrepublik und Westberlin mit einem Kanal mit 90 MHz Bandbreite auf 14/12 GHz.
- Flächendeckende Übertragung von Fernsehprogrammen zur Einspeisung in örtliche Kabelnetze auf 5 Kanälen mit 44 MHz Bandbreite auf 14/12 GHz.
- Schnelle Datenübertragung und Videokonferenzübertragung auf 2 Kanälen mit je 44 MHz Bandbreite auf 14/12 GHz.

Durch die Auflösung der DDR hat sich das Konzept in der Zwischenzeit geändert. Der Satellit wird vorwiegend zur Übertragung von Fernsehprogrammen eingesetzt.

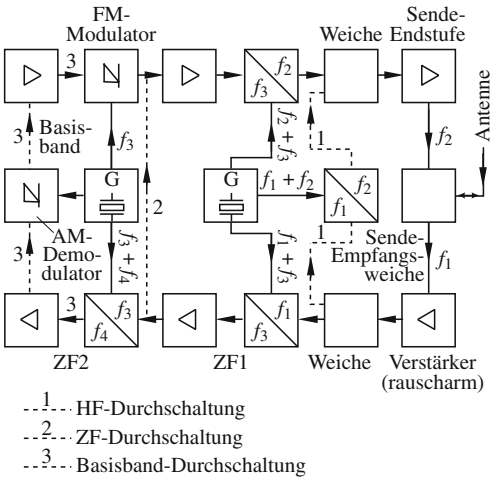


Abb. 102.1 Prinzipschaltbild eines Transponders

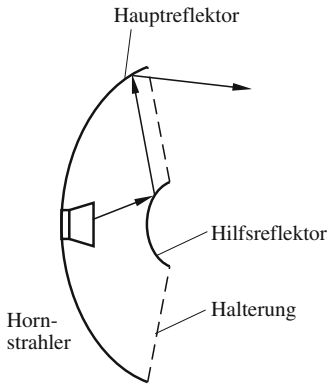


Abb. 102.2 Cassegrain-Antenne

Aus den Einsatzgebieten für den Satellitenfunk seien einige wenige genannt: Der europäische Wettersatellit *Meteosat* gehört zu einem erdumspannenden Netz aus Wettersatelliten, deren Informationen allen Ländern zur Verfügung stehen. Die konventionelle Nachrichtenübertragung zwischen Schiffen und Landstationen über Kurzwelle wird durch Satellitenübertragung ersetzt. Einige weitere Anwendungen folgen unten.

Geostationäre Satelliten haben einige Vorteile:

- Ein erdumspannendes Kommunikationssystem kann mit nur drei Satelliten bei einer gegenseitigen Verschiebung von 120° erreicht werden.
- Nachrichtenverbindungen können mit erdfesten und mobilen Stationen aufgebaut werden.
- Es können auch größere Regionen flächendeckend erreicht werden.

Allerdings gibt es auch einige Nachteile:

- Das Erreichen der Polgebiete ist wegen des geringen Neigungswinkels nicht möglich.
- Die Signallaufzeit zwischen zwei Erdefunkstellen liegt zwischen 240 und 275 ms. Wendet man das *Raumdiversity-Verfahren* (Kap. 101 „Richtfunktechnik“) an, kann es zu Synchronisationsproblemen kommen.
- Während des Durchlaufes durch den Erdschatten muss die Stromversorgung der Satelliten von Batterien übernommen werden.

- *Transponder*: Der interne elektronische Aufbau von Satelliten ist u. a. so ausgelegt, dass er zwischen drei unterschiedlichen Betriebsarten umschalten kann, Abb. 102.1.
  - *Hochfrequenzdurchschaltung*: Die von der Erdefunkstelle empfangene Frequenz (z. B. 6 GHz) wird in einem Frequenzumsetzer in eine *niedrigere* Frequenz (z. B. 4 GHz) umgesetzt und zur Erde zurückgesendet. Im Satelliten lassen sich niedrigere Sendefrequenzen mit höherem Wirkungsgrad verarbeiten. Bei niedrigeren Frequenzen nimmt das Freiraumdämpfungsmaß ab, das Antennengewinnmaß ebenso, sodass das Gesamtdämpfungsmaß konstant bleibt (siehe Beispiel 1: nur λ ändert sich).
  - *Zwischenfrequenzdurchschaltung*: Das empfangene Signal wird in eine deutlich niedrigere Frequenz, die erste Zwischenfrequenz (ZF1) umgesetzt und verstärkt. Das ist vorteilhaft, weil die niedrigere ZF1 einfacher zu handhaben ist.
  - *Basisband-Durchschaltung*: Das Ausgangssignal der 1. ZF-Stufe wird erneut umgesetzt (ZF2), verstärkt und demoduliert, sodass man das Basisbandsignal zur Verfügung hat. Dieses Verfahren ist dann erforderlich, wenn z. B. die Modulationsart geändert werden soll: Empfang eines einseitenbandmodulierten Signals mit unterdrücktem Träger und Aussendung eines frequenzmodulierten Signals. Neuere Satelliten haben mehr als 10 Transponder eingebaut, die die Sendestufe gemeinsam ansteuern.
- *Antennen*: Verwendet werden vorwiegend *Cassegrain-Antennen* (Doppelreflektorantennen), bei denen das Signal nach Abb. 102.2 über einen kleinen Hornstrahler in der Mitte des Hauptreflektors in Richtung auf einen hyperbolischen Hilfsreflektor eingespeist wird, der das Signal zum Hauptreflektor zurückwirft, von wo aus es endgültig die Antenne verlässt. Durch diesen Aufbau werden Antennenwirkungsgrade von über 65 % erreicht.
- *Polarisation*: Durch horizontale bzw. vertikale Polarisation lassen sich auf der gleichen Trägerfrequenz zwei getrennte Signale übertragen. Ebenso kann man bei der zirkularen Polarisation zwischen links- und rechtsdrehend unterscheiden.
- *Energieversorgung*: Die Energieversorgung geschieht überwiegend mit Solarzellen, wobei Akkumulatoren die Speisung während des Durchlaufes durch den Erdschatten übernehmen. Über die Sonnenstrahlung steht eine Leistung von 25 … 50 W je m^2 Solarzellenfläche zur Verfügung. Die für den Betrieb des Satelliten erforderliche Leistung kann je nach Typ bei maximal 2 … 5 kW liegen, sodass Flächen von bis zu 80 m^2 benötigt werden. Das lässt sich durch ausklappbare Paddel erreichen, die so ausgerichtet werden, dass sie stets optimal zur Sonne stehen.
- *DSR – **D**igitales **S**atelliten**r**adio*: Übertragung von Hörfunkprogrammen in CD-Qualität durch Übertragung der Information in digital kodierter Form.

Daten: Frequenzbereich: 11,7 ... 12,5 GHz; 40 Kanäle mit je 27 MHz Bandbreite; Kanalabstand etwa 19,2 MHz; durch zirkulare kanalweise wechselnde Polarisation keine Störung benachbarter Kanäle; Frequenzmodulation.

- *Footprint* eines Satelliten: Darstellung der vom Satelliten auf die einzelnen Punkte der Erdoberfläche abgestrahlten Leistung einschließlich des Antennengewinns. Es werden Linien gleicher Leistung eingetragen. Damit kann die Leistungsabnahme vom Zentrum zu den Außenzonen abgelesen werden. Die Beschriftung dieser Linien kann z. B. in *EIRP*, *equivalent isotropically radiated power*, in logarithmischem Maßstab mit der „Einheit" dBW erfolgen. Für den Satelliten Astra-1D gilt im Zentrum ein Wert von 51 dBW. Daraus lässt sich die Verfügbarkeit eines Programmes bzw. die erforderliche Größe der Empfangsantenne bestimmen. Alternativ dazu kann das Diagramm auch mit der auf der Erdoberfläche herrschenden Leistungsflussdichte beschriftet werden (*PFD*), da EIRP und PFD in einem festen Verhältnis zueinander stehen. Dieser ebenfalls logarithmische Wert wird in dB(W/m^2) angegeben.

Für eine gute Fernsehbildqualität ist ein Signal-Rausch-Verhältnis $S/N \geq 38$ dB erforderlich. Bezieht man den Gewinn durch Frequenzmodulation, G_{FM}, mit ein, folgt für das Träger-Rausch-Verhältnis C/N:

$$\frac{C}{N} = \frac{S}{N} - G_{\mathrm{FM}} \tag{102.1}$$

Setzt man als Näherungswert

$$G_{\mathrm{FM}} \approx 26\,\mathrm{dB} \tag{102.2}$$

ein, ergibt sich für eine gute Fernsehbildqualität:

$$\frac{C}{N} \geq 12\,\mathrm{dB}. \tag{102.3}$$

Es gilt:

$$\frac{C}{N} = \frac{G}{T} \cdot P_0 \cdot \frac{\lambda^2}{4 \cdot \pi \cdot k \cdot B}$$

bzw.

$$\frac{C}{N} = 10 \cdot \lg\{\frac{G_e}{T} \cdot P_0 \cdot \frac{\lambda^2}{4 \cdot \pi \cdot k \cdot B}\} \quad \text{in dB} \tag{102.4}$$

P_0 Leistungsflussdichte in W/m^2; λ Wellenlänge in m; k Boltzmann-Konstante: $1{,}38 \cdot 10^{-23}$ W s/K; B Bandbreite in Hz (hier meist 27 MHz).

Mit der Güte G_e/T der Empfangsanlage in „Gewinn pro Kelvin":

$$\frac{G_e}{T} = \frac{G \cdot a}{a \cdot T_A + (1-a) \cdot T_u + (F-1) \cdot T_u} \quad \text{in 1/K} \tag{102.5}$$

G Antennengewinn; a Wirkungsgrad, $a \approx 0{,}95 \ldots 0{,}97$; T_A Rauschtemperatur der Antenne in K, $T_A \approx 50$ K; T_u Umgebungstemperatur in K, $T_u \approx 293$ K; F Rauschzahl des Empfängers (LNB); G siehe (102.6).

In (102.4) und (102.5) sind lediglich die Faktoren G und F beeinflussbar. Danach kann ein niedriges Signal-Rausch-Verhältnis letztlich nur über einen entsprechend großen Antennenradius r und eine niedrige Rauschzahl des Empfängers (mit LNB, low noise block, bezeichnet) erreicht werden (101.2):

$$G = q \left(\frac{2 \cdot \pi \cdot r}{\lambda}\right)^2 \tag{102.6}$$

r Radius der Antenne in m; q Flächenausnutzungsfaktor, $q \approx 0{,}6$; λ Wellenlänge in m.

- *GPS*: **G***lobal* **P***ositioning* **S***ystem*. Dabei handelt es sich um ein satellitengestütztes System zur Positionsbestimmung. Weitere Einzelheiten dazu sind im Kap. 104 „Funkmesstechnik" aufgeführt.
- *Fernsehbild- und Rundfunk-Empfang über Satelliten der ASTRA-Serie*. Für die Übertragung von Rundfunk- und Fernsehprogrammen werden u. a. Satelliten („Fernsehsatelliten") der ASTRA-Serie eingesetzt, die von SES ASTRA S. A. in Betzdorf, Luxemburg, betrieben werden. Über diese Satelliten werden die Programme der deutschen Öffentlich-Rechtlichen Rundfunk- und Fernsehanstalten vorwiegend übertragen. Begonnen wurde mit den Satelliten 1A (1988 bis 2004) und 1B (1991 bis 2006), die inzwischen abgeschaltet worden sind. Als Nachfolger sind u. a. die Satelliten 1H und 1M auf der geostationären Position 19,2° Ost aktiv. Diese Position ist zur Versorgung von Westeuropa besonders gut geeignet. Übertragen wird im Frequenzbereich von 10,7 bis 12,75 GHz, der in 7 Bänder (A bis G) aufgeteilt ist: A bis D → Low-Band, E bis G → High-Band. Zusätzlich wurde das Ka-Band im Frequenzbereich 18,3 bis 18,8 GHz) eingerichtet, auf dem die Satelliten der neueren Generation (z. B. 1H und 1L) nicht nur, aber auch, senden. Jedem Satelliten sind aus diesem Spektrum 3 bis 5 Bänder zugeteilt. Die Satelliten haben ein Gewicht im Bereich von etwa 3500 bis 6500 kg, eine Ausgangsleistung bis 200 W, Leistungsverbrauch der Satelliten intern in der Größenordnung 6 bis 10 kW, erwartete Lebensdauer ca. 15 Jahre. Die Satelliten haben Antriebsaggregate, die sie einerseits genau auf der geostationären Position halten, andererseits das Einschwenken in eine neue Position ermöglichen. Aktuelle Daten zu den einzelnen Satelliten kann man erfahren über die Internetseite www.ses-astra.com der oben genannten Betreiberfirma.
- *Intelsat*: Im Jahr 1965 wurde Intelsat gegründet mit der Zielsetzung, durch den Einsatz von Satelliten die Datenkommunikation im Sinne von Daten-, Bild- und Ton-Übertragung weltweit ohne merkliche Verzögerung zu ermöglichen. Der erste medienwirksame Einsatz erfolgte mit der direkt übertragenen Mondlandung der drei amerikanischen Astronauten im Jahr 1969. Heute werden Datenkanäle vermietet, um z. B. Videokonferenzen weltweit in Echtzeit durchzuführen oder Korrespondentenberichte zeitnah und aktuell zu den entsprechenden Fernsehsendungen zu übermitteln. Auch Fernsehprogramme werden übertragen, wenn auch nicht unbedingt die Sendungen der Öffentlich-Rechtlichen Fernsehanstalten.

Nachrichtenübertragung über Lichtwellenleiter (LWL)

103

Teil XIII

Wilfried Plaßmann



103.1 Physikalische Grundlagen 1270
103.2 Grundmodelle von Lichtwellenleitern 1271
103.3 Technische Ausführung von Lichtwellenleitern 1273
103.4 Lichtsender 1274
103.5 Lichtempfänger 1274
103.6 Verbinden von Lichtleitern 1275
103.7 Optische Schalter 1275

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_103

Die optische Nachrichtenübertragung gewinnt an Bedeutung, weil sie nicht durch elektromagnetische Störfelder beeinflusst wird. Sender- und Empfängerkreis sind außerdem galvanisch voneinander getrennt, und es stehen Impulsraten zur Verfügung, die bei Kupferkabeln nur mit wesentlich höheren Kosten erreicht werden können. Zur Übertragung steht theoretisch der Bereich des sichtbaren Lichtes mit dem sich anschließenden nahen Infrarotbereich zur Verfügung. Praktisch aber wird dieser Bereich auf drei Fenster eingeschränkt, siehe Punkt „Dämpfung“.

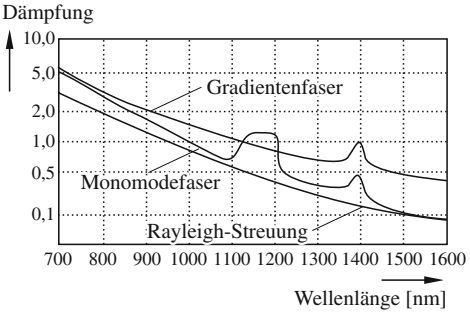


Abb. 103.2 Dämpfung als Funktion der Wellenlänge für Quarzglas [25]

103.1 Physikalische Grundlagen

- *Reflexion* und *Brechung*: Der Lichtwellenleiter besteht nach Abb. 103.1 aus dem Kern mit der Brechzahl n_1 und dem Mantel mit einer etwas geringeren optischen Dichte und der Brechzahl n_2. Das Licht tritt aus dem Medium mit der Brechzahl n_0 (z. B. Luft) unter dem Winkel α in den Kern ein. Der Winkel β lässt sich bestimmen nach dem *Brechungsgesetz von Snellius*:

$$\sin\beta = \frac{n_0}{n_1}\cdot\sin\alpha \tag{103.1}$$

Dieser Strahl trifft auf die Grenzfläche Kern–Mantel und wird wiederum reflektiert. Dabei gibt es drei Möglichkeiten:

1. Der Strahl dringt in den Mantel ein und ist für die Übertragung verloren (Strahl 1, Brechung).
2. Der Strahl wird so reflektiert, dass er genau parallel zur Grenzfläche Kern–Mantel verläuft (Strahl 2).
3. Er wird total reflektiert (Strahl 3, Totalreflexion).

- *Akzeptanzwinkel, numerische Apertur*: Der Grenzfall (Strahl 2), dass das Licht nicht in den Mantel eindringt und damit für die Nachrichtenübertragung zur Verfügung steht, ist erreicht, wenn $d = 90°$ ist:

$$(\sin\gamma)_{\mathrm{gr}} = \frac{n_2}{n_1}\cdot\sin\delta = \frac{n_2}{n_1}\cdot\sin(90°) = \frac{n_2}{n_1}$$

Der maximal zulässige Winkel α_A (Abb. 103.1), unter dem das Licht in den Lichtleiter eindringen darf, errechnet sich damit zu:

$$\sin\alpha_A = \frac{n_1}{n_0}\cdot\sin\beta_{\mathrm{gr}} = \frac{n_1}{n_0}\cdot\cos\gamma = \frac{n_1}{n_0}\cdot\sqrt{1-\sin^2\gamma}$$
$$= \frac{n_1}{n_0}\sqrt{1-\frac{n_2^2}{n_1^2}} = A_N \tag{103.2}$$

n_0, n_1, n_2 Brechzahlen (siehe Abb. 103.1), α_A ist der *Akzeptanzwinkel*, $\sin\alpha_A = A_N$ die *numerische Apertur*.

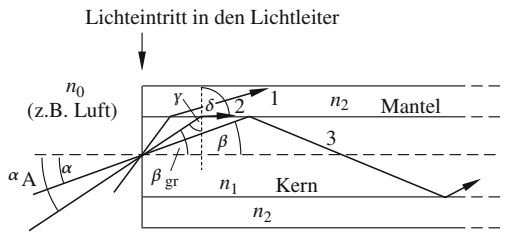


Abb. 103.1 Reflexion und Brechung von Licht bei einem Lichtwellenleiter

- *Näherung für die numerische Apertur mit $n_0 = 1$ (Luft) und $n_1 \approx n_2$*:

$$\frac{n_1^2-n_2^2}{2\cdot n_1^2} = \frac{(n_1+n_2)\cdot(n_1-n_2)}{2\cdot n_1^2}$$
$$\approx \frac{2\cdot n_1\cdot(n_1-n_2)}{2\cdot n_1^2} \approx \frac{n_1-n_2}{n_1} = \Delta \tag{103.3}$$
$$A_N \approx n_1\sqrt{2\cdot\Delta} \approx n_1\sqrt{2\cdot\frac{n_1-n_2}{n_1}}$$

Relative Brechzahldifferenz: $\Delta = \frac{n_1-n_2}{n_1}$, siehe (103.3). Bei Quarzglas-Lichtwellenleitern gilt: $n_1 \approx 1{,}5$, $\Delta \approx 0{,}01$.

- *Moden*: Diejenigen Lichtwellen, die im Lichtleiter ausbreitungsfähig sind. Ihre Zahl hängt ab vom Kerndurchmesser, von der Lichtwellenlänge und der numerischen Apertur. (Siehe auch Beispiel 3).
- *Modendispersion, Impulsverbreiterung*: Zeitliche Verbreiterung des Ausgangssignals gegenüber dem Eingangssignal durch die unterschiedlichen Laufzeiten der Lichtwellen bei den möglichen Eintrittswinkeln α in den Lichtleiter ($0 \le \alpha \le \alpha_A$, (103.2)). Bei impulsförmigem Lichtsignal am Lichtleitereingang spricht man auch von *Impulsverbreiterung*.
- *Bandbreite*: Die bei Lichtwellenleitern *theoretisch* zur Verfügung stehende Bandbreite lässt sich aus Tab. 103.1 bestimmen, indem die jeweilige Kanalbreite errechnet wird:

$$(3{,}66-3{,}53)\cdot10^{14}\,\mathrm{Hz} = 0{,}13\cdot10^{14}\,\mathrm{Hz} = 13\,\mathrm{THz}$$
$$(2{,}46-2{,}26)\cdot10^{14}\,\mathrm{Hz} = 0{,}2\cdot10^{14}\,\mathrm{Hz} = 20\,\mathrm{THz}$$
$$(1{,}97-1{,}91)\cdot10^{14}\,\mathrm{Hz} = 0{,}06\cdot10^{14}\,\mathrm{Hz} = 6\,\mathrm{THz}$$

Auch wenn die Impulsverbreiterung (siehe unten) diese theoretischen Werte verringert, steht dennoch ein vergleichsweise großer Übertragungsbereich zur Verfügung.

- *Dämpfung*: Im Wesentlichen tragen drei Mechanismen zur Dämpfung von Lichtwellen im Lichtwellenleiter bei:
  - Rayleighstrahlung: Streustrahlung durch mikroskopisch kleine Dichte- und damit Brechzahlschwankungen.
  - Resonanzabsorption durch Verunreinigungen.
  - Grunddämpfung durch Verluste bei der Lichtübertragung.

Trägt man die Summe dieser Dämpfungen als Funktion der Frequenz auf, ergibt sich für Quarzglas der in Abb. 103.2 gezeigte Verlauf [25]. Für die technische Ausnutzung ergeben sich daraus die in Tab. 103.1 aufgelisteten drei Fenster zur Signalübertragung, weil dort die Dämpfung relativ niedrig ist. Für die Rayleigh-Streuung gilt: Dämpfung $\sim 1/l^4$.

Tab. 103.1 Frequenzbereiche für Lichtwellenleiter

Farbe	Wellenlänge	Frequenz	Bemerkungen
Violett bis	0,38 µm bis	$7{,}89 \cdot 10^{14}$ Hz = 789 THz bis	Für Lichtwellenleiter möglicher Bereich
Infrarot	(0,78 ... 1,6) µm	$(3{,}85 \ldots 1{,}87) \cdot 10^{14}$ Hz = (385 ... 187) THz	
Infrarot	≈ (0,82 ... 0,85) µm	$(3{,}66 \ldots 3{,}53) \cdot 10^{14}$ Hz	Für Lichtwellenleiter ausgenutzte Bereiche
Infrarot	≈ (1,22 ... 1,33) µm	$(2{,}46 \ldots 2{,}26) \cdot 10^{14}$ Hz	
Infrarot	≈ (1,52 ... 1,57) µm	$(1{,}97 \ldots 1{,}91) \cdot 10^{14}$ Hz	

103.2 Grundmodelle von Lichtwellenleitern

Für die Abmessungen des Lichtleiters und die Abhängigkeit der Brechzahlen n_1, n_2 innerhalb des Lichtleiters vom Radius haben sich drei Modelle als brauchbar erwiesen, Abb. 103.3.

103.2.1 Mehrmoden-Stufenindex mit Totalreflexion

Nach Abb. 103.3a geht der Mantel mit der Brechzahl n_2 abrupt in den Kern mit der Brechzahl n_1 über. Der Lichteintritt ist nach (103.2) so ausgelegt, dass Totalreflexion stattfindet. Diejenigen Lichtstrahlen, die weniger oft reflektiert werden, erreichen eher das Ende des Lichtleiters als die anderen. Deshalb wird die in Abb. 103.3 skizzierte Lichtleistung $P_{L\,\mathrm{ein}}$ am Eingang entsprechend als $P_{L\,\mathrm{aus}}$ am Ausgang empfangen. Form und Dauer haben sich geändert. Deshalb werden vorzugsweise digital kodierte Signale übertragen, bei denen begrenzte Kurvenformänderungen zulässig sind. Allerdings wird die zur Verfügung stehende Bandbreite der Lichtleiter durch die Signalverlängerung herabgesetzt. Die Zahl M_S der ausbreitungsfähigen Moden kann mit der folgenden Formel abgeschätzt werden:

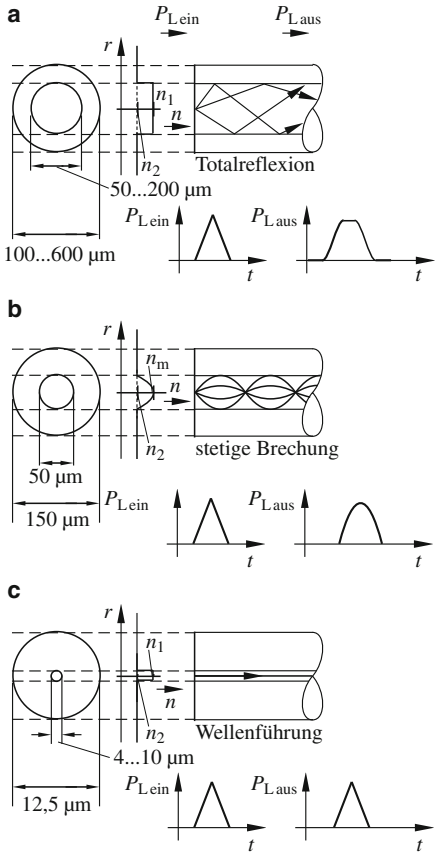


Abb. 103.3 Lichtwellenleitertypen. **a** Mehrmoden-Stufenindex, **b** Mehrmoden-Gradientenindex, **c** Einmoden-Stufenindex

Ausbreitungsfähige Moden:

$$M_S \approx 2 \left(\frac{\pi \cdot r_0}{\lambda} \right)^2 (n_1^2 - n_2^2) \tag{103.4}$$

r_0 größter Kernradius in m; λ Lichtwellenlänge im Vakuum in m; n_1 Brechzahl des Kerns; n_2 Brechzahl des Mantels.

Der größtmögliche Laufzeitunterschied Δt_S eines Lichtleiters der Länge l ergibt sich aus der Differenz zwischen Grundmode und dem Mode höchster Ordnung:

Impulsverbreiterung:

$$\Delta t_S \approx \frac{l}{c} \cdot n_1 \cdot \Delta \tag{103.5}$$

n_1 Brechzahl des Kerns; c Lichtgeschwindigkeit im Vakuum in m/s; l Lichtleiterlänge in m; Δ siehe (103.3).

103.2.2 Mehrmoden-Gradientenindex

Abb. 103.3b stellt das Prinzip dar. Die Brechzahl n im Kern hängt vom Radius r ab:

$$n(r) = n_M \left[1 - 2 \cdot \Delta \left(\frac{r}{R} \right)^x \right]^{1/2},$$

häufig auch

$$n(r) = n_M \left[1 - \Delta \cdot \left(\frac{r}{r_0} \right)^2 \right] \tag{103.6}$$

n_M Brechzahl in Kernmitte; R Konstante in m; x Konstante; r_0 größter Kernradius in m; Δ siehe (103.3).

Die Zahl M_G der ausbreitungsfähigen Moden ist halb so groß wie beim Mehrmoden-Stufenindex:

Ausbreitungsfähige Moden:

$$M_G = \frac{M_S}{2} \tag{103.7}$$

M_S siehe (103.4).

Tab. 103.2 Daten von Lichtwellenleitern

Leiterart	Kerndurchmesser in µm	Manteldurchmesser außen in µm	Dämpfung bei 850 nm in dB/km	Dispersion in ns/km	Bitrate bei 1 km Länge in Mbit/s	Impulsverbreiterung bei 1 km Länge in ns
Mehrmoden-Stufenindex	50 … 200	100 … 600	5 … 30	40 … 50	≈ 60	10 … 100
Mehrmoden-Gradientenindex	50	150	3 … 10	1 … 2	≈ 600	0,1 … 1
Einmoden-Stufenindex	4 … 10	12,5	2 … 5	0,1 … 0,2	≈ 1000	< 0,005

Man erhält für alle Moden geringere Laufzeitunterschiede, sodass die Modendispersion geringer ist als bei Lichtleitern mit Mehrmoden-Stufenindex:

Impulsverbreiterung:

$$\Delta t_G = \frac{I}{c} \cdot n_1 \cdot \frac{\Delta^2}{2} \tag{103.8}$$

n_1 Brechzahl Kern; c Lichtgeschwindigkeit im Vakuum in m/s; l Lichtleiterlänge in m; Δ siehe (103.3).

Damit ist dieser Lichtleiter besonders geeignet für einen hohen Informationsfluss oder große Entfernungen. Nachteilig ist das aufwendige Herstellungsverfahren, da die Brechzahl des Quarzglaskernes durch das Aufbringen einer Vielzahl von Glasschichten übereinander variiert wird.

103.2.3 Einmoden-Stufenindex

Es kann sich nur ein Modus längs der Faserachse ausbreiten, sodass die Dispersion und damit die Impulsverbreiterung theoretisch entfällt (Abb. 103.3c). Für den Kernradius r muss dann folgende Forderung erfüllt sein:

$$\textit{Kernradius} \quad r \leq \frac{\lambda}{2\pi} \cdot \frac{2{,}405}{\sqrt{n_1^2 - n_2^2}} \tag{103.9}$$

λ Lichtwellenlänge im Vakuum in m; n_1 Brechzahl des Kerns; n_2 Brechzahl des Mantels.

Die Impulsverbreiterung Δt_E ist theoretisch null, in der Praxis kann man mit folgendem Wert rechnen:

Impulsverbreiterung (praktisch):

$$\Delta t_E = M_\lambda \cdot \Delta\lambda \cdot l \tag{103.10}$$

M_l Dispersion in s/(m m), hängt von λ ab; $\Delta\lambda$ Linienbreite der Lichtquelle in m; l Lichtleiterlänge in m.

Mit diesen Lichtleitern lassen sich sehr große Entfernungen bei sehr großer Signalbandbreite überbrücken. Nachteilig sind der kleine Kerndurchmesser (siehe Beispiel 2) sowie die damit verbundenen Probleme beim Aneinanderfügen von zwei Kabeln. Die Stoßstelle muss auf Bruchteile von Mikrometern genau justiert werden.

Tab. 103.2 listet einige typische Eigenschaften von Lichtleitern der drei beschriebenen Leiterarten auf.

Beispiel 1

Gegeben sind die Brechzahlen $n_1 = 1{,}51$, $n_2 = 1{,}50$. Dann folgt: $\Delta = \frac{1{,}51-1{,}50}{1{,}50} = 0{,}067$. Die Impulsverbreiterung für einen Lichtwellenleiter von 1 km Länge errechnet sich zu:

a) Stufenprofilfaser:

$$\Delta t_S = \frac{I}{c} n_1 \cdot \Delta = \frac{1\,\text{km}}{3 \cdot 10^5\,\text{km/s}} 1{,}51 \cdot 0{,}0067$$
$$= 34\,\text{ns} \quad \text{bei } l = 1\,\text{km}.$$

b) Gradientenfaser:

$$\Delta t_S = \frac{I}{c} n_1 \cdot \Delta = \frac{1\,\text{km}}{3 \cdot 10^5\,\text{km/s}} 1{,}51 \cdot \frac{0{,}0067^2}{2}$$
$$= 0{,}11\,\text{ns} \quad \text{bei } l = 1\,\text{km}$$

c) Monomodefaser mit der Dispersion 3 ps/(nm km) und einer Laserlichtquelle mit der Linienbreite $\Delta\lambda = 5\,\text{nm}$:

$$\Delta t_E = (3\,\text{ps/(nm km)}) \cdot 5\,\text{nm} \cdot 1\,\text{km}$$
$$= 15\,\text{ps} \quad \text{bei } l = 1\,\text{km}$$

Für die ausnutzbare Bandbreite infolge der Impulsverbreiterung ergeben sich die in Tab. 103.3 aufgeführten Näherungswerte. ◀

Um die theoretisch verfügbare Bandbreite in der Praxis optimal zu nutzen, werden nur geringe Leiterlängen von einigen km verwendet und Zwischenverstärker eingefügt. Bei den Einmoden-Lichtleitern ist die Impulsverbreiterung durch die Linienbreite der Lichtquelle gegeben. Es werden deshalb Laserlichtquellen mit einem entsprechend niedrigen Wert eingesetzt.

Beispiel 2

Ein Monomode-Lichtleiter soll die Lichtwellenlänge $\lambda = 0{,}85\,\mu\text{m}$ übertragen. Für die Brechzahlen gilt: $n_1 = 1{,}51$, $n_2 = 1{,}50$. Nach (103.9) folgt daraus:

$$r \leq \frac{0{,}85\,\mu\text{m}}{2\pi} \cdot \frac{2{,}405}{\sqrt{1{,}51^2 - 1{,}50^2}} = 1{,}88\,\mu\text{m}$$

Tab. 103.3 Bandbreite nach Beispiel 1 bei 1 km Lichtleiterlänge

Wellenlänge (technisch ausgenutzter Bereich)	Frequenz (technisch ausgenutzter Bereich)	Theoretisch verfügbare Bandbreite bei 1 km Länge	Impulsverbreiterung, Beispiel 1 bei 1 km Länge	Bandbreite mit Impulsverbreiterung bei 1 km Länge
$\approx (0{,}82 \ldots 0{,}85)$ mm	$(3{,}66 \ldots 3{,}53) \cdot 10^{14}$ Hz	13 THz	34 ns	$\approx 14{,}5$ MHz
$\approx (1{,}22 \ldots 1{,}33)$ mm	$(2{,}46 \ldots 2{,}26) \cdot 10^{14}$ Hz	20 THz	0,11 ns	$\approx 4{,}4$ GHz
$\approx (1{,}52 \ldots 1{,}57)$ mm	$(1{,}97 \ldots 1{,}91) \cdot 10^{14}$ Hz	6 THz	15 ps	≈ 33 GHz

Tab. 103.4 Daten von Lichtwellenleiterkabeln

LWL-Typ	Mehrmoden-Gradientenkabel		Einmoden-Stufenindex-Kabel
Kabeltyp	G50/125	G62,5/125	
Kerndurchmesser in µm	50 ± 3	$62{,}5 \pm 3$	$9 \ldots 10 \pm 1$
Manteldurchmesser in µm	125 ± 3	125 ± 3	125 ± 3
Faserdurchmesser in µm	250 ± 3	250 ± 3	250 ± 25
Dämpfung bei 850 nm in dB/km	2,5 … 3,5	3,2	0,33 … 0,45 bei 1300 nm
Dispersion in ps/(nm km)			3,5 … 5 bei 1300 nm
Zugfestigkeit			Maximal 2 … 3 kN
Biegeradius in mm			
Innenraumkabel	≥ 150		
Außenraumkabel	≥ 400		
Biegeradius in mm			≥ 400

Der Radius kommt in die Größenordnung der Lichtwellenlänge, und bei diesen Lichtleiterabmessungen werden an die Herstellung hohe Anforderungen gestellt. ◀

Beispiel 3

Für einen Lichtleiter mit Mehrmoden-Stufenindex sind gegeben: Brechzahlen $n_1 = 1{,}51$, $n_2 = 1{,}50$, zu übertragende Wellenlänge $l = 0{,}85$ mm, Radius des Lichtleiters $r_0 = 25$ mm. Dann errechnet sich die Zahl der ausbreitungsfähigen Moden nach (103.4) zu:

$$M_S \approx 2 \cdot \left(\frac{\pi^2 \cdot 25^2\,\mu\text{m}^2}{0{,}85^2\,\mu\text{m}^2} \right) \cdot (1{,}51^2 - 1{,}50^2) \approx 512$$

Für den Mehrmoden-Gradientenindex-Lichtleiter folgt daraus mit (103.7):

$$M_G \approx \frac{M_S}{2} = 256. \quad \blacktriangleleft$$

103.3 Technische Ausführung von Lichtwellenleitern

Lichtwellenleiter werden aus Quarzglas (SiO_2) hergestellt. Der Kern wird mit Germaniumdioxid, Fluor oder Bortrioxid dotiert. Auf den Mantel wird eine Kunststoffschicht aufgebracht, die neben einem mechanischen Schutz auch sicherstellt, dass in den Mantel eindringende Lichtstrahlen in diese Schutzschicht übergehen und absorbeiert werden. Bei Mehrmoden-Gradientenleitern bringt man etwas 40 bis 50 einzelne Schichten aufeinander, um die Brechzahl in Abhängigkeit vom Kernradius zu ändern. Für einen ausreichenden mechanischen Schutz der Lichtleiterkabel entsprechend dem von Kupferkabeln wird es in eine Aderhülle aus Kunststoff lose eingelegt, und zwar entweder eine einzelne Ader pro Aderhülle („Hohlader“) oder mehrere Lichtleiterkabel in eine Aderhülle („Bündelader“). Zusätzlich wird eine Stützader aus glasfaserverstärktem Kunststoff mitgeführt, die die geforderte mechanische Zugfestigkeit sicherstellt. Es werden weiterhin Kupferadern integriert, die sowohl zur Fehlerortung als auch zur Stromversorgung von Zwischenverstärkern eingesetzt werden. Die verbleibenden Hohlräume füllt man mit weichem Kunststoff aus, um einen Schutz gegen das Eindringen von Wasser zu erreichen.

Lichtleiterkabel mit Mehrmoden-Gradientenindex für Außenverlegung haben genormte Maße: Kerndurchmesser 50 mm, Manteldurchmesser 125 mm.

Als Beispiel für eine Übertragung über Lichtwellenleiter wird das Video-System „OPAL-3“ der Fa. Raynet angeführt, das zur Programmversorgung vornehmlich ländlicher Gebiete ausgelegt ist. Es ermöglicht die Übertragung von 36 amplitudenmodulierten Fernsehkanälen und 30 frequenzmodulierten UKW-Rundfunkkanälen. Der Frequenzbereich erstreckt sich von 47 bis 446 MHz, wobei die Kanalrasterung im Fernsehübertragungsbereich eingehalten wird. Der Teilnehmer kann bis zu 25 km von der Einspeisestelle entfernt sein. Die optische Sendeleistung beträgt etwa 7 dBm ($\approx 5{,}6$ mW), die Wellenlänge 1310 nm, die empfangene Leistung liegt bei 0,5 mW, die optische Sendequelle ist eine Laserdiode, der Empfänger eine PIN-Fotodiode.

Tab. 103.4 listet einige Eigenschaften von LWL-Kabeln auf. DIN VDE 0888 und IEC 793 enthalten weitere Daten.

In der Typenbezeichnung von LWL-Kabeln sind Angaben über deren Aufbau und Einsatzbereich enthalten. Tab. 103.5 gibt eine Übersicht.

Tab. 103.5 Kennzeichnung von Lichtwellenleiterkabeln

Buchstabenkennung	Bedeutung	Buchstabenkennung	Bedeutung
A-	Kabel für außen	2Y	PE-Hülle
I-	Kabel für innen	Y	PVC-Hülle
V	Vollader	(L)2Y	Schichtenmantel
W	Hohlader, gefüllt	B	Bewehrung
H	Hohlader, ungefüllt	E	Einmodenfaser
D	Bündelader, gefüllt	G	Gradientenfaser
B	Bündelader, ungefüllt	.../...	Kern-/Manteldurchmesser in mm (Mehrmoden-Gradkabel); Feld-/Manteldurchmesser in mm (Einmoden-Stufenkabel)
S	Metallelement in der Kabelseele	...B...	für Bereich um 850 nm[a]
F	Füllung der Verseilhohlräume (Kabelseele)	...F...	für Bereich um 1300 nm[a]
(ZN)	Zugentlastungselemente, nichtmetallisch	...H...	für Bereich um 1550 nm[a]

[a] siehe Tab. 103.1

Das bei der Telekom eingesetzte Kabel A-DSF(L)(ZN)2Y, 10×4 E9/125 ist ein Einmoden-Stufenindexkabel mit 9/125 µm Feld-/Manteldurchmesser und 10 Bündeln mit je 4 Fasern. Es ist für Außenverlegung vorgesehen und enthält einen Schichtenmantel (Al, PE) und nichtmetallische Zugentlastungselemente. Der Typenbezeichnung können noch weitere Einzelheiten, z. B. über die Füllung, entnommen werden.

103.4 Lichtsender

Für Lichtsender werden LEDs (**l***ight* **e***mitting* **d***iode, Leuchtdiode)* im Infrarotbereich sowie LASER-Dioden (**l***ight* **a***mplification by* **s***timulated* **e***mission of* **r***adiation)* verwendet. Wichtige Kenngrößen sind zum Vergleich in Tab. 103.6 als typische Größen industriell gefertigter Dioden zusammengestellt. Es ergibt sich eine Impulsverbreiterung durch die Linienbreite (Breite des ausgesendeten Lichtspektrums) des Lichtsenders.

Die Modulation der Lichtsender geschieht durch Intensitätsmodulation. LEDs können sowohl mit analogen als auch mit digitalen Signalen moduliert werden. Bei Analogmodulation sind die Verfahren

- Amplituden-Intensitätsmodulation
- Frequenz-Intensitätsmodulation
- Phasen-Intensitätsmodulation

möglich. Bei digitaler Modulation können die im Abschn. 96.4 „Pulsträger unkodiert“ beschriebenen Verfahren eingesetzt werden. LASER-Dioden sind für die Analogmodulation weniger gut geeignet, da keine Proportionalität zwischen dem Modulationsstrom und der Lichtleistung besteht. Sie werden deshalb vorwiegend mit digital kodierten Signalen moduliert. Für beide Lichtsender sind die Anstiegs- (t_r) und die Abfallzeit (t_f) wichtige Kenngrößen. Sie liegen bei Leuchtdioden in der Größenordnung 3 ... 15 ns, sodass Informationsflüsse von 30 bis 150 Mbit/s möglich sind, sofern die Impulsverbreiterung diese Werte zulässt. Laserdioden haben wesentlich kürzere Zeiten von etwa 0,5 ns, und damit sind Werte von 1 bis 2 Gbit/s erreichbar. Mit speziellen Strukturen und geeignetem Aufbau lassen sich noch höhere Werte, bis zu 50 GBit/s, erreichen.

Tab. 103.6 Typische Daten von Lichtsendern

Lichtsender	Wellenlänge in nm	Linienbreite in nm	Dauerleistung in mW	Impulsverbreiterung in ns/km
LED	830	45	0,06	5
LASER	880	3 ... 5	10	0,015

103.5 Lichtempfänger

- *PIN-Photodioden*: Als Lichtempfänger werden PIN-Photodioden und Lawinen-(Avalanche-)Fotodioden eingesetzt. Die PIN-Fotodiode hat ihren Namen von der Schichtenfolge: Zwischen *P*- und *N*-Schicht ist eine schwach oder undotierte **I***ntrinsic-Schicht* eingefügt. Hier werden die Photonen absorbiert, sodass die Diodenparameter kaum von der angelegten äußeren Spannung abhängen. Durch diesen Aufbau verringert sich die Kapazität zwischen den Anschlüssen der Diode, was eine erhöhte Grenzfrequenz zur Folge hat. Für den Einsatz als Lichtempfänger sind die Anstiegs- und Abfallzeiten beider Diodentypen wichtige Kenngrößen. Für PIN-Dioden werden Werte von 1 bis 2 ns erreicht, Lawinendioden erreichen etwa 200 ps. Damit liegen diese Werte in der gleichen Größenordnung wie die der Lichtsender, sodass insgesamt für eine Übertragungsstrecke mit LED und Fotodiode Informationsflüsse von bis zu 150 Mbit/s erreicht werden können, für eine Strecke mit Laserdiode und Lawinendiode bis 1 Gbit/s. Das setzt allerdings voraus, dass die Gesamt-Lichtleiterlänge bei größeren Übertragungsstrecken durch Zwischenverstärker geeignet unterteilt wird, damit die Impulsverbreiterung den Informationsfluss nicht wesentlich verringert. Bei sorgfältigem kapazitätsarmem Aufbau können Datenübertragungen bis zu 10 GBit/s erreicht werden, durch besondere Schaltungstechniken und spezielle Dioden auch bis zu 50,GBit/s. PIN-Dioden haben weiterhin folgende Vorteile: 1. Linearität zwischen eingestrahlter Lichtintensität und elektrischem Ausgangssignal über etwa 7 Zehnerpotenzen. 2. Geringes Eigenrauschen. 3. Betriebsspannung im Bereich von 5 V.
- *Lawinen-Photodioden*: Hier wird der von den Photovervielfachern bekannte Vervielfachungseffekt eingesetzt. Damit können auch sehr lichtschwache Signale ausgewertet werden. Allerdings haben diese Dioden ein erhöhtes Eigenrauschen und nur einen eingeschränkten Dynamikbereich.

103.6 Verbinden von Lichtleitern

Lichtleiterkabel werden in Längen bis zu 10 km gefertigt. Um Kabel zu verbinden oder Kabel an Lichtverstärker anzuschließen, benutzt man *Spleiß-* und *Steckverbindungen*.

- *Spleißverbindung*: Durch Ritzen und Brechen der Aderenden mit entsprechenden Werkzeugen sind die Stirnflächen plan und verlaufen rechtwinklig zur Achse. Nachdem ein Kleber aufgebracht worden ist, werden die Enden mit Mikromanipulatoren voreinander justiert und dann verklebt. Eine andere Möglichkeit besteht darin, die Stirnflächen auf Stoß zu bringen und diese Stoßstelle einem Lichtbogen mit einer Temperatur von 2000 °C auszusetzen. Innerhalb von wenigen Sekunden verschmelzen die beiden Glasfasern. Diese Methode eignet sich besonders für Verbindungen „vor Ort“, da entsprechende Schweißgeräte mit den Justiereinrichtungen in handlicher Form von der Industrie angeboten werden und zu dauerhaften Verbindungen mit nur geringer Dämpfung ($< 0{,}5$ dB) führen. Eine besonders große Dämpfung (0,4 ... 0,6 dB) ergibt sich, wenn beide LWL einen Achsenversatz aufweisen oder die Durchmesser beider LWL-Fasern voneinander abweichen. Weniger kritisch ist die Dämpfung, wenn eine Stirnfläche nicht ganz plan ist.
- *Steckverbindung*: Die Faser-Stirnflächen werden poliert, in Präzisions-Führungshülsen eingebettet und mit einer Gewindehülle umgeben. Durch Zusammenschrauben oder Zusammenstecken werden die Stirnflächen aufeinandergedrückt. Die mechanische Führung muss äußerst präzise (auf einige µm oder weniger) arbeiten, und deshalb bringt man in kritischen Anwendungsfällen an den Steckern zusätzlich Feinjustiereinrichtungen an. Es besteht eine Vielzahl von praktisch eingesetzten Verbindungssystemen, die untereinander nicht kompatibel sind. Hier einige ausgewählte Modelle:
  - Die Telekom hat ein System angeregt, dass genormt wurde und im Wesentlichen von ihr eingesetzt wird: Version LSA nach DIN 47256 und Version LSB nach DIN 47217.
  - MIC: Für Computernetze. Stecker auch unter FDDI-Stecker bekannt. Teilstandard genormt nach ISO 9314-3. Sender- und Empfängerstrecke nebeneinander verpolungssicher eingebaut.
  - BFOC(ST): Von AT&T entwickelt. Bajonettverriegelung, dadurch einfache Handhabung.
  - SC: Rastet beim Einschieben ein und wird durch Ziehen ausgerastet.

Die Dämpfung an der Stoßstelle (Einfügedämpfung) liegt zwischen 0,2 dB und 0,5 dB.

103.7 Optische Schalter

Der verbreitete Einsatz der optischen Signalübertragung erfordert auch das Verteilen von optischen Kanälen auf andere beliebig wählbare Kanäle. Hier seien 2 Verfahren angeführt:

- *Elektro-Optischer Schalter*: Ausgenutzt wird der Pockels-Effekt. Das optische Signal wird in zwei (oder mehr) Lichtstrahlen aufgeteilt (Y-Verteiler), und jeder Strahl durchläuft einen Flüssig-Kristall, an den entweder in Längs- oder in Querrichtung zum durchlaufenden Strahl eine Spannung angelegt wird. Damit ändert sich die Polarisationsebene, und nur bei Übereinstimmung von Strahlpolarisation und Kristall-Polarisationsebene wird der Strahl durchgelassen. Die Einfügedämpfung ist relativ hoch und liegt im Bereich 6 bis 10 dB. Die Schaltzeiten liegen im Bereich von ns.
- *Mikrospiegel-Schalter*: Der Lichtstrahl fällt auf einen Spiegel, der diesen Strahl auf einen anderen frei wählbaren Lichtleiter umlenkt. Zur Auslenkung des Spiegels können Piezokristalle eingesetzt werden. Die Einfügedämpfung liegt im Bereich 1 bis 5 dB, die Schaltzeit wegen der Mechanik im Bereich ms bis s.

Funkmesstechnik

Wilfried Plaßmann

104

Teil XIII



104.1 Radar 1278
104.2 Funkortungssystem OMEGA 1280
104.3 GPS – Global Positioning System 1280

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_104

104.1 Radar

104.1.1 Grundlagen, Kenngrößen

Ein Hauptgebiet der Funkmesstechnik ist durch den Begriff RADAR (*ra*dio *d*etection *a*nd *r*anging gekennzeichnet. Bei diesem Verfahren werden impulsförmige elektromagnetische Wellen von einer Antenne ausgesendet und an Körpern oder Stoffverteilungen (Wolken) reflektiert. Die Sendeantenne wird auf Empfang umgeschaltet, und anhand des Echos sind Rückschlüsse auf die Lage und die Beschaffenheit der Körper oder der Stoffverteilungen möglich. Angewendet wird die Radartechnik bei der Kontrolle und der Sicherung des Land-, Wasser- und Flugverkehrs, in der Meteorologie zur Wetterprognose, in der Astrologie und im militärischen Bereich.

Die Antenne sendet einen scharf gebündelten Strahl aus, dessen Dauer kürzer ist als die Laufzeit zum nächstgelegenen Körper und zurück. Man kann daher aus der Laufzeit auf die Entfernung und aus der Antennenstellung auf die Richtung des Körpers schließen. Überschreitet der Körper eine bestimmte Mindestgröße, können eventuell sogar Aussagen über die Körperform und dessen Beschaffenheit gemacht werden. Wenn ein räumlicher Bereich erfasst werden soll, dreht sich die Antenne und bildet die Umgebung während einer Umdrehung sektorweise ab. Im folgenden Umlauf wird das vorhergehende Bild aktualisiert.

- *Radarquerschnitt* (auch *Rückstreu-* oder *Echoquerschnitt*): Die am Körper ankommende Strahlungsdichte der vom Sender ausgestrahlten Welle ist S_e (siehe auch Abschn. 95.2 „Kenngrößen“):

$$S_e = \frac{g \cdot P_S}{4 \cdot \pi \cdot r^2} \quad \text{in} \quad \frac{\text{W}}{\text{m}^2} \tag{104.1}$$

  g Antennengewinn; P_S Senderleistung in W; r Abstand Sender–Körper in m.

Davon wird ein Teil reflektiert, und es kommt an der Radar-Empfangsantenne die Strahlungsdichte S_S an:

$$S_S = \frac{A_e \cdot S_e}{4 \cdot \pi \cdot r^2} \quad \text{in} \quad \frac{\text{W}}{\text{m}^2} \tag{104.2}$$

S_e nach (104.1); r Abstand Sender–Körper in m, A_e Radarquerschnitt in m^2.

Die Größe A_e ist definiert und heißt *Radarquerschnitt*. Nimmt man als reflektierenden Körper eine ebene elektrisch gut leitende Platte der Fläche A_{pl} an, deren Abmessungen viel größer sind als die Wellenlänge λ des Radarsignales und auf die der Radarstrahl senkrecht auftritt, ergibt sich $A_e = 4 \cdot \pi \cdot A_{\text{pl}}^2/\lambda^2$. Der Radarquerschnitt ist das Verhältnis von reflektierender Fläche zu der aus der Wellenlänge des Radarstrahles gebildeten Fläche. Die an der Empfängerantenne herrschende Strahlungsdichte ist um so größer, je größer die Fläche des reflektierenden Körpers und je kleiner die Wellenlänge des Radarsignales ist. In der Praxis kann man auch bei beliebig geformten Körpern häufig mit seiner ebenen Fläche rechnen und erhält brauchbare Näherungswerte.

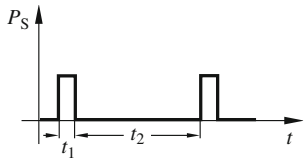


Abb. 104.1 Tastgrad der Senderimpulse

- *Reichweite eines Radargerätes*: Der maximale Abstand $r_{\max}$ eines Körpers lässt sich berechnen, das Ergebnis wird hier angegeben.

$$\textit{Reichweite} \quad r_{\max} = \sqrt[4]{\frac{P_S A_e A^2}{4\pi\lambda^2 k T B F S_r}} \tag{104.3}$$

  P_S Senderleistung in W; A_e Radarquerschnitt in m^2; A Wirkfläche der Empfangsantenne in m^2; λ Radar-Wellenlänge in m; k Boltzmann-Konstante; T absolute (thermodynamische) Temperatur in K; B Bandbreite des Radarsignals in Hz; F Rauschzahl des Empfängers; S_r Störabstand.
- *Primärradar*: Der Körper reflektiert *passiv* den Radarstrahl. Das empfangene Echosignal hat eine geringe Amplitude und wird durch Rauschen und Störungen verfälscht. Störungen sind z. B. auch feststehende Gebäude und Berge. Es sind trotzdem sehr genaue Ortungen möglich.
- *Sekundärradar*: Ein z. B. im Flugzeug eingebauter Transponder antwortet auf ein Codewort des Senders mit einem eigenen Sender und der Übertragung von Informationen (z. B. Identität, Flughöhe). Dieses Verfahren garantiert eine sehr sichere Signalerkennung, ist aber von der Ortungsgenauigkeit dem Primärradar unterlegen.

Beide Verfahren werden oft parallel eingesetzt.

- *Tastgrad beim Impulsradar*: In der Regel wird die gleiche Antenne für das Senden und das Empfangen des Radarsignales verwendet. Es werden deshalb kurze Impulse hoher Leistung zur Erzielung einer großen Reichweite ausgesendet (Antenne im Sendebetrieb) und anschließend deren Echo empfangen (Antenne im Empfangsbetrieb). Abb. 104.1 zeigt die an der Antenne anstehende Sendeleistung P_s. Für die Impulsdauer t_1 muss gelten, dass sie kürzer ist als die Laufzeit der Welle zum nächstgelegenen Körper und zurück, für t_2 folgt, dass erst das Echo vom entferntesten Körper die Antenne erreicht haben muss, bevor ein neuer Impuls ausgesendet werden kann.
- *CW-Radar* (**C***ontinuous* **W***ave-Radar*, *Dauerstrich-Radar*): Radarverfahren, bei dem die Sendeimpulsdauer mehr als 10 % der Pausendauer beträgt, d. h. $t_1 > 0{,}1 \cdot t_2$ in Abb. 104.1. Meist ist die Pausendauer sogar Null. Dann lässt sich z. B. die Geschwindigkeit eines Einzelkörpers wie bei der Verkehrskontrolle nach dem Doppler-Prinzip messen.
- *Entfernungsauflösung*: Die Impulsdauer t_1 nach Abb. 104.1 bestimmt die Entfernungsauflösung. Abb. 104.2 stellt dar, dass sich der Antennenstrahl trotz Bündelung geringfügig aufweitet. Zuerst trifft das Echo von Strahl 1 am Empfänger ein, da dieser Strahl senkrecht auf die Fläche A_1 auftrifft. Der Strahl 2 entsteht durch die nichtideale Bündelung der Antenne. Er hat eine größere Strecke zurückzulegen, da er nicht

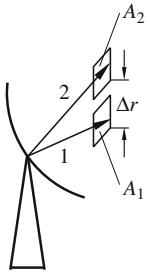


Abb. 104.2 Entfernungsauflösung

senkrecht auf A_2 auftrifft. Er erreicht den Empfänger mit der Zeitverzögerung $2 \cdot \Delta r/c$. Damit beide Echos noch voneinander unterschieden werden können, muss gelten: $2 \cdot \Delta r/c > t_1$. Weicht Strahl 2 maximal von Strahl 1 ab, ergibt sich der Grenzfall:

$$\textit{Entfernungsauflösung} \quad \Delta r = \frac{t_1 \cdot c}{2} \tag{104.4}$$

c Lichtgeschwindigkeit $3 \cdot 10^8$ m/s; t_1 in s, siehe Abb. 104.1. Demnach muss die Impulszeit so kurz wie möglich gewählt werden, um eine hohe Entfernungsauflösung zu bekommen. Kurze Impulse haben aber eine größere Bandbreite B (siehe (104.3)), sodass die Reichweite verkürzt wird.

- *Minimale Tastpause*: Handelt es sich um großflächige Körper der Ausdehnung r_K, darf der nächste Radarimpuls erst nach der Zeit

$$t_{2\,\min} = \frac{2 \cdot r_k}{c} \tag{104.5}$$

(c Lichtgeschwindigkeit $3 \cdot 10^8$ m/s) folgen.

- *Drehgeschwindigkeit*: Soll ein größerer Bereich erfasst werden, muss die Antenne gedreht werden. Da die von der Antenne empfangenen Signale durch Rauschen und Störungen verfälscht sind, müssen für ein brauchbares Signal mindestens 10 bis 20 Echos empfangen und überlagert werden, siehe dazu auch „Zielextraktor“ weiter unten.
  Dann gilt:
  Maximale Drehgeschwindigkeit

$$\Omega_{\max} = \frac{\Phi}{n \cdot t_2} \tag{104.6}$$

Φ Winkel, um den der Radarstrahl aufgeweitet wird; n Zahl der Echos; t_2 Impulspause nach Abb. 104.1 in s.

- *Zielextraktor*: Die endliche Strahlbreite der Radarantenne ergibt beim Überstreichen eines reflektierenden Objektes mehrere Echos, die in der Regel gestört sind. Abb. 104.3a stellt vier solcher gestörten Signale dar. Beim ersten Echo (U_E) wird im Bereich t_x bis $t_x + \Delta t_x$ ein erhöhtes Empfangssignal festgestellt und die Schwelle mit der Spannung U_s (Erfahrungswert) überschritten. Das gilt auch für den Bereich t_y bis $t_y + \Delta t_y$. Beim Eintreffen des zweiten Echos wird die Spannung U_s wiederum näherungsweise im Bereich t_x bis $t_x + \Delta t_x$ überschritten, während im Bereich t_y bis $t_y + \Delta t_y$ die Schwellspannung nicht erreicht wird. Auch beim dritten und vierten Echo ist die empfangene Signalspannung wieder näherungsweise im Bereich t_x bis $t_x + \Delta t_x$ größer als die Spannung U_s, während in allen anderen Fällen gilt, dass die Spannung U_s nur einmal überschritten wurde. Damit ist das Signal im Bereich t_x bis $t_x + \Delta t_x$ einem Körper zuzuordnen. Bei den anderen die Schwellspannung überschreitenden Werten handelt es sich um Störungen. Die beschriebene Auswertung erfolgt in einem binären Wanderfensterdetektor nach Abb. 104.3b. Das Echosignal wird laufend digitalisiert (insgesamt k Werte) und digital mit der Schwellspannung U_s verglichen. Entsprechend wird 0 oder 1 zugeordnet. Diese Werte werden in ein Schieberegister mit k Stellen eingeschoben und am Ende in k Addierern (für jede der k Stellen ein Addierer) gespeichert. Beim zweiten Echo geschieht das gleiche, wobei am Ende die im Schieberegister enthaltenen Digitalwerte zu denen des ersten Echos hinzuaddiert werden. Nach Eintreffen des vierten Echos und Addition enthalten die Zellen, die den Digitalwerten im Bereich t_x bis $t_x + \Delta t_x$ zugeordnet sind, die Zahl 4, da für jedes Echo die Schwellspannung überschritten wurde. Alle anderen Zellen enthalten nur eine 1 (Schwellspannung einmal überschritten) oder nur eine 0 (Schwellspannung nie überschritten). Mit der Festlegung Summe $\geq 3 \mathrel{\hat{=}}$ Ziel, Summe $< 3 \mathrel{\hat{=}}$ kein Ziel (Abb. 104.3b: $i \geq m \Rightarrow 1 \mathrel{\hat{=}}$ Ziel, $i < m \Rightarrow 0 \mathrel{\hat{=}}$ kein Ziel) wird die Entscheidung über das Vorhandensein eines Zieles getroffen.

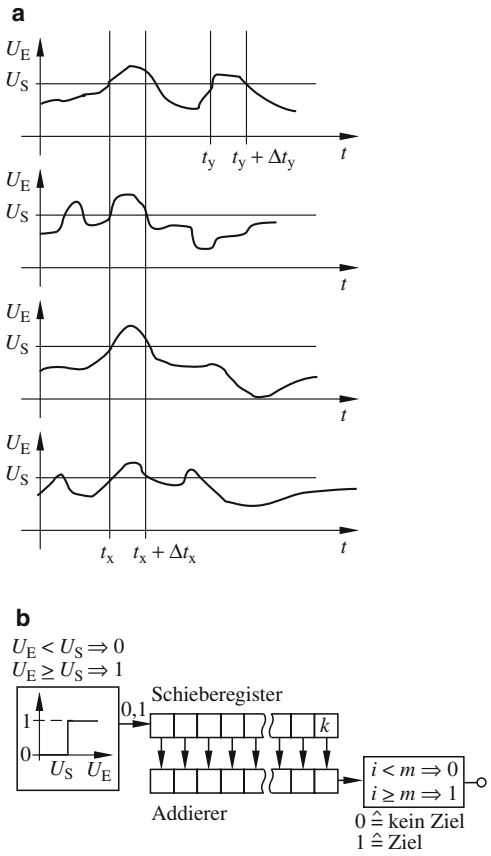


Abb. 104.3 Echosignale vier aufeinanderfolgender Sendersignale. **a** Signalverläufe, **b** Wanderfensterdetektor

104.1.2 Daten von Radaranlagen

Die für die Flugüberwachung eingesetzten Radaranlagen haben beispielsweise die folgenden technischen Daten:

- Umlaufzeiten zwischen 3 und 15 s
- Hochfrequenz-Spitzenleistung bis 10 MW
- Impulsdauer 10 µs
- Pausendauer 10 ms
- Antenne horizontal scharf gebündelt, vertikal weniger scharf gebündelt
- Frequenzen zwischen 400 MHz bis 75 GHz in dafür vorgesehenen Bereichen

Für eine Rundsichtanlage (Allzweckgerät, z. B. Radaranlage auf kleineren Schiffen) werden die folgenden typischen Werte angegeben:

- Frequenzband: 9,320 . . . 9,480 GHz
- Impulslänge: Nahbereich 0,1 µs, Fernbereich 0,2 µs
- Impulsfolgefrequenz: 1000 Hz
- Impulsspitzenleistung: 10 kW
- Horizontaler Antennen-Öffnungswinkel: 1,75°
- Vertikaler Antennen-Öffnungswinkel: 20°
- Nahauflösung: 30 m
- Abstandsauflösung: 20 m
- Entfernungsbereiche: 0,5–1–3–8–15–30 Seemeilen
- Antennendrehzahl: 24 pro Minute
- Leistungsbedarf: 200 W

104.2 Funkortungssystem OMEGA

Unter einer Vielzahl von Funkortungssystemen hat das System *OMEGA* Bedeutung, weil es erdumspannend ist und voraussichtlich noch einige Zeit bestehen bleiben wird. Es gibt 8 Bodenstationen in Japan, Australien, Norwegen, Trinidad, USA, Argentinien und auf Madagaskar und Hawaii. Jeder Sender sendet 3 Frequenzen im VLF-Bereich aus: $f_1 = 10{,}2\,\text{kHz}$ (Hauptträger), $f_2 = 13{,}60\,\text{kHz}$ (1. Hilfsträger) und $f_3 = 11{,}33\,\text{kHz}$ (2. Hilfsträger). Wegen der großen Reichweite dieser Wellen ist jeder Punkt der Erde von mehreren Sendern erreichbar. Die einzelnen Sender senden die drei Frequenzen synchron mit einer für jeden Sender typischen Zeitdauer aus. Sie können damit identifiziert werden. In den zur Positionsbestimmung eingesetzten Empfängern werden die zwei Differenzfrequenzen $f_0 = f_2 - f_1 = 3{,}40\,\text{kHz}$ und $f_4 = f_3 - f_1 = 1{,}13\,\text{kHz}$ gebildet, und damit wird eine grobe Ortsbestimmung vorgenommen. Da gilt: $3 \cdot f_0 = f_1$ und $4 \cdot f_0 = f_2$, kann aus den Signalen zweier Sender (S_1, S_2) die Phasendifferenz $\Delta\varphi_1$ gemessen und daraus eine Linie für $\Delta\varphi_1 = \text{const}$ gezeichnet werden (Abb. 104.4). Diese Linie ist eine Hyperbel, deswegen gehört dieses Verfahren zu den Hyperbelortungsverfahren:

$$\Delta\varphi_1 = \frac{2 \cdot \pi}{c} \cdot 3 \cdot 4 \cdot f_0 \cdot (PS_1 - PS_2) \qquad (104.7)$$

Das gleiche Verfahren wird mit einem weiteren Sender, S_3, durchgeführt. Es ergibt sich eine zweite Hyperbel für $\Delta\varphi_2 = \text{const}$. Der Schnittpunkt beider Hyperbeln ergibt den Standort P.

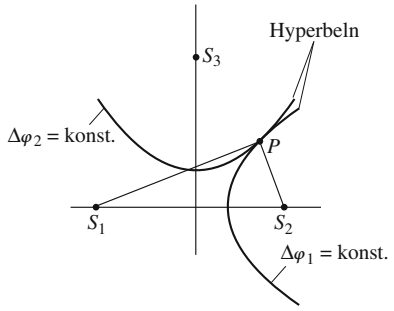


Abb. 104.4 Standortbestimmung mit dem Ortungssystem OMEGA

104.3 GPS – Global Positioning System

GPS ist ein satellitengestütztes Navigationssystem. Es wurde 1995 für militärische Zwecke errichtet und wird vom amerikanische Verteidigungsministerium betrieben. Es dient zur Zeitübertragung und zur Ortung. Insgesamt umkreisen etwa 30 Satelliten den Erdball in einer Höhe von 20.183 km. Sie sind also nicht geostationär (stationär: ortsfest, bezogen auf die Erdoberfläche, wie etwa die Fernsehsatelliten der ASTRA-Serie). Die Umlaufbahnen sind genau vorgegeben, und zwar so, dass an jedem Punkt der Erdoberfläche die Signale von mindestens 3 Satelliten stets gleichzeitig empfangen werden können. Damit ist es möglich, die absolute Lage eines Punktes, d. h. geografische Länge, geografische Breite und Höhe über Normal-Null (NN), zu bestimmen. Das Prinzip beruht darauf, dass die genaue Position des Satelliten übermittelt wird und aus der Laufzeit der Signale Satellit-Empfänger die Position des Empfängers bestimmt werden kann. Der Fehler der Ortung hängt u. a. auch ab von der Synchronisation der Zeit des Empfängers auf der Erdoberfläche mit der Zeit des GPS-Systems. Deshalb ist es wünschenswert, dass ein vierter Satellit für diese Synchronisation empfangen werden kann. In der Regel sind sogar 5 oder 6 Satellitensignale gleichzeitig zu empfangen, was bei entsprechender Auswertung durch z. B. Mittelwertbildung zu einer Verringerung des Messfehlers führt. Das von den Satelliten ausgesendete Signal ist sehr schwach und wird vom Rauschen überdeckt. Es sind deshalb statistische Verfahren zur Signalerkennung erforderlich (siehe Kap. 108 „Optimierte Nachrichten- und Datenübertragung“, Abschn. 108.3 „Anwendung der Korrelation bei gestörten Signalen“).

Bis zum Jahr 2000 wurde das von den Satelliten ausgesendete Signal künstlich nach einem geheim gehaltenen Schlüssel „verrauscht“, was zu einem erhöhten Fehler bei der Ortsbestimmung führte. Nur das Militär konnte durch die entsprechende Entschlüsselung dieses Verrauschen rückgängig machen. Auf der Erdoberfläche hat man dieses „Verrauschen“ dadurch korrigiert, dass man ortsfeste Punkte genau vermessen und die Entfernungen dieser Punkte zueinander gespeichert hat. Abweichungen der aktuell gemessenen Werte zu den tatsächlichen korrekten Entfernungen dieser Punkte untereinander waren dann auf die „Verrauschung“ zurückzuführen. Das Prinzip der

Ausnutzung ortsfester Bezugspunkte hat zum *D*ifferential-*GPS* (*DGPS*) geführt. Damit wird eine genauere Ortsbestimmung erreicht, weil z. B. atmosphärische Störungen weniger Bedeutung haben. Wünschenswert ist eine möglichst große Anzahl solcher Referenzpunkte. In der Praxis allerdings hat man den Weg der virtuellen Referenzpunkte gewählt, deren Daten z. B. von Rundfunksendern der ARD mit dem RDS-Signal übertragen werden.

Seit dem Jahr 2000 wird das GPS in den Navigationssystemen der Kraftfahrzeuge mit Erfolg eingesetzt. Auch in Handys wird es zunehmend eingebaut, zumal es in die vorhandenen elektronischen Bausteine problemlos zusätzlich integriert werden kann.

Der Fehler der Ortsbestimmung auf der Erdoberfläche hängt ab u. a. von der momentanen Entfernung der Satelliten zueinander, der Synchronisation der Zeiten von Sender und Empfänger, vom Standort des Empfängers (Hindernisse, Reflexionen) und von Störungen. Die Praxis hat gezeigt, dass der Ort in der Regel mit vertretbarem Aufwand (Navigationssysteme im Kraftfahrzeug und im Handy) bis auf eine absolute Abweichung von maximal 2 bis 10 m bestimmt werden kann.

Da über die weitere Nutzung der Satelliten für zivile Zwecke von den USA jährlich neu entschieden wird, planen die europäischen Staaten ein eigenes System mit dem Namen Galileo.

Elektroakustische Wandler

105

Teil XIII

Wilfried Plaßmann



105.1 Definitionen, Kenngrößen 1284
105.2 Schallempfänger 1285
105.3 Schallsender 1289

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_105

Um Sprache und Musik über größere Entfernungen zu übertragen, wird der *Hörschall* mit einem *Mikrofon* in ein proportionales elektrisches Signal umgewandelt und auf drahtlosem Wege (Beispiel Rundfunkübertragung) oder drahtgebunden (Beispiel Telefon, Kabelrundfunk, oft in Kombination mit dem drahtlosen Weg) zum Empfänger übermittelt. Am Empfangsort setzt z. B. ein Lautsprecher das elektrische Signal wieder in ein akustisches Signal um.

105.1 Definitionen, Kenngrößen

Im folgenden sind einige Kenngrößen der Akustik zusammengestellt, weitere können z. B. DIN 1320 entnommen werden.

- *Schall*: Mechanische Energie in Form von Schwingungen und Wellen eines elastischen Mediums.
- *Hörschall*, *Infraschall*, *Ultraschall*:
  Hörschall: Schall im Frequenzbereich des menschlichen Hörens, ca. 16 Hz bis 16 kHz.
  Infraschall: Schall unterhalb ca. 16 Hz.
  Ultraschall: Schall oberhalb ca. 16 kHz.
- *Geräusch*: Schallsignal, das meistens ein nicht zweckbestimmtes Schallereignis charakterisiert (Beispiel Maschinengeräusch).
- *Schallkennimpedanz* (kurz *Kennimpedanz*): Widerstand, den das Medium der Schallwelle entgegensetzt (siehe auch Erläuterung zu (105.3)):
  Schallkennimpedanz:

$$W_0 = c \cdot \rho \quad \text{in } \frac{\text{kg}}{\text{m}^2\,\text{s}} \tag{105.1}$$

  c Schallgeschwindigkeit im m/s, ρ Dichte in kg/m^3.
  In Luft bei Normaldruck und 20 °C ist die Kennimpedanz 413 N s/m^3.
- *Wellenwiderstand*: Ist das Medium verlustfrei, wird die Schallkennimpedanz reell und als *Wellenwiderstand* bezeichnet.
- *Schalldruck*: Durch die Schallschwingung hervorgerufener Wechseldruck p in N/m^2 (siehe auch Erläuterung zu (105.3)). Der Praxis angepasst gibt man ihn häufig in µbar an:

$$1\,\mu\text{bar} = 10^{-1}\,\text{N/m}^2 = 10^{-1}\,\text{Pa} \tag{105.2}$$

- *Schallschnelle*: Geschwindigkeit, mit der sich die Atome bzw. Moleküle durch den Schalldruck um ihre Ruhelage bewegen:

$$v = \omega \cdot s = \frac{p}{W_0} \quad \text{in } \frac{\text{m}}{\text{s}} \tag{105.3}$$

  ω Kreisfrequenz in s^{-1}, s Schwingweg der Schallquelle in m, p Schalldruck in N/m^2 bzw. kg/(s^2 m), W_0 Schallkennimpedanz in kg/(s m^2).
  Diese Gleichung legt die Analogie zum elektrischen Feld nahe: Mit den Entsprechungen $v \leftrightarrow \underline{I}$, $p \leftrightarrow \underline{U}$, $W_0 \leftrightarrow \underline{Z}$ folgt daraus das Ohm'sche Gesetz, angewendet auf den Wechselstromkreis. Die äquivalente Definition lautet dann: „Die Stromschnelle ist die Geschwindigkeit, mit der die Elektronen durch die Spannung ihre Lage ändern."
- *Schallgeschwindigkeit*: Geschwindigkeit, mit der sich die Schallwelle fortpflanzt. Sie berechnet sich aus den Materialeigenschaften zu:

$$c = \sqrt{\frac{E}{\rho}} \quad \text{in m/s} \tag{105.4}$$

  E Elastizitätsmodul in N/m^2, ρ Dichte in kg/m^3.
  Sie beträgt in Luft bei 20 °C und Normalluftdruck 340 m/s, in Wasser ca. 1500 m/s und in Stahl 5050 m/s.
- *Dämpfungsmaß*: Logarithmisches Verhältnis der Amplituden einer Feldgröße an zwei hintereinanderliegenden Punkten.
- *Schallintensität*: Produkt aus Schallschnelle und Schalldruck oder Quotient aus Schalleistung und Fläche:

$$J = v \cdot \rho = \frac{p^2}{W_0} \quad \text{in } \frac{\text{Pa m}}{\text{s}},\ \frac{\text{N}}{\text{m s}} \text{ oder } \frac{\text{W}}{\text{m}^2} \tag{105.5}$$

  Formelzeichen siehe (105.1)–(105.4).
  Es handelt sich dabei um eine Leistung pro Fläche, wie die letzte Einheitenangabe zeigt.
- *Schalleistung*: Produkt aus Schallintensität J und Fläche A:

$$P_{\text{ak}} = J \cdot A \quad \text{in W oder } \frac{\text{N m}}{\text{s}} \tag{105.6}$$

- *Hörschwelle*: Die Hörschwelle ist der Schalldruck, den das Ohr gerade noch wahrnehmen kann, er beträgt $p_0 = 2 \cdot 10^{-4}\,\mu\text{bar} = 2 \cdot 10^{-5}\,\text{Pa}$.
- *Schallpegel*: Logarithmisches Verhältnis des tatsächlichen Schalldruckes p_1 zu einem *Bezugs-Schalldruck* p_b. Die Bezugsgröße p_b ist anzugeben.
- *Schalldruckpegel*: Logarithmisches Verhältnis des tatsächlichen Schalldruckes p_1 zum *Schalldruck der Hörschwelle* $p_0 = 2 \cdot 10^{-4}$ mbar (siehe auch „Lautstärkepegel", (105.13)):

$$L_p = 20 \cdot \lg \frac{p_1}{p_0} \quad \text{in dB} \tag{105.7}$$

- Elektroakustischer Übertragungsfaktor für Schallsender (Lautsprecher), im folgenden kurz Übertragungsfaktor genannt:

$$B_S = \frac{p}{U} \quad \text{in } \frac{\text{Pa}}{\text{V}} \tag{105.8}$$

  p Schalldruck in Pa, U elektrische Spannung in V.
- Elektroakustischer Übertragungsfaktor für Schallempfänger (Mikrofone), im folgenden kurz Übertragungsfaktor genannt:

$$B_E = \frac{U}{p} \quad \text{in } \frac{\text{V}}{\text{Pa}} \tag{105.9}$$

  p Schalldruck in Pa, U elektrische Spannung in V.
- *Absoluter elektroakustischer Übertragungsfaktor*:

$$B_{S0} = 1\,\text{Pa/V} \quad \text{für Schallsender (Lautsprecher)} \tag{105.10a}$$

$$B_{E0} = 1\,\text{V/Pa} \quad \text{für Schallempfänger (Mikrofone)} \tag{105.10b}$$

- *Elektroakustisches Übertragungsmaß für Schallsender* (Lautsprecher):

$$G_S = 20 \cdot \lg \frac{B_S}{B_{S0}} \quad \text{in dB} \tag{105.11}$$

  Formelzeichen siehe (105.8), (105.10a).

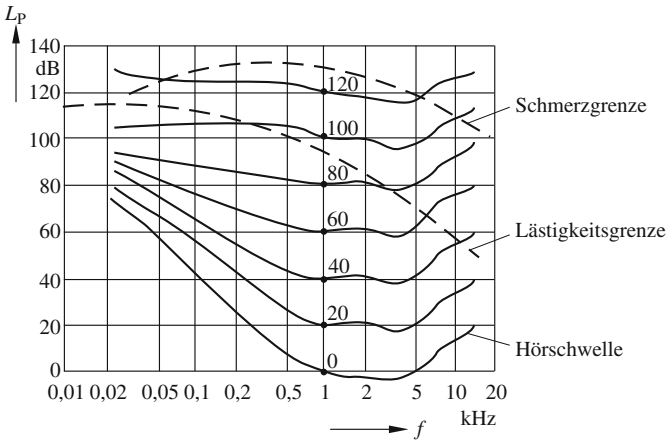


Abb. 105.1 Schalldruckpegel als Funktion der Frequenz (Kurven gleicher Lautstärkeempfindung), nach [28]

- *Elektroakustisches Übertragungsmaß für Schallempfänger* (Mikrofone):

$$G_E = 20 \cdot \lg \frac{B_E}{B_{E0}} \quad \text{in dB} \qquad (105.12)$$

  Formelzeichen siehe (105.9), (105.10b).
- *Lautstärkepegel*: Er ist gleich dem Schalldruckpegel (in dB, siehe oben) des als gleichlaut beurteilten Normschalles (1 kHz):

$$L = 20 \cdot \lg \frac{P_1}{P_0} \quad \text{in phon} \qquad (105.13)$$

  p_1 Schalldruck des Normalschalles, $p_0 = 2 \cdot 10^{-4}$ mbar Schalldruck des Normalschalles an der Hörschwelle.
  Da Schalldruckpegel und Lautstärkepegel nicht zwingend den gleichen Wert ergeben, wird die Einheit „phon“ verwendet. Abb. 105.1 zeigt den Zusammenhang zwischen dem Schalldruckpegel L_p und der Frequenz f. Aufgetragen sind Kurven gleicher Lautstärkeempfindung. Danach werden Frequenzen im Bereich 2 ... 5 kHz vom Ohr am besten aufgenommen.
- *Lautheit*: Zusammenhang zwischen der *tatsächlichen Zunahme* des Lautstärkepegels und der *menschlichen Empfindung über das Maß dieser Zunahme*. Die Abhängigkeit ist experimentell ermittelt worden und ist nichtlinear, siehe Abb. 105.2. Gibt man die gleiche *Zunahme* des Lautstärkepegels bei kleinen und bei großen Lautstärkepegeln vor, wird die Zunahme bei kleinen Pegeln als größer empfunden als bei großen Pegeln. Zur Unterscheidung zum Lautstärkepegel wurde die Einheit „sone“ gewählt.
- *Schmerzschwelle*: Sie liegt bei etwa $(10^6 \ldots 10^7) p_0 = (20 \ldots 200)\,\mathrm{N/m^2}$.
- *Klirrfaktor*: Dieser in Abschn. 92.4.12.3 behandelte Begriff bekommt hier eine praktische Bedeutung. Durch Messungen mit einer Vielzahl von Versuchspersonen hat man festgestellt, dass das menschliche Ohr Klirrfaktoren > 1 % wahrnimmt. Für eine Übertragung muss demnach gelten, dass dieser Wert insgesamt nicht überschritten werden darf, wenn die Wiedergabe unverzerrt klingen soll.

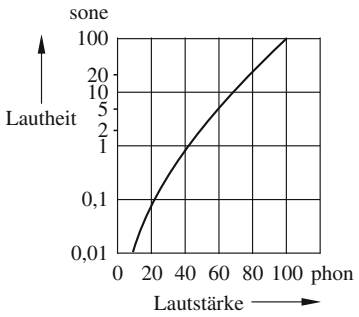


Abb. 105.2 Zusammenhang von Lautstärke und Lautheit, nach [28]

- *Schallwerte aus der Praxis*: Für den vom Menschen verursachten Schall werden die folgenden drei Werte angegeben:
  - Direkt am Mund herrscht ein Schalldruck von etwa 0,5 ... 1,5 Pa.
  - Die Schalleistung der menschlichen Stimme liegt in der Größenordnung 10^{-5} W.
  - Die in der Praxis auftretende maximale Schallleistung liegt bei einigen mW.

105.2 Schallempfänger

105.2.1 Kenngrößen für Mikrofone

Für die als Schallempfänger eingesetzten Mikrofone sind u. a. die folgenden Kenngrößen eingeführt worden:

- *Übertragungsbereich*: Das ist der vom Mikrofon gewandelte Frequenzbereich, der im Idealfall den gesamten Hörschall umfasst.
- *Richtcharakteristik*: Richtungsabhängigkeit des Übertragungsfaktors B_E (105.9). In der Hauptrichtung (0-Achse) wird $B_E = 1$ gesetzt, und für abweichende Richtungen wird der zugehörige Übertragungsfaktor auf den Wert in der Hauptrichtung bezogen. Die Richtcharakteristik ist von der Frequenz abhängig und wird häufig für eine bestimmte Frequenz angegeben. Abb. 105.3 zeigt die bei Mikrofonen am häufigsten auftretenden Richtcharakteristiken.

Die Mikrofone mit Kreis- und Achtercharakteristik sind nach Abschn. 99.1, Rundfunk-Stereoübertragung, für Stereoaufnahmen bestimmt. Allgemein eignet sich das Mikrofon mit Kreischarakteristik für Rundumaufnahmen. Wird ein solches Mikrofon mit Verstärkeranlage und Lautsprecher(n) betrieben (Raum- oder Freiplatzbeschallung), kommt es leicht zum Rückkopplungspfeifen, wenn das verstärkte Schallsignal das Mikrofon mit großer Amplitude erreicht und erneut verstärkt wird. Deshalb konstruiert man Mikrofone, die durch die Form des Gehäuses, durch Schallabschirmung zu bestimmten Seiten, durch Luftöffnungen auf der Membranrückseite und durch Wölbung der Membran die verschiedenen Cardioide-Charakteristiken erhalten. Durch diese Maßnahmen soll eine Richtcharakteristik erreicht werden, die im Wesentlichen nur den Schalldruck von vorn verarbeitet, den von der Rückseite aber nicht. Soll eine extreme Richtwirkung erreicht werden, kann das Mikrofon

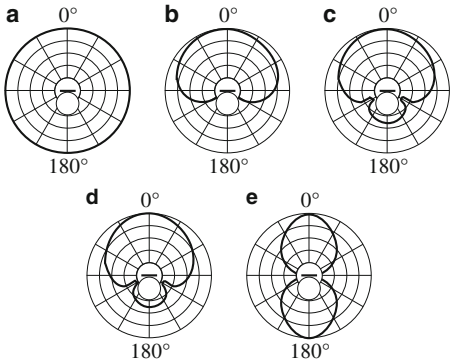


Abb. 105.3 Richtcharakteristiken von Mikrofonen mit folgenden Formen: **a** Kreis, **b** Niere oder Cardioide, **c** Supercardioide, **d** Hypercardioide, **e** Acht

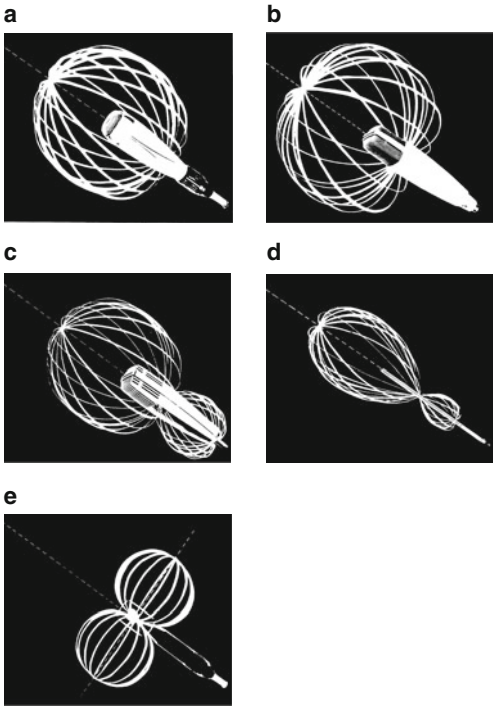


Abb. 105.4 Richtcharakteristiken nach Abb. 105.3, räumlich dargestellt (Unterlagen Fa. Sennheiser)

im Brennpunkt eines Parabolspiegels angeordnet werden. Die Richtcharakteristik wird als ebenes Diagramm dargestellt, weil die Mikrofone in dieser Hinsicht rotationssymmetrisch bezüglich der Längsachse und damit der Hauptachse sind. Abb. 105.4 stellt die Richtdiagramme aus Abb. 105.3 räumlich dar.

- *Elektrische Impedanz, minimale Abschlussimpedanz*: Das Mikrofon ist als Spannungsquelle mit Innenwiderstand (*Elektrische Impedanz*) zu sehen, an die ein ausreichend großer Lastwiderstand (*minimale Abschlussimpedanz*) angeschlossen werden muss, damit der Spannungsfall am Innenwiderstand des Mikrofons nicht zu groß wird und außerdem die Eigenschaften des Mikrofons nicht durch einen zu kleinen Lastwiderstand nachteilig beeinflusst werden (Klirrfaktor, Amplitudengang).

Tab. 105.1 Wandlerprinzipien für Mikrofone und deren Anwendungen

Wandlerprinzip	Name des Mikrofons	Anwendung
Piezo-Prinzip		
a) piezoresistiv	Kohlemikrofon	Im Telefon älterer Bauart
b) piezoelektrisch	Kristallmikrofon	Billiges Gebrauchsmikrofon
Elektromagnetisches Prinzip	Magnetisches Mikrofon	u. a. in Hörgeräten
Elektrodynamisches Prinzip	Dynamisches Mikrofon	u. a. als Studiomikrofon
Elektrostatisches Prinzip		
a) Normalausführung	Kondensatormikrofon	u. a. als Studiomikrofon
b) Sonderausführung mit Elektret	Elektret-Kondensatormikrofon	Als Gebrauchsmikrofon

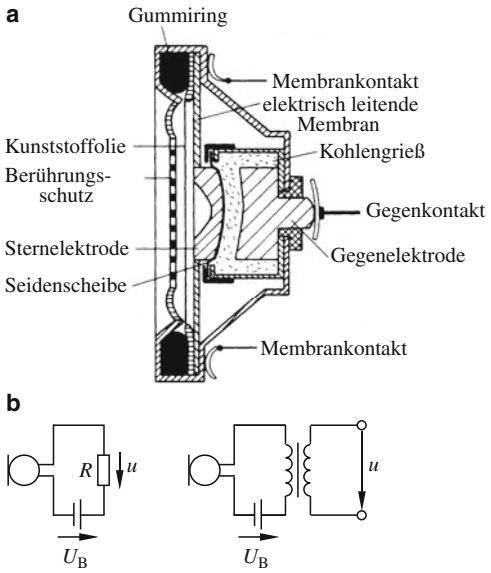


Abb. 105.5 Kohlemikrofon. **a** Aufbau, nach [28], **b** praktische Schaltung mit Lastwiderstand bzw. Übertrager

105.2.2 Mikrofonsysteme

Die am häufigsten eingesetzten Mikrofonarten lassen sich nach dem Wandlerprinzip einteilen (Tab. 105.1).

105.2.2.1 Kohlemikrofon

Abb. 105.5a stellt den Aufbau dar. Durch den Schalldruck bewegt sich die Membran in horizontaler Richtung und drückt die Kohlekörner mehr oder weniger stark zusammen. Dadurch ändert sich der Widerstand. Die Widerstandsänderung, die in etwa dem Schalldruck proportional ist, kann nach Abb. 105.5b in eine proportionale Spannungsänderung u umgewandelt werden. Dabei hat die Lösung mit Übertrager den Vorteil der galvanischen Trennung zwischen Mikrofon- und Signalkreis. Das Kohlemikrofon benötigt eine Spannungsversorgung. Der Klirrfaktor kann bis zu 25 % betragen, deshalb wurde es für Sprachübertragung im Telefon eingesetzt, wo es auf die Verständlichkeit

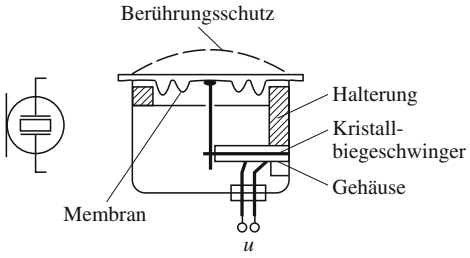


Abb. 105.6 Kristallmikrofon

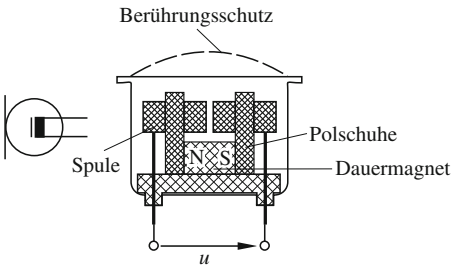


Abb. 105.7 Elektromagnetisches Mikrofon

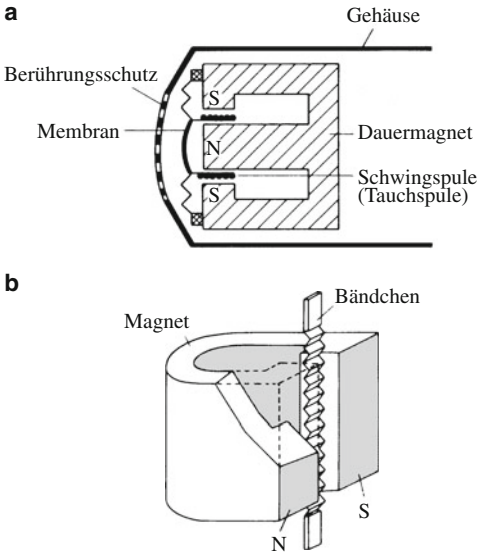


Abb. 105.8 Elektrodynamisches Mikrofon, nach [28]. **a** Tauchspulmikrofon, **b** Bändchenmikrofon

und nicht die originalgetreue Übertragung ankommt. Es ist das einzige Mikrofon, das, bedingt durch die erforderliche Spannungsversorgung, eine Leistungsverstärkung bewirkt. Nimmt man an, dass die Schallleistung der Sprache bei 10^{-5} W liegt und dass davon 1/10 vom Mikrofon verarbeitet wird, erhält man am Mikrofonausgang eine elektrische Leistung von etwa 1 mW. Die Leistungsverstärkung ist dann $1\,\text{mW}/1\,\mu\text{W} = 1000$. Der Übertragungsfaktor (105.9) beträgt $B_E \approx 100\,\text{mV}/\mu\text{bar} = 1000\,\text{mV/Pa} = 1\,\text{V/Pa}$.

105.2.2.2 Kristallmikrofon

Es wird der piezoelektrische Effekt ausgenutzt, nach dem bestimmte Materialien (Quarz, Seignettesalz, Piezokeramik) bei Druck eine proportionale Spannung abgeben, Abb. 105.6. Eine Membran aus Metall oder Kunststoff ist mechanisch mit dem Kristall verbunden. Membranbewegungen erzeugen Druck- und Zugspannungen im Kristall und können als Spannung u an den Elektroden abgegriffen werden. Da der Effekt durch Ladungsverschiebungen hervorgerufen wird, ist die elektrische Impedanz sehr groß (MΩ-Bereich), sodass zunächst ein Verstärker als Impedanzwandler nachgeschaltet werden muss. Der Klirrfaktor liegt bei einigen %, der Übertragungsfaktor bei etwa 5 mV/Pa.

105.2.2.3 (Elektro-)Magnetisches Mikrofon

Nach Abb. 105.7 besteht dieses Mikrofon aus einem feststehenden Dauermagneten, auf dessen Polschuhen eine ebenfalls feststehende Spule sitzt. Als Joch befindet sich vor den Polschuhen die durch den Schalldruck bewegte Membran, die den magnetischen Widerstand des Kreises und damit den magnetischen Fluss ändert. Damit wird in der Spule eine Spannung u induziert. Das Mikrofon ist robust und billig, der Klirrfaktor liegt bei einigen %. Durch das physikalische Prinzip bedingt steigt die Ausgangsspannung mit steigender Schallfrequenz. Diese lineare Verzerrung muss durch einen entsprechend ausgelegten Verstärker beseitigt werden.

105.2.2.4 (Elektro-)Dynamisches Mikrofon als Tauchspul- oder Bändchenmikrofon

An der Membran ist eine Spule befestigt, die in das Feld eines feststehenden Dauermagneten eintaucht, Tauchspulmikrofon, Abb. 105.8a. Die Ausgangsspannung steigt mit steigender Schallfrequenz, der Übertragungsfaktor ist in der Größenordnung 1 mV/Pa. Deshalb ist ein Vorverstärker erforderlich, der die niedrige Ausgangsspannung verstärkt und den Amplitudenfrequenzgang korrigiert. Der Klirrfaktor ist gering ($\leq 1\,\%$), deshalb wird es auch als Studiomikrofon eingesetzt. Eine Sonderausführung ist das Bändchenmikrofon, Abb. 105.8b, bei dem die Membran als geripptes Bändchen ausgeführt ist, das sich direkt im Feld des Dauermagneten bewegt. Auch hier steigt die Ausgangsspannung mit steigender Schallfrequenz, wobei bei hohen Schallfrequenzen die Ausgangsspannung durch die geringe Masse des Bändchens der Schalländerung praktisch trägheitslos folgt. Der Übertragungsfaktor ist, da es sich um eine Spule mit nur einer Windung handelt, in der Größenordnung $10 \ldots 100\,\mu\text{V/Pa}$. Der Klirrfaktor liegt unter 1 %.

105.2.2.5 Kondensatormikrofon

Wie in Abb. 105.9a dargestellt, bildet die Membran die eine Platte eines Plattenkondensators, die andere Platte ist eine im Abstand von etwa 10 µm angebrachte feste Elektrode. Ändert sich der Schalldruck, ändert sich der Abstand der Elektroden und damit die Kapazität ($C \pm \Delta C$). Der Kondensator erhält nach Abb. 105.9b über eine Spannungsquelle mit der Spannung U_B und zwei hochohmige Widerstände R ($10 \ldots 100\,\text{M}\Omega$) eine Ladung Q, die dann konstant ist, wenn die Periodendauer der niedrigsten Schallfrequenz viel kleiner ist als die Zeitkonstante

Tab. 105.2 Qualitative Eigenschaften von Mikrofonsystemen

	Empfindlichkeit gegenüber Erschütterungen	Übertragungsfaktor	Klirrfaktor	Verarbeitung tiefer Frequenzen	Verarbeitung hoher Frequenzen	Preis, Schaltungsaufwand
Kohlemikrofon	−	+	−−	−	◯	+
Kristallmikrofon	◯	◯	◯	+	+	+
Magnetisches Mikrofon	+	◯	−	−	+	+
Dynamisches Mikrofon	+	+	+	+	+	+
Bändchenmikrofon	◯	◯	+	◯	+	+
Kondensatormikrofon	+	+	++	+	+	◯
Elektretmikrofon	+	+	◯	◯	+	+

Eigenschaften: −− schlecht, − weniger gut, ◯ mittelmäßig, + gut, ++ sehr gut

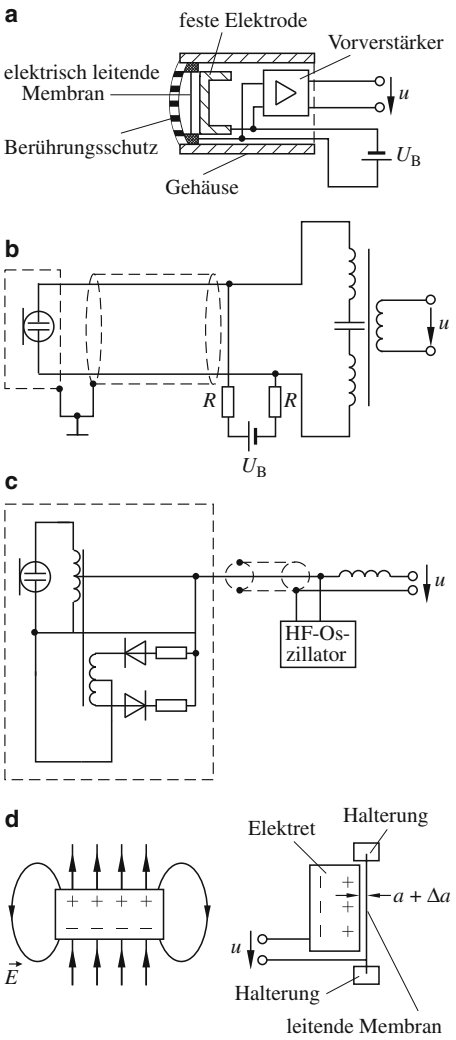


Abb. 105.9 Kondensatormikrofon. **a** Aufbau, nach [28], **b** Tonaderspeisung (DIN 45595), **c** Hochfrequenzschaltung, **d** Elektretmikrofon

$2RC$. Dann gilt: $u = U \pm \Delta C/C$. C liegt in der Größenordnung 500 … 1000 pF, sodass für R bei $f = 16$ Hz und $C = 1$ nF gilt: $R \gg 10\,\mathrm{M\Omega}$. Die zwei Leitungen zum Mikrofon führen sowohl die Signalspannung als auch die Spannung zur Ladungserzeugung, und damit ist nur eine zweiadrige abgeschirmte Leitung erforderlich (*Tonaderspeisung* nach DIN 45595). Mit dem Übertrager werden Speisespannung und Nutzspannung voneinander getrennt. Der Klirrfaktor ist wesentlich kleiner als 1 %, sodass dieses Mikrofon als Studiomikrofon eingesetzt wird. Allerdings empfiehlt es sich, in das Mikrofon einen Vorverstärker mit einem Eingangswiderstand $> 1\,\mathrm{G\Omega}$ zu integrieren, um Leitungskapazitäten nicht wirksam werden zu lassen und Störeinstrahlungen in den hochohmigen Signalkreis zu verringern. Ohne Verstärker erreicht man Übertragungsfaktoren von etwa 1 mV/Pa. Eine weitere Möglichkeit zur Mikrofonspeisung, die *Hochfrequenzschaltung*, zeigt Abb. 105.9c. Das Mikrofon erhält keine Gleichspannung, sondern liegt als Kondensator in einem Oszillatorkreis, dessen Frequenz sich schallproportional ändert (Frequenzmodulation). In einem Demodulator wird das Signal demoduliert und zusammen mit dem Hochfrequenzsignal über die Leitung übertragen. Die Spule trennt Hochfrequenz- und Nutzsignal. Der Vorteil dieser Schaltung ist, dass nur eine einadrige abgeschirmte Leitung erforderlich ist.

105.2.2.6 Elektret-Kondensatormikrofon

Eine zunehmend eingesetzte Variante ist das Elektret-Kondensatormikrofon. Ein Elektret hat auf zwei gegenüberliegenden Stirnflächen ortsfeste elektrische Ladungen, sodass ein elektrisches Feld auch ohne äußeres Anlegen einer Spannung dauerhaft vorhanden ist, Abb. 105.9d. Man beschichtet die Membran eines Kondensatormikrofons mit einem Elektret, und damit bewirkt die als konstant angenommene elektrische Feldstärke $\vec{E}$ bei einer Lageänderung Δa der Membran eine Spannungsänderung Δu an den Kon-densatorplatten: $\Delta u = E \cdot \Delta a$. Diese Mikrofone werden u. a. in Kassettenrecordern und Anrufbeantwortern eingesetzt, weil sie unempfindlich gegen Körperschall und magnetische Streufelder sind.

105.2.3 Daten und Eigenschaften verschiedener Mikrofonsysteme

Tab. 105.2 stellt die Eigenschaften von Mikrofonsystemen qualitativ einander gegenüber.

Tab. 105.3 listet typische Werte der verschiedenen Mikrofonsysteme auf.

Tab. 105.3 Typische Werte für verschiedene Mikrofonsysteme

	Übertragungsfaktor in mV/Pa	Elektrische Impedanz in Ω	Klirrfaktor in %	Frequenzbereich in Hz	Dynamik in dB[a]
Kohlemikrofon	1000	50 … 500	25	600 … 4000	
Kristallmikrofon	5	1 … 5 MΩ	2	30 … 10.000	60
Magnetisches Mikrofon	10	2000	10	300 … 6000	
Dynamisches Mikrofon	1	200	1	40 … 14.000	60
Bändchenmikrofon	0,1	0,1	0,5	40 … 18.000	50
Kondensatormikrofon	1	100 MΩ	0,1	20 … 20.000	80

[a] Die Dynamik ist das logarithmische Verhältnis von der größten noch einwandfrei verarbeitbaren Schallamplitude zur kleinstmöglichen verarbeitbaren Schallamplitude (siehe auch DIN 40146)

105.3 Schallsender

Unter den Begriff Schallsender fallen Lautsprecher und Hörer. Es wurden bzw. werden die folgenden Systeme verwendet.

105.3.1 Lautsprecher- und Hörer-Systeme

105.3.1.1 Elektrodynamisches System

Im Luftspalt eines *Elektromagneten* bewegt sich eine Spule, deren Strom einen angenähert proportionalen Schalldruck erzeugt. Dieses System wurde angewendet, als noch keine leistungsfähigen Dauermagnete zur Verfügung standen.

105.3.1.2 (Elektro-)Magnetisches System

Eine Membran bewegt sich im Luftspalt eines Dauermagneten, Abb. 105.10. Auf dem Dauermagneten bzw. seinen Polschuhen sind zusätzlich zwei in Reihe geschaltete Spulen angebracht, durch die der zum Schalldruck proportionale Wechselfluss erzeugt wird und der sich dem Gleichfluss des Dauermagneten überlagert. Dieses System wird u. a. im Telefon (auch als Mikrofon möglich) und in Kopfhörern eingesetzt. Die auf die Membran ausgeübte Kraft F ist proportional dem Quadrat der Flussdichte B. Ist B_0 die vom Dauermagneten erzeugte Flussdichte und nimmt man die Flussdichte des Nutzsignales sinusförmig an, $B_S = \hat{B} \cdot \sin \omega t$, folgt für die Kraft:

$$F \sim B^2 = (B_0 + \hat{B} \cdot \sin \omega t)^2$$
$$= B_0^2 + 2 \cdot B_0 \cdot \hat{B} \cdot \sin \omega t + \hat{B}^2 \cdot \sin^2 \omega t \qquad (105.14)$$

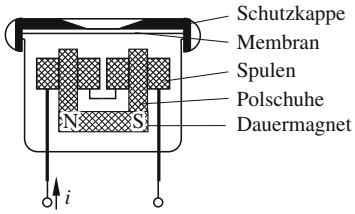


Abb. 105.10 (Elektro-)Magnetisches System

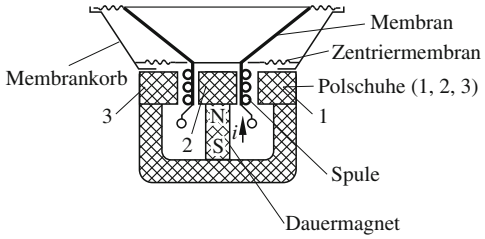


Abb. 105.11 Aufbau eines dynamischen Lautsprechers

Der Summand $\hat{B}^2 \cdot \sin^2 \omega t$ führt zu nichtlinearen Verzerrungen, sodass für die Konstruktion des Systems gelten muss: $B_0 \gg \hat{B}$.

105.3.1.3 Dynamisches System

Da keine Verwechslung mehr mit dem elektrodynamischen System (Abschn. 105.3.1.1) gegeben ist, spricht man vom dynamischen System anstelle vom permanentdynamischen System. Nach Abb. 105.11 bewegt sich eine vom Signalstrom i durchflossene Spule im Feld eines Dauermagneten. Mit der Spule mechanisch verbunden ist eine trichterförmige Membran, die die umgebende Luft zu Schallschwingungen anregt. Die auf die Spule ausgeübte Kraft berechnet sich nach der *Lorentz-Kraft*:

$$F \sim B_0 \cdot i \qquad (105.15)$$

B_0 vom Dauermagneten erzeugte Flussdichte im Luftspalt in $\mathrm{V\,s/m^2}$, i Signalstrom in A.

Man erhält damit theoretisch einen linearen Zusammenhang zwischen der Membranauslenkung und dem Signalstrom.

105.3.1.4 Elektrostatisches System

Es handelt sich hierbei um die Umkehrung des beim Kondensatormikrofon eingesetzten Prinzips. Nach Abb. 105.12 ist in geringem Abstand zu einer festen Elektrode FE die Membran M als bewegliche Elektrode angebracht. Beide sind elektrisch voneinander isoliert und bilden damit die Platten eines Kondensators. Beim Anlegen einer Spannung an die Elektroden wird

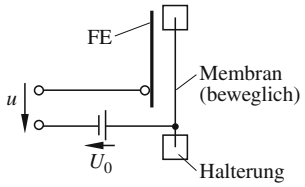


Abb. 105.12 Prinzip eines elektrostatischen Lautsprechers

die Membran M von FE angezogen, wölbt sich durch und überträgt diese Bewegung an die umgebende Luft. Da die Kraft F proportional zum Quadrat der Spannung ist, muss auch hier durch eine entsprechend große Gleichspannung U_0 eine Linearisierung erreicht werden. Es gilt die Ableitung zu (105.14), wenn man die Flussdichte durch die Spannung ersetzt:

$$F \sim u^2 = (U_0 + \hat{u} \cdot \sin \omega t)^2 = U_0^2 + 2 \cdot U_0 \cdot \hat{u} \cdot \sin \omega t + \hat{u}^2 \cdot \sin^2 \omega t \qquad (105.16)$$

Für eine nahezu lineare Abhängigkeit muss gelten: $U_0 \gg \hat{u}$.

Durch die geringe Masse der bewegten Teile ergibt sich besonders bei hohen Frequenzen ein geringer Klirrfaktor. Diese Lautsprecher können allerdings nur wenig Leistung abgeben, und die Schallabstrahlung ist stark gerichtet.

105.3.1.5 Piezoelektrisches System

Es handelt sich um die Umkehrung des beim Kristallmikrofon verwendeten Prinzips. Abb. 105.13 zeigt das Kristallelement, das einseitig fest eingespannt ist und mit dem anderen Ende spannungsproportional schwingt und die Lautsprechermembran bewegt. Es sind zwei Kristallplättchen so aufeinandergeklebt, dass sich die entsprechende Biegung analog zu einem Bimetallsystem ergibt. Die mechanische Auslenkung ist nur bedingt proportional zur angelegten Spannung, sodass diese Lautsprecher einen nicht zu vernachlässigenden Klirrfaktor haben. Sie werden zur Wiedergabe von mittleren und hohen Frequenzen eingesetzt.

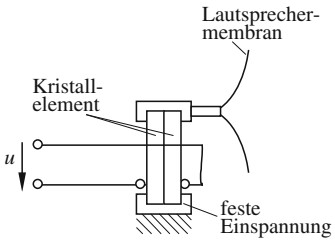


Abb. 105.13 Kristall-Lautsprecher

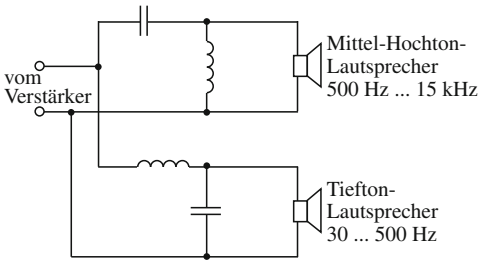


Abb. 105.14 Trennung von Tiefton- und Hochtonlautsprechern durch Filter (Frequenzweichen)

105.3.2 Kenngrößen, Daten

Es folgen noch einige Begriffe im Zusammenhang mit Lautsprechern und Hörern.

- *Akustischer Kurzschluss*: Befindet sich eine Lautsprechermembran frei im Raum, d. h. sowohl Vorder- als auch Rückseite strahlen eine Schallwelle ab, können sich beide Schallwellen im Raum so überlagern, dass sie sich ungünstigenfalls genau auslöschen. Man spricht dann von *akustischem Kurzschluss*. Das ist besonders kritisch, wenn die Wellenlänge λ der Schallschwingung groß ist gegen den Membrandurchmesser. Fügt man dagegen den Lautsprecher in eine passende Öffnung einer starren, sehr großen Wand ein, wird dieser Kurzschluss vermieden. Eine alternative Lösung mit weniger großen Abmessungen sind Lautsprechergehäuse, die allseitig geschlossen sind und nur vorn eine Öffnung in Lautsprechergröße haben.
- *Richtcharakteristik*: Kommt die abgestrahlte Schallwellenlänge in die Größenordnung des Membrandurchmessers, setzt eine Richtwirkung ein, die mit weiter sinkender Wellenlänge zunimmt. Es gilt:

Schallfrequenz	10 Hz	100 Hz	1000 Hz	10 kHz	16 kHz
Schallwellenlänge	34 m	3,4 m	34 cm	3,4 cm	2,125 cm

(bei 20 °C in Luft und 1000 hPa).
Demnach ist es nur unvollkommen möglich, mit *einem* Lautsprecher den gesamten Hörschall zu übertragen. Als Obergrenze zwischen höchster und niedrigster von einem Lautsprecher übertragbarer Frequenz gilt der Wert 100. Man setzt deshalb zwei oder drei Lautsprecher ein und optimiert jeden für einen Teilfrequenzbereich. Bei zwei Lautsprechern wählt man z. B. einen Tieftonlautsprecher für den Frequenzbereich 30 bis 500 Hz und einen Mittel-Hochton-Lautsprecher für den Frequenzbereich 500 Hz bis 15 kHz (Abb. 105.14). Bei drei Lautsprechern kann die Unterteilung so aussehen: 30 bis 500 Hz, 500 Hz bis 4 kHz, 4 bis 20 kHz. Die einzelnen Lautsprecher werden über Frequenzweichen mit dem zugeordneten Frequenzbereich angesteuert.

Tab. 105.4 Typische Kenndaten verschiedener Schallsendersysteme

	Nennscheinwider-stand bzw. Kapazität	Frequenzbereich	Leistung bis max.	Übertragungs-faktor	Verzerrungen	Anwendungen
Magnet. Lautsprecher	200 … 2000 Ω	300 Hz … 5 kHz	3 W	5 Pa/V	Groß	Billige Systeme
Magnet. Kopfhörer	200 … 2000 Ω	100 Hz … 6 kHz	50 mW	5 Pa/V	Groß	Telefon, Kleinhörer
Dynam. Breitband-Lautspr.	3 … 25 Ω	30 Hz … 18 kHz	200 W	0,5 Pa/V	Klein	Universell
Dynam. Tiefton-Lautspr.	2 … 8 Ω	30 … 500 Hz	500 W	0,7 Pa/V	Klein	HiFi-Anlagen
Dynam. Hochton-Lautspr.	3 … 25 Ω	500 Hz … 20 kHz	80 W	0,2 Pa/V	Klein	HiFi-Anlagen
Dynam. Kopfhörer	4 … 200 Ω	16 Hz … 20 kHz	100 mW	0,1 Pa/V	Sehr klein	Höchste Ansprüche
Piezoel. Lautsprecher	1 … 5 nF	1 … 20 kHz	2 W	0,5 Pa/V	Mittel	Allgemeine Anwendung
Elektrostatischer Lautspr.	100 … 500 pF	1 … 20 kHz	5 W	0,1 Pa/V	Klein	In kleineren HiFi-Anlagen

- *Hörer*, *Kopfhörer*: Hörer sind Schallsender, die den Schall direkt am Ohr abstrahlen. Deshalb heißen sie auch Kopfhörer. Prinzipiell sind für Hörer alle genannten Systeme brauchbar, allerdings hat sich für die Wiedergabe hochwertiger Musikaufnahmen der dynamische Hörer durchgesetzt, der seinen Namen vom verwendeten dynamischen System hat. Die bei 1 kHz gemessene Impedanz liegt bei 4, 8 oder 16 Ω zum direkten Anschluss an die Lautsprecherausgänge eines Verstärkers oder bei 200 Ω. Es gibt sie als *offene Hörer*, bei denen die Membran nach vorn und hinten abstrahlt, und als *geschlossene Hörer*, bei denen der Hörer nur nach vorn abstrahlt.
  Tab. 105.4 listet einige Eigenschaften der verschiedenen Systeme auf.
- *Motional Feedback-System*: Vor allem bei Tieftonlautsprechern gibt es durch die relativ große Masse der Membran mit Schwingspule Signalverzerrungen. Man bringt deshalb auf der Membran einen Weg- oder Beschleunigungsaufnehmer an, der den tatsächlichen Weg bzw. die momentane Beschleunigung ermittelt und mit dem Sollwert der ansteuernden elektrischen Größe vergleicht. Weichen beide voneinander ab, wird der Spulenstrom entsprechend eingestellt. Man erreicht auf diese Weise kleinere Klirrfaktoren.

Vermittlungstechnik

106

Teil XIII

Wilfried Plaßmann



106.1 Grundbegriffe 1294
106.2 Vermittlung 1294
106.3 Verkehrstheorie 1295
106.4 Endgerät 1296

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_106

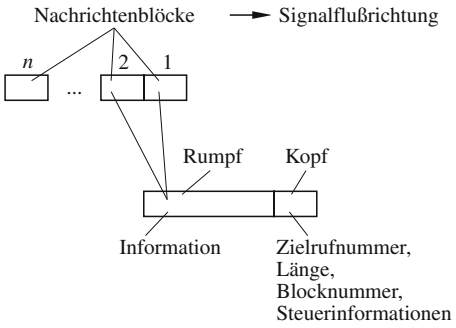


Abb. 106.1 Paketvermittlung

106.1 Grundbegriffe

In einem vorhandenen Nachrichtennetz sorgt die *Vermittlungstechnik* dafür, dass Informationen zu jeder Zeit von einem beliebigen *Zugangspunkt* zu einem beliebigen anderen Zugangspunkt übertragen werden können. Der gesamte Vorgang wird mit *Vermittlung* bezeichnet. Das größte weltweit vorhandene Nachrichtennetz ist das kombinierte Internet-Telefonnetz. Die *Zugangspunkte* sind die Anschlusspunkte (*ortsfeste* Anschlussdosen) für die *Endgeräte*, wie z. B. Rechner, Telefonapparate, Faxgeräte oder Modems, oder aber ein *ortsveränderliches* Endgerät wie ein Laptop, ein schnurloses Telefon oder ein Funktelefon („Handy“). Die Informationen werden vorwiegend als elektrische Signale übertragen, allerdings gewinnt die optische Signalübertragung an Bedeutung, siehe Kap. 103. Die Verbindung zweier oder mehrerer Zugangspunkte kann entweder nur für eine bestimmte Zeit (Telefongespräch, Senden eines Faxes) oder dauernd erfolgen („Standleitung“ als Telefon- und/oder Datenleitung zur Verbindung von z. B. Zweigwerken einer Firma oder als Internetverbindung über DSL, siehe Abschn. 107.2.2.2). Bisher wurden überwiegend *geschaltete* Kanäle eingesetzt, d. h., beide Endgeräte wurden über einen für diese Verbindung reservierten Kanal verbunden: *Leitungs-* oder *Durchschalte-Vermittlung*. Im Gegensatz dazu spricht man von *virtuellen* Kanälen, wenn ein Kanal für mehrere Verbindungen „gleichzeitig“ eingesetzt wird: Bei Telefongesprächen kann der Kanal in den Sprechpausen anderweitig genutzt werden. Das setzt voraus, dass die Übertragung in zeitlich begrenzten Paketen geschieht: *Paketvermittlung*. Nach Abb. 106.1 besteht die gesamte Nachricht aus einzelnen *Nachrichtenblöcken*, die sich aus dem *Kopf* mit den zugehörigen Zuordnungs- und Steuerinformationen und dem *Rumpf* mit der eigentlichen Information zusammensetzen.

Die Paketvermittlung wird im digitalen Datex-P-Netz der Deutschen Telekom angewendet, siehe auch Abschn. 107.2.2.2.

106.2 Vermittlung

Um das oben genannte Ziel einer Vermittlung zu erreichen, zeigt Abb. 106.2 eine naheliegende Lösung für deren Aufbau. *Jeder* der *N* Teilnehmer kann *jederzeit* mit jedem anderen Teilnehmer verbunden werden. Das setzt voraus, dass die *Teilnehmer-Anschlussleitung* (TA) jedes Teilnehmers bis zur *Ortsvermittlungsstelle* (OVST) (kurz auch *Vermittlungsstelle*) geführt wird. Diese Lösung mag bei maximal einigen tausend Teilnehmern (Telefonanschlüssen) durchführbar sein, für die gesamte Bundesrepublik ist dieser Lösungsweg nicht praktikabel. Die Vermittlungstechnik für das Festnetz ist historisch gewachsen und gibt deshalb gewisse Rahmenbedingungen vor.

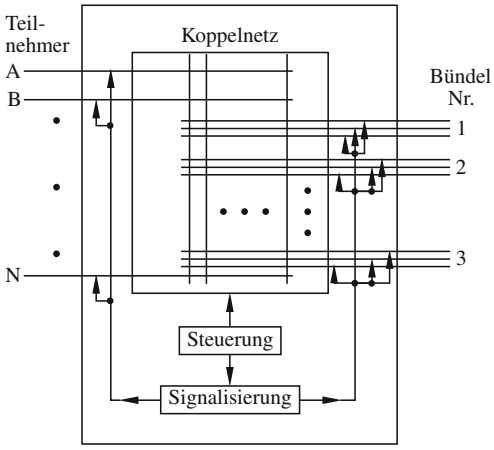


Abb. 106.2 Prinzip einer Vermittlungsstelle mit maximal möglicher Zahl von Verbindungen

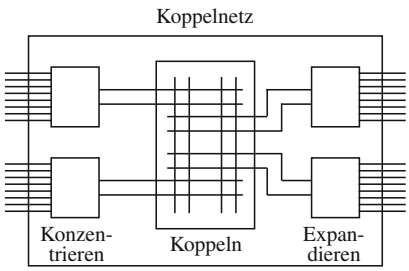


Abb. 106.3 Koppelnetz mit Konzentration und Expansion

Dazu gehört, dass 1. die Rufnummer in der Regel noch die Struktur der Vorgängerlösung aufweist und 2. die Teilnehmer-Anschlussleitungen noch als 2-Draht-Kupferleitungen vorliegen und häufig bis zur Vermittlungsstelle geführt worden sind. Weiterhin hat die Erfahrung gezeigt, dass von den *möglichen* Verbindungen im Mittel lediglich etwa 10 bis 15 % gleichzeitig belegt sind. Deshalb hat man das im Abb. 106.3 gezeigte Verfahren der Konzentration-Expansion angewendet. Das war nicht nur auf höheren Ebenen der Fall, auch im Bereich der OVST wurde bei der Erschließung von Neubaugebieten die Ankopplung an das bestehende Netz zwecks Erweiterung vorgenommen. Zum besseren Verständnis der Begriffe sei hier kurz die Vorgängerlösung erläutert, und zwar an der Rufnummer 05723 71xx bzw. 71xx (xx steht für 2 Ziffern). Bei der Rufnummer 71xx handelt es sich um einen Teilnehmer aus Bad Nenndorf (westlich von Hannover). Die unterste Ebene der Vermittlung war die Ortsvermittlungsstelle (OVST), die räumlich die durch die Ortsgrenzen vorgegebenen Teilnehmer erfasste.

Angrenzende kleinere Orte wurden größeren Orten zugeordnet. Um einen Teilnehmer aus dem eigenen Ortsnetz anzurufen, wurde und wird die Rufnummer gewählt (71xx), die allerdings nie mit einer Null beginnen darf, denn daran erkennt die OVST, dass es sich um ein Ortsgespräch handelt und ihr Bereich nicht verlassen wird. Befindet sich der anzurufende Teilnehmer nicht im Ortsbereich des Anrufenden, muss die Vorwahl des Anzurufenden der Rufnummer vorausgeschickt werden, im Beispiel 05723. Jeder Ortsbereich hat seine eigene eindeutige Vorwahl. An der 0, der *Verkehrsausscheidungsziffer Inland*, erkennt die OVST, dass es sich um ein Ferngespräch handelt, also ein Gespräch, dass den Bereich der eigenen OVST verlässt und in das sog. *Nationale Fernnetz* übergeht. An den Ziffern 5723 zeigt sich die Hierarchie des Telefonnummernaufbaus: Die OVST stellt zunächst eine Verbindung mit der höchsten Ebene, einer von 8 Zentralvermittlungsstellen (ZVST), hier Hannover anhand der Ziffer 5, her. Dann folgt die nächstniedrigere Ebene, die Hauptvermittlungsstelle (HVST), anhand der 57 in Minden/Westf., weitergehend über die Knotenvermittlungsstelle (KVST) 572 in Stadthagen zur Ortsvermittlungsstelle (OVST) 5723 in Bad Nenndorf. Dieser Weg war auch von dem Anrufenden in umgekehrter Reihenfolge zu beschreiten. Dieser sog. *Kennzahlweg* war streng strukturiert aber nicht immer der optimale Weg (längerer Verbindungsaufbau, Belegen längerer Leitungen/Umweg und damit erhöhte Leitungsblockierungen). Es wurden deshalb im Laufe der Zeit schon viele Quer- oder Direktverbindungen eingerichtet (OVST A direkt zu OVST B, OVST A direkt zu HVST D), um den Fernsprechverkehr zu optimieren. Der Kennzahlweg wurde zur Ausnahme. Soll ein Anruf ins Ausland erfolgen, muss der OVST dies mitgeteilt werden. Vor der eigentlichen Rufnummer wird deshalb 00yy gewählt, wobei die 00 die *Verkehrsausscheidungsziffer Ausland* ist, gefolgt von der *Länderkennzahl* yy (49 ist für Anrufer aus dem Ausland die Länderkennzahl für Deutschland). Der Verbindungsaufbau erfolgt dann über die Auslandsvermittlungsstelle (AVST) in Frankfurt.

Allerdings ist heute eine eindeutige Zuordnung zwischen Rufnummer und Vermittlungsstelle nicht mehr in jedem Fall möglich. Zum einen kann die Rufnummer bei Umzug mitgenommen werden, zum anderen werden Rufnummern blockweise an Telefonanbieter vergeben.

Mit der Einführung von ISDN (siehe Abschn. 107.2.2) und der damit verbundenen Digitalisierung der Sprachsignale fand ein mehr oder weniger gleitender Übergang von der alten zur neuen Vermittlungstechnik statt. Die strenge hierarchische Struktur ist nicht mehr vorhanden. Es gibt letztlich nur noch 2 Ebenen, die sehr stark vernetzt sind. Es wird unterschieden zwischen:

- Vermittlungseinheit Ortsnetz (VE:O),
- Vermittlungseinheit Fernnetz (VE:F),
- Vermittlungseinheit mit Netzübergangsfunktion (VE:N),
- Vermittlungseinheit Ausland (VE:A),
- Durchgangsvermittlungsstelle.

- *Vermittlungseinheit Ortsnetz (VE:O)*: Sie sind Bestandteil der unteren Ebene, sind gegenüber der früheren OVST wesentlich größer und können mehr als 100.000 Anschlüsse verwalten. Mehrere VE:O lassen sich zusammenschalten, auch Teilvermittlungsstellen als Zusammenfassung kleinerer Einheiten werden eingesetzt, die dann mit einer VE:O verbunden werden. Die VE:Os können direkt mit anderen Vermittlungseinheiten Verbindung aufnehmen.
- *Vermittlungseinheit Fernnetz (VE:F)*: National gehören sie zur oberen Ebene. Sie sind sehr intensiv untereinander vernetzt und Durchgangsvermittlungsstellen. Mehrere geografisch angrenzende bzw. benachbarte VE:O sind mit ihrer VE:F verbunden. Sie haben i. Allg. keine direkten Teilnehmeranschlüsse.
- *Vermittlungseinheit mit Netzübergangsfunktion (VE:N)*: Hier werden Verbindungen aus dem eigenen Netz in das Netz einer anderen Telefongesellschaft weitergeleitet. Sie sind häufig in die VE:F integriert.
- *Vermittlungseinheit Ausland (VE:A)*: Sie übernehmen den Telefonverkehr mit dem Ausland. Es sind reine Durchgangsvermittlungsstellen.
- *Durchgangsvermittlungsstelle*: Sie wickeln nur den Verkehr zwischen Vermittlungsstellen ab und sind sehr stark mit anderen Vermittlungsstellen vernetzt.

Für die Orts- und Fernvermittlung werden die Systeme EWSD (Siemens, Bosch, Telekom) und S 12 (Alcatel) eingesetzt. Ein Unterschied besteht darin, dass Hin- und Rückweg bei einer Telefonverbindung bei EWSD gemeinsam geschaltet werden, bei S 12 getrennt. Technisch gesehen sind die Vermittlungsanlagen riesige Schränke, die gefüllt sind mit elektronischen Baugruppen. Mechanische Schalteinrichtungen sind nicht mehr enthalten. Die gesamte Vermittlung wird über Rechner gesteuert. Das hat den Vorteil, dass Verschleißelemente entfallen, die Vermittlungsstellen vollautomatisch und teilweise ohne anwesendes Personal betrieben werden und die Kontrolle und Fehlersuche über Fernzugriff möglich ist. Von den Telefongesellschaften angebotene Dienste können bei Bedarf über Knopfdruck zu- oder abgeschaltet werden.

106.3 Verkehrstheorie

Betrachtet man die im Mittel auftretende Belegung von Vermittlungseinrichtungen an Werktagen über einen Zeitraum von 24 Stunden, ergeben sich Schwankungen, die u. a. von den Geschäftszeiten, den Mittagspausen, dem Beginn des ermäßigten Tarifs und dem Eintritt der Nachtruhe abhängen. An Sonn- und Feiertagen erhält man eine andere Verteilung, außerdem ist sie unterschiedlich für den Orts- und den Fernverkehr. Um ein Beurteilungskriterium für den Ausbau einer Anlage zu bekommen, ermittelt man eine *repräsentative* Stunde, die *Hauptverkehrsstunde*. Die in ihr festgelegte Belegung liegt zwischen dem Mittelwert über 24 Stunden und dem Maximalwert und ergibt sich u. a. aus einer Wirtschaftlichkeitsberechnung.

Verkehrsmenge:

$$Y = c \cdot T_m \quad \text{in Erlangstunden (Erl h)} \tag{106.1}$$

c Zahl der Belegungen, T_m mittlere Belegungsdauer in s.

Verkehrswert:

$$y_v = Y/T = c \cdot T_m/T \quad \text{in Erlang (Erl)} \tag{106.2}$$

T Beobachtungsdauer in s.

1 Erl bedeutet, dass die Einrichtung dauernd durch Belegung genutzt wird. Dies ist damit der Höchstwert, der erreicht werden kann. Für den Teilnehmeranschluss an das öffentliche Netz gilt ein Mittelwert von etwa 0,1 Erl.

- *Angebot*: Verkehrswert, der einer Anlage zugeführt wird, unabhängig davon, ob er bearbeitet werden kann oder nicht.

$$\textit{Angebot} \quad A = C_a \cdot T_m \tag{106.3}$$

C_a Zahl der Belegungen bzw. Belegungsversuche, T_m mittlere Belegungsdauer in s.

- *Leistung*: Maximaler Verkehrswert einer Anlage.

$$\textit{Leistung} \quad y = C_y \cdot T_m \tag{106.4}$$

C_y Zahl der maximal möglichen Belegungen, T_m mittlere Belegungsdauer in s.

- *Verlustsystem*: Das Angebot übersteigt die Leistung. Dann gilt:

$$\textit{Restverkehr} \quad R = A - y = (C_a - C_y) \cdot T_m \tag{106.5}$$

und für den Verlust erhält man:

$$\textit{Verlust} \quad V = R/y \quad \text{(bezogen auf die Leistung)} \tag{106.6}$$

- *Wartezeitsystem*: Eine momentan nicht herstellbare Belegung wird solange in einer Warteschleife gespeichert, bis sie durch eine freiwerdende Einrichtung durchgeführt werden kann. Daraus folgt die

$$\textit{mittlere Wartedauer} \quad T_w = T_{w\,\mathrm{ges}}/C_w \tag{106.7}$$

$T_{w\,\mathrm{ges}}$ Wartezeit aller wartenden Belegungen zusammen, C_w Anzahl der verzögerten Belegungen.

106.4 Endgerät

Ein häufig eingesetztes Endgerät ist der *Fernsprechapparat*, abgekürzt FeAp (Telefonapparat). Abb. 106.4a zeigt den Innenaufbau einer älteren Ausführung mit mechanischen Kontakten. Dieses Modell ist heute durch vollelektronische Geräte ersetzt worden, Abb. 106.4b, kann aber manchmal noch verwendet werden (siehe unten). Es werden 2 Leitungen, die a- und die b-Ader, benötigt. Bei aufgelegtem Hörer sind die Kontakte des Gabelumschalters GU geöffnet. Ein eingehender Ruf (Klingelzeichen) wird als 25- oder 50-Hz-Wechselspannung eingespeist und gelangt über C1 an den Wecker W. Ist die Wählscheibe in Ruhestellung, sind die Kontakte nsi und nsr geschlossen, nsa ist geöffnet. Beim Abheben des Hörers schließen die Kontakte des GU, und die Sprechwechselspannung des Anrufers gelangt über den Übertrager Ü an den Fernhörer F (Hörkapsel). Die zwei antiparallelen Dioden (GG) begrenzen Störgeräusche mit sehr hohen Amplituden auf ein erträgliches Maß. Spricht der Angerufene in das Mikrofon M, wird die Sprechwechselspannung zum Anrufer übertragen. In der Wicklung w1 des Übertragers Ü fließt der Sprech-Wechselstrom entgegengesetzt zu dem in

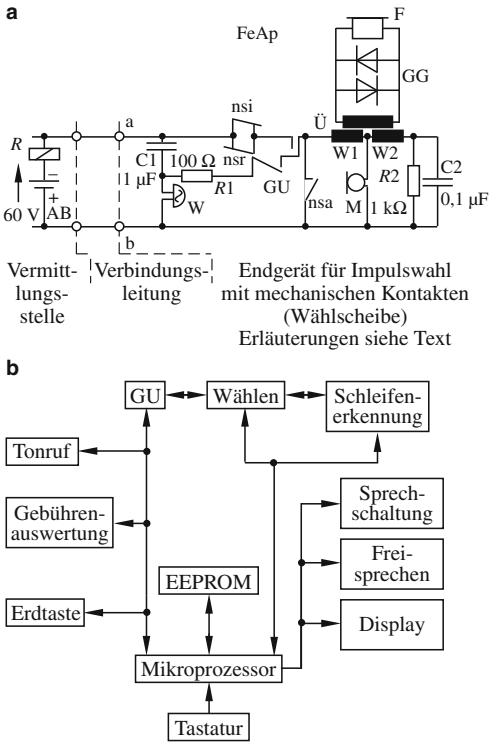


Abb. 106.4 Fernsprechapparate. **a** Innenschaltung des FeAp 48/61 für Impulswahl (ältere Ausführung), **b** Prinzip eines Komforttelefons für Mehrfrequenzwahl

Wicklung w2, sodass sich die Flüsse auf der Sekundärseite des Übertragers teilweise aufheben. Damit hört der Sprechende seine eigene Sprache nur noch mit geringer Amplitude im Hörer, was als angenehm empfunden wird. Eine vollständige Kompensation ist übrigens nicht erwünscht, weil dann beim Sprechenden der Eindruck eines defekten Telefons entsteht. Die Kombination aus R2 und C2 hat auch die Aufgabe des Leitungsabschlusses. Wird der Hörer abgehoben, fließt über nsi/nsr, GU, w1, w2, R2 und das in der Vermittlungsstelle vorhandene Relais R ein Gleichstrom, sodass das Relais anzieht und den Verbindungswunsch erkennt und weiterleitet. Wird die Wählscheibe aufgezogen, Abb. 106.5, schließt der Kontakt nsa, und der Kontakt nsr öffnet. Durch Überbrücken von R2 durch nsa kann in der Vermittlungsstelle ein noch höherer Strom fließen, der für die Steuerung eines Wählers benutzt wird (Änderung des Schleifenstromes). Lässt man die Wählscheibe nach Wahl der Ziffer N ablaufen, öffnet nsi $(N + 2)$-mal. Die letzten 2 zusätzlichen Öffnungen werden aber durch nsr wieder unterdrückt, sodass die Zahl der Unterbrechungen tatsächlich der gewählten Ziffer entspricht. Abb. 106.5 stellt den Ablauf bei Wahl der Ziffer 4 dar. Der Wähler in der Vermittlungsstelleerkennt eine neue gewählte Ziffer daran, dass zwischen zwei Unterbrechungen eine Zeit von mindestens 400 ms vergeht. Mit den zwei zusätzlichen unterdrückten Impulsen lässt sich diese Zeit auch bei sehr schnellem Wählen zweier aufeinanderfolgender Ziffern mit der Wählscheibe erreichen. Bei Telefonapparaten mit Zifferntasten wird die Ziffernfolge bei zu schneller Tastenbetätigung intern zwischengespeichert und dann zur Wähleinrichtung entsprechend den Vorgaben übertragen.

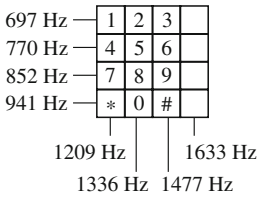


Abb. 106.5 Zuordnung der Frequenzen zu den Tasten beim Mehrfrequenzwahlverfahren

Die in Abb. 106.4b im Prinzip dargestellte Innenschaltung neuer Telefonapparate setzt einen Mikroprozessor zur Steuerung ein, um den wesentlich erweiterten Funktionsumfang zu bearbeiten. Dazu gehören die Ansteuerung eines LCD-Displays und die Bereitstellung der Wählfrequenzen. Weiterhin müssen Funktionseinheiten wie die Freisprecheinrichtung, die Gebührenerfassung, besondere Tastenbelegungen und der Wählvorgang bedient werden. In einem EEPROM werden das Steuerprogramm und z. B. vom Anwender häufig benutzte Telefonnummern oder Nummern für die Wahlwiederholung abgespeichert.

Für die Übermittlung der Ziffern einer Rufnummer gibt es zwei Verfahren:

- *Impulswahlverfahren*: Dieses Verfahren wurde oben beschrieben. Die einzelnen Ziffern werden durch eine entsprechende Anzahl von Änderungen des Schleifenstromes übermittelt. Dieses Wahlverfahren wird nicht mehr verwendet. Nur in privaten Telefonanlagen ist es noch manchmal einstellbar, wird dann aber intern in das Mehrfrequenzwahlverfahren umgesetzt.
- *Mehrfrequenzwahlverfahren*: Die Ziffern werden durch gleichzeitiges Senden von zwei unterschiedlichen Frequenzen übertragen. Abb. 106.5 zeigt die Zuordnung. Die 4 Tasten in der Spalte ganz rechts sind für den Verbindungsaufbau nicht erforderlich und werden, falls vorhanden, für Sonderfunktionen eingesetzt (z. B. Wahlwiederholung der zuletzt gewählten Telefonnummer).
- *Netzkennzahl*: Diese Zahlenkombination wird der eigentlichen vollständigen Rufnummer vorangestellt und legt fest, mit welcher Telefongesellschaft die folgende Verbindung durchgeführt und abgerechnet werden soll.

Kommunikations- und Datennetze

107

Teil XIII

Wilfried Plaßmann



107.1 Strukturen lokaler Kommunikations- und Datennetze 1300
107.2 Öffentliche Kommunikations- und Datensysteme 1300
107.3 Öffentliche Funknetze 1305
107.4 Private Funknetze 1307

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_107

Die zur Zeit vorhandenen Kommunikations- und Datennetze werden ständig weiterentwickelt und dem Stand der Technik angepasst. Von den momentan zur Verfügung stehenden Netzen sollen einige mit ihren wesentlichen Eigenschaften dargestellt werden. Weitere Einzelheiten siehe auch Kap. 55 „Datenkommunikation“.

Bei Kommunikations- und Datennetzen unterscheidet man drei Betriebsarten:

- *Simplex-Betrieb*: Die Übertragung von Informationen oder Daten ist nur in *einer* Richtung möglich. Sender und Empfänger können ihre Funktion nicht tauschen.
- *Halbduplex-Betrieb*: Informationen können nicht gleichzeitig in beiden Richtungen übertragen werden. Beide Seiten einigen sich, wer momentan der Sender und wer der Empfänger ist. Bekannt ist das Verfahren unter dem Begriff *Gegensprechverkehr*.
- *Vollduplex-Betrieb*: Der Informationsaustausch geschieht in beiden Richtungen gleichzeitig, entweder auf getrennten Kanälen (Leitungen), für jede Richtung ein Kanal, oder über nur einen Kanal (Leitung), wobei beide Informationsflüsse ineinander geschachtelt werden, was bei einer Digitalübertragung einfach realisierbar ist und bei entsprechend hoher Bitrate einer „quasi gleichzeitigen“ Übertragung in beiden Richtungen entspricht.

107.1 Strukturen lokaler Kommunikations- und Datennetze

Diese Netze, mit LAN (*l*ocal *a*rea *n*etwork) bezeichnet, überschreiten streng genommen nicht die Grundstücksgrenzen. Sie werden sowohl gewerblich in Firmen als auch im privaten häuslichen Bereich eingesetzt. Abb. 107.1 zeigt verschiedene Möglichkeiten der Vernetzung der einzelnen Stationen.

Bei der Sternform sorgt eine Zentraleinheit für einen geordneten und damit sicheren Datenaustausch. Allerdings ist das gesamte System bei Ausfall der Zentraleinheit nicht mehr funktionsfähig. Das Ringsystem erfordert einen relativ geringen Verdrahtungsaufwand. Bei Ausfall einer Station kann das gesamte System nur eingeschränkt weiter benutzt werden. Der Datentransfer ist problematischer, da er über mehrere Zwischenstationen zur Zielstation (Zielrechner) weitergeleitet wird. Das Senden und Empfangen von Daten wird durch das *Token-(Ring-)Verfahren* geregelt: Erst wenn eine Station ein bestimmtes Bitmuster, das Token erhält, sendet sie die Daten einschließlich Empfängeradresse. Diejenigen Stationen, die nicht an sie adressierte Daten erhalten, leitet sie an die Folgestation weiter. Wenn die Zielstation den Empfang der Daten quittiert hat, gibt die Sendestation den Token an die Nachbarstation weiter.

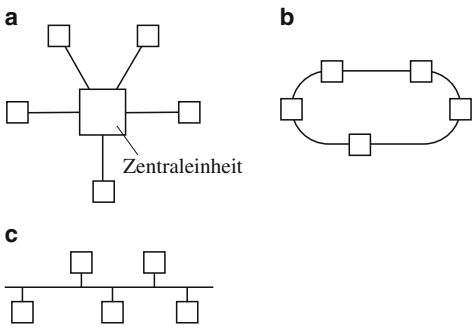


Abb. 107.1 LAN-Netzstrukturen. **a** Stern, **b** Ring, **c** Bus

Die Organisation auf der Busleitung geschieht nach der *Ethernet*-Spezifikation. Eine Station sendet nur dann, wenn auf dem Bus keine Aktionen stattfinden. Senden mehrere Stationen gleichzeitig, weil sie zum gleichen Zeitpunkt starten, wird der Sendevorgang von den beteiligten Stationen abgebrochen und nach zufallsbestimmten und damit unterschiedlichen Verzögerungszeiten erneut gestartet. Die auf dem Bus übertragenen Daten sind mit einer Zieladresse versehen und werden auch nur von der adressierten Station aufgenommen und quittiert. Dieses Organisationsform wird mit *CSMA/CD-Verfahren* bezeichnet (**c***arrier-***s***ense* **m***ultiple* **a***ccess/***c***ollision* **d***etection)*.

Tab. 107.1 fasst wesentliche Eigenschaften der drei Netzstrukturen zusammen.

107.2 Öffentliche Kommunikations- und Datensysteme

Aus der Vielzahl der Systeme zur Informations- und Datenübertragung sind hier einige Systeme herausgegriffen worden, die im allgemeinen Sprachgebrauch mit den Begriffen Telefon, Handy, Internet, VoIP, Telefax und PC (Rechner) in Verbindung gebracht werden.

107.2.1 Einführung

Die Vorteile der Digitaltechnik, hier speziell bezogen auf den Telefonverkehr, sind: höhere Sprachqualität, geringere Fehlerquote, schnellerer Verbindungsaufbau, die Möglichkeit der Verschlüsselung und effektivere Ausnutzung der vorhandenen Leitungen. Über den Telefonverkehr hinaus waren, begünstigt durch den technischen Fortschritt, neue Dienste möglich und auch von den Anwendern gewünscht. Das deshalb neu zu entwickelnde System hatte folgende Zielvorgaben:

- Bestehende Dienste sind in einem einheitlichen System zusammenzufassen.
- Neu hinzukommende Dienste müssen sich einfach integrieren lassen.
- Mehrere Dienste müssen parallel genutzt werden können.
- Die genannten Vorteile der Digitaltechnik müssen für alle Dienste nutzbar gemacht werden können.
- Die vorhandenen Telefonleitungen sollen für alle Dienste gemeinsam weiter verwendet werden können.

Vorgegeben war, dass die vorhandenen Telefonleitungen mit dem Telefonverkehr nicht ausgelastet waren (Datenrate

Tab. 107.1 Eigenschaften von Netzstrukturen

Sternstruktur	
Kennzeichen	Zentralrechner (Server) und Einzelrechner
Vorteile	Relativ hohe Übertragungsrate – keine Systemstörung, wenn Einzelrechner oder Leitung ausfällt – mit wenig Aufwand um weitere Einzelrechner erweiterbar
Nachteile	Bei Ausfall des Zentralrechners fällt gesamte Anlage aus – der Zentralrechner muss bei großen Netzen sehr leistungsfähig sein
Max. Übertragungsrate	2 . . . 100 Mbit/s, auch abhängig vom verwendeten Verbindungskabeltyp
Max. Zahl der Teilnehmer	$\gg 100$, hängt von der Leistungsfähigkeit des Zentralrechners ab
Verbindungskabel	Koaxialkabel; UTP; STP; Lichtwellenleiter[a]
Steuerung	Zentralrechner

Ringstruktur	
Kennzeichen	Einzelrechner in geschlossener Ringform miteinander verbunden
Vorteile	Kein Zentralrechner erforderlich – Zugriff auf andere Einzelrechner relativ einfach
Nachteile	Störung des Betriebes bei Ausfall eines Einzelrechners oder Kabels (lässt sich mit einigem Aufwand minimieren)
Max. Übertragungsrate	ca. 16 Mbit/s (theoretisch), 1 Mbit/s praktisch (kritisch: Übergang Ring auf Rechner)
Max. Zahl der Teilnehmer	ca. 100
Verbindungskabel	Koaxialkabel; UTP; STP[a]
Steuerung	Token-Ring-Struktur

Busstruktur ohne File-Server	
Kennzeichen	Anschluss der Einzelrechner an einen Systembus, kein Zentralrechner
Vorteile	Zusätzliche Einzelrechner leicht einfügbar
Nachteile	Zunehmende Zahl von Einzelrechnern verringert die Übertragungsrate – Bei Kabeldefekt fällt gesamte Anlage aus
Max. Übertragungsrate	10 Mbit/s (10 Base T); 100 Mbit/s (100 Base TX), 1000 Mbit/s (1000 BaseT)
Max. Zahl der Teilnehmer	ca. 100
Verbindungskabel	UTP; STP[a]; Koaxialkabel[b]
Steuerung	Einzelrechner untereinander. Ethernet-Spezifikation

Busstruktur mit File-Server	
Kennzeichen	Anschluss der Einzelrechner an einen Systembus, mit File-Server (Zentralrechner)
Vorteile	Zusätzliche Einzelrechner leicht einfügbar – File-Server stellt häufig benötigte Daten zur Verfügung, d. h. Anlage einfach erweiterbar mit neuen Daten
Nachteile	Zunehmende Zahl von Einzelrechnern verringert die Übertragungsrate – Bei Kabel- und Server-Defekt fällt gesamte Anlage aus
Max. Übertragungsrate	10 Mbit/s (10 Base T); 100 Mbit/s (100 Base TX), 1000 Mbit/s (1000 BaseT)
Max. Zahl der Teilnehmer	ca. 100
Verbindungskabel	UTP; STP[a]; Koaxialkabel[b]
Steuerung	File-Server und Einzelrechner. Ethernet-Spezifikation

[a] UTP: Verdrillte Doppelader, nicht abgeschirmt, bis 100 Mbit/s (Kategorie 5; evtl. Kategorie 6, 7); STP: Verdrillte Doppelader, abgeschirmt, bis 100 Mbit/s (Kategorie 5).
[b] 50-W-Kabel UG-274 (Thinnet-Cable), doppelt abgeschirmt, und RG 58 (Cheapernet Cable), einfach abgeschirmt

144 kBit/s bei ISDN). Die erreichbaren Datenraten hängen wesentlich von der Kabellänge zwischen Vermittlungsstelle und Kundenstandort ab und liegen in der Größenordnung 1 MBit/s bei etwa 8 bis 10 km bis zu 16 MBit/s (übliche Angabe 16.000 kBit/s) bei etwa 1 km.

107.2.2 ISDN

Mit der Einführung des ISDN (*I*ntegrated *S*ervices *D*igital *N*etwork) wurde dieses neue System geschaffen. Inzwischen ist das zunächst in der Bundesrepublik verwendete nationale ISDN durch das europaweit gültige Euro-ISDN (DSS1, *D*igital *S*ubscriber *S*ignaling System 1) abgelöst worden. Wenn im Folgenden von ISDN gesprochen wird, ist Euro-ISDN gemeint.

Das ISDN ist ein digitales Netz. Es ermöglicht die Übertragung von Informationen und Daten von unterschiedlichen Anwendungen, z. B. Telefon (Festnetz), Internet, Telefax, VoIP (Voice over IP, Internettelefonie), Datex-Netz und Bild-Telefonie. Für die Verbindung vom Endteilnehmer zur Vermittlungsstelle (häufig Telekom) dient die vorhandene 2-Draht-Leitung. Da unterschiedliche Dienste über die gleiche Leitung übertragen werden, müssen die beim Kunden angeschlossenen Geräte erkennen können, ob ein eingehender Ruf für sie bestimmt ist, sie ihn also annehmen oder aber ablehnen sollen. Man hat dazu

Tab. 107.2 ISDN-Anschlussarten, weitere Erläuterungen siehe Text

Basisanschluss		Primärmultiplex-Anschluss (PM × As)
Mehrgeräte-Anschluss	Anlagen-Anschluss	Sonderform des Anlagen-Anschlusses
8 Endgeräte anschließbar	Zahl der Endgeräte gleich Zahl der Durchwahlnummern	Für mittlere bis große TK- oder DV-Anlagen
Telefonanlage (TK) anschließbar	Erfordert Telefonanlage (TK)	
3 (max. 10) Rufnummern (*M*ultiple *S*ubscriber *N*umber, MSN)	Eine Rufnummer mit angehängter Durchwahlnummer	
2 B-Kanäle (Nutzkanäle)	2 B-Kanäle (Nutzkanäle)	30 B-Kanäle
1 D-Kanal (Steuerkanal)	1 D-Kanal (Steuerkanal)	1 D-Kanal

die *Dienstekennung* eingeführt, eine Zahl, die den Dienst kennzeichnet, u. a. Sprache, Fax Gruppe 2 oder 3, Fax Gruppe 4, E-Mail oder Daten.

Insgesamt bietet ISDN folgende Dienste:

- Telefonie
- Telefax (Gruppe 4)
- Datenübertragung (64 kBit/s pro B-Kanal)
- Teletex
- Bildübertragung

Die folgende Gliederung orientiert sich an den Begriffen und Anwendungen in der Praxis.

107.2.2.1 Telefon, Telefax, Internet ($<$ 64 kBit/s)

In diesem Kapitel werden diejenigen Dienste genannt, die im Frequenzbereich bis 144 kHz (theoretisch) bzw. 120 kHz (praktisch) übertragen werden.

Die von der Vermittlungsstelle kommende 2adrige Leitung wird beim Kunden mit dem NT (Netzabschluss, *N*etwork *T*erminator) abgeschlossen (mit vorhandenem Internetanschluss über DSL siehe weiter unten). Je nach Anforderung des Kunden gibt es mehrere Ausführungsformen des Telefonanschlusses nach Tab. 107.2.

Ein ISDN-Gerät muss für den Mehrgeräte- oder Anlagen-Anschluss geeignet sein. Zwar wird für beide das gleiche Protokoll verwendet, aber es sind evtl. unterschiedliche Funktionen zu erfüllen. Es gibt Geräte, die die Art des Anschlusses automatisch erkennen. Wird ein Basisanschluss verwendet, wird der zugeordnete Netzabschluss mit NTBA bezeichnet. Bei einem Anlagenanschluss muss der Kunde eine Telefonanlage (TK *Tele*kommunikationsanlage) installieren, weil der Provider beim Erkennen der Rufnummer die von den Anrufenden angehängte Durchwahlnummer zum Endkunden durchreicht. Die Telefonanlage muss diese Nummern dann den vorgesehenen Apparaten zuordnen. Da bei einem Basisanschluss zwei B-Kanäle zur Verfügung stehen, können gleichzeitig zwei voneinander unabhängige Aktionen durchgeführt werden, z. B. zwei Telefonate, ein Telefonat und ein Faxversand oder ein Telefonat und eine Internetverbindung. Im Folgenden wird ein Basisanschluss-Mehrgeräteanschluss mit NTBA angenommen.

POTS

Mit POTS (*P*lain *O*ld *T*elephone *S*ervice) wird der Telefon-Dienst bezeichnet, der Sprache analog überträgt. Damit hat der Begriff hier eigentlich keine Bedeutung, er wird aber häufig zu Vergleichszwecken herangezogen.

Teilnehmeranschlussleitung

Die *T*eilnehmer*an*schluss*l*eitung (TAL) ist die Verbindungsleitung, die die Vermittlungsstelle mit dem Endkunden verbindet. Sie ist als 2adrige Kupferleitung ausgeführt und wird auch mit „Letzte Meile“ bezeichnet. Beim Endkunden wird sie mit dem NTBA (s. u.) abgeschlossen. Sie hat die Schnittstellenbezeichnung U_{k0} (k für Kupfer, 0 für Basisanschluss). Die Länge kann maximal etwa 8 km betragen, ohne dass Zwischenverstärker erforderlich sind. Für die auf der Leitung übertragenen Signale wird der 4B3T-Code (auch mit MMS43 bezeichnet) verwendet. Das codierte Signal ist ternär, hat also 3 Zustände: $+2$ V, -2 V und 0. Es werden jeweils 4 Bits des Digitalsignals in 3 Schritte des ternären Signals umgesetzt. Damit stehen $2^4 = 16$ Digitalwerten $3^3 = 27$ Ternärwerte gegenüber. Durch geschickte Zuordnung lässt sich eine Datenreduktion auf maximal 120 kBit/s und eine Gleichstromfreiheit der übertragenen Signalfolge erreichen. Letzteres ist erforderlich, weil im NTBA Transformatoren eingesetzt werden. Siehe dazu NTBA.

NTBA

Der NTBA wandelt die U_{k0}-Schnittstelle um in einen 4adrigen S_0-Bus. Damit benötigt der NTBA für seine aktiven elektronischen Komponenten eine eigene externe Stromversorgung, in der Regel aus dem 230-V-Netz. Er kann und muss allerdings auf diese externe Versorgung verzichten können für den Fall, dass das 230-V-Netz ausfällt und trotzdem noch der Telefonverkehr für z. B. einen Notruf funktionsfähig gehalten werden muss. Er ist demnach so ausgelegt, dass er *ein* Telefon, das keine eigene externe Versorgung hat (Netzteil, Steckernetzteil) oder eine Telefonanlage (nur Steuerung des Einganges, nicht der gesamten Anlage) versorgen kann. Die Leistung dazu bekommt er über die Spannung auf der U_{k0}-Schnittstelle (etwa 40 V, maximal etwa 0,4 W) und einen eingebauten Spannungswandler. Mit angeschlossenem 230-V-Netz am NTBA kann er 4 Telefone (Endgeräte), die über kein eigenes externes Netzteil (Steckernetzteil) verfügen, mit Strom versorgen. Bei mehr als 4 Geräten am S_0-Bus müssen diese ein eigenes Netzteil haben. Für eine evtl. angeschlossene Telefonanlage ist wichtig, dass sie bei Ausfall des 230-V-Netzes *ein* Telefon aus der Anlage herauslöst und dieses einzeln und direkt an den S_0-Bus anschaltet. An diesen S_0-Bus können maximal 12 parallel geschaltete Steckdosen angeschlossen werden. Daran sind maximal 8 gleichzeitig aktive Endgeräte zu betreiben. Die Kabellänge kann zwischen 100 und

1000 m liegen, je nach Anschlussart und Qualität der Leitungsverlegung. Zu den verschiedenen Auslegungen des Busses gibt es Unterlagen der Netzbetreiber. Der NTBA nimmt auch die Trennung der Signale für Hin- und Rückleitung (Sprechsignal und Hörsignal) vor. Dazu ist ein Transformator mit Leitungsnachbildung erforderlich, siehe Abb. 106.4.

Digitalisierung

Jedes Endgerät muss die Signale in digital codierter Form nach den Vorgaben des S_0-Busses zur Verfügung stellen. Damit benötigt jeder Telefonapparat (Fernsprechapparat) einen Analog-Digital-Umsetzer. Es wird deshalb zwischen analogen und ISDN-Telefonen unterschieden. Analoge Telefone sind nur dann noch verwendbar, wenn ein zwischengeschalteter *Terminaladapter* die notwendige Anpassung vornimmt. Das gilt auch für Modems und Faxgeräte, die man aber häufig von analog auf digital umschalten kann. Telefonanlagen haben den Terminaladapter in der Regel eingebaut. Für den beim Telefonieren übertragenen Frequenzbereich ist die *Sprachverständlichkeit* entscheidend (300 Hz bis 3,4 kHz, bewährte Werte aus der Analog-Übertragungstechnik) und nicht die originalgetreue Übertragung der Sprache (etwa 50 Hz bis 10 kHz). Zur Anwendung kommt der Code G.711, der die höchste Frequenz mit 3,5 kHz vorgibt. Dieser Wert ist in der Vermittlungsstelle fest eingestellt, deshalb ist die Anwendung der HD-Telefonie mit der höchsten Frequenz 7 kHz im Festnetz nicht ohne weiteres möglich. Die Abtastfrequenz beträgt 8 kHz. Codiert wird zunächst mit 12 Bit, anschließend wird umcodiert, indem für die großen Lautstärken (große Signalamplituden) weniger Bit verwendet werden als für leise Töne. Das Ohr empfindet das eher als Vorteil denn als Nachteil. Die Umcodierung geschieht über eine quasilogarithmische Kennlinie (a-Law). Daraus folgt ein Digitalsignal mit 8-Bit-Codierung. Für jeden der zwei B-Kanäle ergeben sich daraus 64 kBit/s, für den D-Kanal werden 16 kBit/s benötigt. Insgesamt ergibt das eine Bitrate von 144 kBit/s am S_0-Bus.

Telefax (Fax)

Es können Faxgeräte der Gruppen 3 (digitale Daten im analogen Festnetz, max. 14,4 kBit/s) und 4 (digitale Daten im ISDN-Netz, max. 64 kBit/s) angeschlossen werden. Häufig werden Faxgeräte in Multifunktionsgeräte eingebaut, in denen die Funktionen Faxen, Drucken, Scannen und Fotokopieren vereint sind. Die Verbindung zum ISDN-Telefonnetz geschieht entweder über einen Terminaladapter oder häufig über ein Kombigerät, das die Funktionen Modem und Router für die Einwahl ins Internet über DSL gestattet, sowie eine Kleintelefonanlage mit dem Übergang analog $\leftrightarrow$ S_0-Bus enthält.

Ein Fax wird erstellt, indem eine Vorlagenseite punktweise über Fotodiodenzeilen abgetastet wird und die jeweiligen Helligkeitswerte übertragen werden. Das hat den Vorteil, dass auch beliebige Grafiken übertragen werden können. Sender und Empfänger sind nicht mehr an einen gemeinsamen Zeichensatz gebunden. Bei „feiner“ Auflösung ergeben sich 7,7 Zeilen pro mm („mittlere“ Auflösung 3,85 Zeilen pro mm, Grad der Auflösung am Fax-Gerät einstellbar) und pro Zeile 1728 Punkte. Das ergibt ca. 3,9 Millionen Bildpunkte und erfordert bei einer Übertragungsrate von 10.000 Bit/s etwa 400 s Übertragungszeit. Um diese Zeit zu verringern, wählt man folgende Codierung: Haben z. B. 1728 aufeinanderfolgende Bildpunkte den Helligkeitswert „weiß“, was bei Schriftübertragung zwischen den einzelnen Zeilen auftritt, werden nur diese Zahl und der Wert „weiß“, beide entsprechend kodiert, übertragen. Damit kann eine Seite im Mittel in einer Minute gesendet werden. Die hier genannten Zahlenwerte sind Beispiele, weil bei den Faxgeräten mehrere unterschiedliche Abtastparameter möglich sind.

Teletex, Telex

Mit diesem Dienst verbindet sich der Begriff *Fernschreiber*. Der Teletex-Dienst ist eine Weiterentwicklung des Telex-Dienstes. Im internationalen Verkehr ist der Telexdienst (noch) weit verbreitet, sodass auch in der Bundesrepublik die Teletex-Geräte telex-fähig sein sollten. Die wesentlichen Unterschiede beider Dienste sind:

Telex-Format	Teletex-Format
5 Bit pro Zeichen	8 Bit pro Zeichen
Übertragung mit 50 bit/s	Übertragung mit 2400 bit/s
Internationales Telegraphenalphabet Nr. 2, ca. 64 Zeichen	ca. 310 Zeichen

Im internationalen Verkehr geschieht die Übertragung im Telex-Format. Dieses kann national in das Teletex-Signal umgewandelt werden. Als Endgeräte sind neben dem Fernschreiber auch Schreibmaschinen und Rechner mit Drucker- und Bildschirmausgabe möglich. Im nationalen Verkehr hat dieser Dienst nur noch eine untergeordnete Bedeutung.

Leistungsmerkmale

Der Einsatz der Digitaltechnik ermöglicht die Bereitstellung nützlicher Funktionen, mit *Leistungsmerkmalen* bezeichnet. Sie sind teilweise providerabhängig und wechseln zeitweise. Deshalb sind hier nur einige angeführt, bei Bedarf ist der aktuelle Stand zu erfragen.

- Halten einer Verbindung: Während eines Telefongespräches ein zweites Gespräch annehmen und den ersten Anrufer in eine Warteschleife legen, ohne die Verbindung zu unterbrechen.
- Umstecken am Bus: Ein Telefon kann aus einer Telefonbuchse ausgesteckt werden (Raum 1) und innerhalb von 3 min an einem anderen Ort in eine andere Telefonbuchse (Raum 2) wieder eingesteckt werden, ohne dass das Gespräch unterbrochen wird.
- Übermittlung der Rufnummer.
- Anrufweiterschaltung: Der Anrufer kann vom Angerufenen mit jedem Telefonanschluss weltweit weiterverbunden werden, auch mit Mobiltelefonen des D- und E-Netzes.
- Anklopfen: Während eines Gespräches wird durch Töne angezeigt, dass ein weiterer Anruf ansteht.
- Dreierkonferenz: Drei Personen mit unterschiedlichen Rufnummern (Aufenthaltsorten) können gemeinsam miteinander sprechen und sich hören.

Internetzugang

Die Verbindung vom Rechner zum Internet erfolgt über ein ISDN-Modem oder eine ISDN-Schnittstellenkarte im Rechner (CAPI-Schnittstelle). Beide nehmen die entsprechende Umcodierung mit dem Protokoll des S_0-Busses vor. Die Daten werden über einen der B-Kanäle mit maximal 64 kBit/s übertragen. Dabei besteht die Möglichkeit, ein Telefonat und die Internetverbindung gleichzeitig auszuführen. Die Datenrate ist auf maximal 64 kBit/s begrenzt. Durch geeignete Software können beide B-Kanäle zusammengefasst werden, sodass sich eine Datenrate von 128 kBit/s ergibt.

107.2.2.2 Internet über DSL (> 64 kBit/s), VoIP, IPTV

Der auf den vorhandenen 2adrigen Leitungen noch verfügbare Frequenzbereich (siehe Einführung) wird für DSL (*D*igital *S*ubscriber *L*ine) genutzt. Unter DSL werden unterschiedliche Trägersysteme verstanden, die gemäß der Umsetzung geeigneter Protokollvorgaben moduliert werden. Für die Organisation der Übertragung wird das Protokoll IP (*I*nternet *P*rotocol) in der Version 4 verwendet, dass durch die Version 6 abgelöst wird, weil in absehbarer Zeit nicht mehr genügend Geräteadressen zur Verfügung stehen (siehe auch Teil VII „Datentechnik“, Kap. 55 „Datenkommunikation“). Unter Berücksichtigung der Anwenderpraxis ist das asymmetrische Verfahren ADSL eingerichtet worden, bei dem für den Download (Herunterladen) wesentlich mehr Frequenzbereich zur Verfügung gestellt wird als für den Upload (Heraufladen). Dazu ein Beispiel für die Verwendung des Frequenzbereiches 0 bis 1,1 MHz:

- Bis 120 kHz verwendet für ISDN-Telefonie,
- 138 bis 276 kHz verwendet für Upload,
- 276 kHz bis 1,1 MHz verwendet für Download.

Als obere Frequenzgrenze werden von den Providern unterschiedliche Werte abgeboten, abhängig von der Leitungslänge zwischen dem Endkunden und der Vermittlungsstelle und dem Übertragungsverfahren. Begonnen wurde mit dem Verfahren ADSL (8 MBit/s Download, 640 kBit/s Upload), das zunächst von ADSL2 abgelöst wurde (12 MBit/s Download, 1,2 MBit/s Upload) und dann durch ADSL2+ ersetzt worden ist (25 MBit/s Download, 3,5 MBit/s Upload). Moduliert wird mit dem DMT-Verfahren (*D*iscrete *M*ulti*t*on). Bei ADSL z. B. wird der Frequenzbereich bis 1,1 MHz mit 256 Trägern mit dem Abstand 4,3125 KHz belegt. Die Träger 32 bis 64 (ca. 138 bis 276 kHz) werden dem Uplink zugeordnet, die Träger 65 bis 256 dem Downlink (ca. 276 bis 1104 kHz). Mit zunehmender Frequenz werden die Übertragungseigenschaften der Leitung schlechter. Es ist vorgesehen, dass man den einzelnen Trägern wahlweise zwischen 2 und 15 Bits aufmodulieren kann. Damit erhält man eine angepasste Übertragung dadurch, dass den Trägern mit steigender Frequenz eine abnehmende Bit-Zahl aufmoduliert wird. Ziel ist VDSL1 mit etwa 50 MBit/s für den Download und etwa 2,5 MBit/s für den Upload, VDSL2 soll dann etwa 100 MBit/s bereitstellen. Hier wird überlegt, ob unterschiedliche Datenraten für Up- und Downlink noch sinnvoll sind. Ein Problem dabei ist die geringe Reichweite von unter einem Kilometer. Es haben sich einige Übertragungsraten mit den dazu maximal erreichbaren Entfernungen als Standardwerte herausgebildet: 1 MBit/s bei maximal 8 bis 10 km, 6 MBit/s bei maximal 4 bis 6 km und 16 MBit/s bei maximal 1 bis 2 km. Es hat sich eingebürgert, von 1000, 2000, 6000 bzw. 16.000 kBit/s zu sprechen (auch „1000er Anschluss„, “2000er Anschluss„, “6000er Anschluss„ bzw. “16.000er Anschluss„). Höhere Werte, 25 bis 50 MBit/s, sind im Aufbau.

Die Trennung der Frequenzbereiche von Telefon und ADSL wird erreicht durch eine passive Filterkombination, den sog. Splitter (oder auch BBAE), mit einem Tiefpass von 0 bis 120 kHz und einem Hochpass ab 138 kHz, so festgelegt in „Annex B“. („Annex A“ nur im Ausland, Grenzfrequenz für beide Filter 25 kHz, für analoge Telefonübertragung). Das auf der Teilnehmeranschlussleitung beim Endkunden ankommende Signal wird zunächst auf den Eingang des Splitters gegeben. Er hat 2 Ausgänge. An den einen wird der NTBA angeschlossen, an den anderen z. B. ein Router, der wiederum die Verbindung zum Internet herstellt. Der zur Verfügung gestellte Frequenzbereich ab etwa 138 kHz kann vielfach genutzt werden, hier werden 3 häufige Anwendungen vorgestellt.

Begriffe

Modem: Ein Modem wandelt analoge Signale (Sprache vom Telefon) in digitale Signale, die dann zusammen mit Steuersignalen moduliert und übertragen werden. Der Empfänger demoduliert die empfangenen Signale und wandelt sie in analoge zurück, daher der Name **MO**dulation-**DEM**odulation. Für den Internetverkehr werden höhere Übertragungsraten zur Verfügung gestellt, bis zu 56.000 Bit/s.

Router: Er dient dazu, die Verbindung zwischen 2 unterschiedlichen Netzwerken mit unterschiedlichen Protokollen und Architekturen herzustellen. In der Praxis wird er häufig eingesetzt, um ein lokales Netzwerk, ein LAN (**L**ocal **A**rea **N**etwork) mit dem Internet (WAN, **W**ide **A**rea **N**etwork) zu verbinden.

*LAN (**L**ocal **A**rea **N**etwork)*: Ein Netzwerk, dass in einem in der Regel auf die Grundstücksgrenzen beschränkten Gebiet elektronische Geräte miteinander zur gemeinsamen Nutzung (z. B. Drucker, Massenspeicher, Scanner, ...), zum Informationsaustausch (Rechner, ...) oder zur Wartung und Programmierung (Messgeräte, Anlagenprogrammierung, ...) verbindet. Jeder kann mit jedem verbunden werden und kommunizieren, wobei Zugriffsrechte vergeben werden können. Neben Firmen nutzen auch zunehmend Privathaushalte diese Netzwerke, sofern mehr als ein Rechner genutzt wird. Rechner und Geräte haben in der Regel bereits eine LAN-Schnittstelle eingebaut. Preislich ist der Aufbau dieser Netzwerke erschwinglich.

WLAN: Siehe Abschn. 107.3 „Öffentliche Funknetze“.

*WAN (**W**ide **A**rea **N**etwork)*: Weitbereichsnetzwerk. Ein Beispiel für ein WAN ist das Internet.

Switch: Er ist der „Sternpunkt“, mit dem die einzelnen Rechner und Geräte zum LAN zusammengeschaltet werden. Er verbindet gemäß Anforderung die in Frage kommenden Rechner und Geräte miteinander. Es sind *mehrere* Verbindungen *gleichzeitig* möglich. Die früher eingesetzten *Hubs* ermöglichten nur *eine* Verbindung *gleichzeitig*. Alle Geräte, die zum LAN zusammengeschaltet werden sollen, werden am Switch in eine

der vorhandenen Anschlussbuchsen eingesteckt, ein beliebiger Anschluss wird bei Bedarf mit dem Router verbunden. Damit können alle Rechner gleichzeitig und jeder völlig selbstständig mit dem Internet verbunden werden. Die Zuordnung der Pakete für die einzelnen LAN-Teilnehmer übernimmt der Router durch entsprechende Adressierung.

Internetzugang

Für den Internetzugang muss zunächst eine persönliche eindeutige Zugangskennung (Ziffern, Zeichen, Sonderzeichen) mit Passwort für den sicheren Zugang vom Provider bereitgestellt werden. Im Rechner wird ein Internetzugangsprogramm gestartet, dass die Signale aufbereitet und dem Router zur Codierung übergibt. Diesen Router hat der Kunde zwischen Rechner und Splitter bereitzustellen, entweder als Schnittstellenkarte im Rechner (weniger gebräuchlich) oder als getrenntes Gerät. Jeder Rechner, jeder Router, der sich im Internet anmeldet, hat eine eindeutige einmalige Adresse, die IP-Adresse, unter der er zweifelsfrei zu identifizieren ist. So können ihm auch die angeforderten Informationen, versehen mit seiner Adresse, zugesandt werden. Das Internet hat sich zu einem weltweiten Verbund entwickelt, über den u. a. wesentliche Teile des Informationsaustausches und des geschäftlichen Briefverkehrs in Form von E-Mails (elektronischer Post, bei Bedarf mit eingescannten Vorlagen als Anhang), die Informationsbeschaffung (Produktbeschreibungen, Datenblätter, Nachschlagewerke, ...) und ein Teil des Zahlungsverkehrs (Online-Banking) stattfinden.

Triple Play

Die oben angeführten Datenraten von bis zu 16 bzw. 25 MBit/s benötigt der normale Haushalt nicht für den Internetverkehr, weil die Internetseiten-Betreiber die dazu erforderlichen Serverkapazitäten nicht bereitstellen (können). Die Idee ist vielmehr, den Kunden über das Internet hinaus weitere Anwendungen anzubieten, und zwar über *eine* Leitung in gebündelter Form die folgenden drei Dienste: Telefon in Form von VoIP, Internet und Fernsehen. Es bleibt festzuhalten, dass die Verbindung vom Kunden zur Vermittlungsstelle selbst dann, wenn nur das Internet genutzt wird, bereits eine „Standleitung" ist und sich damit gut für Internettelefonie (schnelle Einwahl, stete Bereitschaft sowohl für den Anrufer als auch für den Angerufenen) und Fernsehprogrammübertragung (stete Bereitschaft, Übertragung ohne Unterbrechungen) eignet. Voraussetzung ist, dass der Kunde eine sog. *Flatrate* mit dem Provider vereinbart hat, er also eine monatliche Pauschale zahlt, die den permanenten Verbleib im Internet beinhaltet und unabhängig von der übertragenen Informationsmenge ist.

VoIP

VoIP (**V**oice **o**ver **IP**, Internettelefonie) ermöglicht die Sprachübertragung (Telefonieren) über das Internet. Es wird das paketorientierte IP-Protokoll verwendet. Ein weiteres Protokoll ist H.323. Es ist übergreifend für die Videokommunikation über TCP/IP geschaffen. Für die Sprachübertragung sind Einzelheiten in H.245 und H.450 geregelt. Als Signalisierungsprotokoll wird SIP verwendet.

Die Telefonapparate müssen VoIP-geeignet sein, denn sie haben bei der Sprachübertragung und der Signalisierung die entsprechenden Protokolle umzusetzen. Auf dem Markt werden aber Geräte angeboten, die die Signale von gewöhnlichen Analogtelefonen entsprechend umsetzen. Für VoIP werden keine geografisch festen Anschlüsse mit zugeordneten Telefonnummern vergeben, sondern die Identifizierung erfolgt (noch vorwiegend) über die IP-Adresse, die beim Provider hinterlegt ist. Anhand der Telefonnummer wird entschieden, ob es sich um ein Gespräch beim gleichen Provider handelt (meist gebührenfrei) oder zu einem anderen Provider oder ins Festnetz oder ins Funknetz führt. (die letzten 3 Möglichkeiten sind u. U. gebührenpflichtig). Eine Voraussetzung für einen sinnvollen Einsatz ist eine dauernde Verbindung ins Internet. Der Vorteil ist, dass jeder, der eine solche Verbindung hat, weltweit jederzeit über Telefon erreicht werden kann, allerdings nicht in jedem Fall gebührenfrei. Probleme treten auf durch die paketorientierte Übermittlung, bei der sowohl prinzipbedingte konstante Laufzeiten (Analog-Digital-Analog-Umsetzung, Protokollumsetzungen, Wegsuche) als auch schwankende Laufzeiten, sog. Jitter, durch sich ändernde Übertragungswege auftreten. Laufzeiten oberhalb von etwa 300 ms werden als unangenehm empfunden. Für eine VoIP-Verbindung werden etwa 100 kBit/s benötigt.

Internetfernsehen IPTV

Beim Digitalfernsehen ist die Übertragung von etwa 4 MBit/s für ein einwandfreies Bild einschließlich dazugehörendem Ton ausreichend (siehe Kap. 99 „Ton- und Bildübertragung", Abschn. 99.2 „Digitale Farbfernsehtechnik"). Bei einer Internetverbindung ab 6 MBit/s kann deshalb *ein* Fernsehprogramm über das Internet übertragen werden (Television over IP), wobei in der Regel noch ausreichend Bandbreite für den Internetverkehr zur Verfügung steht. Der Fernsehzuschauer wählt über Internet aus einer theoretisch riesigen, praktisch je nach Anbieter auf 100 ... 300 begrenzten Anzahl von Fernsehprogrammen eines aus, dass dann über Internet einem Receiver übergeben wird, der das demodulierte Bild- und Tonsignal über einen SCART-Anschluss in den Fernseher einspeist. Zusätzlich können, je nach Anbieter, weitere Leistungen kostenlos oder gebührenpflichtig zur Verfügung gestellt werden (z. B. Spezialsendungen, Wiederholungen, Filme). Auch hier ist eine ununterbrochen bestehende Leitung zur Vermittlungsstelle unbedingt erforderlich. Bei Leitungen mit 16 MBit/s oder höher ist die Übertragung mehrerer Programme gleichzeitig möglich, sofern der Provider diese Möglichkeit vorsieht.

107.3 Öffentliche Funknetze

Bei den Funknetzen wird die Kommunikation bzw. der Datenaustausch beim Endkunden über Funkwellen, also drahtlos, ausgeführt.

107.3.1 Mobilfunk mit GSM-Technik

Der aktuell angebotene Mobilfunk läuft im allgemeinen Sprachgebrauch auch unter dem Begriff *Handy*. Er gehört technisch

Tab. 107.3 Frequenzbereiche des D- und E-Netzes

System	Frequenz	Verwendung im
GSM-900 **G**lobal **S**ystem for **M**obile Communications	890 … 915 MHz (Uplink) 935 … 960 MHz (Downlink)	D-Netz
DCS-1800 **D**igital **C**ommunication **S**ystem	1710 … 1785 MHz (Uplink) 1805 … 1880 MHz (Downlink)	E-Netz

Tab. 107.4 Kennwerte UMTS

Name	Modulationsverfahren	Frequenzbereich
FDD	**F**requency **D**ivision Multiplex (Frequenzmultiplexverfahren)	1,920 … 1,980 GHz (Uplink) 2,110 … 2,170 GHz (Downlink)
TDD	**T**ime **D**ivision Multiplex (Zeitmultiplexverfahren	1,900 … 1,920 GHz 2,010 … 2,025 GHz

Verwendung des IP-Protokolles; paketvermittelte Übertragung

zur Generation 2G und verwendet für die Übertragung die GSM-Technik (**G**lobal **S**ystem for **M**obile Communications) mit den Eigenschaften: digital, leitungsvermittelt, 9,6 kBit/s, Übertragung in TDMA (**T**ime **D**ivision **M**ultiple **A**ccess). In der Bundesrepublik werden das D- und das E-Netz verwendet. Die Frequenzbereiche zeigt Tab. 107.3. Mit Uplink ist die Informationsflussrichtung von A nach B gekennzeichnet, mit Downlink die Gegenrichtung. Da die einzelnen Provider den Mobilfunk sowohl im D- als auch im E-Netz betreiben, wurden die *Dualband-Handys* entwickelt, die für beide Netze geeignet sind.

Das Handy des Teilnehmers A nimmt Kontakt mit einem nahegelegenen Funkturm auf, der die Verbindung zum Teilnehmer B herstellt und damit den Informationsaustausch einleitet. Teilnehmer B kann ein Handy-Teilnehmer sein oder er kann einen VoIP- oder Festnetz-Anschluss besitzen. Es bleibt festzuhalten, dass das Mobilfunknetz mit seiner vergleichsweise niedrigen Datenrate (9,6 kBit/s. s. o.) nur beschränkt für den Datenaustausch geeignet ist.

Hier folgen häufig verwendete Begriffserklärungen.

Basisstation: Die Basisstation besteht aus einer Sender-Empfänger-Kombination, die die Verbindung zum Handy drahtlos herstellt. Die Basisstation erhält und versendet ihr Signal in der Regel über Kabel, es sind aber auch Richtfunkstrecken möglich.

Funkzelle: Die Funkzelle gibt geografisch das Gebiet an, in dem der Empfänger (Handy) die Signale der Basisstation ausreichend gut empfangen kann. Ihre Größe hängt ab von der Sendeleistung und den vorgegebenen Hindernissen (Berge, Bäume, hohe Gebäude). Die Größe liegt zwischen einigen m^2 und maximal etwa 4 km^2. Eine Funkzelle kann bis maximal etwa 900 Gespräche gleichzeitig durchführen. Soll ein Empfang über den Bereich einer Funkzelle hinaus ohne Unterbrechung möglich sein, müssen sich die Funkzellen an den Grenzen überlappen.

Netzabdeckung: Sie gibt an, wie viel Prozent der Bevölkerung den entsprechenden Funkdienst nutzen können.

Roaming: Beim Roaming ist es möglich, das eigene Land (z. B. Bundesrepublik) zu verlassen und trotzdem problemlos mit einem Teilnehmer im eigenen Land weiter zu kommunizieren (telefonieren). Das setzt voraus, dass das fremde Land die gleichen Systeme einsetzt wie das eigene Land und dass ein Abkommen zwischen beiden Ländern über die Handhabung (z. B. Gebührenabrechnung) besteht.

Handover: Dies wird angewendet, wenn ein Empfänger (Handy) den Bereich einer Basisstation verlässt und in den Bereich einer benachbarten Basisstation wechselt (Wanderung, Autofahrt). Das Gerät des Empfängers erkennt die abnehmende Empfangsfeldstärke und sucht laufend nach Basisstationen mit größeren Feldstärken. Bei erfolgreicher Suche schaltet der Empfänger auf die Basisstation mit der größeren oder, sofern vorhanden, der größten Feldstärke um. Bei der Sprachübertragung ist dieses Verfahren problemlos anwendbar, weil kurzzeitige Aussetzer oder Fehler, die bei dem Übergang nicht zu vermeiden sind, meist in Kauf genommen werden können. Problematisch wird es bei der Datenübertragung, wenn diese u. U. mehr oder weniger kurzfristig unterbrochen wird. Es gibt Provider, die den Übergang unter bestimmten Bedingungen unterbrechungsfrei vornehmen können.

SIM-Karte: Diese Karte muss in den Sockel des Handys eingesteckt werden. Sie enthält einen Chip, auf dem eine weltweit einmalige Kennung, eine PIN (**P**ersonal **I**dentity **N**umber) und vom Handy-Besitzer eingespeicherte Daten (z. B. Notizbuch, Telefonbuch) eingetragen sind.

*SMS (**S**hort **M**essage **S**ervice)*: Eine SMS ist eine Kurznachricht, die vorwiegend von Handy zu Handy übertragen wird. Sie kann aus maximal 160 Zeichen (Buchstaben, Zahlen, Sonderzeichen) oder maximal 140 Byte bestehen. Aber auch dazu geeignete Festnetz-Telefone können am SMS-Austausch teilnehmen.

107.3.2 Mobilfunk mit UMTS-Technik

Die Übertragungsrate der GSM-Technik mit 9,6 kBit/s reicht für eine schnelle Datenübertragung in der Regel nicht aus. Deshalb wurde mit UMTS eine neue Generation des Funknetzes geschaffen, das eine Übertragungsrate von theoretisch etwa 2 MBit/s vorsieht. In der Praxis wird aber mit einem wesentlich kleineren Wert gerechnet, der bei etwa (200 … 400) kBit/s liegt. Da dieses Netz im Aufbau ist, lassen sich gesicherte Werte dafür noch nicht angeben. Einige Kennwerte sind in Tab. 107.4 aufgelistet. Für den Vollduplexverkehr werden getrennte Kanäle für beide Richtungen verwendet, hier mit Uplink und Downlink gekennzeichnet.

Mit UMTS sollen zunächst folgende Kommunikationsdienste zur Verfügung gestellt werden, weitere können folgen:

- Internet
- Bild- und Videotelefonie

- E-Mail-Verkehr
- Datenübertragung
- Rundfunk- und Fernsehempfang
- Navigation

UMTS bietet die Möglichkeit, mehrere unterschiedliche Dienste gleichzeitig zu nutzen, also z. B. zu telefonieren und parallel dazu Daten zu versenden.

Zunächst werden GSM und UMTS parallel betrieben, um den Übergang kontinuierlich zu gestalten (eingeschränkte Kompatibilität) und durch den Aufbau bedingte Überlastungen der UMTS-Zellen durch GSM abfangen zu können. Langfristig aber soll GSM abgelöst werden.

107.4 Private Funknetze

Diese Funknetze werden so eingesetzt, dass sie die Grundstücksgrenzen in der Regel nicht überschreiten.

107.4.1 Schnurlostelefone

Die Schnurlostelefone stellen eine Verbindung über Funk zu einer Basisstation her. Die Basisstation kann mehrere Telefone miteinander verbinden, sie erfüllt dann die Funktion einer Telefonanlage. Die Basisstation kann aber auch zusätzlich mit einem Festnetzanschluss (analog oder ISDN) oder dem VoIP-Anschluss, meist über Kabel, verbunden werden. Die Reichweite beträgt im Freien bis zu 300 m, in Gebäuden ist sie wesentlich geringer und liegt bei 10 bis 30 m. Es können mehrere Basisstationen betrieben werden, die sich in überlappenden Funkzellen organisieren lassen, entsprechend dem Zellenaufbau bei Handys. Es ist Handover und damit die Abdeckung größerer Areale möglich. Die technische Grundlage ist der DECT-Standard (**D**igital **E**nhanced **C**ordless **T**elecommunications). Der Hochfrequenz-Übertragungsbereich liegt bei 1880 bis 1900 MHz. Er wird in 12 Bereiche eingeteilt, je ein Bereich an der oberen und unteren Frequenzgrenze wird als Schutzband ausgespart. Damit bleiben 10 Bereiche mit je einem Träger. Als Modulationsverfahren für diese Träger wird GFSK (**G**aussian **F**requency **S**hift **K**eying) eingesetzt. Jeder Träger hat 24 Zeitfenster pro Block (Rahmen), je 12 für beide Richtungen. Das Nutzsignal wird mit ADPCM (**A**daptive **D**ifferential **P**ulse **C**ode **M**odulation) versehen und aufmoduliert. Damit ergibt sich eine Rate von 24 kBit/s pro Kanal. Für die Datenübertragung lassen sich 23 Kanäle softwaremäßig bündeln (ein Kanal muss für die entgegengesetzte Richtung bereitstehen), und es stehen 552 kBit/s zur Verfügung. Die in die Schnurlostelefone eingebauten technischen Möglichkeiten sind denen von Festnetztelefonen und Handys gleichwertig. Hier seien einige genannt:

- Gebührenfreie interne Gespräche
- Gleichzeitiger Betrieb mehrerer Mobiltelefone
- Gute Tonqualität
- Eigenschaften wie leistungsfähige Festnetztelefone
- Geringer Stromverbrauch
- Telefone unterschiedlicher Hersteller verwendbar
- Minimale Verzögerung
- Verschlüsselung der übertragenen Informationen

Ein Kriterium bei den Schnurlostelefonen ist die strahlungsreduzierte bzw. strahlungsarme Auslegung. Die Sendeleistung ist im Laufe der Jahre stetig reduziert worden. Es gibt Anlagen, bei denen sich die Sendeleistung automatisch nach der Entfernung zur Basisstation einstellt. Außerdem wird die Sendeleistung auf ein Minimum reduziert, wenn das Telefon nicht benutzt wird.

107.4.2 Bluetooth

Mit Bluetooth wird eine Funkverbindung über kurze Entfernungen bezeichnet. Die Reichweite hängt ab von der Sendeleistung und wird in 3 Klassen eingeteilt:

Klasse	Maximale Senderleistung	Reichweite bei optischer Sicht etwa
1	100 mW	100 m
2	2,5 mW	20 m
3	1,0 mW	10 m

Diese Funkverbindung wird häufig im häuslichen Bereich eingesetzt, wobei Hindernisse (Mauern, Einrichtungsgegenstände) die Reichweite verringern. Damit können Kabelverlegungen vereinfacht werden oder ganz entfallen. Hier einige Beispiele für den Einsatz:

- Übertragung von gespeicherten Bildern der Digitalkamera auf die Festplatte eines Rechners
- Verbindung von kabelloser Maus und Tastatur mit dem Rechner
- Verbindung eines CD- oder DVD-Laufwerkes mit dem Rechner
- Übertragung der Tonsignale vom Mischpult/Equalizer/Receiver zur Leistungs-Endstufe
- Übertragung von auf dem Handy gespeicherten Daten auf den Rechner

Gesendet wird auf einer Frequenz zwischen 2,40 und 2,48 GHz. Das ist das ISM-Band (**I**ndustrial, **S**cientific, **M**edical)), das von vielen Anwendern benutzt wird, u. a. auch von Mikrowellengeräten und vom WLAN (siehe 4.3 unten). Um eine ausreichende Störsicherheit zu bekommen, springt die Frequenz 1600mal pro Sekunde. Es können Daten mit einer Rate von etwa 720 kBit/s (Download) und etwa 55 kBit/s (Upload) übertragen werden bzw. etwa 430 kBit/s für beide Richtungen bei symmetrischem Betrieb. Durch die Festlegung von Profilen lassen sich die Verbindungen unabhängig von den Herstellern der Geräte betreiben. So gibt z. B. das Profil BIP vor, dass es sich um eine Bildübertragung handelt. Besondere Bedeutung kommt dem Profil HDP zu, das die Übertragung medizinischer Daten festlegt. Bluetooth ist für die Medizintechnik einsetzbar. Die übertragenen Daten können mit einem 128-Bit-Wort verschlüsselt werden. Der Stromverbrauch ist sehr gering, was für den Einsatz in Kameras und Handys von Bedeutung ist.

107.4.3 WLAN

Das WLAN (**W**ireless **L**ocal **A**rea **N**etwork) ist ein drahtloses LAN (siehe Abschn. 107.2.2.2). Ihm liegen die Standards IEEE 802.11a, IEEE 802.11b, IEEE 802.11g und IEEE 802.11n zugrunde, die für verschiedene Frequenzbänder vorgesehen sind. In der Praxis lassen sich die folgenden Werte erreichen:

Standard IEEE	Frequenzband	Brutto-Übertragungsrate max.	Netto-Übertragungsrate (für Nutzdaten) max.
802-11a	5,15–5,725 GHz	54 MBit/s	25 MBit/s
802-11b	2,4–2,4835 GHz	11 MBit/s	5 MBit/s
802-11g	2,4–2,4835 GHz	54 MBit/s	25 MBit/s
802-11n	2,4–2,4835 GHz 5,15–5,725 GHz	300 MBit/s	100 MBit/s

Im 2,4-GHz-Band stehen 13 Kanäle im Frequenzbereich von 2,412 GHz bis 2,472 GHz zur Verfügung. Da eine WLAN-Verbindung einen Frequenzbereich von 20 MHz benötigt, können überlappungsfrei nur 3 WLAN-Verbindungen in einem gemeinsam genutzten Bereich betrieben werden, und zwar auf den Kanälen 1, 7 und 13.

Die zulässige effektive Strahlungsleistung (EIRP) beträgt 100 mW bei 2,4 GHz bzw. 200 bis 1000 mW (frequenz- und verfahrensabhängig) bei 5,4 GHz. Damit lassen sich Reichweiten von 10 bis 100 m erreichen.

Die Übertragung im 5-GHz-Band ist weniger gestört als im 2,4-GHz-Band. Viele Geräte unterstützen alle 4 Standards und suchen den jeweils optimalen Übertragungsstandard aus. Der Standard IEEE 802-11n kann nur eingesetzt werden, wenn das Gerät das Verschlüsselungsverfahren WPA2 (verwendet AES, Advanced Encryption Standard) unterstützt und dieses auch vom Anwender eingestellt ist. Das Verschlüsselungsverfahren WEP ist veraltet und unsicher, WPA ist weniger sicher als WPA2.

Das an das Netzwerk anzuschließende Gerät erhält einen Sender-Empfänger, häufig in Form eines USB-Sticks, der in einen USB-Port eingesteckt wird. Er nimmt über eine Funkverbindung Kontakt auf mit einem speziellen Sender-Empfänger, dem sog. *Access-Point*, der wiederum direkt mit dem Netzwerk verbunden ist.

Optimierte Nachrichten- und Datenübertragung

108

Teil XIII

Wilfried Plaßmann



108.1 Codierung 1310
108.2 Optimalfilter 1316
108.3 Anwendung der Korrelation bei gestörten Signalen 1317
108.4 Verschlüsselung von Daten 1318
108.5 Fensterung 1321
Literatur 1322

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_108

Bei der Nachrichten- und Datenübertragung gibt es drei wesentliche Vorgaben:

1. Die Wirtschaftlichkeit des Übertragungskanals;
2. die begrenzte Bandbreite der zur Verfügung stehenden Übertragungskanäle und
3. sehr stark gestörte Übertragungskanäle, bei denen ein Ausweichen auf weniger gestörte Kanäle nicht möglich ist (Bildübertragung aus dem Weltraum). Diesen Vorgaben sind die vom Anwender häufig gestellten Forderungen nach einer möglichst preiswerten, schnellen und fehlerfreien Übertragung anzupassen. Im folgenden werden einige Wege zur Realisierung aufgezeigt. Es wird vorausgesetzt, dass die Informationen in digital codierter Form vorliegen, was auch in der Technik angestrebt wird.

108.1 Codierung

Bei der Codierung der Nachrichten bzw. Daten unterscheidet man zwei Bereiche: Die *Quellen-* und die *Kanalcodierung*.

Quellencodierung: Die Information der Quelle enthält neben der „eigentlichen" Information noch zwei weitere Anteile:

1. Redundanz (siehe auch Abschn. 92.4.4);
2. Irrelevante Information. Sie entsteht als Nebenprodukt und hat keinerlei Verknüpfung mit der eigentlichen Information.

Aufgabe der Quellencodierung ist zunächst die Beseitigung der irrelevanten Information. Anschließend wird die vorhandene Redundanz so weit wie möglich reduziert, weil sie in der Regel zur Fehlererkennung und -korrektur nicht *optimal* geeignet ist.

Kanalcodierung: Ist der Übertragungskanal gestört, wird das von der Quelle gelieferte „quellencodierte" Signal so umcodiert und mit neuer Redundanz versehen, dass eine optimale Fehlererkennung und -korrektur möglich wird.

108.1.1 Quellencodierung

108.1.1.1 Grundlagen

Es wird vorausgesetzt, dass die irrelevante Information bereits entfernt worden ist. Außerdem wird von dem praxisnahen Fall ausgegangen, dass die zu codierenden k Zeichen (Einzelzeichen oder Zeichenfolgen) nicht alle mit gleicher Wahrscheinlichkeit auftreten und dass nicht *genau* 2^n ($n = 1, 2, 3, \ldots$) unterschiedliche Zeichen zu codieren sind, d. h. $k \neq 2^n$. Die Aufgabe der Codierung besteht darin, die mittlere Codewortlänge l_m so klein wie möglich zu machen, weil dadurch die Übertragungszeit bzw. die benötigte Kanalbandbreite gering gehalten werden kann (Abschn. 92.4.7):

Mittlere Codewortlänge

$$l_m = \sum_{i=1}^{k} l_i \cdot p(x_i) \quad \text{in bit} \tag{108.1}$$

l_i Codewortlänge (Anzahl der Binärstellen) des Zeichens bzw. der Zeichenfolge x_i in bit, $p(x_i)$ Wahrscheinlichkeit für das Auftreten des Zeichens x_i.

Die naheliegende Lösung, nämlich alle Zeichen mit der gleichen Anzahl von Stellen zu codieren ($l_1 = l_2 = \ldots = l_k$), führt zu einer mittleren Codewortlänge $l_{\text{mg}} = l_i$, d. h. die mittlere Codewortlänge ist genau so groß wie die Codewortlänge der einzelnen Zeichen. Eine Reduzierung der mittleren Codewortlänge ($l_m < l_{\text{mg}}$) lässt sich dadurch erreichen, dass

1. die Zeichen mit den größeren Wahrscheinlichkeiten kürzere Codewortlängen erhalten als die mit den geringeren (Beispiel 1), und
2. nicht *einzelne* Zeichen, sondern Zeichen*gruppen* codiert werden (Beispiel 2).

Für die Wahrscheinlichkeit des Auftretens einer Zeichen*gruppe* aus zwei Zeichen (x_1, x_2) gilt die bedingte Wahrscheinlichkeit beider Zeichen, $p(x_1, x_2)$, d. h. die Wahrscheinlichkeit für das Auftreten des zweiten Zeichens (x_2) unter der Voraussetzung, dass das erste Zeichen (x_1) aufgetreten ist. Sie wird berechnet, indem die Wahrscheinlichkeiten der beiden Zeichen miteinander multipliziert werden:

Bedingte Wahrscheinlichkeit für zwei Zeichen

$$p(x_1, x_2) = p(x_1) \cdot p(x_2) \tag{108.2}$$

Für die Codierung gelten zwei wichtige Sätze:

1. *Eindeutigkeit*: Den einzelnen Zeichen oder Zeichengruppen dürfen nur solche Codierungen zugeordnet werden, die *Endpunkte* im Codebaum sind. Andernfalls kann die empfangene Bitfolge nicht eindeutig decodiert werden (Beispiel 1).
2. *Existenz*: Ein decodierbarer Code existiert genau dann, wenn die Ungleichung von *Kraft* erfüllt ist:
3. *Ungleichung von Kraft*

$$\sum_{i=1}^{k} 2^{-l_i} \leq 1 \quad \text{für Binärcode} \tag{108.3}$$

l_i Zahlenwert der Codewortlänge (Anzahl der Binärstellen) des Zeichens bzw. der Zeichenfolge x_i.

Beide Sätze sind inhaltlich ähnlich, wobei die Ungleichung von Kraft noch nichts über die Eindeutigkeit der Decodierung aussagt (Beispiel 1).

Shannon hat eine Beziehung für die mittlere Codewortlänge angegeben:

$$\textit{Satz von Shannon} \quad H \leq l_m \leq H + 1 \text{ bit} \tag{108.4}$$

H Entropie nach (92.2) in bit, l_m mittlere Codewortlänge nach (108.1) in bit.

Tab. 108.1 Auswahl von Möglichkeiten zur Codierung von drei Zeichen (l_i: Zahlenwert)

Zeichen		Codierung 1		Codierung 2		Codierung 3		Codierung 4	
x_i	$p(x_i)$	Codewort	l_i	Codewort	l_i	Codewort	l_i	Codewort	l_i
x	0,5	00	2	0	1	0	1	0	1
y	0,3	01	2	01	2	10	2	10	2
z	0,2	10	2	10	2	11	2	110	3
Mittlere Länge		$l_m = 2$ bit		$l_m = 1{,}5$ bit		$l_m = 1{,}5$ bit		$l_m = 1{,}7$ bit	
Ungleichung von *Kraft*		$\sum_{i=1}^{k} 2^{-l_i} = \frac{3}{4} < 1$		$\sum_{i=1}^{k} 2^{-l_i} = 1$		$\sum_{i=1}^{k} 2^{-l_i} = 1$		$\sum_{i=1}^{k} 2^{-l_i} = \frac{7}{8} < 1$	
Bemerkungen		„Standardcodierung“		Nicht eindeutig decodierbar		Optimalcode nach *Huffman/Fano*		Kommacode	

Tab. 108.2 Zu Beispiel 2

Zeichengruppe	xx	xy	xz	yy	yx	yz	zz	zx	zy
Wahrscheinlichkeit[a]	0,25	0,15	0,1	0,09	0,15	0,06	0,04	0,1	0,06
Optimalcode nach *Fano*	00	01	101	1101	100	1110	11111	1100	11110
Ungleichung von *Kraft*	$\sum_{i=1}^{k} 2^{-l_i} = \frac{16+8+6+2}{32} = 1$								
Satz von *Shannon*	$H = 2{,}97$ bit $\Rightarrow 2{,}97$ bit $\leq l_m \leq 3{,}97$ bit								
Mittlere Codewortlänge	$l_m = 3{,}05$ bit								

[a] mit (108.2)

Tab. 108.3 Codierverfahren nach *Fano* (zu Beispiel 2)

Zeichengruppe	xx	xy	yx	xz	zx	yy	yz	zy	zz
Wahrscheinlichkeit, sortiert	0,25	0,15	0,15	0,1	0,1	0,09	0,06	0,06	0,04
1. Teilstrich	$\sum p = 0{,}4$		$\sum p = 0{,}6$						
1. Bit	0	0	1	1	1	1	1	1	1
2. u. 3. Teilstrich			$\sum p = 0{,}25$		$\sum p = 0{,}35$				
2. Bit	0	1	0	0	1	1	1	1	1
4. u. 5. Teilstrich					$\sum p = 0{,}19$		$\sum p = 0{,}16$		
3. Bit			0	1	0	0	1	1	1
6. u. 7. Teilstrich								$\sum p = 0{,}1$	
4. Bit					0	1	0	1	1
8. Teilstrich									
5. Bit								0	1
Codierung	00	01	100	101	1100	1101	1110	11110	11111

Beispiel 1

Es sollen die Zeichen x, y und z codiert werden, die mit den Wahrscheinlichkeiten $p(x) = 0{,}5$; $p(y) = 0{,}3$ und $p(z) = 0{,}2$ auftreten. Tab. 108.1 gibt verschiedene Möglichkeiten der Codierung wieder, wenn die Zeichen einzeln, also noch keine Zeichengruppen, codiert werden.

Codierung 1 als naheliegende Lösung ergibt eine vergleichsweise große mittlere Codewortlänge und ist deshalb weniger geeignet. Völlig ungeeignet ist Codierung 2, weil z. B. bei der Übertragung der Bitfolge 010 diese entweder als yx (01-0) oder als xz (0-10) interpretiert werden kann. Im Codebaum ist die codierte 0 kein „Endpunkt“, denn es folgt noch die Codierung 01. Codierung 3 ist eine optimierte Lösung und wird als *Optimalcode* bezeichnet. Er ist eindeutig decodierbar und hat die kleinste Codewortlänge. Die Codierung 4 ist ebenfalls eindeutig, hat aber eine größere Codewortlänge als der Optimalcode. Sie besitzt den Vorteil, dass die 0 das Ende jedes einzelnen Codewortes angibt und sofort mit der Decodierung begonnen werden kann. Dieser Code wird auch als Kommacode bezeichnet: Die Null steht für ein Trennzeichen, so wie das Komma bei Aufzählungen. ◀

Beispiel 2

Zur weiteren Verringerung der mittleren Codewortlänge werden Zeichen*gruppen* aus zwei (oder mehr) aufeinanderfolgenden Zeichen gebildet und codiert. Tab. 108.2 listet die vorkommenden Kombinationen und deren Wahrscheinlichkeiten nach (108.2) auf. Weiterhin ist der Optimalcode nach Fano gebildet und eingetragen (Tab. 108.3). Zu Kontrollzwecken sind (108.1), (108.3) und (108.4) ausgewertet und die Zahlenwerte eingefügt. Der Code ist nahezu optimal, weil l_m nur geringfügig größer ist als H. ◀

Tab. 108.4 Weitere Codiermöglichkeit nach *Fano* (zu Beispiel 2)

Zeichengruppe	*xx*	*xy*	*xz*	*yy*	*yx*	*yz*	*zz*	*zx*	*zy*
Optimalcode nach *Fano*	00	010	100	1100	011	1101	1111	101	1110
Daraus folgt: $l_m = 3{,}00\,\text{bit}$; $\sum_{i=1}^{k} 2^{-l_i} = 1$									

108.1.1.2 Optimalcodes

Optimalcodes zeichnen sich dadurch aus, dass die mittlere Codewortlänge l_m nur geringfügig über der Entropie H liegt. Dazu werden häufig auftretende Zeichen oder Zeichengruppen mit weniger Bit, selten auftretende entsprechend mit mehr Bit codiert. Die dazu notwendigen Codierungsverfahren sind von *Huffman* und *Fano* entwickelt worden. Der im folgenden dargestellte Fano-Algorithmus (siehe dazu auch Tab. 108.3) wird nach folgenden Regeln gebildet:

1. Die Zeichen oder Zeichengruppen werden nach abnehmender Wahrscheinlichkeit sortiert.
2. Die Liste wird durch einen ersten Teilstrich so in zwei Teile zerlegt, dass die Summe der Wahrscheinlichkeiten je Teil etwa 0,5 ist. Die Elemente des einen Listenteils erhalten eine 0 als erstes Bit, die anderen eine 1.
3. Die Elemente des Listenteils mit der 0 werden wiederum durch einen zweiten Teilstrich in zwei Teile geteilt, sodass beide etwa die gleiche Wahrscheinlichkeitssumme erhalten. Der eine Teil erhält als zweite Codierung eine 0, der andere eine 1. Mit dem Teil, der als erstes Bit eine 1 erhalten hat, wird genauso verfahren.
4. Das Prinzip wird solange fortgesetzt, bis nach einer Teilung nur noch Einzelelemente übrig sind.

Tab. 108.3 zeigt das Verfahren mit den Zeichengruppen und Wahrscheinlichkeiten aus Beispiel 2 und Tab. 108.2.

Man kann durchaus zu unterschiedlichen Lösungen kommen, wenn man z. B. den ersten Teilstrich so setzt, dass links davon die Zeichengruppen *xx*, *xy* und *yx* stehen.

Je nach Lage der anderen Teilstriche kann das zu einer in Tab. 108.4 dargestellten weiteren Möglichkeit der Codierung führen.

108.1.1.3 Datenreduktion

Die zur Datenreduktion verwendeten Verfahren sind unter Abschn. 96.4.4.7 und 96.4.4.8 aufgeführt. Angewendet werden sie auch beim „digitalen Fernsehen“, wobei dort der Sonderfall besteht, dass kurzzeitige geringe Signalverfälschungen in Kauf genommen werden, weil das Auge als Empfänger sie nicht oder als nur wenig störend wahrnimmt.

108.1.2 Kanalcodierung

Bei der Kanalcodierung wird dem quellencodierten und damit nahezu redundanzfreien Signal neue Redundanz hinzugefügt mit dem Ziel, Übertragungsfehler so weit wie möglich zu erkennen und zu korrigieren.

108.1.2.1 Grundlagen

Für die Kanalcodierung sind die folgenden Begriffe von Bedeutung:

- *Hamming-Distanz h*: Sie gibt an, in wieviel Stellen sich die einzelnen (gleichlangen) Codewörter *minimal* unterscheiden.
- Gewicht eines Codewortes: Anzahl der Einsen im Codewort.
- *Gleichgewichtige Codes*: Jedes Codewort enthält die gleiche Anzahl von Einsen.
- *Fehlererkennende Codes*: Es wird erkannt, dass ein übertragenes Bitmuster falsch ist, es kann aber nicht ermittelt werden, welches Bit falsch ist. Es muss deshalb eine erneute Übertragung gestartet werden.
- *Fehlerkorrigierende Codes*: Das falsch übertragene Bitmuster wird als falsch erkannt und korrigiert. Eine erneute Übertragung ist nicht erforderlich.

Die Anzahl der erkennbaren bzw. korrigierbaren Fehler ergibt sich aus der Hamming-Distanz h:

- *Zahl der erkennbaren Fehler (ohne Korrektur)*:

$$h - 1 \tag{108.5a}$$

- *Zahl der selbständig korrigierbaren Fehler*:

$$\begin{aligned} &\frac{h-1}{2} && \text{für } h = 1, 3, 5, \ldots \\ &\frac{h}{2} - 1 && \text{für } h = 2, 4, 6, \ldots \end{aligned} \tag{108.5b}$$

  Bei einem Code mit $h = 3$ kann man entweder zwei Fehler erkennen, ohne sie zu korrigieren, oder einen Fehler erkennen und korrigieren (Beispiel 3).
- *Blockcodes*: Alle Codewörter sind gleich lang.
- *Faltungscodes*: Die Kontrollbits werden nicht zu jedem einzelnen Codewort hinzugefügt, sondern während der Übertragung der Daten gebildet und in geeigneter Form eingebaut (siehe Abschn. 108.1.2.3). Die Eigenschaften dieser Codes werden in der Regel durch Simulation erfasst, weil die mathematische Darstellung umfangreich ist.
- *Systematische Codes*: Die ersten Stellen sind stets die Informationsstellen, die restlichen enthalten die Prüfbits (siehe Codierung im Beispiel 3). Diese Codes sind vom Aufbau her übersichtlich.
- *Zyklische Codes*: Die Codewörter werden nach Rechenvorschriften mit sogenannten *Generatorpolynomen* gebildet. Sie sind nicht übersichtlich, haben aber den Vorteil, dass besonders bei langen Codewörtern die Codierung und Decodierung durch die Verwendung mathematischer Algorithmen mit relativ niedrigem Aufwand möglich ist.

Beispiel 3

Übertragen werden sollen 4 Daten, die mit 00, 01, 10 und 11 codiert werden (x_1, x_2). Es soll *ein* Fehler erkannt und selbständig korrigiert werden. Nach (108.5a) und (108.5b) ist eine Hamming-Distanz $h = 3$ erforderlich. Es wird nach Tab. 108.5 mit 5 Bit codiert.

a) Es wird angenommen, dass *ein* Fehler auftritt, d. h. die fehlerhafte Codierung 0100**1** übertragen wird. Darin ist die unterstrichene Null für x_4 falsch. Die Bedingungen für die Fehlererkennung und -korrektur nach Tab. 108.5 liefern:
Erste Bedingung erfüllt $\Rightarrow$ x_1, x_3, x_5 ohne Fehler.
Zweite Bedingung nicht erfüllt $\Rightarrow$ x_2, x_4, x_5 enthalten Fehler.
Dritte Bedingung erfüllt $\Rightarrow$ x_1, x_3 ohne Fehler.
Vierte Bedingung nicht erfüllt $\Rightarrow$ x_2, x_4 enthalten Fehler.
Fünfte Bedingung erfüllt $\Rightarrow$ x_1, x_2, x_5 ohne Fehler.
Die 2. und 5. Bedingung ergeben die eindeutige Aussage, dass x_4 falsch ist, sodass die Codierung richtig lautet: 01011. Wichtig ist die Voraussetzung, dass nur *ein* Fehler auftritt, der mit $h = 3$ erkannt und korrigiert werden kann!

b) Das übertragene Codewort 00**11**0 enthält *zwei* Fehler (x_3, x_4). Die Auswertung der Bedingungen nach Tab. 108.5 ergibt:
Erste Bedingung nicht erfüllt $\Rightarrow$ x_1, x_3, x_5 enthalten Fehler.
Zweite Bedingung nicht erfüllt $\Rightarrow$ x_2, x_4, x_5 enthalten Fehler.
Dritte Bedingung nicht erfüllt $\Rightarrow$ x_1, x_3 enthalten Fehler.
Vierte Bedingung nicht erfüllt $\Rightarrow$ x_2, x_4 enthalten Fehler.
Fünfte Bedingung erfüllt $\Rightarrow$ x_1, x_2 und x_5 ohne Fehler.
Mit der 5. Bedingung, dass x_1, x_2, x_5 richtig sind, ergibt sich, dass x_3, x_4 falsch sein müssen, was in diesem Spezialfall eine Fehlerkorrektur ermöglichen könnte. Sind dagegen x_2, x_5 falsch, erhält man aus den 5 Bedingungen, dass zwei von den Bits x_1, x_2, x_5 falsch sind. Damit sind 2 Fehler erkannt, aber wegen fehlender Eindeutigkeit ist eine Korrektur nicht möglich. Gleichung (108.5a) gilt also für den allgemeinen Fall. ◀

Tab. 108.5 Codierung zu Beispiel 3. Die Prüfbits ergänzen auf eine gerade Anzahl von Einsen

x_1	x_2	x_3	x_4	x_5
0	0	0	0	0
0	1	0	1	1
1	0	1	0	1
1	1	1	1	0
Zu übertragende Daten		Prüfbit für x_1	Prüfbit für x_2	Prüfbit für x_1 und x_2

Bedingungen 1 . . . 5 für die Fehlererkennung und -korrektur:
1: x_1 falsch $\Rightarrow$ x_3 und x_5 nicht erfüllt
2: x_2 falsch $\Rightarrow$ x_4 und x_5 nicht erfüllt
3: x_3 falsch $\Rightarrow$ x_1 nicht erfüllt
4: x_4 falsch $\Rightarrow$ x_2 nicht erfüllt
5: x_5 falsch $\Rightarrow$ x_1 und x_2 nicht erfüllt

Tab. 108.6 Wahrscheinlichkeiten für das Auftreten des Zeichens y_j, wenn x_i gesendet wurde. Erläuterungen siehe Text

a)

$p(y_j \mid x_i)$	y_1	y_2	y_3
x_1	0,7[a]	0,2	0,1
x_2	0,1	0,8	0,1
x_3	0,2	0,2	0,6

b)

$p(y_j \mid x_i)$	y_1	y_2	y_3
x_1	0,7	0,2	0,1
x_2	0,2	0,2	0,6
x_3	0,4	0,2	0,4

108.1.2.2 Maximum-Likelihood-Verfahren

Die Quelle sendet Zeichen x_i ($i = 1 \ldots k$) aus ihrem Zeichenvorrat aus, die die Senke (Empfänger) einem Zeichen y_j ($j = 1 \ldots k$) ihres Zeichenvorrates zuordnet. Es wird hier angenommen, dass Quelle und Senke den gleichen Zeichenvorrat besitzen. Diese Zeichen werden bei der Übertragung über einen gestörten Kanal teilweise so verändert, dass man der Senke ein Entscheidungskriterium zur Verfügung stellen muss. Dazu wird die Wahrscheinlichkeit für das Eintreffen eines Zeichens y_j am Empfänger verwendet, wenn das Zeichen x_i gesendet wurde. Tab. 108.6 zeigt zwei Beispiele.

Der mit [a] gekennzeichnete Wert 0,7 in Tab. 108.6a ist die Wahrscheinlichkeit dafür, dass der Empfänger das Zeichen y_1 decodiert, wenn x_1 gesendet wurde. Die übrigen Tabellenwerte sind entsprechend zu interpretieren. Die Summe der Wahrscheinlichkeiten einer Zeile ergibt 1, weil angenommen wird, dass der Empfänger mit Sicherheit eines seiner drei Zeichen dem empfangenen Zeichen zuordnet. Die aufgelisteten Wahrscheinlichkeiten als wesentliche Eigenschaft des gestörten Kanals müssen bekannt sein. Dann lautet die Likelihood-Bedingung:

Likelihood-Bedingung

$$p(y_j \mid x_{i\text{-opt}}) \geq p(y_j \mid x_{i \neq i\text{-opt}}) \tag{108.6}$$

Der Empfänger empfängt das Zeichen y_i und ordnet ihm das Zeichen $x_i = x_{i\text{-opt}}$ mit der größten vorhandenen Wahrscheinlichkeit zu. In Tab. 108.6a ist die Zuordnung eindeutig, denn bei empfangenem y_2 wird das Zeichen x_2 zugeordnet, weil es mit der größten Wahrscheinlichkeit auftritt. Ebenso wird y_1 zu x_1 und y_3 zu x_3 gehörig erkannt. Die Wahrscheinlichkeiten in Tab. 108.6b lassen diese eindeutige Zuordnung nicht mehr für alle Werte zu. Wird z. B. y_2 empfangen, kann es x_1, x_2 oder x_3 zugeordnet werden, da alle drei mit gleicher Wahrscheinlichkeit auftreten. Die Auswahl geschieht zufällig, was darauf hindeutet, dass der Kanal nach Tab. 108.6b weniger gut zur Datenübertragung geeignet ist.

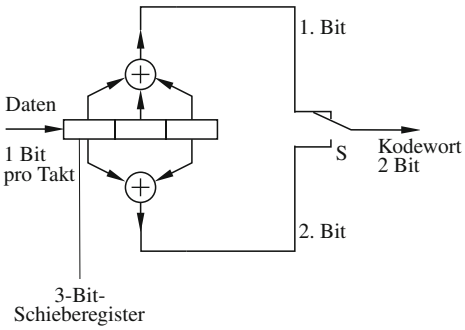


Abb. 108.1 Faltungscodierer

Tab. 108.7 Codewort des Faltungscodierers nach Abb. 108.1

Inhalt Schieberegister (von links)			Änderung des Codewortes	
Alt	Eingeschobenes Datenbit	Neu	Alt 1. Bit/2. Bit (Abb. 108.1)	Neu 1. Bit/2. Bit (Abb. 108.1)
000	0	000	00	00
000	1	100	00	11
001	0	000	11	00
001	1	100	11	11
010	0	001	10	11
010	1	101	10	00
011	0	010	01	11
011	1	101	01	00
100	0	010	11	10
100	1	110	11	01
101	0	010	00	10
101	1	110	00	01
110	0	011	01	01
110	1	111	01	10
111	0	011	10	01
111	1	111	10	10

108.1.2.3 Faltungscodierer mit Likelihood-Viterbi-Decodierer

Abb. 108.1 zeigt ein einfaches Beispiel für einen Faltungscodierer. Mit jedem Schiebetakt wird ein Datenbit in das 3-Bit-Schieberegister geschoben. Der obere Modulo-2-Addierer erzeugt das erste Bit, der untere das zweite Bit des Codewortes, wobei der Schalter S als Parallel-Seriell-Umsetzer dient. Das Codewort wird übertragen und ist eine Kombination aus Daten- und Korrekturbit. Tab. 108.7 zeigt das übertragene Codewort in Abhängigkeit vom Zustand des Schieberegisters und dem eingeschobenen Datenbit.

Zur Erklärung der prinzipiellen Wirkungsweise des Decodierers wird angenommen, dass der Inhalt des Schieberegisters 000 ist. In der Praxis wird das durch die Übertragung geeigneter Bitmuster erreicht. Es werde die Codewortfolge 00 11 10 00 01 01 11 00 00 (Codeworte 1 . . . 9) empfangen. Da der Decodierer das Funktionsprinzip des Faltungscodierers kennt, ermittelt er für jeden empfangenen Codewort*wechsel* das zugehörige Datenbit. Tab. 108.8 zeigt den Vorgang.

Der Übergang vom ersten auf das zweite Codewort in der Codewortfolge (zweite Spalte Tab. 108.8) von 00 auf 11 wurde, sofern die Übertragung fehlerfrei erfolgte, durch Einschieben des Datenbits „1“ in das Schieberegister hervorgerufen, wie sich aus der vierten Zeile der Tab. 108.7 ablesen lässt (000 → 100). Der nächste Codewortübergang von 11 auf 10 ergibt sich, wenn sich das im Schieberegister enthaltene Wort 100 (Tab. 108.7, Zeile 11) durch Einschieben einer „0“ in 010 ändert. Das Verfahren wird weiter fortgesetzt, wobei die übertragene Codewortfolge und die daraus ermittelten Datenbits zunächst bis zur Auswertung des letzten Codewortes zwischengespeichert werden. Da kein Widerspruch zwischen der Aufeinanderfolge der Schieberegisterinhalte und den Codewörtern besteht, sind die in der letzten Zeile von Tab. 108.8 aufgeführten Datenbits in dieser Reihenfolge auch tatsächlich ausgesendet und fehlerfrei empfangen worden und können vom Decoder freigegeben werden.

Die fehlerfreie Codewortfolge von oben, 00 11 10 00 01 01 11 00 00, wird jetzt bei der Übertragung verfälscht und als 00 11 10 1̲0 01 01 11 00 00 empfangen. Die unterstrichene „1“ kennzeichnet die fehlerhafte Stelle.

Nach Eintreffen des dritten Codewortes (10) schließt der Empfänger auf den Schieberegisterinhalt 010 (3. Spalte, Tab. 108.9, fett gedruckt). Trifft das folgende falsche Codewort 10 ein, ist das ein Übergang 10 auf 10, was dem „Wechsel“ des Schieberegisterinhaltes von 111 (Spalte 4, Tab. 108.9, fett gedruckt) auf 111 gemäß der letzten Zeile in Tab. 108.7 entspricht. Spalte 3 bestimmt den Schieberegisterinhalt zu 010, Spalte 4 zu 111. Dieser Widerspruch deutet auf einen Übertragungsfehler hin. Um das Verfahren einfach darstellen zu können, wird der Schieberegisterinhalt 010 nach Eintreffen des dritten Codewortes (10) als richtig angenommen. Damit kann der nächste Schieberegisterinhalt entweder nur 001 sein, wenn eine 0 eingeschoben wird, oder 101, wenn das Datenbit 1 ist. Im ersten Fall ergibt sich für 010 → 001 das neue Codewort 11 (7. Zeile Tab. 108.7), im zweiten Fall 010 → 101 das neue Codewort 00. Nach Empfang des

Tab. 108.8 Decodiervorgang ohne Übertragungsfehler

Empfangenes Codewort alt	00	11	10	00	01	01	11	00	00	Codeworte 1 . . . 9
Empfangenes Codewort neu	11	10	00	01	01	11	00	00		Codeworte 2 . . . 9
Schieberegister-Inhalt alt	000	100	010	101	110	011	001	000	000	
Schieberegister-Inhalt neu	1̲00	0̲10	1̲01	1̲10	0̲11	0̲01	0̲00	0̲00		Eingeschobenes Datenbit unterstrichen
Datenbit	1	0	1	1	0	0	0	0		Zunächst zwischengespeichert

Ausgegebene Datenbitfolge: 10110000

Tab. 108.9 Decodiervorgang mit einem Übertragungsfehler, erster Durchlauf

Empfangenes Codewort alt	00	11	10	10	01	01	11	00	00	Fehlerhafte Stelle unterstrichen
Empfangenes Codewort neu	11	10	10	01	01	11	00	00		
Schieberegister-Inhalt alt	000	100	**111**	111	**110**	011	001	000	000	Hervorgehobene Stellen: Widerspruch
Schieberegister-Inhalt neu	100	**010**	111	**011**	011	001	000	000		
Datenbit	(1)	(0)	?	?	?	?	?	?		Noch nicht angebbar

Tab. 108.10 Decodiervorgang mit einem Übertragungsfehler, zweiter Durchlauf

Empfangenes Codewort alt	00	11	10	11	01	01	11	00	00	Eingesetztes Codewort unterstrichen
Empfangenes Codewort neu	11	10	11	01	01	11	00	00		
Schieberegister-Inhalt alt	000	100	010	**100**	110	011	001	000	000	Widerspruch fett gekennzeichnet
Schieberegister-Inhalt neu	100	010	**001**	110	011	001	000	000		
Datenbit	(1)	(0)	(0)	?	?	?	?	?		Noch nicht angebbar

letzten Codewortes wird in einem zweiten Durchlauf das Codewort, das zum Widerspruch führte, nämlich das vierte (10, hier das fehlerhafte), durch das eben bestimmte Codewort 11 ersetzt und der Zyklus nach Tab. 108.9 bis zum Schluss durchlaufen. Tab. 108.10 stellt den Vorgang bis zum erneuten Widerspruch dar.

In einem dritten Durchlauf wird das andere Codewort, 00, eingesetzt. Es folgt der in Tab. 108.8 aufgelistete Sachverhalt. Da es keinen Widerspruch gibt, wird die in der letzten Zeile dargestellte Datenbitfolge ausgegeben.

Der Vorgang der Decodierung lässt sich anschaulich im *Trellis-Diagramm* darstellen. In Abb. 108.2a ist der Codebaum für den Faltungscodierer nach Abb. 108.1 abgebildet. Der Ausgangszustand des Schieberegisterinhaltes ist 000. Je nachdem, welches Informationsbit eingeschoben wird, verzweigt sich der Codebaum nach oben (Informationsbit 0) oder unten (Informationsbit 1). Im Codebaum sind die Codeworte aufeinanderfolgend eingetragen. Es zeigt sich, dass der Codebaum innerhalb des Wiederholungsintervalls (Linien I, II) alle möglichen Zustände bzw. Übergänge enthält (Verzweigungspunkte A, B, C, D). Man kann deshalb den unteren Teil mit dem oberen zusammenfassen und alle weiteren, rechts der Linie II auftretenden Codierungen weglassen. Es ergibt sich das Trellis-Diagramm nach Abb. 108.2b.

Abb. 108.2c zeigt vereinfacht den Entscheidungsvorgang, wenn die (fehlerhafte) Codewortfolge 00 11 11 11 00 empfangen wurde. Nur für den Fall, dass im Entscheidungspunkt E1 das wahrscheinlichere Codewort 10 eingesetzt wird, gelangt man zur Ebene A (durchgezogene Linie). Die anderen Möglichkeiten (gestrichelte Linien) haben eine größere Hamming-Distanz zur wahrscheinlichsten Codewortfolge und scheiden aus. Es wird deshalb die Codewortfolge 0011 10 11 00 decodiert. Wird eine längere Codewortfolge gesendet, benötigt man letztlich nur den im Abb. 108.2c dargestellten Teil des Diagramms zur Decodierung, weil sich die Entscheidungsvorgänge durch die Codewort*wechsel* in gleicher Art wiederholen *(Viterbi-Algorithmus)*.

In der Praxis werden Schieberegister mit bis zu acht Stellen eingesetzt. Damit können mehrere Fehler erkannt und korrigiert werden. Dieses Verfahren wird u. a. auch bei der Datenübertragung über Satelliten eingesetzt.

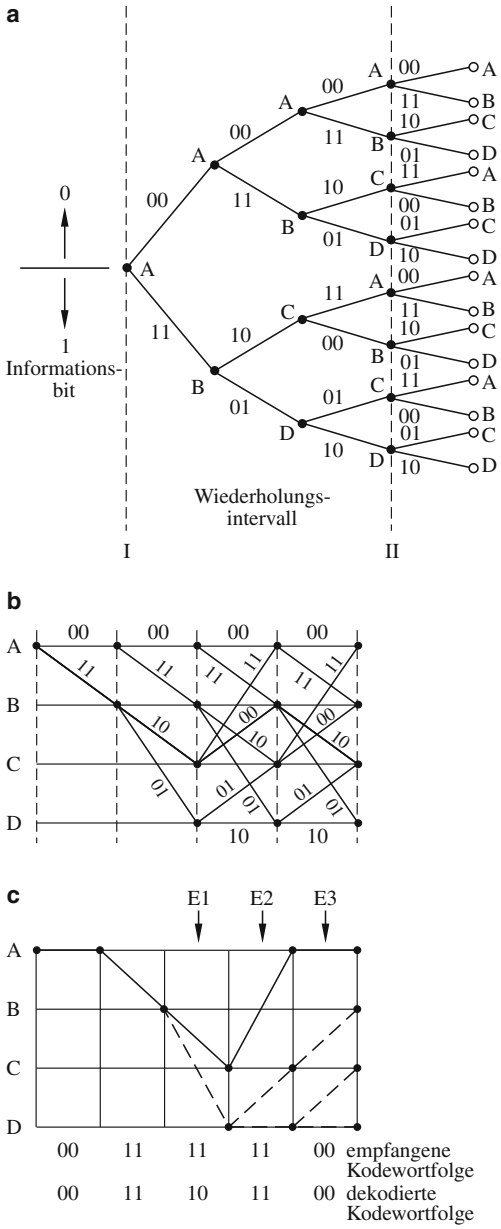


Abb. 108.2 Decodierung eines Faltungscodes. **a** Codebaum für die Codeworte, **b** Trellis-Diagramm, **c** Prinzip der Decodierung mit Likelihood-Viterbi-Decoder

108.2 Optimalfilter

Ein Optimalfilter hat die Aufgabe, ein Nutzsignal, dem ein Störsignal überlagert ist, möglichst wenig zu verändern und das Störsignal so weit wie möglich zu unterdrücken. Das Prinzip wird an dem folgenden Modell dargestellt, Abb. 108.3a.

Hinweis: Um Verwechslungen zu vermeiden, wird die Laplace-Transformierte mit p und nicht mit s bezeichnet.

Gegeben sind ein ungestörtes Signal $u_S(t)$ mit der spektralen Leistungsdichte $F_S(f)$ (Abschn. 92.4.11.2) und ein Störsignal $u_N(t)$ mit der spektralen Leistungsdichte $F_N(f)$, die sich additiv überlagern. Dieses Signal wird auf den Eingang eines Filters mit der Übertragungsfunktion $\underline{H}(f)$ (Abschn. 92.4.9.3) gegeben.

Am Ausgang erscheint das Signal $u_a(t)$, Abb. 108.3a. Für Nutz- und Störsignal werden spektrale Leistungsdichten, wie in Abb. 108.3b eingezeichnet, angenommen, wobei $F_N(f)$ „weißem“ Rauschen entspricht (Abschn. 92.4.11.1 „Rauschen“). Die eigentliche Aufgabe besteht darin, eine optimale Filter-Übertragungsfunktion $\underline{H}(f) = \underline{H}_{\text{opt}}(f)$ zu finden, die das Signal $u_S(t)$ wenig verändert und die Störung ausreichend unterdrückt. Hier wird zunächst angenommen, $\underline{H}(f) = \underline{H}_{\text{opt}}(f)$ sei berechnet worden und habe einen „idealen“ Tiefpass ergeben, dessen Grenzfrequenz f_g aber für unser Modell noch einstellbar ist. „Ideal“ heißt hier, dass bis f_g alle Frequenzanteile mit unveränderter Amplitude durchgelassen werden und oberhalb f_g alle Frequenzanteile vollständig unterdrückt werden. Die Grenzfrequenz f_g ist in Abb. 108.3b an beliebiger Stelle eingetragen. Es ergeben sich zwei Fehler:

1. $\overline{\delta}_S^2$: Durch die Begrenzung der Signalfrequenzen auf den Bereich $0 \ldots f_g$ wird das ungestörte Signal $u_S(t)$ verfälscht, und zwar um so mehr, je kleiner f_g gewählt wird.

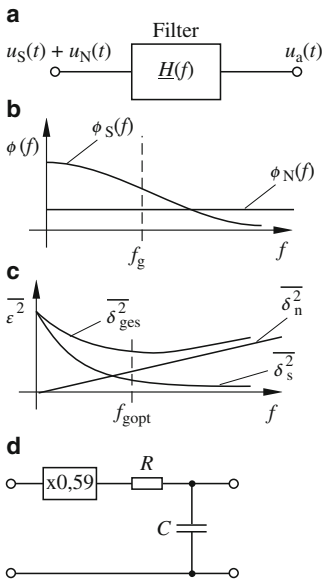


Abb. 108.3 Optimalfilter. **a** Prinzip, **b** Spektrale Leistungsdichte von Nutz- und Störsignal, **c** Gesamtfehler, **d** Realisierung des Beispiels

2. $\overline{\delta}_N^2$: Gleichzeitig wird der Bereich der Störfrequenzen auch auf den Bereich $0 \ldots f_g$ begrenzt. Je kleiner f_g gewählt wird, desto geringer sind die Verfälschungen von $u_S(t)$ durch das Störsignal.

Es gibt damit eine optimale Grenzfrequenz $f_g = f_{g\,\text{opt}}$, bei der die Signalverfälschungen noch vertretbar und die Störungen ausreichend klein sind. Um das zu zeigen, werden die zwei Fehler berechnet:

$$\overline{\delta}_S^2 = 2 \cdot \int_{f_g}^{\infty} \Phi_S(f) df; \quad \overline{\delta}_N^2 = 2 \cdot \int_0^{f_g} \Phi_N(f) df \tag{108.7}$$

Das Nutzsignal $u_S(t)$ habe die spektrale Leistungsdichte

$$\Phi_S(f) = \frac{1}{1 + 4\pi^2 \frac{f^2}{f_0^2}} \tag{108.8}$$

und für das Störsignal soll gelten

$$\Phi_N(f) = 0{,}2 = \text{const} \tag{108.9}$$

abgeleitet vom „weißen“ Rauschen.

Setzt man diese Ausdrücke in (108.7) ein, ergeben sich die Fehler in Abhängigkeit von f_g wie in Abb. 108.3c dargestellt. Da beide Fehler unkorreliert sind, werden sie für den Gesamtfehler $\overline{\delta}^2_{\text{ges}}$ addiert. Ziel der Optimierung ist, den Gesamtfehler zu minimieren:

$$\overline{\delta}^2_{\text{ges}} = \overline{\delta}_S^2 + \overline{\delta}_N^2 = \text{Minimum} \tag{108.10}$$

Setzt man (108.8) und (108.9) in (108.7) ein und bildet die Summe nach (108.10), so folgt:

$$2 \int_{f_g}^{\infty} \frac{1}{1 + 4\pi^2 \frac{f^2}{f_0^2}} df + 2 \int_0^{f_g} 0{,}2 df$$
$$= \frac{f_0}{2} - \frac{f_0}{\pi} \cdot \arctan\left[\frac{f_g}{f_0} 2\pi\right] + 0{,}2 \cdot f_g = \text{Minimum}$$

Daraus ergibt sich das Minimum zu $f_g \approx 0{,}47 \cdot f_0$.

Die bisher als bekannt vorausgesetzte optimale Übertragungsfunktion $\underline{H}(f) = \underline{H}_{\text{opt}}(f)$ lässt sich berechnen nach der Beziehung

$$\underline{H}_{\text{opt}}(f) = \frac{\Phi_S(f)}{\Phi_S(f) + \Phi_N(f)} \tag{108.11}$$

Werden die Ausdrücke aus (108.8) und (108.9) in (108.11) eingesetzt, erhält man:

$$\underline{H}_{\text{opt}}(f) = \frac{\frac{1}{1+4\pi^2 \frac{f^2}{f_0^2}}}{\frac{1}{1+4\pi^2 \frac{f^2}{f_0^2}} + \frac{1}{5}} = \frac{5 \cdot f_0^2}{6 \cdot f_0^2 + 4\pi^2 f^2} \tag{108.12}$$

Diese Übertragungsfunktion lässt sich so nicht realisieren, weil sie Polstellen in der rechten p-Halbebene hat. Man kann sie aber dadurch annähern, dass die Polstellen in der rechten Halbebene vernachlässigt werden. Zunächst zerlegt man $F_S(f) + F_N(f)$ in der p-Ebene sowohl im Zähler als auch im Nenner in ein Produkt aus konjugiert komplexen Null- bzw. Polstellen:

$$\frac{1}{1+4\pi^2\frac{f^2}{f_0^2}} + \frac{1}{5} = \frac{6\cdot f_0^2 + 4\pi\cdot f^2}{5(f_0^2+4\pi\cdot f^2)}$$

$$\Rightarrow \quad \frac{(p+\sqrt{6}\cdot f_0)\cdot(p-\sqrt{6}\cdot f_0)}{5\cdot(p+f_0)\cdot(p-f_0)} \quad p \Leftrightarrow j2\pi\cdot f$$

Der Ausdruck wird wie folgt zerlegt und bezeichnet:

$$\frac{(p+\sqrt{6}\cdot f_0)\cdot(p-\sqrt{6}\cdot f_0)}{5\cdot(p+f_0)\cdot(p-f_0)} = \underbrace{\frac{(p+\sqrt{6}\cdot f_0)}{5\cdot(p+f_0)}}_{\Phi_+}\cdot\underbrace{\frac{(p-\sqrt{6}\cdot f_0)}{(p-f_0)}}_{\Phi_-}$$

Man bildet

$$\begin{aligned}\frac{\Phi_S(p)}{\Phi_-(p)} &= \frac{f_0^2}{f_0^2-p^2}\cdot\frac{p-f_0}{p-\sqrt{6}\cdot f_0}\\ &= \frac{-f_0^2}{(p+f_0)\cdot(p-f_0)}\cdot\frac{p-f_0}{p-\sqrt{6}\cdot f_0}\\ &= \frac{-f_0^2}{(p+f_0)\cdot(p-\sqrt{6}\cdot f_0)}\\ &= \underbrace{\frac{\frac{f_0}{\sqrt{6}+1}}{(p+f_0)}}_{\text{realisierbar}} - \underbrace{\frac{\frac{f_0}{\sqrt{6}+1}}{(p-\sqrt{6}\cdot f_0)}}_{\text{nicht realisierbar}}\end{aligned}$$

$$\Rightarrow \quad \left.\frac{\Phi_S(p)}{\Phi_-(p)}\right|_{\text{real}} = \frac{\frac{f_0}{\sqrt{6}+1}}{(p+f_0)}$$

Der vorletzte Ausdruck wurde durch Partialbruchzerlegung in den letzten umgewandelt. Davon nimmt man den ersten Summanden, weil er realisierbar ist, den zweiten lässt man als nicht realisierbar weg. Damit folgt:

$$\begin{aligned}\underline{H}_{\text{real}}(p) &= \frac{1}{\Phi_+(p)}\cdot\left.\frac{\Phi_S(p)}{\Phi_-(p)}\right|_{\text{real}}\\ &= \frac{5(p+f_0)}{p+\sqrt{6}\cdot f_0}\cdot\frac{1}{(p+f_0)}\cdot\frac{f_0}{\sqrt{6}+1}\\ &= \frac{5f_0}{p+\sqrt{6}\cdot f_0}\cdot\frac{1}{\sqrt{6}+1}\\ \underline{H}_{\text{real}}(f) &= \frac{5f_0}{(\sqrt{6}+1)\,j2\pi f + (\sqrt{6}+1)\sqrt{6}\cdot f_0}\\ &= \frac{0{,}59}{1+\frac{j2\pi f}{2{,}45 f_0}}\end{aligned}$$

Das ist ein Tiefpassfilter erster Ordnung, das nach Abb. 108.3d mit $R\cdot C = 1/(2{,}45\cdot f_0)$ und einem Abschwächer mit dem Faktor 0,59 realisiert werden kann.

108.3 Anwendung der Korrelation bei gestörten Signalen

Den hier betrachteten Nutzsignalen ist ein Störsignal in Form von Rauschen additiv überlagert. Nach Abschn. 92.4.11.2 und Beispiel 8 in Kap. 92 gilt für die Autokorrelationsfunktion von Rauschen

$$\Phi_{SS(N)}(\tau) = \lim_{T\to\infty}\frac{1}{2\cdot T}\int_{-T}^{T} u_N(t)\cdot u_N(t-\tau)dt = 0 \qquad (108.13)$$

$u_N(t)$ Rauschspannung in V.

Für das Nutzsignal erhält man:

$$\Phi_{SS(S)}(\tau) = \lim_{T\to\infty}\frac{1}{2\cdot T}\int_{-T}^{T} u_S(t)\cdot u_S(t-\tau)dt \mathrel{\hat{=}} (u_S(t))^2$$

$$\text{für } \tau\to 0 \qquad (108.14)$$

$u_S(t)$ Nutzsignalspannung in V.

Der zu übertragende Datenblock wird inhaltlich nicht verändert und mehrfach übertragen. Der Empfänger addiert die Amplituden der empfangenen Datenblöcke, und zwar so, dass die Bedingung $t \to 0$ nach (108.14) erfüllt ist. In der Praxis heißt das, dass jeweils nur solche Amplitudenwerte addiert werden, die stets zu den gleichen Zeitpunkten des Nutzsignals gehören. Dazu muss man wiederum den Beginn des im gestörten übertragenen Signal enthaltenen Nutzsignales kennen. Das aber ist in den meisten Anwendungen wenigstens näherungsweise gegeben. Damit geht $u_N(t)$ mit wachsender Zahl von Additionen gegen Null, während das Nutzsignal $u_S(t)$ unverändert bleibt, weil immer die gleichen Amplitudenwerte addiert werden. Abb. 108.4 zeigt ein Beispiel dazu. n gibt die Anzahl der Überlagerungen an. In dem verwendeten Rechnerprogramm wird n stets um 1 erhöht, wobei zu jedem neuen n ein neues zufälliges Rauschen erzeugt wird. Für $n = 1$ sind zunächst nur das Nutzsignal und das Rauschen der ersten zwei Diagramme überlagert. Von den 600 ausgegebenen Diagrammen sind einige ausgesucht, die den Rückgang der Amplitude der resultierenden Rauschspannung und damit die zunehmende Erkennbarkeit des Nutzsignals zeigen (im Programm: Addition von n Amplitudenwerten und anschließende Division durch n). Dem Programm ist in diesem Fall der jeweilige Beginn eines neuen gestörten Signalverlaufes bekannt, sodass die Auswertung mit minimalem Aufwand möglich ist. Sind nach Abb. 108.4 die Scheitelwerte der Rauschspannungen etwa 10mal so groß wie die des Nutzsignals, muss das gestörte Signal zwischen 10- und 25mal übertragen werden, damit ein digital codierter 1-0-Wechsel sicher erkannt werden kann. Es zeigt sich weiterhin, dass bei der Übertragung analoger Signale eine sehr große Anzahl von Wiederholungen notwendig ist, um einigermaßen fehlerfreie Werte zu erhalten ($n = 600$ oder mehr).

Angewendet wird dieses Verfahren z. B. bei der Übertragung von Bildern der Oberfläche weit entfernter Planeten.

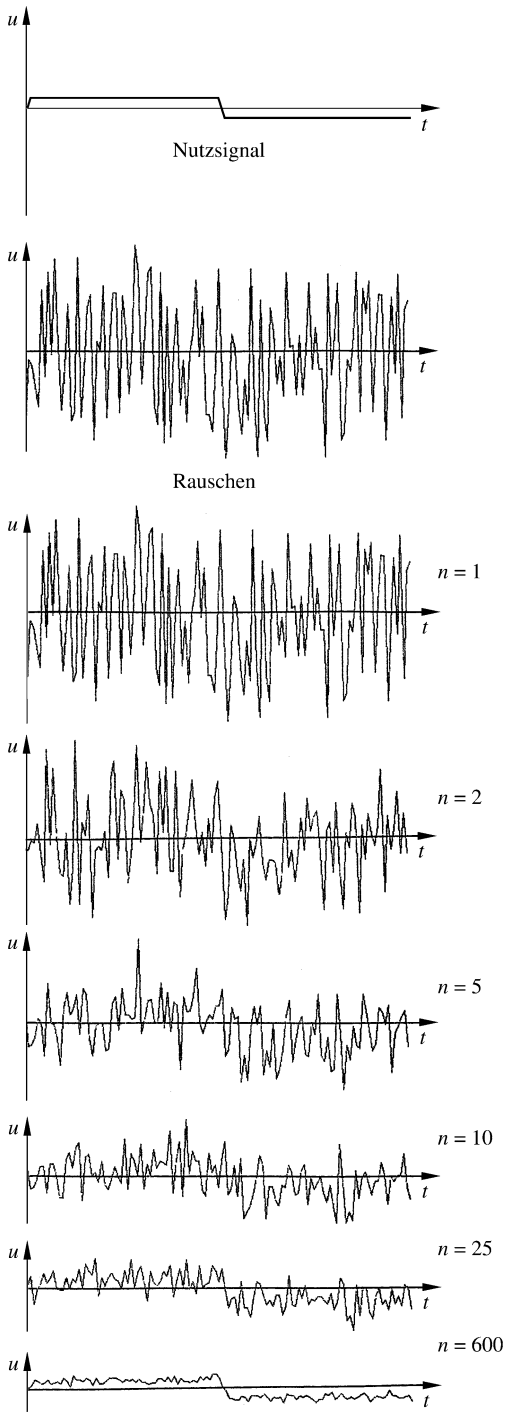


Abb. 108.4 Signalformen von korrelierten gestörten Signalen, *n*: Zahl der Überlagerungen

108.4 Verschlüsselung von Daten

Die Verschlüsselung von Daten ist zu einer Notwendigkeit geworden. Als Beispiel soll hier der Geldverkehr in Form von Online-Banking oder das Abheben von Geld am Geldautomaten mit der erforderlichen Identifikation des Abhebenden genannt werden. Im Folgenden wird neben allgemeinen Anmerkungen die Datenverschlüsselung an dem aktuellen RSA-Verfahren (benannt nach den Erfindern *R*ivest, *S*hamir und *A*dleman) praktisch gezeigt. Grundlage für diese Art der Verschlüsselung ist die Primzahlrechnung. Weitergehende Informationen bekommt man in der Literatur unter den Begriffen *Euler-Funktion*, *Kleiner Satz von Fermat* und *Satz von Euler-Fermat*.

Für die Entwicklung von Verschlüsselungsverfahren gilt: Eine hundertprozentig sichere Verschlüsselungsmethode gibt es nicht. Man kann allerdings die Verfahren so auswählen, dass der apparative Aufwand und vor allem der Zeitaufwand für „Codeknacker“ in keinem lohnenden Verhältnis mehr zum Nutzen stehen.

Forderungen:

1. Die Verschlüsselung soll möglichst so sicher sein, dass nur befugte Personen oder Einrichtungen an die Information der verschlüsselten Datei gelangen können. Beispiele: Übertragung wichtiger Dokumente im Internet; Übertragung der Daten (PIN, Kontonummern, Geldbetrag) zur Bank beim Bezahlvorgang mit Scheckkarte im Geschäft.
2. Eine berechtigte Person soll nachweisen können, dass genau sie und nur sie die Berechtigung zu bestimmten Handlungen besitzt. Beispiel: Geld abheben am Geldautomaten. Hierbei muss eine sichere Verbindung zu einem Zentralrechner hergestellt werden, der die notwendigen Schritte zur Identifizierung durchführt.

Die Verschlüsselung (Codierung) geschieht häufig über ein *Schlüsselwort* (kurz *Schlüssel*) oder *Passwort*, mit dessen Hilfe die Information so verändert wird, dass sie keinen Sinn ergibt. Nur das korrekte Passwort ergibt bei der Umkehrung des Verschlüsselungsvorganges, auch Entschlüsselung oder Decodierung genannt, die Originalinformation.

Am Vorgang des Verschlüsselns sind zwei Gruppen beteiligt:

Die erste Gruppe sind die Entwickler, die ein geeignetes Verfahren bereitstellen. Hierzu sind allgemeine mathematische und insbesondere Statistik-Kenntnisse erforderlich. Der verschlüsselte Text darf z. B. keine Rückschlüsse auf die Buchstaben- oder Silbenhäufigkeit zulassen. Auch an bestimmten Stellen vermutete öfter wiederkehrende Formulierungen (Datum, Anrede, Adressen, Schlüsselwörter) dürfen als solche nicht das Entschlüsseln erleichtern. Die enorme Leistungssteigerung von Rechnern, auch die der Heim-PCs, hat es ermöglicht, dass Verschlüsselungsverfahren, wenn sie nicht ausreichend konsequent eingesetzt werden, mehr oder weniger unterlaufen werden können. Genannt sei der „Brute-Force-Angriff“ (wörtlich „Rohe-Gewalt-Angriff“), bei dem an einer verschlüsselten Datei sehr viele unterschiedliche Passwörter der Reihe nach ausprobiert werden. Entweder man erhält schließlich eine sinnvolle Information oder der Angriff ist fehlgeschlagen. Die verwendete

Sammlung von Passwörtern wird gebildet aus den Eintragungen in Lexika, in Wörterbüchern und in einer „Hitliste“ der weltweit häufig verwendeten Passwörter, die nicht im Lexikon oder im Rechtschreibwörterbuch stehen. Mit Heim-PCs und dem geschickten Einsatz von Grafikkartenprozessoren, die sehr leistungsfähig sind, können einige Hunderttausend Wörter innerhalb von wenigen Sekunden ausprobiert werden. Der „Brute-Force-Angriff“ ist einfach zu automatisieren und erfordert keine speziellen Kenntnisse der Verschlüsselungstheorie. Zwei wichtige Folgerungen aus dem „Brute-Force-Angriff“ sind (siehe auch Anmerkung oben): 1. Es kann keine absolut sichere Verschlüsselung geben. 2. Es nimmt immer mehr Zeit in Anspruch, das richtige Passwort durch Ausprobieren zu bekommen, je mehr unterschiedliche Zeichen verwendet werden (nur Ziffern oder Ziffern + Großbuchstaben oder Ziffern + Großbuchstaben + Kleinbuchstaben + Sonderzeichen) und je länger das Passwort ist (6 oder 10 oder 15 Zeichen). Bei einem 10-stelligen Passwort, bei dem jede Stelle aus einem Element des Zeichenvorrates von insgesamt 128 Zeichen besteht (Großbuchstaben + Kleinbuchstaben + Ziffern + Sonderzeichen) und dieses dann binär mit 7 Bit codiert wird, ergeben sich insgesamt $2^{70} = 1.180.591.620.717.411.303.424$ unterschiedliche Werte. Werden 256 unterschiedliche Zeichen eingesetzt, die mit 8 Bit codiert werden, und wird ein Passwort mit 16 Bit verwendet, ergeben sich insgesamt $2^{128} = 340.282.366.920.938.463.463.374.607.431.768.211.456$ unterschiedliche Werte. Ein Unsicherheitsfaktor bei der Angabe der Zeit bis zur erfolgreichen Entschlüsselung ist die Tatsache, dass man im günstigsten Fall (das ist der ungünstigste Fall für die Verschlüsselung) das gesuchte Passwort beim ersten Versuch erhält (Wahrscheinlichkeit ist $(1/(2^{70}))$ bzw. $(1/(2^{128}))$ in den Beispielen oben) oder im ungünstigsten Fall (das ist der optimale Fall für die Verschlüsselung) beim 2^{70}ten bzw. 2^{128}sten Versuch (Wahrscheinlichkeit ist dann 1).

Mit Großrechnern im Verbund und 10^{15} Tests auf Richtigkeit pro Sekunde werden dafür zurzeit *maximal* 13,7 Tage bzw. 10^{16} Jahre benötigt (Beispiele oben). Irgendwann im Zeitbereich zwischen „praktisch Null“ und 13,7 Tagen bzw. 10^{16} Jahren und einem enormen apparativen Aufwand bekommt man das richtige Ergebnis. Es stellt sich dann die Frage, ob sich das „Codeknacken“ noch lohnt.

Die zweite Gruppe sind die Anwender, die für eine verschlüsselte Übertragung ein Passwort selber erzeugen. Hier liegt eine große Gefahr für die Sicherheit der Verschlüsselung. Der „Erfolg“ eines Angriffes kann nur dadurch erschwert werden, dass für das Passwort 1. völlig sinnlose Zeichen*kombinationen* gewählt werden (Begriffe aus Lexika und Passwort-Hitlisten werden als erstes – und leider häufig auch erfolgreich – zum Codeknacken eingesetzt) und 2. so viel Zeichen wie möglich verwendet werden. Es empfiehlt sich auch, das oder die Codewörter im Abstand von wenigen Wochen zu ändern. Es ist sicher nicht wünschenswert, dass der Code geknackt worden ist und die eigenen Informationen weiter verwendet werden, ohne dass man selber davon weiß. Die Erfahrung, verknüpft mit dem momentanen Stand der Technik, zeigt, dass niemals weniger als 8 bis 10 Zeichen für das Passwort verwendet werden sollten und dass Klein- und Großbuchstaben, Ziffern und Sonderzeichen zum Einsatz kommen. Das Problem für den Anwender ist, dass er sich solche Passwörter kaum merken kann. Er wird sie aufschreiben oder auf dem Rechner speichern. Beides sind, wenn sie nicht sehr sorgfältig gehandhabt werden, zusätzliche Gefahrenquellen.

Bezogen auf den Schlüsselaustausch kann man zwei Verfahren unterscheiden:

1. Symmetrische Verfahren: Der Sender der Nachricht und der Empfänger verwenden das gleiche Schlüsselwort, das sie vor Anwendung der verschlüsselten Übertragung schon ausgetauscht haben müssen, natürlich geheim. Nehmen n Teilnehmer an dem verschlüsselten Datenaustausch teil, und will ein Sender an $(n-1)$ Empfänger eine Nachricht senden, müssen $(n-1)$mal die Schlüsselwörter ausgetauscht bzw. versendet worden sein, wieder geheim. Es besteht die Gefahr, dass eine unerwünschte Person das Passwort mithört oder abfängt. Wollen alle n Teilnehmer mit jedem anderen Teilnehmer verschlüsselte Nachrichten austauschen und jeder sein eigenes Passwort benutzen, müssen zuvor $(n \cdot (n-1))/2$ Passwörter geheim ausgetauscht werden.
2. Asymmetrische Verfahren: Sender und Empfänger müssen *kein gemeinsam* benutztes *geheimes* Schlüsselwort austauschen, sondern nur mehrere nicht geheime Wörter. Beide können daraus ihr *eigenes* Schlüsselwort erzeugen, aber *nicht* das Schlüsselwort des anderen, und ein Mithörer kann daraus weder das Schlüsselwort des Senders noch des Empfängers generieren.

In einem weltweit ausgeschriebenen Wettbewerb hat sich das RSA-Verfahren als zurzeit zuverlässigstes Verfahren herausgestellt. Sein Wirkungsprinzip ist bewusst offengelegt, damit Schwachstellen aufgedeckt werden können. Ganz allgemein formuliert nutzt das Verfahren die Tatsache aus, dass es für die Folge der Primzahlen kein Bildungsgesetz gibt und damit die Zerlegung einer sehr großen Zahl (mehr als 20 ... 50 Stellen) in Primzahlfaktoren ohne jegliche Zusatzinformation sehr zeitintensiv ist.

RSA-Verfahren, Schlüsselaustausch

Der Schlüsselaustausch beim RSA-Verfahren wird hier an vergleichsweise kleinen Primzahlen gezeigt. Vorteil: Das Prinzip ist nachvollziehbar. Nachteil: Eine ausreichende Sicherheit ergibt sich erst mit einer sehr viel größeren Stellenzahl der Primzahlen.

Anna (A) und Bernd (B) wollen eine geheim zu haltende Nachricht austauschen. Als erstes tauschen sie den Schlüssel aus. Dazu wählt einer von beiden eine beliebige Primzahl p und eine Zufallszahl z und teilt dem anderen diese beiden Zahlen mit. Sie werden nicht verschlüsselt übertragen.

Anna wählt eine Zufallszahl a mit der Bedingung $a < p$ und hält diese Zahl geheim.

Bernd wählt eine Zufallszahl b mit der Bedingung $b < p$ und hält diese geheim.

Anna berechnet $A \equiv z^a$ (modulo p); Bernd $B \equiv z^b$ (modulo p).

Anna sendet das Ergebnis A an Bernd, Bernd das Ergebnis B an Anna, unverschlüsselt.

Anna berechnet $S_{ab} \equiv B^a$ (modulo p); Bernd berechnet $S_{ab} \equiv A^b$ (modulo p).

Tab. 108.11 Codier- und Decodiervorgang

Originaltext	1	0	0	0	Leerzeichen	E	U	R
i (fortlaufend nummeriert)	1	2	3	4	5	6	7	8
Zuordnung: Textelement $\leftrightarrow$ Zahl M_i	1 $\leftrightarrow$ 31	0 $\leftrightarrow$ 30	0 $\leftrightarrow$ 30	0 $\leftrightarrow$ 30	Leerz. $\leftrightarrow$ 20	E $\leftrightarrow$ 40	U $\leftrightarrow$ 54	R $\leftrightarrow$ 52
Originaltext verschlüsseln								
$C_i \equiv (M_i)^e$ (modulo n) $\Rightarrow C_i \equiv (M_i)^{27}$ (modulo 55)	$C_1 \equiv 26$	$C_2 \equiv 35$	$C_3 \equiv 35$	$C_4 \equiv 35$	$C_5 \equiv 15$	$C_6 \equiv 50$	$C_7 \equiv 54$	$C_8 \equiv 13$
Nachricht senden								
$C_1 C_2 \ldots C_8$	26 35 35 35 15 50 54 13							
Gesendete Nachricht entschlüsseln								
$M_i \equiv (C_i)^d$ (modulo n) $\Rightarrow M_i \equiv (C_i)^3$ (modulo 55) $\rightarrow$ ergibt Zahl $\leftrightarrow$ ergibt Textelement	$M_1 \equiv 31 \leftrightarrow 1$	$M_2 \equiv 30 \leftrightarrow 0$	$M_3 \equiv 30 \leftrightarrow 0$	$M_4 \equiv 30 \leftrightarrow 0$	$M_5 \equiv 20 \leftrightarrow$ Leerzeichen	$M_6 \equiv 40 \leftrightarrow$ E	$M_7 \equiv 54 \leftrightarrow$ U	$M_8 \equiv 52 \leftrightarrow$ R
Originaltext, formatiert	1000 EUR							

Beide erhalten den gleichen Wert, also das gleiche Schlüsselwort S_{ab}.

Mit Zahlenwerten: Beide einigen sich auf die Primzahl $p = 157$ und die Zufallszahl $z = 13$. Das tauschen sie unverschlüsselt untereinander aus. Anna wählt die Zufallszahl $a = 111$ und hält sie geheim, Bernd wählt die Zufallszahl $b = 14$ und hält sie geheim. Beide Zahlen sind voraussetzungsgemäß kleiner als p.

Anna bildet $A \equiv z^a$ (modulo p) $\equiv 13^{111}$ (modulo 157) $\equiv 156$.

Bernd bildet $B \equiv z^b$ (modulo 157) $\equiv 13^{14}$ (modulo 157) $\equiv 12$.

Anna sendet 156 an Bernd, Bernd sendet 12 an Anna, unverschlüsselt.

Anna berechnet $S_{ab} \equiv 12^{111}$ (modulo 157) $\equiv 1$.

Bernd berechnet $S_{ab} \equiv 156^{14}$ (modulo 157) $\equiv 1$.

Beide Werte sind gleich. Aus den unverschlüsselt übertragenen Zahlen $z = 13$, $p = 157$, $A = 156$ und $B = 12$ lässt sich der Schlüssel nicht bestimmen, und weder Anna noch Bernd können die Zufallszahl b bzw. a des anderen errechnen.

RSA-Verfahren, Prinzip Datenaustausch

Zur Übertragung größerer Datenmengen geht man folgendermaßen vor:

1. Bestimmen der Verschlüsselungsparameter:
   Wählen zweier (beliebiger) Primzahlen p und q mit der Bedingung, dass die Zahl der unterschiedlichen zu übertragenden Zeichen (bei Textübertragung z. B. Zahl der Großbuchstaben + Zahl der Kleinbuchstaben + Zahl der Ziffern + Zahl der Sonderzeichen) kleiner als $p \cdot q$ ist.
   Berechnen des Produktes $n = p \cdot q$ (n ist der RSA-Modul).
   Wählen einer Zahl e aus dem Zahlenbereich $2 \leq e \leq (n-1)$ mit der Bedingung „e ist teilerfremd zu $J(n) = (p-1) \cdot (q-1)$“ ($J(n)$ ist die Euler-Funktion).
   Wählen einer Zahl d aus dem Zahlenbereich $1 \leq d \leq (n-1)$ mit der Bedingung „$e \cdot d = 1$ (modulo $J(n)$)“.
   Die Zahlen p, q und $J(n)$ werden jetzt nicht mehr benötigt.
   $\rightarrow$ (n, e) ist der öffentliche Schlüssel, der unverschlüsselt ausgetauscht wird.
   $\rightarrow$ (n, d) ist der geheime Schlüssel, wobei nur d geheim zu halten ist. Er kann z. B. nach dem oben erläuterten Verfahren ausgetauscht werden.
2. Verschlüsseln:
   Die zu verschlüsselnde Nachricht liegt entweder als eine einzelne Zahl M oder als Zahlenfolge mit den Einzelelementen M_i vor. Bei der Textübertragung könnte das z. B. der ASCII-Code der einzelnen Zeichen sein. Dann werden die einzelnen Zahlen (eindeutig verknüpft mit den Zeichen), in der Reihenfolge wie in der Nachricht, der Operation $C_i \equiv (M_i)^e$ (modulo n) unterzogen. Diese Einzelelemente C_i werden, in der Reihenfolge unverändert, aneinandergereiht und übertragen.
3. Entschlüsseln:
   Die übertragenen Einzelelemente werden, in der Reihenfolge unverändert, der Operation $M_i \equiv (C_i)^d$ (modulo n) unterzogen, und das ergibt, aneinandergereiht, den unverschlüsselten Text.

RSA-Verfahren, einfaches Beispiel zum Datenaustausch

Übertragen eines verschlüsselten Textes nach dem RSA-Verfahren.

Zunächst $p = 5$, $q = 11$ gewählt, ergibt $n = p \cdot q = 55$. $J(n) = (5-1) \cdot (11-1) = 40$. Gewählt $e = 27$. Dazu wird $d = 3$ so gewählt, dass die notwendige Bedingung $e \cdot d = 27 \cdot 3 = 81$, 81 (modulo 40) $= 1$ erfüllt ist. Die Wahl von e und dem dazu geeigneten d kann durchaus mit einigem Aufwand verbunden sein. Es können wegen $n = 55$ insgesamt 54 unterschiedliche Zeichen übertragen werden, also z. B. 26 Großbuchstaben, 10 Ziffern und 18 Sonderzeichen.

Verschlüsselt übertragen werden soll der Text „1000 EUR“ – ohne die Anführungsstriche. Die Text-Elemente (Text-Zeichen) sind wie folgt den Zahlenwerten zugeordnet (nicht ganz konsequent der ASCII-Code): 1 $\leftrightarrow$ 31, 0 $\leftrightarrow$ 30, Leerzeichen $\leftrightarrow$ 20, E $\leftrightarrow$ 40, U $\leftrightarrow$ 54, R $\leftrightarrow$ 52. Diese Zuordnung muss Sender und Empfänger bekannt sein und kann gegebenenfalls vorher öffentlich ausgetauscht werden. Es können damit Zahlen bis maximal 54 verschlüsselt werden. Tab. 108.11 zeigt die Berechnungen zur Ver- und Entschlüsselung.

Ungünstig bei dieser Art der Codierung ist, neben den viel zu kleinen Primzahlen und damit einhergehender unsicherer Verschlüsselung, dass dreimal das Zahlenpaar 30 hintereinander übertragen wird. Das deutet auf gleiche Zeichen im Originaltext hin. Günstiger wäre z. B., mehrere Zeichen zu einer Gruppe zusammenzufassen und diese dann insgesamt zu verschlüsseln. Wenn zwei benachbarte Zeichen zu einer Gruppe zusammengefasst werden, sieht das z. B. so aus: 3130, 3030, 2040, 5452. Der Empfänger muss nach dem Empfang die Gruppe wieder in die Originalzeichen auftrennen. In dem hier dargestellten Beispiel ist diese Gruppenbildung allerdings nicht möglich. Dazu müssten p und q und anschließend e und d neu gewählt werden mit der Bedingung $p \cdot q = n > 5452$ (größter Gruppenwert).

108.5 Fensterung

Häufig ergibt sich die Forderung, die in einem Signal enthaltenen Frequenzanteile zu bestimmen. Ein gängiger Lösungsweg dazu ist die Digitalisierung des Signals und die Anwendung der Fourieranalyse bei periodischen Signalen bzw. der Fouriertransformation in Form der schnellen Fouriertransformation (FFT, siehe Teil XIV „Signal- und Systemtheorie“, Abschn. 117.2 „Schnelle Fouriertransformation“) bei einmaligen Signalen. In der Praxis lassen sich dazu speziell entwickelte und optimierte mathematische Programme einsetzen. Der Anwender oder das Programm wählt aus dem gegebenen Zeitsignal entweder einen Ausschnitt oder das gesamte Signal aus. Dabei wird der Startzeitpunkt $t = 0$ und die Dauer T_{Fe} festgelegt (Abb. 108.5 bis 108.8). Das gesamte Signal oder der Ausschnitt wird digitalisiert und zur weiteren Verarbeitung abgespeichert. Das verarbeitende Programm muss davon ausgehen, dass es sich um ein periodisches Signal mit der Periodendauer T_{Fe} handelt (Voraussetzung für die Anwendung der FFT). Der gespeicherte Signalverlauf wird also mit der Periodendauer T_{Fe} periodisch fortgesetzt. Dabei tritt das Problem auf, dass dem Signalwert zum Zeitpunkt $t = T_{Fe}$ der Signalwert zum Zeitpunkt $t = 0$ folgt. Damit ergibt sich in der Regel ein Sprung im Signalverlauf, der aber in Wirklichkeit nicht vorhanden ist und der im Frequenzspektrum zusätzliche, im Originalsignal nicht vorhandene, Frequenzanteile bewirkt. Die Ergebnisse der Fourierzerlegung bzw. Fouriertransformation sind verfälscht. Dabei bewirkt ein Signalsprung Harmonische mit den Frequenzen $f_n = n/T_{Fe}$ mit $n = 1, 2, 3, \ldots$ und besonders großen Amplituden. An einem einfachen Beispiel mit einem periodischen sinusförmigen Signalverlauf mit der Frequenz $f = 1\,\text{kHz}$ soll das gezeigt werden. Aus diesem Signal wird ein Signalausschnitt (Zeitfenster) mit der Dauer $T_{Fe} = 4{,}3\,\text{ms}$ entnommen, Abb. 108.5. Das Programm zur Berechnung der FFT setzt dieses Signal periodisch fort, Abb. 108.6, und bildet daraus die Fouriertransformierte.

Das Frequenzspektrum zeigt Abb. 108.8. Bei $t = 4{,}3\,\text{ms}$ und Vielfachen davon gibt es einen Sprung im Signalverlauf, der einen Gleichanteil (der Signalanteil von 4,0 bis 4,3 ms wird nicht mehr durch einen gleich großen negativen Signalanteil kompensiert) und ein Spektrum mit abnehmender Amplitude bei steigender Frequenz zur Folge hat. Es werden demnach Frequenzanteile ermittelt, die im Originalsignal nicht vorhanden sind und deren Ursache in dem Signalsprung liegt.

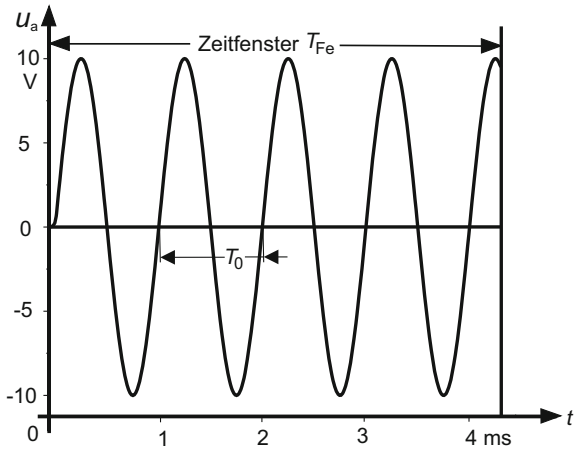


Abb. 108.5 Signalausschnitt zur Bildung der FFT

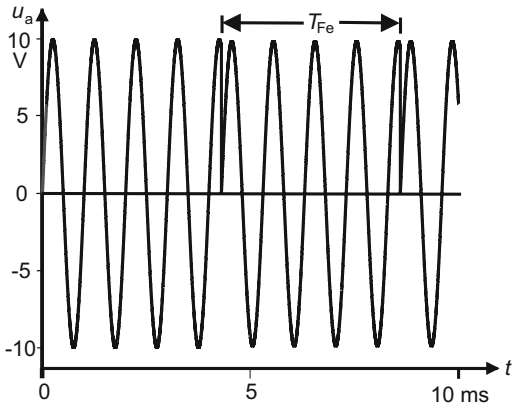


Abb. 108.6 Periodisch fortgesetztes Signal aus Abb. 108.5

Will man diesen Fehler verringern, gibt es zwei Möglichkeiten:

1. Die Zeitdauer T_{Fe} des abgespeicherten Signals wird so gewählt, dass sie ein ganzzahliges Vielfaches der Periodendauer T_0 des gegebenen periodischen Signals ist: $T_{Fe} = n \cdot T_0$ mit $n = 1, 2, \ldots$ Dazu muss T_0 überhaupt erst einmal vorhanden sein (periodisches Signal) und dann auch noch sicher erkannt werden. Diese Maßnahme ist besonders dann kaum durchführbar, wenn es sich um einen automatisch ablaufenden Vorgang handelt, bei dem in der Regel T_{Fe} fest vorgegeben ist.
2. Das abgespeicherte Signal wird zu jedem Zeitpunkt mit einem Faktor, auch *Gewichtungsfunktion* genannt, multipliziert und erst dann ausgewertet. Dieser Faktor ist zu Beginn ($t = 0$) null, geht am Ende ($t = T_{Fe}$) des gespeicherten Signals wiederum gegen Null und ist im überwiegenden Teil des Signalverlaufes 1 oder nahe 1. Damit hat die periodische Fortsetzung des gespeicherten Signals keinen zusätzlichen Sprung, da es zu Beginn und am Ende der „Periode“ den – erzwungenen – Wert 0 hat. Die Gewichtsfunktion wird auch mit Fenster oder Fensterfunktion bezeichnet.

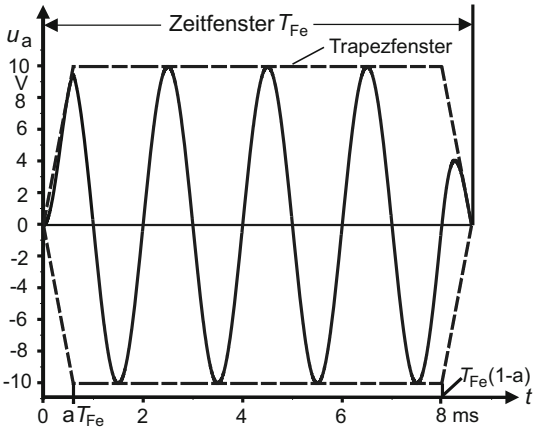


Abb. 108.7 Signal aus Abb. 108.5 nach Passieren eines Trapezfensters

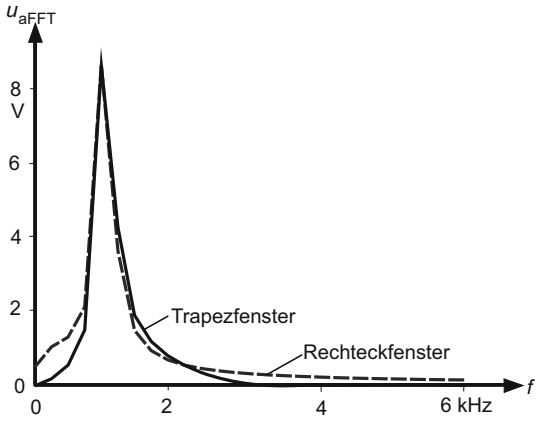


Abb. 108.8 FFT der Signale nach Abb. 108.5 (Rechteckfenster) und Abb. 108.7 (Trapezfenster)

Im Abb. 108.7 ist das abgespeicherte Signal mit der Gewichtsfunktion eines Trapezes multipliziert (*Trapezfenster*). Der Sprung bei $t = T_{Fe}$ ist nicht mehr vorhanden, und der Gleichanteil hat sich verringert. Das zeigt auch der Verlauf in Abb. 108.8, Trapezfenster. Wenn das gespeicherte Signal wie in Abb. 108.5 ohne Einwirken einer Gewichtsfunktion verarbeitet wird, ist das gleichbedeutend mit der Multiplikation mit einer Gewichtsfunktion, die konstant gleich 1 ist. Man spricht dann formal von einem Rechteckfenster, Abb. 108.8.

Für die Gewichtsfunktionen gibt es unterschiedliche Möglichkeiten. Der Anwender muss entscheiden, welche Funktion für seine Anwendung geeignet ist. Einige bewährte Beispiele sind im Folgenden dargestellt. Neben der bereits oben genannten Forderung nach kleinen Werten in der Nähe der Fenstergrenzen sollte die Gewichtsfunktion auch keine Knickstellen enthalten, weil diese ein ausgeprägtes Spektrum bis hin zu höheren Frequenzen besitzen. Insofern ist die hier verwendete Trapezfunktion nur bedingt einsetzbar. Ein Trapezfenster mit „abgerundeten" Knickstellen findet als *Tukey*-Fenster Verwendung.

Es folgen einige Beispiele für häufiger verwendete Gewichtsfunktionen (Fensterfunktionen) f_{Fe}:

1. Rechteckfenster: $f_{Fe}(t) = 1$ für $0 < t < T_{Fe}$, 0 sonst
2. Trapezfenster: $f_{Fe1}(t) = \frac{1}{a \cdot T_{Fe}} \cdot t$ für $0 \le t \le a \cdot T_{Fe}$.
   $f_{Fe2}(t) = 1$ für $a \cdot T_{Fe} \le t \le T_{Fe} \cdot (1-a)$.
   $f_{Fe3}(t) = (1 - \frac{t}{T_{Fe}}) \cdot \frac{1}{a}$ für $T_{Fe}(1-a) \le t \le T_{Fe}$.
   mit $0 < a < 1/2$
3. Tukey-Fenster: Trapezfenster mit abgerundeten Knickstellen
4. Hamming-Fenster: $f_{Fe}(t) = 0{,}54 - 0{,}46 \cdot \cos((2 \cdot \pi \cdot t)/T_{Fe})$ für $0 \le t \le T_{Fe}$.
5. Hanning-Fenster: $f_{Fe}(t) = 1 - \cos^2((\pi \cdot t)/T_{Fe})$ für $0 \le t \le T_{Fe}$.
6. Blackman-Fenster: $f_{Fe}(t) = 0{,}42 - 0{,}5 \cdot \cos((2 \cdot \pi \cdot t)/T_{Fe}) + 0{,}08 \cdot \cos((4 \cdot \pi \cdot t)/T_{Fe})$ für $0 \le t \le T_{Fe}$.
7. Von Hann-Fenster: $f_{Fe}(t) = 0{,}5 \cdot (1 - \cos((2 \cdot \pi \cdot t)/T_{Fe}))$ für $0 \le t \le T_{Fe}$.

In der Nachrichtentechnik treten eigentlich nur stochastische Signalverläufe auf, da nur sie Information enthalten. Die auf die oben geschilderte Art bestimmten Frequenzanteile sind dann „Momentaufnahmen mit mehr oder weniger großen Abweichungen". Anhand von Abb. 108.5 ist zu erkennen, dass sich die Ergebnisse der FFT auch ändern, wenn das Zeitfenster bei gleicher Dauer nur verschoben wird oder wenn die Dauer des Zeitfensters geändert wird. Die erhaltenen Ergebnisse sind deshalb durch geeignete Maßnahmen der Realität anzupassen. Auch hier können die im Kap. 97 „Filter" genannten Entwurfs- und Simulationsprogramme nützlich sein.

Literatur

Literatur allgemein

1. *Herter, E., Lörcher, W.:* Nachrichtentechnik. Verlag Hanser München, Wien. 4,9. Auflage 2003.
2. *Küpfmüller, K.:* Die Systemtheorie der elektrischen Nachrichtenübertragung. Verlag Hirzel Stuttgart. Dritte Auflage 1968.
3. *Bossert, M., Fliege, N., Kammeyer, K. D.:* Nachrichtenübertragung. Verlag Vieweg-Teubner Wiesbaden. 4. Auflage 2008.
4. *Von Grüningen, D.:* Digitale Signalverarbeitung. Fachbuchverlag Leipzig. 4. Auflage 2008.
5. *Werner, M.:* Nachrichten-Übertragungstechnik. Verlag Vieweg-Teubner. 1. Auflage 2006.
6. *Nocker, R.:* Digitale Kommunikationssysteme 2. Verlag Vieweg-Teubner. 1. Auflage 2005.
7. *Butz, T.:* Fouriertransformation für Fußgänger. Verlag Vieweg-Teubner. 6. Auflage 2009.

Literatur zu Kapitel 93

8. *Weißgerber, W.:* Elektrotechnik für Ingenieure 3. Verlag Vieweg-Teubner Wiesbaden. 7. Auflage 2009.

Literatur zu Kapitel 94

9. *Doebke, W.:* Die Berechnung von Leitungsproblemen mit neuen Systemgrößen. AEU, Band 11, 1957, Heft 12.
10. *Geschwinde, H.:* Kreis- und Leitungs-Diagramme. Verlag Franzis München. 1975.
11. *Janssen, W.:* Hohlleiter und Streifenleiter. Verlag Hüthig Heidelberg. 1977.
12. *Krank, W.:* Über die Theorie und Technik des elliptischen Wellrohrleiters. Dissertationa D 82 RWTH Aachen.
13. *Meinke-Gundlach:* Taschenbuch der Hochfrequenztechnik. Hrsg. L. Lange und K. H. Löcherer. Verlag Springer Berlin, Heidelberg, New York, Tokyo. 1986.
14. *Michel, K.-J.:* Zweitoranalyse mit Leistungswellen. Verlag B. G. Teubner Stuttgart. 1981.
15. *Scheffler, E.:* Flexwell-Kabel – Koaxiale Hochfrequenzkabel mit gewelltem Außenleiter. Telefunken-Zeitung. Jahrgang 33, Heft 136.
16. *Stadler, E.:* Hochfrequenztechnik. Verlag Vogel Würzburg. 1973.
17. *Voges, E.:* Hochfrequenztechnik. Band 1, Bauelemente und Schaltungen. Verlag Hüthig Heidelberg. 1986.

Literatur zu Kapitel 95

18. *Unger, H.-G.:* Hochfrequenztechnik in Funk und Radar. Verlag Teubner Stuttgart. 1972.
19. *Schumny, H.:* Signalübertragung. Verlag Vieweg Braunschweig/Wiesbaden. 2. Auflage 1987.

Literatur zu Kapitel 96

20. *Mäusl, R.:* Analoge Modulationsverfahren. Verlag Hüthig Heidelberg. 1988.
21. *Steinbuch, K., Rupprecht, W.:* Nachrichtentechnik, Band 2: Nachrichtenübertragung. Verlag Springer Berlin, Heidelberg, New York. 3. Auflage 1982.
22. *Hölzer, E., Holzwarth, H.:* Pulstechnik, Band 1: Grundlagen. Verlag Springer Berlin, Heidelberg, New York. 1982.
23. *Mäusl, R.:* Digitale Modulationsverfahren. Verlag Hüthig Heidelberg. 3. Auflage 1991.

Literatur zu den Kapiteln 99, 103, 105 und 108

24. *Mäusl, R.:* Fernsehtechnik. Hüthig-Verlag Heidelberg. 4. Auflage 2006.
25. *Opielka, D.:* Optische Nachrichtentechnik. Verlag Vieweg Braunschweig/Wiesbaden. 1995.
26. *Brückner, V.:* Optische Nachrichtentechnik. Verlag Vieweg-Teubner Wiesbaden. 1. Auflage 2003.
27. *Schiffner, G.:* Optische Nachrichtentechnik. Verlag Vieweg-Teubner Wiesbaden. 1. Auflage 2005.
28. *Schumny, H.:* Signalübertragung. Verlag Vieweg Braunschweig/Wiesbaden. 2. Auflage 1987.
29. *Werner, M.:* Information und Codierung. Verlag Vieweg-Teubner Wiesbaden. 2. Auflage 2008.

Signal- und Systemtheorie

XIV



109 Einführung 1327
110 Grundbegriffe 1331
111 Periodische nichtsinusförmige zeitkontinuierliche Signale 1333
112 Nichtperiodische zeitkontinuierliche Signale 1339
113 Spezielle Signale 1355
114 Leistung 1359
115 Faltungsintegral 1363
116 Abtasttheorem 1367
117 Nichtkontinuierliche (zeitdiskrete) Signale 1369
118 Zufällige Signale 1375

Häufig verwendete Formelzeichen

Nähere Erläuterungen dazu finden sich an der angegebenen Stelle im Text

a Verhältnis von Nutz- zu Störsignal allgemein, (118.51)
$\underline{C}_k$ komplexer Fourierkoeffizient, (111.5), (111.7)
$|\underline{C}_k|$ (komplexes) Amplitudenspektrum, (111.8)
E_{s1} Energie des Signals s_1, (118.23)
$\mathbf{F}$ Fouriertransformierte, Abschn. 112.1.1
$\mathbf{F}^{-1}$ inverse Fouriertransformierte, Abschn. 112.1.1
$F(s)$ Laplacetransformierte allgemein, Abschn. 112.2.1
$F_1(s)$ Frequenzabhängige Eingangsgröße des Systems bei der Laplacetransformation, Abb. 109.1
$F_2(s)$ Frequenzabhängige Ausgangsgröße des Systems bei der Laplacetransformation, Abb. 109.1
$f_1(t)$ Zeitabhängige Eingangsgröße des Systems; nur bei der Laplacetransformation wegen Verwechslungsgefahr mit $s_1(t)$ verwendet, Abb. 109.1
$F(j\omega)$ Alternative Schreibweise für $\underline{S}(f)$ in der Literatur, hier nicht verwendet
$H(f)$ Übertragungsfunktion in Abhängigkeit von der Frequenz bei der Fouriertransformation, Abb. 109.1
$H(x)$ Summenhäufigkeitsfunktion, Summenfunktion, Abschn. 118.1
$H(x)^*$ Summenhäufigkeitsfunktion, Summenfunktion, Gaußverteilung mit $\bar{x}$ im Koordinatenursprung, (118.41)
$H(f)_{nT_{ab}}$ Übertragungsfunktion bei abgetasteten Signalen, (117.6)
$H(s)$ Übertragungsfunktion in Abhängigkeit von der Frequenz bei der Laplacetransformation, Abb. 109.1
$H(z)$ Übertragungsfunktion in Abhängigkeit von der Frequenz bei der z-Transformation, Abb. 109.1
$\underline{H}(\omega)$ Alternative Schreibweise für $H(f)$ in der Literatur, hier nicht verwendet
$\underline{H}(j\omega)$ Alternative Schreibweise für $H(f)$ in der Literatur, hier nicht verwendet
$h(t)$ Übertragungsfunktion des Systems in Abhängigkeit von der Zeit, Abb. 109.1
$h(x)$ Wahrscheinlichkeitsdichte(funktion), Häufigkeit, Abschn. 118.1
$h(x)^*$ Wahrscheinlichkeitsdichte(funktion), Häufigkeit, Gaußverteilung mit $\bar{x}$ im Koordinatenursprung, (118.40)
$\mathbf{L}$ Laplacetransformierte, Abschn. 112.2.1
$\mathbf{L}^{-1}$ inverse Laplacetransformierte, Abschn. 112.2.1
$\underline{S}(f)$ Spektralfunktion, Spektraldichte, (110.9)
$S(f)$ alternative Schreibweise für $\underline{S}(f)$ in der Literatur, hier nicht verwendet
$S_1(f)$ Frequenzabhängige Eingangsgröße des Systems bei der Fouriertransformation, Abb. 109.1
$S_1(j\omega)$ alternative Schreibweise für $S_1(f)$ in der Literatur, hier nicht verwendet
$S_2(f)$ Frequenzabhängige Ausgangsgröße des Systems bei der Fouriertransformation, Abb. 109.1
$S_1(z)$ Frequenzabhängige Eingangsgröße des Systems bei der z-Transformation, Abb. 109.1
$S_2(z)$ Frequenzabhängige Ausgangsgröße des Systems bei der z-Transformation, Abb. 109.1
S_{ss} Leistungsspektralfunktion, (110.10)
s Variable bei der Laplacetransformation, Abschn. 112.2.1
$s_1(t)$ Zeitabhängige Eingangsgröße des Systems; bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_1(t)$ bezeichnet, Abb. 109.1
$s_2(t)$ Zeitabhängige Ausgangsgröße des Systems; bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_2(t)$ bezeichnet, Abb. 109.1
$w(x)$ Wahrscheinlichkeit für das Eintreffen des Ereignisses x, (118.1a)
$\bar{x}$ Mittelwert, (118.4), (118.5)
Φ_{SS} Autokorrelationsfunktion, AKF, (118.11)
Φ_{s1s2} Kreuzkorrelationsfunktion der zwei Signale $s_1(t)$ und $s_2(t)$, KKF, (118.21)
Φ_{NN} Kreuzkorrelierte von Rauschen, (118.48a)
$\delta, \delta(t)$ Stoßfunktion, δ-Funktion, Dirac-Impuls, Abschn. 113.1
$\sigma, \sigma(t)$ Sprungfunktion, Abschn. 113.2
σ^2 Varianz, (118.7)

Einführung

109

Wilfried Plaßmann



109.1 Darstellung in der Zeit- und in der Frequenzebene 1328
109.2 Hinweise zur Anwendung 1329

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_109

109.1 Darstellung in der Zeit- und in der Frequenzebene

Abb. 109.1 stellt den Aufgabenbereich der Signal- und Systemtheorie dar. Ein Eingangssignal, das entweder als Funktion der Zeit ($s_1(t)$) oder als Funktion der Frequenz ($\underline{S}_1(f)$, $F_1(s)$, $S_1(z)$) gegeben ist, erfährt durch ein System, dessen Eigenschaften durch $h(t)$ bzw. $\underline{H}(f)$, $H(s)$ oder $H(z)$ gekennzeichnet werden, eine Beeinflussung und erscheint als geändertes Ausgangssignal $s_2(t)$ bzw. $\underline{S}_2(f)$, $F_2(s)$ oder $S_2(z)$. Die Signal- und Systemtheorie erfasst in einem ersten Schritt die Zusammenhänge zwischen den eingezeichneten Größen. Stellt das System z. B. die Übertragungsstrecke für Daten dar, dann sind sehr hohe Anforderungen zu erfüllen: Schnelle und (fast) fehlerfreie Übertragung möglichst vieler unterschiedlicher Daten gleichzeitig, die sich eindeutig wieder trennen lassen müssen, wobei nur eine begrenzte Bandbreite zur Verfügung steht. Deshalb folgt häufig in einem zweiten Schritt die Optimierung aller an der Übertragung beteiligten Größen, d. h., sowohl die Eingangsgrößen als auch die Systemgrößen müssen geeignet gewählt bzw. aneinander angepasst werden. Für die Darstellung im Frequenzbereich werden drei Verfahren angewendet: Die Fouriertransformation mit den Größen $\underline{S}_1(f)$, $\underline{H}(f)$, $\underline{S}_2(f)$, die Laplacetransformation mit $F_1(s)$, $H(s)$, $F_2(s)$ und die z-Transformation mit $S_1(z)$, $H(z)$, $S_2(z)$. Obwohl die Größen in der Frequenzebene komplexe Größen sein können, müssen sie nicht als solche gekennzeichnet werden, siehe (110.6) und (110.7).

Es gelten folgende Bezeichnungen:

- $s_1(t)$, $(f_1(t))$ Zeitabhängige Eingangsgröße des Systems; bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_1(t)$ bezeichnet,
- $h(t)$ Übertragungsfunktion des Systems in Abhängigkeit von der Zeit,
- $s_2(t)$, $(f_2(t))$ Zeitabhängige Ausgangsgröße des Systems; bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_2(t)$ bezeichnet,
- $S_1(f)$, $\underline{S}_1(f)$ Frequenzabhängige Eingangsgröße des Systems, Fouriertransformation,
- $H(f)$, $\underline{H}(f)$ Übertragungsfunktion in Abhängigkeit von der Frequenz, Fouriertransformation,
- $S_2(f)$, $\underline{S}_2(f)$ Frequenzabhängige Ausgangsgröße des Systems, Fouriertransformation,
- $F_1(s)$ Frequenzabhängige Eingangsgröße des Systems, Laplacetransformation,
- $H(s)$ Übertragungsfunktion in Abhängigkeit von der Frequenz, Laplacetransformation,
- $F_2(s)$ Frequenzabhängige Ausgangsgröße des Systems, Laplacetransformation,
- $S_1(z)$ Frequenzabhängige Eingangsgröße des Systems, z-Transformation,
- $H(z)$ Übertragungsfunktion in Abhängigkeit von der Frequenz, z-Transformation,
- $S_2(z)$ Frequenzabhängige Ausgangsgröße des Systems, z-Transformation.

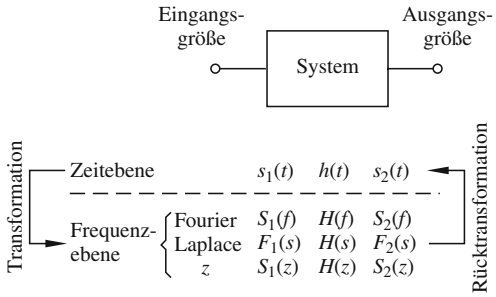


Abb. 109.1 System mit zugeordneten Größen

Nach Abb. 109.1 können die Signale am Ein- und Ausgang sowie die Systemeigenschaften sowohl als Funktion der Zeit (z. B. Spannungsimpuls der Höhe 2 V und der Dauer 2 ms) als auch der Frequenz (z. B. Tiefpass mit der Grenzfrequenz f_g) dargestellt werden. Während $s_1(t)$, $h(t)$ und $s_2(t)$ „anschauliche“ Größen sind, werden die Systemeigenschaften häufig in der Frequenzebene dargestellt. Das hat im wesentlichen zwei Gründe:

1. Zwischen den Ein- und Ausgangsgrößen besteht die einfache Beziehung:

$$\underline{S}_2(f) = \underline{H}(f) \cdot \underline{S}_1(f) \tag{109.1a}$$

bei der Fouriertransformation,

$$F_2(s) = H(s) \cdot F_1(s) \tag{109.1b}$$

bei der Laplacetransformation,

$$S_2(z) = H(z) \cdot S_1(z) \tag{109.1c}$$

bei der z-Transformation.
Die Größen $s_1(t)$, $h(t)$ und $s_2(t)$ dagegen sind über die mathematisch aufwendigere Operation der Faltung verknüpft (siehe Kap. 115):

$$s_2(t) = h(t) * s_1(t) = \int_{-\infty}^{+\infty} h(\tau) \cdot s_1(t-\tau) d\tau. \tag{109.2}$$

$*$ Symbol für „Faltung“.

2. Nahezu jedes System enthält, absichtlich eingesetzt oder vom Aufbau gegeben, die Elemente L und C in einer ihm eigenen spezifischen Kombination. Die Beziehung zwischen der Spannung u und dem Strom i beim Kondensator bzw. bei der Spule wird über je einen Differentialquotienten in der Zeitebene erfasst: $u_L = L \cdot di_L/dt$, $i_C = C \cdot du_C/dt$. Damit führt die Verknüpfung von $s_1(t)$ mit $s_2(t)$ auf eine Differentialgleichung. Für die Berechnung ist es aber i. Allg. günstiger, wenn eine der Gleichungen (109.1a), (109.1b), (109.1c) verwendet werden kann, weil dann, unter Anwendung geeig-

neter Transformationen, keine Differentialgleichung gelöst werden muss und die Signal- und die Systemeigenschaften in getrennten, voneinander unabhängigen Ausdrücken zusammengefasst sind: $\underline{S}_1(f)$ und $\underline{H}(f)$ bzw. $F_1(s)$ und $H(s)$ oder $S_1(z)$ und $H(z)$. Für *sinusförmige* Größen im *eingeschwungenen* Zustand kann die Berechnung analog zu (109.1a), (109.1b) ohne Lösen einer Differentialgleichung in Form der komplexen Rechnung erfolgen. Das führt zu den Impedanzen $\underline{Z}_C = 1/j\omega C$ bzw. $\underline{Z}_L = j\omega L$, die nicht mehr unmittelbar in ihrer Zeit-, sondern in ihrer Frequenzabhängigkeit gegeben sind. Damit liegen auch die Ergebnisse der Berechnung in Abhängigkeit von der Frequenz vor.
Über die Fourier- und die Laplace-Transformation bei zeitkontinuierlichen Signalen bzw. die z-Transformation bei zeitdiskreten Signalen lässt sich eine der komplexen Rechnung analoge Methode auch bei Einschwingvorgängen und nichtsinusförmigen periodischen oder einmaligen Signalverläufen anwenden. Es ergibt sich dann ebenfalls eine Darstellung in Abhängigkeit von der Frequenz.

Für die Darstellung der Signal- und Systemtheorie gilt hier die folgende Zielsetzung: Die gegebenen oder geforderten Größen, die in ihrer Zeitabhängigkeit vorliegen, werden in die Frequenzebene transformiert. Für das System müssen die Eigenschaften ebenfalls in der Frequenzebene vorliegen. Dann kann mit (109.1a), (109.1b), (109.1c) die gesuchte Größe berechnet und anschließend in die Zeitebene zurücktransformiert werden. In der Praxis werden vorwiegend die zeitabhängigen Größen ausgewertet: Bei der Datenübertragung ist entscheidend, in welcher Form die Digitalsignale empfangen werden, denn daraus ergibt sich, ob sie noch fehlerfrei erkannt werden können oder nicht. Auch beim Einsatz von Filtern ist meist nur von Bedeutung, wie sich der zeitliche Signalverlauf ändert. Da sehr umfangreiche Transformationstabellen zur Verfügung stehen, lohnt häufig der „Umweg“ über die Frequenzebene, weil insgesamt die Berechnungen mit weniger Aufwand möglich sind als die direkte Erfassung in der Zeitebene.

Da die Leistungsfähigkeit der Rechner stetig zugenommen hat, wurden in den letzten Jahren Programme entwickelt, die die Systemberechnungen direkt in der Zeitebene durchführen, also $h(t)$, $s_1(t)$ und $s_2(t)$ verwenden.

109.2 Hinweise zur Anwendung

Im folgenden werden Verfahren zur Erfassung von Systemen und Signalen dargestellt. Letztlich bleibt es dem Anwender überlassen, welchen Weg er zur Lösung des gegebenen Problems bevorzugt. Soll eine Transformation verwendet werden, ist diejenige auszuwählen, die prinzipiell geeignet ist und die die geringeren Schwierigkeiten aufweist: Die Fouriertransformation hat eventuell Konvergenzprobleme, die die Laplacetransformation vermeidet. Für abgetastete Signale bietet sich die z-Transformation an. Mit geeigneten Rechner- bzw. Mathematikprogrammen kann auch die direkte Lösung in der Zeitebene mit dem Faltungsintegral (109.2) sinnvoll sein.

Zum besseren Verständnis seien noch die folgenden Bemerkungen angeführt:

1. Bei dem System kann es sich z. B. handeln um
   - ein Energieübertragungssystem (Generator–Transformator–Leitung–Transformator–Verbraucher),
   - ein Nachrichtenübertragungssystem (Fernsehkamera–Sender–Kabel–Empfänger–Bildschirm),
   - ein Datenübertragungssystem PC-Netzwerkkarte–Router–DSL-Leitung–Kontoführungseinrichtung des Geldinstitutes),
   - ein Regelsystem (Konstanthaltung einer vorgegebenen Flughöhe),
   - einen Gleichstromsteller mit angeschlossenem Motor,
   - eine Verstärkerschaltung,
   - eine Filterschaltung (Tiefpass, Bandpass zur Signalselektion),
   - einen Modulator bzw. Demodulator (Rundfunksender, Rundfunkempfänger).
2. Ein System kann eine große Anzahl von Einzelkomponenten enthalten und damit sehr komplex sein, wie in den ersten 3 Beispielen unter Punkt 1, es kann aber auch aus nur wenigen Elementen bestehen und damit überschaubar sein, so wie in den letzten 2 Beispielen unter Punkt 1. Bei komplexen Systemen kann man gegebenenfalls auf Vereinfachungen und Näherungslösungen zurückgreifen, die übersichtlich sind und die die wesentlichen Eigenschaften trotzdem noch erfassen. Die optimale Lösung findet man, falls erforderlich, durch Testreihen mit Parameteränderungen.
3. Bei den im folgenden dargestellten formelmäßigen Zusammenhängen und bei den Transformationsvorschriften stößt man durchaus in mathematische Grenzbereiche vor. Ziel dieses Kapitels ist aber nicht die umfassende theoretische Behandlung der Signal- und Systemtheorie, sondern der Gedanke, das Verständnis für die Zusammenhänge zu fördern und dem Anwender ein Hilfsmittel zur Lösung der ihm gestellten Aufgaben an die Hand zu geben.
4. Die Signal- und Systemtheorie befasst sich auch mit stochastischen, d. h. mit nicht vorhersagbaren Signalen. Sie tut dies, weil zum einen Störsignale (Rauschen) in dieser Form auftreten, zum anderen eine Nachrichten- bzw. Datenübertragung nur in Form von nicht deterministischen Signalen möglich ist, denn nur sie enthalten Information (siehe Teil XIII „Nachrichtentechnik“, Abschn. 92.4).
5. Zur Beurteilung der Eigenschaften des Systems werden spezielle Eingangssignale verwendet. Man kann dann über das gemessene $s_2(t)$ die Größe $h(t)$ bzw. $\underline{H}(f)$ bestimmen. Anschließend lässt sich eine Aussage über den Verlauf von $s_2(t)$ bei beliebigen Funktionen $s_1(t)$ angeben. Zu diesen besonderen Funktionen gehören die Sprungfunktion, der (angenäherte) Dirac-Impuls und (weißes) Rauschen.
6. Der Themenkreis Zweitore (Vierpole) kann sowohl im Teil XIII „Nachrichtentechnik“ als auch in diesem Kapitel behandelt werden. Hier wurde der erste Weg gewählt, weil die Ein- und Ausgangsgrößen sinusförmig sind und nur der eingeschwungene Zustand betrachtet wird.
7. In der Literatur werden unterschiedliche Formelzeichen bzw. Bezeichnungen verwendet, siehe (110.4), (110.5), (110.6), (110.6a), (110.7), (110.8), (110.9).

Grundbegriffe

110

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_110

Für die Begriffe Systemanalyse und -synthese gilt:

1. *Systemanalyse*: Das Eingangssignal ist gegeben, das Ausgangssignal gesucht.
2. *Systemsynthese*: Das Eingangssignal ist gegeben, das Ausgangssignal muss bestimmte Bedingungen erfüllen. Gesucht ist ein geeignetes System.

Für die weiteren Ausführungen wird angenommen, dass das System drei Bedingungen erfüllt:

1. *Stabilität*: Es muss gelten:

$$\lim_{t\to\infty}\{s_2(t)\} = 0 \quad \text{für} \quad \lim_{t\to\infty}\{s_1(t)\} = 0. \tag{110.1}$$

2. *Linearität*:

$$\begin{aligned} &s_{1a}(t) \circ\!\!-\!\!\circ s_{2a}(t), \quad s_{1b}(t) \circ\!\!-\!\!\circ s_{2b}(t); \quad \text{dann folgt} \\ &s_{1a}(t) + s_{1b}(t) \circ\!\!-\!\!\circ s_{2a}(t) + s_{2b}(t) \end{aligned} \tag{110.2}$$

   $\circ\!\!-\!\!\circ$ allgemein für Zuordnung.
3. *Zeitinvarianz*:

$$\begin{aligned} &s_1(t) \circ\!\!-\!\!\circ s_2(t); \quad \text{dann gilt auch} \\ &s_1(t+\tau) \circ\!\!-\!\!\circ s_2(t+\tau) \end{aligned} \tag{110.3}$$

   τ beliebige Zeitverschiebung, $\circ\!\!-\!\!\circ$ allgemein für Zuordnung.

Die folgenden Begriffe und Beziehungen werden in der Signal- und Systemtheorie häufig verwendet (siehe auch DIN 1311).

- *Übertragungsfaktor allgemein*:

$$T = \frac{S_2}{S_1} = |T| \cdot e^{j \arc T} \tag{110.4}$$

  Die Größen S_1 und S_2 sind beliebig, d. h., sie können auch unterschiedliche Dimensionen haben. Soll gekennzeichnet werden, dass es sich um komplexe Größen handelt, kann die folgende Schreibweise gewählt werden:
- *Übertragungsfaktor, wahlweise für komplexe Größen*:

$$\underline{T} = \frac{\underline{S}_2}{\underline{S}_1} = |\underline{T}| \cdot e^{j \arc T} \tag{110.5}$$

  Durch den folgenden Ausdruck wird die Frequenzabhängigkeit des Übertragungsfaktors besonders gekennzeichnet. Er heißt dann Übertragungsfunktion und wird häufig bei Zweitoren (Vierpolen) angewendet:
- *Übertragungsfunktion, Fouriertransformation*:

$$\underline{H}(\omega) = \underline{S}_2(\omega)/\underline{S}_1(\omega) = |\underline{H}(\omega)| \cdot e^{j\cdot\varphi(\omega)} \tag{110.6}$$

  Alternativ werden in der Literatur auch die Schreibweisen

$$\underline{H}(j\omega) = \underline{S}_2(j\omega)/\underline{S}_1(j\omega) \quad \text{bzw.} \quad H(j\omega) = S_2(j\omega)/S_1(j\omega) \tag{110.6a}$$

  oder

$$\underline{H}(f) = \underline{S}_2(f)/\underline{S}_1(f) \quad \text{bzw.} \quad H(f) = S_2(f)/S_1(f) \tag{110.6b}$$

  verwendet. Wird die Laplacetransformation eingesetzt, gilt die folgende Schreibweise:
- *Übertragungsfunktion, Laplacetransformation*:

$$H(s) = \frac{F_2(s)}{F_1(s)} \tag{110.7}$$

  $F_2(s)$ Ausgangsgröße, $F_1(s)$ Eingangsgröße, beide in der symbolischen s-Ebene. Um Verwechslungen zu vermeiden, wird die Schreibweise $F(s)$ anstatt $S(s)$ verwendet.
- *Übertragungsfunktion, z-Transformation*:

$$H(z) = \frac{S_2(z)}{S_1(z)} \tag{110.8}$$

  $S_2(z)$ Ausgangsgröße, $S_1(z)$ Eingangsgröße.
- *Spektralfunktion, Spektraldichte*:

$$S(f) = \int_{-\infty}^{+\infty} s(t) \cdot e^{-j\omega t} dt \tag{110.9}$$

  In der Literatur wird auch $F(j\omega)$ verwendet.
- *Leistungsspektralfunktion*:

$$\lim_{\Delta t\to\infty} \frac{1}{\Delta t} \cdot |S(f)|^2 = S_{ss}(f) \tag{110.10}$$

Periodische nichtsinusförmige zeitkontinuierliche Signale

111

Wilfried Plaßmann



111.1 Reelle und komplexe Fourierreihe 1334
111.2 Beispiele und Anwendungen 1335

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_111

111.1 Reelle und komplexe Fourierreihe

Jeder periodische Signalverlauf $s(t)$ kann nach Fourier dargestellt werden als eine unendliche Summe aus sinus- und cosinusförmigen Signalen, deren Frequenzen ein ganzzahliges Vielfaches der Frequenz des gegebenen Signals sind:

$$s(t) = a_0 + \sum_{k=1}^{\infty}(a_k \cdot \cos k\omega t + b_k \cdot \sin k\omega t) \qquad (111.1)$$

a_0 ist der Gleichanteil (arithmetischer Mittelwert), ω die Kreisfrequenz des gegebenen Signalverlaufes. Für k gilt: $k = 1, 2, 3, \ldots$ a_k und b_k sind die Scheitelwerte der Signale mit der Kreisfrequenz $k \cdot \omega$. Folgende Bezeichnungen werden allgemein verwendet:

- Signal mit der Kreisfrequenz $1 \cdot \omega$:
  Grundschwingung oder erste Harmonische;
- Signal mit der Kreisfrequenz $2 \cdot \omega$:
  Erste Oberschwingung oder zweite Harmonische;
- Signal mit der Kreisfrequenz $3 \cdot \omega$:
  Zweite Oberschwingung oder dritte Harmonische;

usw.

Die Größen a_0, a_k und b_k werden nach folgender Vorschrift berechnet:

$$a_0 = \frac{1}{T}\int_0^T s(t)dt \qquad (111.2a)$$

$$a_k = \frac{2}{T}\int_0^T s(t) \cdot \cos k\omega t dt \qquad (111.2b)$$

$$b_k = \frac{2}{T}\int_0^T s(t) \cdot \sin k\omega t dt \qquad (111.2c)$$

Daneben gibt es noch eine zweite Art der Darstellung für die Fourierreihe:

$$s(t) = \sum_{k=0}^{\infty}\hat{s}_k \cdot \sin(k\omega t + \varphi_k) \qquad (111.3)$$

Die Darstellungsarten nach (111.1) und (111.3) lassen sich ineinander überführen:

$$\hat{s}_k = \sqrt{a_k^2 + b_k^2}; \quad \varphi_k = \arctan\frac{a_k}{b_k} \qquad (111.4)$$

- *Symmetriebedingungen*: Erfüllt der gegebene Signalverlauf $s(t)$ bestimmte Bedingungen, treten nicht alle in (111.1) aufgeführten Frequenzen auf:
  1. Symmetrie: $s(t) = s(-t)$, gerade Funktion,

$$\Rightarrow \quad b_k = 0 \quad \text{für alle } k,$$

$$\Rightarrow \quad a_0 = \frac{2}{T}\int_0^{T/2} s(t)dt, \quad a_k = \frac{4}{T}\int_0^{T/2} s(t) \cdot \cos k\omega t dt$$

2. Symmetrie: $s(t) = -s(-t)$ ungerade Funktion,

$$\Rightarrow \quad a_0 = 0, \quad a_k = 0 \quad \text{für alle } k,$$

$$\Rightarrow \quad b_k = \frac{4}{T}\int_0^{T/2} s(t) \cdot \sin k\omega t dt$$

3. Symmetrie: $s(t + T/2) = -s(t)$,

$$\Rightarrow \quad a_{2k} = 0, \quad b_{2k} = 0 \quad \text{für alle } k,$$

$$\Rightarrow \quad a_{2k+1} = \frac{4}{T}\int_0^{T/2} s(t) \cdot \{\cos(2k+1)\omega t\}dt,$$

$$\Rightarrow \quad b_{2k+1} = \frac{4}{T}\int_0^{T/2} s(t) \cdot \{\sin(2k+1)\omega t\}dt$$

1. und 3. Symmetrie gleichzeitig:

$$\Rightarrow \quad b_k = 0, \quad a_{2k} = 0 \quad \text{für alle } k,$$

$$\Rightarrow \quad a_{2k+1} = \frac{8}{T}\int_0^{T/4} s(t) \cdot \{\cos(2k+1)\omega t\}dt$$

2. und 3. Symmetrie gleichzeitig:

$$\Rightarrow \quad a_0 = 0, \quad a_k = 0, \quad b_{2k} = 0 \quad \text{für alle } k,$$

$$\Rightarrow \quad b_{2k+1} = \frac{8}{T}\int_0^{T/4} s(t) \cdot \{\sin(2k+1)\omega t\}dt$$

Liegen Symmetriebedingungen vor, vereinfacht sich die Berechnung.

- *Sprungstellen*: Besitzt eine Funktion eine Sprungstelle, konvergiert die Fourierreihe gegen den arithmetischen Mittelwert aus links- und rechtsseitigem Grenzwert an dieser Stelle. Das in Abb. 111.2 dargestellte Signal hat zwei Sprungstellen, für $t = 0$ und $t = aT$. Die Fourierreihe konvergiert für beide Sprungstellen gegen den Wert $(0 + U_0)/2 = U_0/2$.
- *Gibbs'sches Phänomen*: Wird eine Funktion $s(t)$ nach (111.1) zusammengesetzt, und enthält die Funktion $s(t)$ Sprungstellen, dann zeigt sich an den Sprungstellen ein geringes Überschwingen, das auch für $k \to \infty$ nicht verschwindet ($\approx 9\,\%$). In Abb. 111.4 ist das zu erkennen. Ursache dafür ist u. a., dass die Funktion an den Sprungstellen nicht definiert ist.

Funktionsverläufe in der Physik bzw. Elektrotechnik sind im gesamten Zeitbereich eindeutig definiert, sodass keine Sprungstellen auftreten und das Gibbs'sche Phänomen nicht wirksam wird.

Mit den Beziehungen

$$\cos\alpha = \frac{e^{j\alpha} + e^{-j\alpha}}{2}, \quad \sin\alpha = \frac{e^{j\alpha} - e^{-j\alpha}}{2j}$$

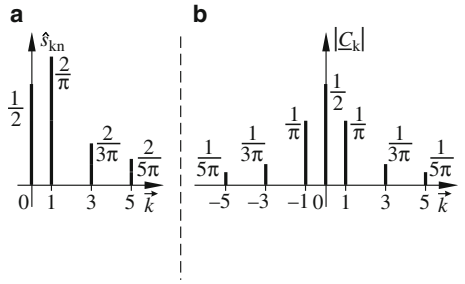


Abb. 111.1 Amplitudenspektrum (siehe Beispiel 1 und Abb. 111.2 für $a = 1/2$) in reeller (**a**) und komplexer Form (**b**)

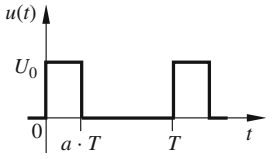


Abb. 111.2 Rechteckförmiger Spannungsverlauf mit veränderlichem Tastgrad a

lässt sich (111.1) wie folgt schreiben:

$$s(t) = a_0 + \sum_{k=1}^{\infty} \left\{ \frac{a_k}{2} \cdot e^{jk\omega t} + \frac{a_k}{2} \cdot e^{-jk\omega t} + \frac{b_k}{2j} \cdot e^{jk\omega t} - \frac{b_k}{2j} \cdot e^{-jk\omega t} \right\}$$

$$= a_0 + \sum_{k=1}^{\infty} \left\{ \underbrace{\left(\frac{a_k}{2} + \frac{b_k}{2j} \right)}_{\underline{C}_k} \cdot e^{jk\omega t} + \underbrace{\left(\frac{a_k}{2} - \frac{b_k}{2j} \right)}_{\underline{C}_{-k}} \cdot e^{-jk\omega t} \right\} \tag{111.5}$$

Es lässt sich zeigen, dass $\underline{C}_{-k}$ durch $\underline{C}_k$ ersetzt werden kann, wenn in der Summe k von $-\infty$ bis $+\infty$ läuft. Dann ergibt sich die Fourierreihe in der komplexen Form:

$$s(t) = \sum_{k=-\infty}^{+\infty} \underline{C}_k \cdot e^{jk\omega t}$$

mit

$$\underline{C}_k = \frac{1}{T} \cdot \int_0^T s(t) \cdot e^{-jk\omega t} dt \tag{111.6}$$

Formal wurden hier negative Frequenzen eingeführt, die physikalisch aber nicht existieren (siehe dazu Abb. 111.1).

In (111.5) kann man schreiben:

$$\underline{C}_k = \frac{a_k}{2} - j\frac{b_k}{2} = |\underline{C}_k| \cdot e^{j\phi_k} \tag{111.7}$$

Daraus leiten sich, verknüpft mit (111.3) und (111.4), zwei für die Anwendung praktische Begriffe ab:

- (*komplexes*) *Amplitudenspektrum*:

$$|\underline{C}_k| = \frac{1}{2} \cdot \sqrt{a_k^2 + b_k^2} = \frac{1}{2} \cdot \hat{s}_k \tag{111.8}$$

  In Abb. 111.1 ist das Amplitudenspektrum in der reellen Form (links) und in der komplexen Form (rechts) für das Beispiel 1 (Abb. 111.2) mit $a = 1/2$ dargestellt. Man gelangt von der komplexen Darstellung zur reellen, wenn die Amplituden der negativen Frequenzen zu den Amplituden der entsprechenden positiven Frequenzen addiert werden. Damit gilt für den Zusammenhang zwischen reellen und komplexen Koeffizienten:

$$a_k = 2 \cdot \mathrm{Re}(\underline{C}_k); \quad b_k = -2 \cdot \mathrm{Im}(\underline{C}_k) \tag{111.9}$$

  Die Amplitudenwerte nach (111.11) (Beispiel 1) sind durch U_0 geteilt. Auf der Abszisse (Abb. 111.1) kann wahlweise die Frequenz, die Kreisfrequenz oder, wie hier, k aufgetragen werden.
- (*komplexes*) Phasenspektrum:

$$\phi_k = \arctan \frac{-b_k}{a_k},$$

  siehe (111.7).
  Andererseits gilt nach (111.4):

$$\varphi_k = \arctan \frac{a_k}{b_k}.$$

Damit lassen sich beide ineinander umrechnen:

$$\phi_k = \arctan \frac{-b_k}{a_k} = -\arctan \frac{b_k}{a_k}$$

$$= -\left(\frac{\pi}{2} - \operatorname{arccot} \frac{b_k}{a_k} \right) = \arctan \frac{a_k}{b_k} - \frac{\pi}{2}$$

oder

$$\phi_k = \varphi_k - \frac{\pi}{2} \tag{111.10}$$

Zu der Darstellung der Spektren (111.8) und (111.10) gehört deshalb die Angabe, ob sie aus der reellen oder der komplexen Form abgeleitet worden sind.

111.2 Beispiele und Anwendungen

Beispiel 1

Soll der Effektivwert einer Gleichspannung verringert werden, wird sie in eine Rechteckspannung mit einstellbarem Tastgrad umgeformt, Abb. 111.2 (Anwendung: Stromrichter, Gleichstromsteller, Helligkeitseinstellung von gepulsten LEDs). Die Gleichspannung wird periodisch ein- und ausgeschaltet. Damit hat dieses Verfahren

Tab. 111.1 Strom-Maximalwerte der Harmonischen im 50-Hz-Stromversorgungsnetz bei Stromverläufen nach Abb. 111.3 (Auszug)

Nummer der Harmonischen	3	5	7	9	11	13	15	...	25	...
Max. Stromwert in A	2,30	1,14	0,77	0,40	0,30	0,21	0,15		0,090	

nur geringe Verluste, weil elektronische Schalter weitgehend ideale Bauelemente sind. Allerdings entstehen gemäß Fourierzerlegung des Signals Wechselspannungen mit Frequenzen nach (111.1), die u. U. als Störsignalquelle unangenehm in Erscheinung treten.

Der Effektivwert (Abb. 111.2) ergibt sich zu:

$$U = \sqrt{\frac{1}{T}\int_0^T \{u(t)\}^2 dt} = \sqrt{\frac{1}{T}\int_0^{aT} U_0^2 dt} = U_0\sqrt{a};$$

Tastgrad: $aT/T = a$; $0 \le a \le 1$.

Für den Gleichanteil a_0 und die Scheitelwerte a_k und b_k folgt mit (111.2a)–(111.2c):

$$a_0 = \frac{1}{T}\int_0^{aT} U_0 dt = U_0 \cdot a$$

$$a_k = \frac{2}{T}\int_0^{aT} U_0 \cdot \cos k\omega t dt = \frac{2U_0}{T}\frac{T}{k2\pi} \cdot \sin k2\pi \cdot \frac{t}{T}\Big|_0^{aT}$$

$$= \frac{U_0}{k\pi} \cdot \sin k \cdot 2 \cdot \pi \cdot a$$

$$b_k = \frac{2}{T}\int_0^{aT} U_0 \cdot \sin k\omega t dt = \frac{2U_0}{T} \cdot \frac{-T}{k2\pi} \cdot \cos k2\pi \frac{t}{T}\Big|_0^{aT}$$

$$= \frac{U_0}{k\pi} \cdot (1 - \cos k \cdot 2 \cdot \pi \cdot a)$$

Die Amplituden der Harmonischen sind abhängig vom Tastgrad. Mit (111.3) und (111.4) ergibt sich:

$$\hat{s}_k = \sqrt{a_k^2 + b_k^2}$$

$$= \frac{U_0}{k\pi}\sqrt{\sin^2(k \cdot 2 \cdot \pi \cdot a) + 1 - 2 \cdot \cos(k \cdot 2 \cdot \pi \cdot a) + \cos^2(k \cdot 2 \cdot \pi \cdot a)}$$

$$= \frac{U_0}{k\pi} \cdot \sqrt{2(1 - \cos k \cdot 2 \cdot \pi \cdot a)}$$

Der Ausdruck unter der Wurzel ist maximal für $\cos k \cdot 2 \cdot \pi \cdot a = -1$, d. h. für $a = 1/2$ und $k = 1, 3, 5, \ldots$:

$$\hat{s}_{k\,\max} = \frac{2 \cdot U_0}{k \cdot \pi} \quad \text{für } a = 1/2 \text{ und } k = 1, 3, 5, \ldots \tag{111.11}$$

Damit sind die Amplituden der Harmonischen für $a = 1/2$ maximal.

Für $k = 2, 4, 6, \ldots$ ist die Amplitude Null, d. h. es entstehen für den Sonderfall $a = 1/2$ keine geradzahligen Harmonischen. ◀

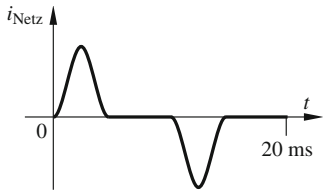


Abb. 111.3 Prinzipieller Stromverlauf bei Geräten mit Brückengleichrichter und Ladekondensator

Beispiel 2

Um die im vorigen Beispiel angesprochenen Störungen ausreichend klein zu halten, ist in IEC 555-2 bzw. VDE 0838 Teil 2 festgelegt worden, dass die Amplituden der Oberschwingungen des Stromes i_{netz} im 50-Hz-Stromversorgungsnetz bestimmte Maximalwerte nicht überschreiten dürfen. Abb. 111.3 stellt den typischen Stromverlauf eines Gerätes mit Zweiweggleichrichter und Ladekondensator dar.

Tab. 111.1 enthält auszugsweise die maximal zulässigen Stromamplituden der Oberschwingungen, bezogen auf eine 16-A-Grundschwingungsamplitude. Infolge der Symmetriebedingung $i(t + T/2) = -i(t)$ (3. Symmetrie, siehe oben) treten nur ungeradzahlige Harmonische auf. ◀

Beispiel 3

Die Grenzfrequenz von Oszilloskopen bezieht sich auf einen sinusförmigen Signalverlauf am Eingang. Werden nichtsinusförmige periodische Signale gemessen, was der Regelfall ist, dann verarbeitet das Gerät nur Harmonische bis zur Grenzfrequenz. Dadurch ist der auf dem Schirm dargestellte Signalverlauf nicht mehr mit dem zu messenden identisch. Nur dadurch, dass die Signalfrequenz wesentlich geringer als die Grenzfrequenz des Oszilloskopes ist (Erfahrungswert $1/10 \ldots 1/20$), kann der Fehler klein gehalten werden. Ein unbekanntes Signal mit der Frequenz 10 MHz erfordert damit ein Oszilloskop mit der Grenzfrequenz 100 ... 200 MHz, soll der Verlauf mit geringem Fehler erfasst werden.

Abb. 111.4 stellt entsprechende Schirmbilder dar. Der verwendete ideale Tiefpass lässt keine Frequenzanteile oberhalb f_g durch, während der reale Tiefpass ab f_g die folgenden Frequenzanteile bis zur angegebenen Frequenz mit 20 dB/Dekade absenkt. Das Überschwingen an den Sprungstellen ist in der Praxis geringer, weil die darzustellenden Signalverläufe keine „echten" Sprungstellen aufweisen (siehe Gibbs'sches Phänomen oben). ◀

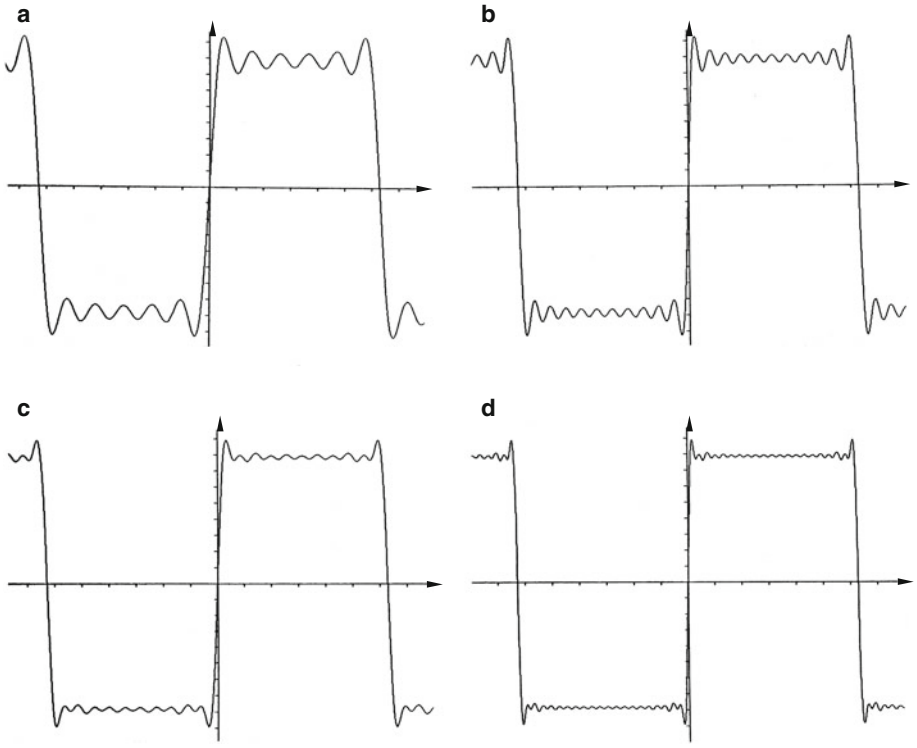


Abb. 111.4 Idealer Rechteckverlauf mit Frequenz f nach Passieren eines idealen (**a**, **b**) bzw. realen (**c**, **d**) Tiefpasses mit der Grenzfrequenz f_g. **a** $f_g = 11f$, **b** $f_g = 21f$, **c** $f_g = 11f$, bis $21f$ Abfall mit 20 dB/Dekade, **d** $f_g = 21f$, bis $41f_g$ Abfall mit 20 dB/Dekade

Beispiel 4

Bei der analogen Stereo-Rundfunkübertragung (siehe Teil XIII „Nachrichtentechnik“, Abschn. 99.1) wird ein Hilfsträger mit 19 kHz übertragen, der im Rundfunkempfänger in einen phasenrichtigen Träger mit 38 kHz umgesetzt werden muss. Das geschieht durch Brückengleichrichtung des sinusförmigen 19-kHz-Trägers. Im neu entstehenden Signal ist die Frequenz 19 kHz nicht mehr enthalten, dafür die benötigte phasenrichtige Frequenz 38 kHz als neue Grundschwingung. Die Fourierreihe berechnet sich zu:

$$u(t) = \frac{2 \cdot \hat{u}}{\pi} - \frac{4 \cdot \hat{u}}{\pi} \sum_k \frac{\cos k\omega t}{k^2 - 1}$$

für $k = 2, 4, 6, \ldots$ und $\omega = 2\pi \cdot 19\,\text{kHz}$.

Die Fourierreihe ist hier noch auf die ursprüngliche Frequenz 19 kHz bezogen. Mit dieser Schaltungstechnik ist die Gewinnung des phasenrichtigen Signals mit der Frequenz 38 kHz ohne großen Aufwand möglich, und die nicht benötigten Oberschwingungen ($k = 4, 6, \ldots$), die u. U. Störungen verursachen könnten, nehmen in ihrer Amplitude durch den Faktor $(k^2 - 1)$ im Nenner relativ stark ab und lassen sich deshalb auf einfache Weise unterdrücken. Die Frequenz 76 kHz tritt nicht auf (Symmetriebedingung), erst 114 kHz ist vorhanden, aber wegen $(k^2 - 1)$ im Nenner mit kleiner Amplitude. ◀

Beispiel 5

Für den Spannungsverlauf nach Abb. 111.2 wird die komplexe Fourierreihe aufgestellt. Nach der Ableitung der reellen Fourierkoeffizienten folgt der Vergleich mit dem Ergebnis aus Beispiel 1.

Für $a = 1/2$ folgt mit (111.6):

$$\begin{aligned}
\underline{C}_k &= \frac{1}{T} \int_0^{T/2} U_0 \cdot e^{-jk\omega t}\, dt \\
&= \frac{U_0}{T} \cdot \frac{e^{-jk\omega t}}{-jk\omega}\bigg|_0^{T/2} \\
&= \frac{U_0 T}{-Tjk2\pi} (e^{-jk\frac{2\pi}{T}\frac{T}{2}} - 1) \\
&= \frac{U_0}{-jk2\pi} (e^{-jk\pi} - 1)
\end{aligned}$$

Mit $e^{-jk\pi} = (-1)^k$ erhält man die komplexe Fourierreihe:

$$\underline{C}_k = \frac{U_0}{jk\pi} \quad \text{für } k = 1, 3, 5, \ldots$$

$$\Rightarrow \quad s(t) = \sum_{k=-\infty}^{+\infty} \frac{U_0}{jk\pi} \cdot e^{jk\omega t}$$

Mit (111.9) lassen sich a_k und b_k der reellen Fourierreihe bestimmen.

$$a_k = 0, \quad \text{da } \underline{C}_k \text{ imaginär.}$$

$$\underline{C}_k = \frac{U_0}{jk\pi} = -j\frac{U_0}{k\pi} \quad \Rightarrow \quad b_k = -2 \cdot \mathrm{Im}(\underline{C}_k) = \frac{2U_0}{k\pi}$$

für $k = 1, 3, 5, \ldots$

Das Ergebnis ist in Übereinstimmung mit (111.11).

Probleme kann es, wie hier, bei der Bestimmung von $\underline{C}_0$ geben. Es ist sinnvoll, in der Berechnungsformel für $\underline{C}_k$ (111.6) im Exponenten der e-Funktion $k = 0$ zu setzen. Dann geht das Integral in die eindeutig lösbare Form nach (111.2a) über, und damit stimmt der Gleichanteil aus der reellen Fourierreihe mit dem aus der komplexen überein. ◀

Beispiel 6

Um ein Maß für die Signalverzerrungen eines Verstärkers zu erhalten, ist der Begriff des Klirrfaktors eingeführt worden. Im Teil XIII „Nachrichtentechnik“, Abschn. 92.4.12.3, sind dazu weitere Einzelheiten dargestellt. Grundlage für das Messverfahren ist das Auftreten von Oberschwingungen, wenn ein sinusförmiges Signal verzerrt wird. Die Amplituden der Oberschwingungen nehmen mit zunehmender Abweichung von der Sinusform, d. h. mit zunehmender Verzerrung, zu und sind damit ein Maß für den Grad der Verzerrung. ◀

Nichtperiodische zeitkontinuierliche Signale

112

Wilfried Plaßmann

Teil XIV



112.1 Fouriertransformation 1340
112.2 Laplacetransformation 1343

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_112

Nichtperiodische Signale haben eine große Bedeutung für die Nachrichten- und Datenübertragung, weil Information nur in nichtdeterministischen Signalen enthalten ist (Teil XIII „Nachrichtentechnik“, Abschn. 92.4.1). Aber auch für die Energie- und Regelungstechnik sind sie von Interesse, weil sie entweder Ein- und Ausschaltvorgänge erfassen oder den Übergang von einem momentan stationären Zustand in einen neuen darstellen (Kurzschluss im Energieversorgungsnetz, Auftreten einer Störgröße im Regelsystem). Die Fourier- und die Laplacetransformation können bei nichtperiodischen zeitkontinuierlichen Signalen eingesetzt werden.

112.1 Fouriertransformation

112.1.1 Transformationsregeln

Für den Übergang von den periodischen zu den nichtperiodischen (aperiodischen, einmaligen) Signalen wird in (111.6)

$$s(t) = \sum_{k=-\infty}^{+\infty} \underline{C}_k \cdot e^{jk\omega t},$$

mit

$$\underline{C}_k = \frac{1}{T} \cdot \int_0^T s(t) \cdot e^{-jk\omega t} dt = \frac{1}{T} \cdot \int_{-T/2}^{+T/2} s(t) \cdot e^{-jk\omega t} dt,$$

der Grenzübergang $T \to \infty$ durchgeführt. Mit $T = 2\pi/\omega$ geht ω entsprechend gegen Null, d. h., das diskrete Spektrum $k\omega$ geht in ein kontinuierliches Spektrum ω über. Verkürzt dargestellt folgt damit:

$$\begin{aligned} s(t) &= \sum_{k=-\infty}^{+\infty} \left\{ \frac{1}{T} \cdot \int_{-T/2}^{+T/2} s(t) \cdot e^{-jk\omega t} dt \right\} \cdot e^{jk\omega t} \\ &= \int_{-\infty}^{+\infty} \underbrace{\left\{ \int_{-\infty}^{+\infty} s(t) \cdot e^{-j\omega t} dt \right\}}_{\underline{S}(f)} \cdot e^{j\omega t} df \end{aligned} \tag{112.1}$$

Die mit $\underline{S}(f)$ bezeichnete Größe ist die *Fouriertransformierte* der Zeitfunktion $s(t)$. Sie existiert nur dann, wenn für das Integral gilt:

$$\int_{-\infty}^{+\infty} |s(t)| dt < \infty.$$

Hinweis: In der Literatur wird anstelle von $\underline{S}(f)$ auch $S(f)$ oder $F(j\omega)$ verwendet. Wird bei der Rücktransformation $F(j\omega)$ eingesetzt, ist der Faktor $1/2\pi$ vor das Integral zu schreiben gemäß $d\omega = 2\pi df$, siehe (112.2b).

Damit erhält man die folgenden zwei Formeln für die Transformation bzw. Rücktransformation:

- *Fouriertransformation, Rücktransformation*:

$$\underline{S}(f) = \int_{-\infty}^{+\infty} s(t) \cdot e^{-j\omega t} dt, \tag{112.2a}$$

symbolische Schreibweise: $\underline{S}(f) \bullet\!\!-\!\!\circ s(t)$.

$$s(t) = \int_{-\infty}^{+\infty} \underline{S}(f) \cdot e^{j2\pi ft} df$$

bzw.

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) \cdot e^{j\omega t} d\omega \tag{112.2b}$$

symbolische Schreibweise: $s(t) \circ\!\!-\!\!\bullet \underline{S}(f)$.
In der Literatur wird auch $S(f)$ und $F(j\omega)$ verwendet, siehe Hinweis oben.
Auch folgende Schreibweise ist üblich:

$$\underline{S}(f) = \mathbf{F}\{s(t)\}; \quad s(t) = \mathbf{F}^{-1}\{\underline{S}(f)\}.$$

$\mathbf{F}$ Fouriertransformierte, $\mathbf{F}^{-1}$ inverse Fouriertransformierte.

- *Spektrum, Spektraldichte*: Die Fouriertransformierte $\underline{S}(f)$ ist eine Funktion der Frequenz und stellt damit das *Spektrum* (*Spektraldichte*) der Funktion $s(t)$ dar.
Die Fouriertransformierte lässt sich auch folgendermaßen darstellen:

$$e^{-j\omega t} = \cos\omega t - j \cdot \sin\omega t \quad \Rightarrow$$

$$\underline{S}(f) = \underbrace{\int_{-\infty}^{+\infty} s(t) \cdot \cos\omega t dt}_{R(f)} + j \cdot \underbrace{(-1) \int_{-\infty}^{+\infty} s(t) \cdot \sin\omega t dt}_{X(f)}$$

$$\underline{S}(f) = R(f) + jX(f) = |\underline{S}(f)| \cdot e^{j\varphi(f)} \tag{112.3}$$

R Realteil, X Imaginärteil von $\underline{S}(f)$

- *Amplitudenspektrum*:

$$|\underline{S}(f)| \quad \text{nach (112.3)} \tag{112.4}$$

- *Phasenspektrum*:

$$\varphi(f) \quad \text{nach (112.3)} \tag{112.5}$$

Es gilt:

$$\begin{aligned} |\underline{S}(f)| &= \sqrt{\{R(f)\}^2 + \{X(f)\}^2}; \\ \varphi(f) &= \arctan \frac{X(f)}{R(f)} \end{aligned} \tag{112.6}$$

$R(f)$ Realteil, $X(f)$ Imaginärteil von $\underline{S}(f)$.

Tab. 112.1 Korrespondenzen zur Fouriertransformation

$s(t)$	$\underline{S}(f)$
$\delta(t)$: Stoßfunktion, δ-Funktion, Dirac-Impuls	1
1	$\delta(f)$ Dirac-Impuls
$\sigma(t)$: Sprungfunktion	$\frac{1}{2}\cdot\delta(f)+\frac{1}{j2\pi\cdot f}$
$\sin\omega_0 t$	$\frac{1}{2j}(\delta(f-f_0)-\delta(f+f_0))$
$\cos\omega_0 t$	$\frac{1}{2}\cdot(\delta(f-f_0)+\delta(f+f_0))$
$\sigma(t)\cdot\sin\omega_0 t$	$\frac{f_0\cdot 2\pi}{(2\pi\cdot f_0)^2-(2\pi\cdot f)^2}+\frac{1}{4j}\cdot(\delta(f-f_0)-\delta(f+f_0))$
$\sigma(t)\cdot\cos\omega_0 t$	$\frac{jf\cdot 2\pi}{(2\pi\cdot f_0)^2-(2\pi\cdot f)^2}+\frac{1}{4}\cdot(\delta(f-f_0)+\delta(f+f_0))$
$\sigma(t)\cdot e^{-t/T}\cdot\sin\omega_0 t$	$T\cdot\frac{2\pi\cdot f_0\cdot T}{(j2\pi\cdot f\cdot T+1)^2+(2\pi\cdot f_0\cdot T)^2}$, $\mathrm{Re}(T)>0$[a]
$\sigma(t)\cdot e^{-t/T}\cdot\cos\omega_0 t$	$T\cdot\frac{j2\pi\cdot f\cdot T+1}{(j2\pi\cdot f\cdot T+1)^2+(2\pi\cdot f_0\cdot T)^2}$, $\mathrm{Re}(T)>0$[a]
$s(t)=\begin{cases}1 & \text{für } \lvert t\rvert<T\\ 0 & \text{für } \lvert t\rvert>T\end{cases}$ Rechteckimpuls	$2\cdot\frac{\sin(2\pi\cdot f\cdot T)}{2\pi\cdot f}$
$s(t)=\begin{cases}1-\frac{\lvert t\rvert}{T} & \text{für } \lvert t\rvert<T\\ 0 & \text{für } \lvert t\rvert>T\end{cases}$ Dreieckimpuls	$\frac{4}{T}\cdot\left\{\frac{\sin(\pi\cdot f\cdot T)}{2\pi\cdot f}\right\}^2$
$s(t)=\begin{cases}\cos^2(\frac{\pi\cdot t}{2\cdot T}) & \text{für } \lvert t\rvert<T\\ 0 & \text{für } \lvert t\rvert>T\end{cases}$ $\cos^2 x$-Impuls	$\frac{\sin 2\pi\cdot f\cdot T}{2\pi\cdot f}\cdot\frac{1}{1-(2\cdot f\cdot T)^2}$
$s(t)=\begin{cases}e^{-t/T} & \text{für } t>0\\ 0 & \text{für } t<0\end{cases}$ Exponentialimpuls	$\frac{T}{1+j2\pi\cdot f\cdot T}$, $\mathrm{Re}(T)>0$[a]

[a] Re Realteil

112.1.2 Eigenschaften

Es folgen einige wichtige Eigenschaften der Fouriertransformation.

1. Multiplikation mit einer Konstanten k:

$$k\cdot s(t)\circ\!\!-\!\!\bullet\, k\cdot\underline{S}(f) \tag{112.7}$$

2. Addition zweier Funktionen:

$$s_1(t)+s_2(t)\circ\!\!-\!\!\bullet\,\underline{S}_1(f)+\underline{S}_2(f) \tag{112.8}$$

3. Zeitverschiebung:

$$s(t)\circ\!\!-\!\!\bullet\,\underline{S}(f)\quad\Rightarrow\quad s(t-t_0)\circ\!\!-\!\!\bullet\,\underline{S}(f)\cdot e^{-j\omega t_0} \tag{112.9}$$

4. Frequenzverschiebung:

$$\underline{S}(f)\bullet\!\!-\!\!\circ\, s(t)\quad\Rightarrow\quad\underline{S}(f-f_0)\bullet\!\!-\!\!\circ\, s(t)\cdot e^{j\omega_0 t} \tag{112.10}$$

5. Faltung im Zeitbereich:

$$\underline{S}_1(f)\cdot\underline{S}_2(f)\bullet\!\!-\!\!\circ\, s_1(t)*s_2(t)=\int_{-\infty}^{+\infty}s_1(\tau)\cdot s_2(t-\tau)d\tau \tag{112.11}$$

$*$ Zeichen für die Operation „Faltung“.

6. Faltung im Frequenzbereich:

$$s_1(t)\cdot s_2(t)\circ\!\!-\!\!\bullet\,\underline{S}_1(f)*\underline{S}_2(f)=\int_{-\infty}^{+\infty}\underline{S}_1(v)\cdot\underline{S}_2(f-v)dv \tag{112.12}$$

$*$ Zeichen für die Operation „Faltung“.

7. Integration im Zeitbereich:

$$\int_{-\infty}^{t}s(\tau)d\tau\circ\!\!-\!\!\bullet\,\frac{\underline{S}(f)}{j2\pi\cdot f} \tag{112.13}$$

8. Differenziation im Zeitbereich:

$$\frac{ds(t)}{dt}\circ\!\!-\!\!\bullet\,(j2\pi\cdot f)\cdot\underline{S}(f) \tag{112.14}$$

In (112.13) und (112.14) wird die Transformierte $\underline{S}(f)\bullet\!\!-\!\!\circ\, s(t)$ verwendet und die Operation Differenzieren durch Multiplikation mit $(j2\pi\cdot f)$ und die Operation Integrieren durch Multiplikation mit $1/(j2\pi\cdot f)$ erfasst.

Parseval'sche Gleichung:

$$\int_{-\infty}^{+\infty}|s(t)|^2dt=\int_{-\infty}^{+\infty}|\underline{S}(f)|^2df \tag{112.15}$$

Mit $s(t)=u(t)$ bzw. $s(t)=i(t)$ steht unter dem Integral, bis auf einen Proportionalitätsfaktor, der Momentanwert der Leistung. Der Proportionalitätsfaktor ist in diesem Fall der Widerstand R: $p=\{i(t)\}^2\cdot R=\{u(t)\}^2/R$. Über den gesamten Zeitbereich integriert ergibt sich die im Signalverlauf enthaltene Energie (Faktor R bzw. $1/R$ eingesetzt, s. o.). Den gleichen Wert erhält man, wenn man die Leistungsbilanz über die im Signal enthaltenen Frequenzanteile bildet.

112.1.3 Korrespondenztabelle

Tab. 112.1 listet einige Korrespondenzen zur Fouriertransformation auf.

Tab. 112.2 Beispiele zur Fouriertransformation: Impulsantwort von idealem Tiefpass, Hochpass und Bandpass

Impulsantwort	$\underline{S}(f)$
$2U_0 \cdot f_g \cdot \frac{\sin 2\pi \cdot f_g \cdot (t-t_0)}{2\pi \cdot f_g \cdot (t-t_0)} = 2U_0 \cdot f_g \cdot \mathrm{si}(2\pi \cdot f_g \cdot \{t-t_0\})$[a]	$\underline{S}(f) = U_0 \cdot e^{-j2\pi \cdot f \cdot t_0}$; $\lvert f\rvert < f_g$ $\underline{S}(f) = 0$; $\lvert f\rvert > f_g$ idealer Tiefpass, Grenzfrequenz f_g
$\delta(t-t_0) - 2U_0 \cdot f_g \cdot \frac{\sin 2\pi \cdot f_g \cdot (t-t_0)}{2\pi \cdot f_g \cdot (t-t_0)} = \delta(t-t_0) - 2U_0 \cdot f_g \cdot \mathrm{si}(2\pi \cdot f_g \cdot (t-t_0))$[a]	$\underline{S}(f) = U_0 \cdot e^{-j2\pi \cdot f \cdot t_0}$; $\lvert f\rvert > f_g$ $\underline{S}(f) = 0$; $\lvert f\rvert < f_g$ idealer Hochpass, Grenzfrequenz f_g
$\frac{2U_0}{\pi \cdot (t-t_0)} \cdot \sin\{\pi \cdot (f_{g2} - f_{g1}) \cdot (t-t_0)\}[\cos\{\pi \cdot (f_{g2} + f_{g1}) \cdot (t-t_0)\}]$	$\underline{S}(f) = U_0 \cdot e^{-j2\pi \cdot f \cdot t_0}$; $f_{g1} < f < f_{g2}$, $-f_{g2} < -f < -f_{g1}$ $\underline{S}(f) = 0$ im übrigen Bereich idealer Bandpass, Grenzfrequenzen f_{g1}, f_{g2}

[a] In der Literatur häufig benutzte Definition: $\mathrm{si}(x) = \sin(x)/x$

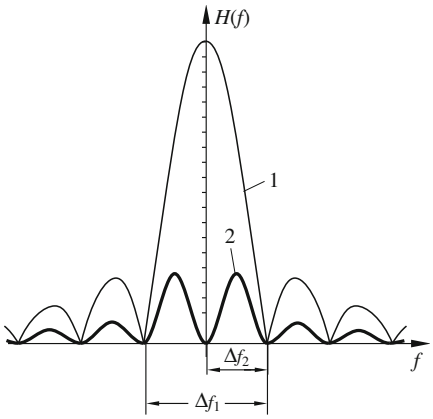


Abb. 112.1 Amplitudenspektrum eines Reckteckimpulses (1) und eines $\sin^2 \omega \cdot T$-Impulses (2)

In Tab. 112.2 sind die Impulsantworten von einem idealen Tiefpass, einem idealen Hochpass und einem idealen Bandpass dargestellt. Für $\underline{S}(f)$ wird ein linearer Phasenverlauf vorausgesetzt. Die Sprungantworten von Tief-, Hoch- und Bandpass beginnen bereits bei $t \to -\infty$, obwohl erst bei $t = 0$ ein Signal am Eingang anliegt. Das System heißt deshalb *nichtkausal*. Es ist technisch nicht realisierbar, weil es im Widerspruch zur physikalischen Wirklichkeit steht. Trotzdem ermöglicht die Auswertung dieses Verhaltens Rückschlüsse auf die realen Eigenschaften des Systems auch bei anderen Eingangssignalen.

- *Unschärferelation*: Multipliziert man die Impulsbreite T eines einmaligen Impulses mit dem Frequenzabstand Δf zweier benachbarter Nulldurchgänge der Spektralfunktion bei $f = 0$ (Abb. 112.1), so gilt die Unschärferelation:

$$\Delta f \cdot T \geq \frac{1}{2\pi} \tag{112.16}$$

Daraus lässt sich folgendes ableiten: 1. Wird bei unveränderter Impulsform nur die Impulsbreite T verkleinert, rücken die Nulldurchgänge der Spektralfunktion weiter auseinander, und damit nehmen die Amplituden zu. 2. Der optimale Wert ergibt sich zu $\Delta f \cdot T = 1/2\pi$, d. h. bei gegebener Impulsbreite sind die Amplituden des Spektrums am kleinsten. Das wird erreicht mit dem Gauß'schen Glockenimpuls:

$$s(t) = \frac{k}{\sqrt{\pi}} \cdot e^{-k^2 \cdot t^2} \tag{112.17}$$

In Abb. 112.1 ist Δf eingezeichnet. Die zugehörige Impulsbreite T kann Abb. 112.7 entnommen werden. Siehe dazu auch die beiden folgenden Beispiele 1 und 2.

112.1.4 Beispiele

Beispiel 1

Berechnet werden soll das Amplitudenspektrum $|\underline{S}(f)|$ und das Phasenspektrum $\varphi(f)$ für einen Rechteckimpuls mit der Amplitude U und der Dauer T. Mit (112.2a) folgt:

$$\begin{aligned}
\underline{S}(f) &= \int_0^T U \cdot e^{-j\omega \cdot t} dt = \frac{U}{j\omega} \cdot (1 - e^{-j\omega \cdot T}) \\
&= U \cdot \frac{1 - \cos(\omega \cdot T) + j \cdot \sin(\omega \cdot T)}{j\omega} \\
&= U \cdot \frac{j \cdot [-1 + \cos(\omega \cdot T)] + \sin[\omega \cdot T]}{\omega} \\
\Rightarrow \quad |\underline{S}(f)| &= \frac{U}{|\omega|} \cdot \sqrt{2 \cdot (1 - \cos(\omega \cdot T))} \qquad (112.18) \\
\Rightarrow \quad \varphi(f) &= \arctan \frac{\cos(\omega \cdot T) - 1}{\sin(\omega \cdot T)} \qquad (112.19)
\end{aligned}$$

◀

Beispiel 2

Gesucht ist das Amplitudenspektrum des Impulses $U \cdot \sin^2(\omega \cdot T)$. Mit $\sin^2(\omega \cdot T) = 0{,}5 \cdot (1 - \cos(2 \cdot \omega \cdot T))$ folgt:

$$\underline{S}(f) = U \cdot \frac{1}{2} \int_{-T/2}^{+T/2} \{1 - \cos(2\omega \cdot T)\} \cdot e^{-j\omega \cdot t} dt$$

$$= \frac{U}{2} \cdot \int_{-T/2}^{+T/2} e^{-j\omega \cdot t} dt - \frac{U}{2} \cdot \int_{-T/2}^{+T/2} \frac{e^{j2\omega \cdot t} + e^{-j2\omega \cdot t}}{2} \cdot e^{-j\omega \cdot t} dt$$

$$= \frac{U}{\omega} \cdot \underbrace{\frac{e^{j\omega \cdot T/2} - e^{-j\omega \cdot T/2}}{2j}}_{\sin(\frac{\omega \cdot T}{2})} - \frac{U}{2} \cdot \int_{-T/2}^{+T/2} \frac{e^{j\omega \cdot t}}{2} dt$$

$$- \frac{U}{2} \cdot \int_{-T/2}^{+T/2} \frac{e^{-j\omega \cdot 3 \cdot t}}{2} dt$$

$$= \frac{U}{\omega} \cdot \sin\left(\frac{\omega \cdot T}{2}\right) - \frac{U}{2\omega} \cdot \underbrace{\frac{e^{j\omega \cdot T/2} - e^{-j\omega \cdot T/2}}{2j}}_{\sin(\frac{\omega \cdot T}{2})}$$

$$- \frac{U}{6\omega} \cdot \underbrace{\frac{e^{j3\omega \cdot T/2} - e^{-j3\omega \cdot T/2}}{2j}}_{\sin(\frac{3\omega \cdot T}{2})} \qquad (112.20)$$

$$= \frac{U}{\omega} \cdot \left\{ \frac{1}{2} \cdot \sin\left(\frac{\omega \cdot T}{2}\right) - \frac{1}{6} \cdot \sin\left(\frac{3\omega \cdot T}{2}\right) \right\} \quad \blacktriangleleft$$

In Abb. 112.1 sind $|\underline{S}(f)|$ aus Beispiel 1 und 2 gleichzeitig dargestellt. Die Amplituden des Spektrums eines $\sin^2 \omega \cdot T$-Impulses sind deutlich geringer als die eines Rechteckimpulses der Dauer T. Besteht bei der gleichzeitigen Übertragung mehrerer digital kodierter Signale die Gefahr der gegenseitigen Beeinflussung, ist der $\sin^2 \omega \cdot T$-Impuls vorteilhafter. Im Beispiel 7, Abschn. 112.2.6 „Beispiele“ zur Laplacetransformation, wird gezeigt, wie ein Rechteck- und ein $\sin^2 \omega \cdot T$-Impuls durch einen Tiefpass im zeitlichen Verlauf geändert werden und in wieweit eine fehlerfreie Signalerkennung noch möglich ist.

112.2 Laplacetransformation

112.2.1 Einführung

Bei der Anwendung der Laplacetransformation treten die bei der Fouriertransformation angesprochenen Probleme mit der Kausalität nicht auf, was im wesentlichen daran liegt, dass gilt:

$$s(t) \equiv 0 \quad \text{für } t \leq 0_- \qquad (112.21)$$

0_- linksseitiger Grenzwert von $s(t)$, siehe Erläuterung weiter unten.

Sie ist damit für die Berechnung der in der Praxis auftretenden Vorgänge besonders geeignet.

- *Verwendete Bezeichnungen*:
  Variable: s (in der Literatur wird auch p benutzt).
  Funktion in der Zeitebene: allgemein $f(t)$; speziell $u(t)$, $i(t)$.
  Transformierte in der s- oder Bildebene: allgemein $\mathbf{L}\{f(t)\} = F(s)$; speziell $\mathbf{L}\{u(t)\} = U(s)$; $\mathbf{L}\{i(t)\} = I(s)$.
  Symbolische Schreibweise: $f(t) \circ\!\!-\!\!\bullet F(s)$.
  Rücktransformation aus der s- in die t-Ebene: $\mathbf{L}^{-1}\{F(s)\} = f(t)$.
  $\mathbf{L}$ Laplacetransformierte, $\mathbf{L}^{-1}$ inverse Laplacetransformierte.

Hier wird bewusst $f(t)$ anstatt $s(t)$ verwendet, um Verwechslungen mit der Variablen s zu vermeiden.

Für die Variable s gilt: $s = a + j\omega$; $a > 0$. Der Realteil wurde eingeführt, um einen endlichen Wert für die Transformationsintegrale zu erhalten (112.22), (112.23a). Damit können für $f(t)$ die in der Praxis vorkommenden Funktionen eingesetzt werden. Auch der Dirac-Impuls $\delta(t)$ und die Sprungfunktion $\sigma(t)$ besitzen eine Laplacetransformierte und lassen sich, obwohl es sich um physikalisch nicht realisierbare Vorgänge handelt, erfolgreich als Näherungslösungen bei der Systemanalyse oder bei Schaltvorgängen einsetzen. Die Einheit von s ist $1/\mathrm{s}$.

112.2.2 Transformation

Die Transformation bzw. Rücktransformation erfolgt durch das Lösen der folgenden Integrale:

Transformation in die Bildebene:

$$F(s) = \int_{0_+}^{\infty} f(t) \cdot e^{-s \cdot t} dt \qquad (112.22)$$

Integrationsgrenze 0_+ siehe „Linksseitiger und rechtsseitiger Grenzwert“ (weiter unten).

Rücktransformation in die Zeitebene:

$$f(t) = \frac{1}{2j\pi} \cdot \int_{\alpha - j\infty}^{\alpha + j\infty} F(s) \cdot e^{s \cdot t} ds \qquad (112.23a)$$

Die Integrationsgrenze 0_+ kennzeichnet den rechtsseitigen Grenzwert der Funktion $f(t)$ für $t = 0$. Das Integral (112.23a) wird zweckmäßigerweise über eine Partialbruchzerlegung und die Anwendung des Residuensatzes gelöst. Dazu zerlegt man die Funktion $F(s)$ in eine Summe, wobei der Nenner der einzelnen Summanden die Form s^k, $(s \pm a_i)^m$ oder $(s^2 \pm b_r^2)^p$, $i, k, m, p, r = 0, 1, 2, \ldots$, hat und $0, a_i, b_r$ die Nullstellen des Nennerpolynoms sind. Der zum jeweiligen Nennerausdruck gehörende Zählerwert kann mit dem Residuensatz bestimmt werden:

$$\operatorname*{Res}_{s=s_0}\{F(s)\} = \frac{1}{(n-1)!} \left\{ \frac{d^{n-1}}{ds^{n-1}} [(s - s_0)^n \cdot F(s)] \right\} \Big|_{s=s_0} \qquad (112.23b)$$

für n-fache Polstelle.

Tab. 112.3 Korrespondenzen zur Laplacetransformation

Nr.	$F(s)$	$f(t \geq 0_+)$ (Hinweis: $f(t \leq 0_-) \equiv 0$
1	1	$\delta(t)$: Stoßfunktion, δ-Funktion, Dirac-Impuls
2	$e^{-a \cdot s}$, $a > 0$	$\delta(t-a)$
3	$\frac{1}{s}$	1, d. h. Sprungfunktion $\sigma(t)$
4	$\frac{1}{s^2}$	t
5	$\frac{1}{s^{n+1}}$, $n = 0, 1, 2, \ldots$	$\frac{t^n}{n!}$
6	$\frac{1}{s+a}$	e^{-at}
7	$\frac{1}{(s+a)^2}$	$t \cdot e^{-at}$
8	$\frac{1}{(s+a)^{n+1}}$	$\frac{t^n \cdot e^{-at}}{n!}$
9	$\frac{1}{(s+a)(s+b)}$, $a \neq b$	$\frac{1}{b-a}(e^{-at} - e^{-bt})$
10	$\frac{1}{(s+a)(s+b)(s+c)}$, $a \neq b \neq c$	$\frac{e^{-a \cdot t}}{(b-a)\cdot(c-a)} + \frac{e^{-b \cdot t}}{(a-b)\cdot(c-b)} + \frac{e^{-c \cdot t}}{(a-c)\cdot(b-c)}$
11	$\frac{1}{s(s+a)}$	$\frac{1}{a} \cdot (1 - e^{-at})$
12	$\frac{1}{s^2(s+a)}$	$\frac{1}{a^2} \cdot (e^{-a \cdot t} + a \cdot t - 1)$
13	$\frac{1}{s(s+a)^2}$	$\frac{1 - e^{-at} - a \cdot t \cdot e^{-at}}{a^2}$
14	$\frac{1}{s^2(s+a)^2}$	$\frac{t}{a^2} - \frac{2}{a^3} + \frac{t \cdot e^{-a \cdot t}}{a^2} + \frac{2 \cdot e^{-a \cdot t}}{a^3}$
15	$\frac{1}{s\,(s+a)(s+b)}$, $a \neq b$	$\frac{1}{ab} + \frac{b \cdot e^{-at} - a \cdot e^{-bt}}{a \cdot b(a-b)}$
16	$\frac{1}{s^2+a^2}$	$\frac{\sin a \cdot t}{a}$
17	$\frac{1}{s^2-a^2}$	$\frac{\sinh a \cdot t}{a}$
18	$\frac{1}{s \cdot (s^2+a^2)}$	$\frac{1 - \cos a \cdot t}{a^2}$
19	$\frac{1}{(s+a)\cdot(s^2+b^2)}$	$\frac{1}{a^2+b^2}[e^{-a \cdot t} + \frac{a}{b} \cdot \sin b \cdot t - \cos b \cdot t]$
20	$\frac{1}{(s^2+a^2)\cdot(s^2+b^2)}$, $a \neq b$	$\frac{1}{b^2-a^2}[\frac{\sin a \cdot t}{a} - \frac{\sin b \cdot t}{b}]$
21	$\frac{1}{(s+a)^2+b^2}$	$\frac{e^{-a \cdot t} \cdot \sin b \cdot t}{b}$
22	$\frac{1}{(s+a)(s+b)^2}$, $a \neq b$	$\frac{e^{-a \cdot t}}{(a-b)^2} + \frac{(a-b) \cdot t - 1}{(a-b)^2} \cdot e^{-b \cdot t}$
23	$U \cdot \frac{e^{-a \cdot s}}{s}$, U Konstante	$U(t-a)$
24	$\frac{s}{s^2+a^2}$	$\cos a \cdot t$
25	$\frac{s}{s^2-a^2}$	$\cosh a \cdot t$
26	$\frac{s}{(s^2+a^2)(s^2+b^2)}$, $a \neq b$	$\frac{\cos a \cdot t - \cos b \cdot t}{b^2-a^2}$
27	$\frac{s}{(s^2+a^2)^2}$	$\frac{t}{2a} \cdot \sin at$
28	$\frac{s+d}{s(s+a)}$	$\frac{d - (d-a) \cdot e^{-a \cdot t}}{a}$
29	$\frac{s+d}{(s+a)(s+b)}$, $a \neq b$	$\frac{(d-a) \cdot e^{-a \cdot t} - (d-b) \cdot e^{-b \cdot t}}{b-a}$
30	$\frac{s+d}{s(s+a)(s+b)}$, $a \neq b$	$\frac{d}{a \cdot b} + \frac{(a-d) \cdot e^{-a \cdot t}}{a \cdot (b-a)} + \frac{(b-d) \cdot e^{-b \cdot t}}{b \cdot (a-b)}$
31	$\frac{s+d}{(s+a)(s^2+b^2)}$	$\frac{(d-a) \cdot e^{-a \cdot t}}{a^2+b^2} + \sqrt{\frac{(b^2+d^2)}{(a^2 \cdot b^2+b^4)}} \cdot \sin(b \cdot t + \Phi)$, $\Phi = \arctan \frac{a}{b} - \arctan \frac{d}{b}$
32	$\frac{s}{(s+a)\cdot(s+b)}$, $a \neq b$	$\frac{a \cdot e^{-a \cdot t}}{a-b} - \frac{b \cdot e^{-b \cdot t}}{a-b}$
33	$\frac{s}{(s+a)^2}$	$(1 - a \cdot t) \cdot e^{-a \cdot t}$

Dabei hat $F(s)$ an der Stelle $s = s_0$ eine n-fache Polstelle. Für $n = 1$ vereinfacht sich (112.23b) zu:

$$\operatorname*{Res}_{s=s_0}\{F(s)\} = (s - s_0) \cdot F(s)\Big|_{s=s_0} \tag{112.23c}$$

für einfache Polstelle.

Die Rücktransformation geschieht dann auf einfache Weise mit den Korrespondenzen Nr. 5, 8, 16, 17, 24, 25 und 27 in Tab. 112.3.

- *Linksseitiger* ($t = 0_-$) *und rechtsseitiger* ($t = 0_+$) *Grenzwert*: Die untere Grenze des Integrals in (112.22) ist mit 0_+ bezeichnet und soll darauf hinweisen, dass für die Funktion $f(t)$ der rechtsseitige Grenzwert für $t = 0$ einzusetzen ist.

Für die in Abb. 111.2 dargestellte Rechteckfunktion gilt: $u(t = 0_-) = 0$, $u(t = 0_+) = U_0$. Da laut Voraussetzung $f(t < 0) \equiv 0$ ist, ergibt sich der linksseitige Grenzwert zu

$$f(t = 0_-) \equiv 0 \tag{112.24}$$

Tab. 112.3 enthält häufig benötigte Transformationen und kann sowohl für die Transformation in die s-Ebene als auch für die Rücktransformation in die Zeitebene benutzt werden. Für die Funktion $f(t)$ gilt: $f(t) = f(t \geq 0_+)$, $f(t \leq 0_-) \equiv 0$. In der Literatur wird manchmal eine alternative Schreibweise gewählt, indem $f(t)$ mit einer Sprungfunktion $s(t)$ zum Zeitpunkt $t = 0$ multipliziert wird. Es sei darauf hingewiesen, dass einige Transformationen aufgenommen wurden, deren Darstellung in der Zeitebene nur für einen begrenzten Zeitbereich möglich ist (Transformationen Nr. 4, 5, 12, 14, 17, 25, 27 mit der Eigenschaft: $f(t) \to \infty$ für $t \to \infty$). Als Beispiel dient ein elektronischer Integrierer mit Operationsverstärker, der nur bis zum Erreichen der Betriebsspannung die Übertragungsfunktion $1/s$ hat.

Durch die Anwendung der Laplacetransformation wird eine Differentialgleichung auf eine algebraische Gleichung zurückgeführt.

112.2.3 Sätze zur Laplacetransformation

Für die Anwendung der Laplacetransformation werden notwendige und nützliche Regeln aufgeführt. Wenn nicht besonders darauf hingewiesen wird, gelten die Gleichheits- und Verknüpfungszeichen in beiden Richtungen.

1. Multiplikation mit einer Konstanten K:

$$K \cdot f(t) \circ\!\!-\!\!\bullet\, K \cdot F(s) \tag{112.25}$$

   Die Begründung liegt darin, dass in den Transformationsgleichungen (112.22) und (112.23a) die Konstante K vor das Integralzeichen gesetzt werden kann.
2. Addition zweier Zeitfunktionen:

$$f_1(t) + f_2(t) \circ\!\!-\!\!\bullet\, F_1(s) + F_2(s)$$

   mit

$$F_1(s) = \mathbf{L}(f_1(t)), \quad F_2(s) = \mathbf{L}(f_2(t)) \tag{112.26}$$

   In einer Summe können die einzelnen Summanden getrennt transformiert und zur Gesamtsumme vorzeichenrichtig addiert werden. Das gilt für die Bild- und die Zeitebene und folgt aus der Regel:

$$\int \{f_1(t) + f_2(t)\}dt = \int f_1(t)dt + \int f_2(t)dt$$

3. Transformierte einer in der Zeitebene abgeleiteten Funktion:

$$\begin{aligned} & \mathbf{L}\{f(t)\} = F(s) \\ \Rightarrow \quad & \mathbf{L}\{\frac{df(t)}{dt}\} = s \cdot \mathbf{L}\{f(t)\} - f(0_+) \\ & = s \cdot F(s) - f(0_+) \end{aligned} \tag{112.27}$$

Es wird die Laplacetransformierte der *nicht abgeleiteten* Funktion in der Zeitebene gebildet und mit s multipliziert, d. h. die Multiplikation mit s kennzeichnet die Differentiation. Die Konstante $f(0_+)$ wurde hier eingeführt, weil es in der Elektrotechnik Funktionsverläufe gibt, die stetig verlaufen müssen. Da die Laplacetransformation aber die Funktionen $f(t)$ für $t \leq 0_-$ zu Null festlegt, muss ein von Null abweichender Wert, der sich nur stetig ändern kann, auf diese Weise berücksichtigt werden (Bezugsgröße in (112.27) ist die nicht abgeleitete Funktion $F(s)$). Diese Konstante hat eine anschauliche Bedeutung, wie noch gezeigt wird (112.37), (112.39).

4. Transformierte der n-ten Ableitung in der Zeitebene:

$$\begin{aligned} \mathbf{L}\left\{\frac{d^n f(t)}{dt^n}\right\} &= s^n \cdot F(s) - s^{n-1} \cdot f(0_+) \\ &\quad - s^{n-2} \cdot \frac{df(t)}{dt}\bigg|_{0_+} - \ldots - \frac{df^{n-1}}{dt^{n-1}}\bigg|_{0_+} \end{aligned} \tag{112.28}$$

5. Transformierte des Integrals in der Zeitebene

$$\begin{aligned} & \mathbf{L}\{f(t)\} \circ\!\!-\!\!\bullet F(s) \\ \Rightarrow \quad & \mathbf{L}\left\{\int_0^t f(t)dt\right\} = \frac{1}{s} \cdot \mathbf{L}\{f(t)\} - \frac{f(0_+)}{s} \end{aligned} \tag{112.29}$$

Die *nicht integrierte* Zeitfunktion wird transformiert und durch s dividiert. Auch hier muss eine Größe $f(0_+)$ aus Kontinuitätsgründen (siehe Erläuterungen zu (112.27) oben) eingeführt werden. Für $f(0_+)/s$ folgt noch eine anschauliche Interpretation (112.37), (112.39).

6. Faltung:

$$F_1(s) \cdot F_2(s) = \mathbf{L}\{f_1(t) * f_2(t)\} = \mathbf{L}\left\{\int_0^t f_1(\tau) \cdot f_2(t-\tau)d\tau\right\} \tag{112.30}$$

$*$ Faltung.

Diese Regel kann nützlich sein, wenn die Rücktransformation von $F(s)$ schwierig oder aufwendig ist, die Zerlegung in ein Produkt $F_1(s) \cdot F_2(s)$ aber möglich ist und deren Rücktransformierten $f_1(t)$ und $f_2(t)$ bekannt sind.

7. Differenzieren in der Bildebene:

$$\begin{aligned} & F(s) \bullet\!\!-\!\!\circ f(t) \\ \Rightarrow \quad & \frac{dF(s)}{ds} = -\mathbf{L}\{t \cdot f(t)\} \end{aligned} \tag{112.31}$$

Differenzieren in der s-Ebene entspricht der Multiplikation von $f(t)$ mit dem Faktor $-t$.

8. Integrieren in der Bildebene:

$$\begin{aligned} & F(s) \bullet\!\!-\!\!\circ f(t) \\ \Rightarrow \quad & \int F(s)ds = \mathbf{L}\left\{\frac{1}{t} \cdot f(t)\right\} \end{aligned} \tag{112.32}$$

Integrieren in der s-Ebene ist gleichbedeutend mit der Division von $f(t)$ durch t.

9. Dämpfung in der Zeitebene:

$$\begin{aligned} & \mathbf{L}\{f(t)\} = F(s) \\ \Rightarrow \quad & \mathbf{L}\{f(t) \cdot e^{-a \cdot t}\} = F(s+a) \end{aligned} \tag{112.33}$$

10. Zeitliche Verschiebung um T_0:

$$\begin{aligned} & \mathbf{L}\{f(t)\} = F(s) \\ \Rightarrow \quad & \mathbf{L}\{f(t-T_0)\} = e^{-T_0 \cdot s} \cdot F(s) \end{aligned} \tag{112.34}$$

Dieser Satz findet Anwendung bei periodischen Funktionen $f(t)$, siehe Beispiel 8.

11. Endwert:

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} s \cdot F(s) \tag{112.35}$$

Voraussetzung ist, dass $s \cdot F(s)$ keine Polstellen im Bereich $\mathrm{Re}(s) \geq 0$ besitzt, d. h. $F(s)$ darf bei $s = 0$ nur eine einfache Polstelle haben. Gilt diese Voraussetzung nicht, ergibt sich bei den aufgeführten Faktoren im Nenner von $F(s)$ für $f(t) \to \infty$ folgendes Verhalten:

a) $1/s^k, k = 2, 3, \ldots \Rightarrow f(t) \to \infty$ für $t \to \infty$ (Transformation Nr. 5, Tab. 112.3)).
b) $(s^2 + a^2)^k, k = 1 \Rightarrow f(t)$ endlich, oszillierend, sin- bzw. cos-förmiger Verlauf (Transformationen Nr. 16 und 24); $k \geq 2 \Rightarrow f(t) \to \infty$ für $t \to \infty$.
c) $(s^2 - a^2)^k, k = 1, 2, \ldots \Rightarrow f(t) \to \infty$ für $t \to \infty$ (Transformationen Nr. 17 und 25).

12. Anfangswert:

$$\lim_{t\to 0} f(t) = \lim_{s\to\infty} s \cdot F(s) \tag{112.36}$$

112.2.4 Anwendung der Laplacetransformation bei bekanntem $H(s)$

Es wird davon ausgegangen, dass die Ausgangsspannung eines Systems mit der Übertragungsfunktion $H(s)$ bei einer vorgegebenen Eingangsspannung berechnet werden soll. Die Übertragungsfunktion $H(s)$ ist bereits bekannt. Dann wird die Eingangsspannung $u_1(t)$ in die s-Ebene transformiert, und zwar entweder durch Lösen des Transformationsintegrals (112.22) oder Auswählen des entsprechenden Ausdruckes aus Tab. 112.3. Über $U_2(s) = U_1(s) \cdot H(s)$ und Rücktransformation in die Zeitebene erhält man $u_2(t)$. Beispiel 5 zeigt diesen Weg.

112.2.5 Bestimmung von $H(s)$

Das System enthält die Elemente R, L, M und C sowie Spannungs- und Stromquellen. Die Übertragungsfunktion $H(s)$ lässt sich auf zwei Arten bestimmen:

112.2.5.1 Bestimmung von $H(s)$ mit Differenzialgleichung

Für das System wird die Differenzialgleichung des Verhältnisses $(u_2/u_1) = f(t)$ in der Zeitebene aufgestellt und mit (112.27) und (112.28) in die s-Ebene transformiert.

112.2.5.2 Bestimmung von $H(s)$ durch direkte Transformation der Einzelelemente

Die im System enthaltenen Komponenten (R_i, L_i, M_i, C_i, u_i und i_i, $i = 1, 2, \ldots$) werden direkt in die s-Ebene übertragen und das Verhältnis $(U_2/U_1) = H(s)$ berechnet. Grundlage dazu sind die Sätze entsprechend (112.25) und (112.26). Daraus folgt, dass ein Widerstand R unverändert in die s-Ebene transformiert wird. Das zeitliche Verhalten eventuell vorhandener Spannungs- und Stromquellen wird über die Transformationsgleichung (112.22) oder mit Tab. 112.3 in die s-Ebene übertragen.

Es muss noch gezeigt werden, wie eine Spule und ein Kondensator in die Bildebene übertragen werden und wie der Strom durch eine Spule zum Zeitpunkt $t = 0$ und die Spannung eines zum Zeitpunkt $t = 0$ aufgeladenen Kondensators erfasst werden. Für gekoppelte Spulen kann im folgenden L durch M ersetzt werden.

Die Beziehung $i_C(t) = C \cdot \frac{du_C(t)}{dt}$ wird transformiert:

$$\mathbf{L}\{i_C(t)\} = \mathbf{L}\left\{C \cdot \frac{du_C(t)}{dt}\right\} = C \cdot \mathbf{L}\left\{\frac{du_C(t)}{dt}\right\}$$
$$\Rightarrow \quad I_C(s) = C \cdot s \cdot \mathbf{L}\{u_C\} - C \cdot f(0_+).$$

Dabei wurden die Bezeichnungen nach Abschn. 112.2.1 und (112.25) und (112.27) verwendet. (Hinweis: Kettenregel $\rightarrow dC/dt = 0$). Die Größe $f(0_+)$ gibt die zum Zeitpunkt $t = 0$ auf dem Kondensator vorhandene Spannung an, die sich nur stetig ändern kann. Damit erhält man die folgende Gleichung:

$$I_C(s) = s \cdot C \cdot U_C(s) - C \cdot u_C(t = 0) \qquad (112.37)$$

Beide Seiten werden durch $s \cdot C$ geteilt und die Gleichung nach $U_C(s)$ aufgelöst:

$$U_C(s) = I_C(s) \cdot \frac{1}{s \cdot C} + \frac{u_C(t = 0)}{s} \qquad (112.38)$$

Diese Gleichung ergibt das in Abb. 112.2a dargestellte Ersatzschaltbild eines auf die Spannung $u_C(t = 0)$ aufgeladenen Kondensators in der s-Ebene. Die Richtung der Spannung $u_C(t = 0)/s$ in der s-Ebene ist gleich der Richtung der Kondensatorspannung $u_C(t = 0)$ in der Zeitebene. Wichtig ist, dass für den physikalisch realen Kondensator in der Zeitebene die eingezeichneten Anschlüsse 1 und 2 gültig sind und damit die „Gesamtspannung“ $U_C(s) \bullet\!\!-\!\!\circ u_C(t)$ nach (112.38) maßgebend ist. Die Unterteilung von $U_C(s)$ in die zwei Summanden $I_C(s) \cdot \frac{1}{s \cdot C}$ und $\frac{u_C(t=0)}{s}$ ergibt sich zunächst rein formal aus (112.37), erleichtert aber die Berechnung, weil beide Größen eine anschauliche Interpretation zulassen. Wie wichtig es ist, dass der physikalisch reale Kondensator in der Zeitebene die besagten Anschlüsse 1 und 2 hat und die zwei Einzelgrößen nur „formale“ (aber praktische) Größen in der s-Ebene sind, sieht man an folgendem Beispiel: Die Spannung $u_C(t = 0)/s$ eines aufgeladenen Kondensators in der s-Ebene bleibt für den Zeitbereich $0_+ \leq t < \infty$ unverändert. Wird ein aufgeladener Kondensator über einen Widerstand entladen, geht die Spannung am Kondensator für $t \rightarrow \infty$ aber gegen Null. Rechnet man $U_C(s)$ nach Gleichung (112.38) zwischen den Anschlüssen 1 und 2 aus,

$$U_C(s) = -\frac{u_C(t = 0)}{s} \cdot \frac{1}{s \cdot C} \cdot \frac{1}{R + 1/s \cdot C} + \frac{u_C(t = 0)}{s},$$

und transformiert in die Zeitebene zurück, erhält man die korrekte Lösung $u_C(t) = u_C(t = 0) \cdot e^{-t/RC}$. Anschaulich kann das in der s-Ebene so erklärt werden, dass sich der Kondensator über den externen (Entlade)-Widerstand auf $-u_C(t = 0)$ auflädt, sodass die von außen über die Anschlüsse 1 und 2 gemessene Summenspannung am physikalisch realen Kondensator für $t \rightarrow \infty$ tatsächlich gegen Null geht. Auch der Strom durch den Widerstand geht rechnerisch für $t \rightarrow \infty$ gegen Null, was in der s-Ebene anschaulich daraus folgt, dass die Differenz der zwei Spannungen, die an den Anschlüssen 1 und 2 gemessen wird und die den Strom durch den Widerstand verursacht, ebenfalls für $t \rightarrow \infty$ gegen Null geht.

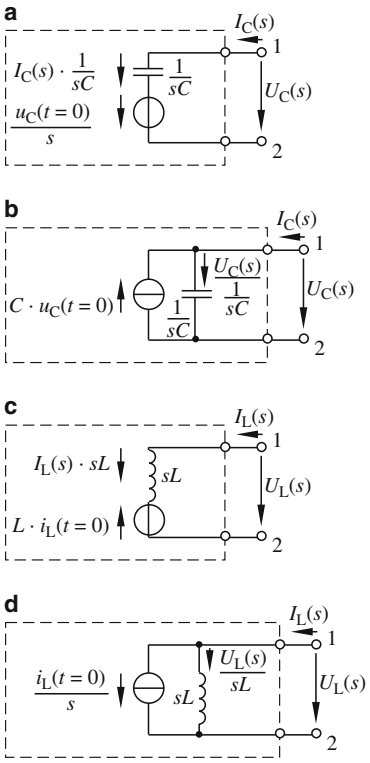


Abb. 112.2 Ersatzschaltbild von Kondensator (**a**, **b**) und Spule (**c**, **d**) in der s-Ebene mit Spannungsquelle (**a**, **c**) oder Stromquelle (**b**, **d**)

Wird (112.37) in folgender Weise umgeformt,

$$\frac{U_C(s)}{\{1/s \cdot C\}} = I_C(s) + C \cdot u_C(t=0),$$

ergibt sich die in Abb. 112.2b gezeigte zweite Möglichkeit für das Ersatzschaltbild eines zum Zeitpunkt $t = 0$ aufgeladenen Kondensators in der s-Ebene. Für den physikalisch realen Kondensator in der Zeitebene sind wiederum die Anschlüsse 1 und 2 maßgebend, d. h. die zwei Teilströme sind lediglich Hilfsgrößen in der s-Ebene, die aber anschaulich interpretiert werden können (siehe Ausführungen im vorigen Absatz).

In gleicher Weise erhält man das Ersatzschaltbild einer Spule in der s-Ebene mit einem zum Zeitpunkt $t = 0$ fließenden Strom. Die Differenzialgleichung für die Spule

$$u_L(t) = L \cdot \frac{di_L(t)}{dt}$$

wird transformiert

$$\mathbf{L}\{u_L(t)\} = \mathbf{L}\left\{L \cdot \frac{di_L(t)}{dt}\right\} = L \cdot \mathbf{L}\left\{\frac{di_L(t)}{dt}\right\}$$
$$\Rightarrow \quad U_L(s) = L \cdot s \cdot \mathbf{L}\{i_L\} - L \cdot f(0_+)$$
$$U_L(s) = L \cdot s \cdot I_L(s) - L \cdot f(0_+)$$

Der Ausdruck $f(0_+)$ stellt den zum Zeitpunkt $t = 0$ durch die Spule fließenden Strom dar. Auch diese Größe muss bei der Transformation der Differenzialgleichung in die s-Ebene berücksichtigt werden, weil der Strom durch eine Spule nur stetig verlaufen kann. Damit ergibt sich:

$$U_L(s) = L \cdot s \cdot I_L(s) - L \cdot i_L(t=0) \tag{112.39}$$

Die Auswertung dieser Gleichung führt zum Ersatzschaltbild einer zum Zeitpunkt $t = 0$ stromdurchflossenen Spule in der s-Ebene nach Abb. 112.2c. Die Polarität der Spannung $L{\cdot}i_L(t=0)$ in der s-Ebene und die Stromflussrichtung von $i_L(t = 0)$ im Zeitbereich sind entgegengesetzt, weil die Spule einen (mathematisch) stetigen Stromfluss erzeugen muss und damit zum Generator wird. Die erforderliche Energie ist im Magnetfeld enthalten. Wird (112.39) umgeformt, erhält man die zweite Möglichkeit eines Ersatzschaltbildes einer zum Zeitpunkt $t = 0$ stromdurchflossenen Spule, Abb. 112.2d.

$$I_L(s) = \frac{U_L(s)}{s \cdot L} + \frac{i_L(t=0)}{s} \tag{112.40}$$

Auch für die Spule gelten die Ausführungen zu den Anschlüssen 1 und 2 des physikalisch realen Kondensators sinngemäß.

Hinweise zur Anwendung

Für die Anwendung der Laplacetransformation nach diesem Verfahren sollen noch einige Hinweise gegeben werden.

1. Die Berechnung der Übertragungsfunktion geschieht mit den gleichen Regeln, mit denen Netzwerkberechnungen durchgeführt werden (Ohm'sches Gesetz, Kirchhoff'sche Regeln 1 und 2, Spannungsteiler, Überlagerungssatz, ...).
2. Die Einheit der „Spannung" in der s-Ebene ist V s, die des „Stromes" A s.
3. Hat ein zurückzutransformierender Ausdruck in der s-Ebene die Eigenschaft „Zählergrad = Nennergrad", kann sich in der Zeitebene ein Dirac-Impuls ergeben. Es sollte geklärt werden, inwieweit diese Lösung realistisch ist. Es kann auch sein, dass sich der Ausdruck durch algebraische Umformung in die Form Zählergrad < Nennergrad umwandeln lässt (Hinweis darauf: Minuszeichen bei einem Summanden im Zähler).
4. Bei der Berechnung einer Schaltung sollte bedacht werden, dass für die Rücktransformation eine Nennerdarstellung der Art

$$\frac{1}{s \cdot s \ldots (s+a) \cdot (s+b) \ldots (s^2 + c^2)}$$

erforderlich ist. Deshalb empfiehlt es sich in der Regel, Klammern *nicht* auszumultiplizieren.

112.2.6 Beispiele

Um die Anwendung der Laplacetransformation zu zeigen, folgen einige Beispiele. Dabei wird durchgehend der im Abschn. 112.2.5.2 aufgezeigte Lösungsweg verwendet. Bei der Berechnung werden elementare Umformungen nicht in Einzelschritten durchgeführt. Für die mathematische Darstellung der gesuchten Größe im Zeitbereich wurde in der Regel berücksichtigt, dass trotz punktweise nicht definierter Eingangssignale die Ausgangsgröße für diese Zeitpunkte eindeutig existiert (Kondensatorspannung bzw. Spulenstrom können sich nur stetig ändern, Abb. 112.3).

Beispiel 3

Anschalten eines Reihenschwingkreises an eine Gleichspannung, Abb. 112.3.

Für die Spannung $u_s(t)$ gilt: $u_s(t < 0) = 0$; $u_s(t > 0) = U_0$, d. h. in der s-Ebene wird sie als Spannung U_0/s (Transformation Nr. 3, Tab. 112.3) eingesetzt. Es wird vorgegeben, dass zum Zeitpunkt $t = 0$ die Spannung am Kondensator und der Strom durch die Spule Null sind, d. h. die in Abb. 112.2a bzw. 112.2c eingetragenen Spannungsquellen haben ebenfalls den Wert Null und sind deshalb nicht eingezeichnet. Die Spannung $U_C(s)$ berechnet sich nach der Spannungsteilerregel zu:

$$\underbrace{\frac{U_C(s)}{U_2(s)}}_{} = \underbrace{\frac{U_0}{s}}_{U_1(s)} \cdot \underbrace{\frac{\frac{1}{sC}}{\frac{1}{sC} + sL + R}}_{H(s)}$$
$$= \frac{U_0}{LC} \cdot \frac{1}{s \cdot \{s^2 + \frac{R}{L}s + \frac{1}{LC}\}}$$

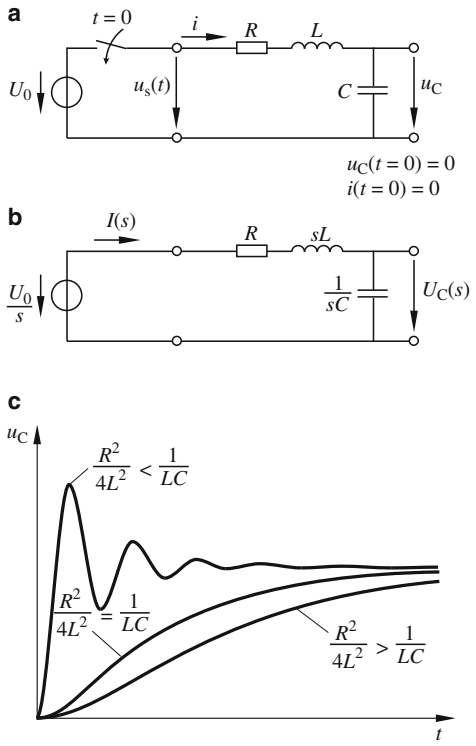


Abb. 112.3 Schaltung zu Beispiel 3. **a** Zeitebene, **b** s-Ebene, **c** Kondensatorspannung

Dabei kann man der Schaltung die Bezeichnungen nach (109.1a), (109.1b) und (109.1c) zuordnen. Um die zugehörige Zeitfunktion zu bekommen, muss der Nenner die Form $s(s+c)(s+d)$ erhalten. Dazu werden die Nullstellen s_{01}, s_{02} des in der geschweiften Klammer enthaltenen Ausdruckes bestimmt:

$$s^2 + \frac{R}{L}s + \frac{1}{LC} = 0$$

$$\Rightarrow \quad s_{01,02} = -\underbrace{\frac{R}{2L}}_{a} \pm \underbrace{\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}}_{b}$$

$$\Rightarrow \quad s_{01} = -a + b; \quad s_{02} = -a - b$$

Es gilt:

$$s^2 + \frac{R}{L}s + \frac{1}{LC} = (s - s_{01})(s - s_{02}).$$

Damit folgt:

$$U_C(s) = \frac{U_0}{LC} \cdot \frac{1}{s(s - \{-a + b\})(s - \{-a - b\})}$$

$$= \frac{U_0}{LC} \cdot \frac{1}{s(s + \underbrace{a - b}_{c})(s + \underbrace{a + b}_{d})}$$

$$= \frac{U_0}{LC} \cdot \frac{1}{s(s + c)(s + d)}; \quad c \neq d$$

1. Fall: $c \neq d$, d. h. $b \neq 0$:

Mit der Transformation Nr. 15 (Tab. 112.3) ergibt sich die gesuchte Größe $u_C(t)$:

$$u_C(t) = \frac{U_0}{LC}\left[\frac{1}{cd} + \frac{d \cdot e^{-ct} - c \cdot e^{-dt}}{cd(c-d)}\right]$$

Wird der Ausdruck cd im Nenner ausgeklammert und durch a und b ersetzt, folgt:

$$u_C(t) = \frac{U_0}{LC} \cdot \frac{1}{a^2 - b^2}\left[1 + \frac{d \cdot e^{-ct} - c \cdot e^{-dt}}{-2b}\right]$$

$$= \frac{U_0}{LC} \cdot \underbrace{\left\{\frac{1}{a^2 - b^2}\right\}}_{LC}$$

$$\cdot \left[1 + \frac{(a+b) \cdot e^{-(a-b)t} - (a-b) \cdot e^{-(a+b)t}}{-2b}\right]$$

$$= U_0\left[1 + \frac{(a+b) \cdot e^{-(a-b)t} - (a-b) \cdot e^{-(a+b)t}}{-2b}\right]$$

$$= U_0\left[1 + e^{-at} \cdot \frac{a \cdot e^{bt} + b \cdot e^{bt} - a \cdot e^{-bt} + b \cdot e^{-bt}}{-2b}\right]$$

$$u_C = U_0\left[1 - e^{-at} \cdot \left\{\frac{a}{b} \cdot \frac{e^{bt} - e^{-bt}}{2} + \frac{e^{bt} + e^{-bt}}{2}\right\}\right] \tag{112.41}$$

mit $a = \frac{R}{2L}$, $b = \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$

Die Werte von R, L und C legen fest, ob der Ausdruck unter der Wurzel positiv oder negativ ist und damit b reell oder imaginär.

a) $\frac{R^2}{4L^2} > \frac{1}{LC} \Rightarrow b$ reell: Es ergibt sich die Lösung nach (112.41). Man kann sie auch mit Hyperbelfunktionen ausdrücken:

$$u_C = U_0\left[1 - e^{-at} \cdot \left\{\frac{a}{b} \cdot \underbrace{\frac{e^{bt} - e^{-bt}}{2}}_{\sinh b \cdot t} + \underbrace{\frac{e^{bt} + e^{-bt}}{2}}_{\cosh b \cdot t}\right\}\right]$$

$$= U_0\left[1 - e^{-at} \cdot \left\{\frac{a}{b} \cdot \sinh bt + \cosh bt\right\}\right]$$

b) $\frac{R^2}{4L^2} < \frac{1}{LC} \Rightarrow b$ imaginär. In (112.41) eingesetzt, folgt:

$$u_C = U_0\left[1 - e^{-at} \cdot \left\{\frac{a}{b} \cdot \underbrace{\frac{e^{jbt} - e^{-jbt}}{2j}}_{\sin b \cdot t} + \underbrace{\frac{e^{jbt} + e^{-jbt}}{2}}_{\cos b \cdot t}\right\}\right]$$

$$= U_0\left[1 - e^{-at} \cdot \left\{\frac{a}{b} \cdot \sin bt + \cos bt\right\}\right]$$

a, b siehe (112.41).

2. Fall: $\frac{R^2}{4L^2} = \frac{1}{LC} \Rightarrow b = 0$

Dann erhält man eine doppelte Nullstelle des Nenners bei $s_{01} = s_{02} = -R/2L$. Daraus folgt für $U_C(s)$:

$$U_C(s) = \frac{U_0}{LC} \cdot \frac{1}{s(s+a)^2} \quad \text{mit } a = \frac{R}{2L}$$

Mit der Korrespondenz Nr. 13 aus Tab. 112.3 und $a = R/2L$ ergibt sich:

$$\begin{aligned} u_C(t) &= \frac{U_0}{LC} \cdot \frac{1 - e^{-a\cdot t} - a \cdot t \cdot e^{-a\cdot t}}{a^2} \\ &= U_0 \cdot \frac{4L}{R^2 C} \cdot \{1 - e^{\frac{-R}{2L}t} - \frac{R}{2L} \cdot t \cdot e^{-\frac{R}{2L}t}\} \end{aligned}$$ ◀

Beispiel 4

Die Reihenschaltung aus Spule und Widerstand wird zu einem beliebigen Zeitpunkt an eine sinusförmige Spannung $u(t)$ geschaltet, Abb. 112.4. Es soll gelten: $i_L(t = 0) = 0$. Der Verlauf des Stromes $i(t)$ ist zu berechnen und daraus abzuleiten, ob es einen günstigen Einschaltzeitpunkt gibt, damit der Momentanwert des Stromes $i(t)$ ab dem Einschaltzeitpunkt so klein wie möglich bleibt.

Die Spannung $u_1(t)$ wird zum (willkürlich gewählten) Zeitpunkt $t = 0$ eingeschaltet und hat sinusförmigen Verlauf mit einem beliebigen Anfangsphasenwinkel:

$$\begin{aligned} u_1(t) &= \hat{u} \cdot (\sin\{\omega \cdot t + \varphi_0\}) \\ &= \hat{u} \cdot \{(\sin\varphi_0)(\cos\omega \cdot t) + (\cos\varphi_0)(\sin\omega \cdot t)\} \end{aligned}$$

Dabei wurde ein Additionstheorem verwendet. Mit den Korrespondenzen Nr. 16 und Nr. 24 aus Tab. 112.3 ergibt sich $U_1(s)$:

$$U_1(s) = \hat{u} \cdot \frac{(\sin\varphi_0) \cdot s + \omega \cdot (\cos\varphi_0)}{s^2 + \omega^2}.$$

Für $I(s)$ folgt:

$$I(s) = \frac{U_1(s)}{R + sL} = \hat{u} \cdot \frac{\sin\varphi_0}{L} \cdot \frac{s + \frac{\omega\cdot(\cos\varphi_0)}{\sin\varphi_0}}{(s + \frac{R}{L})(s^2 + \omega^2)}.$$

Dieser Ausdruck lässt sich unmittelbar mit der Korrespondenz Nr. 31 in die Zeitebene zurücktransformieren:

$$\begin{aligned} i(t) = \hat{u} \cdot \frac{\sin\varphi_0}{L} \cdot \Bigg[& \frac{\{\omega \cdot \frac{\cos\varphi_0}{\sin\varphi_0} - \frac{R}{L}\} \cdot e^{-\frac{Rt}{L}}}{\frac{R^2}{L^2} + \omega^2} \\ & + \sqrt{\frac{\omega^2 + \{\omega \cdot \frac{\cos\varphi_0}{\sin\varphi_0}\}^2}{\frac{R^2}{L^2} \cdot \omega^2 + \omega^4}} \cdot \sin(\omega \cdot t + \Phi) \Bigg] \end{aligned}$$

$$\Phi = \arctan\frac{R}{L \cdot \omega} - \arctan\frac{\cos\varphi_0}{\sin\varphi_0}$$

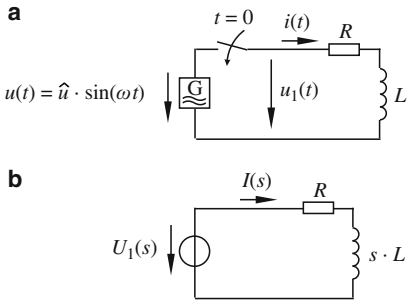


Abb. 112.4 Zu Beispiel 4. **a** Zeitebene, **b** s-Ebene

Der Strom ergibt sich als Summe aus einem periodischen Anteil und einem mit der Zeit abnehmenden Anteil. Der Momentanwert von $i(t)$ ist für $t \geq 0$ dann so klein wie möglich, wenn der zeitlich abnehmende Anteil (erster Summand in der eckigen Klammer) Null ist:

$$\omega \cdot \frac{\cos\varphi_0}{\sin\varphi_0} - \frac{R}{L} = 0$$

oder, umgeformt:

$$\tan\varphi_0 = \frac{\omega \cdot L}{R} \quad \Rightarrow \quad \varphi_0 = \arctan\frac{\omega \cdot L}{R}$$

Bei dieser Dimensionierung gibt es keinen Einschwingvorgang!

Ist $R = 0$, muss der Schalter bei $\varphi_0 = \pm 90°$, d. h. bei dem positiven oder negativen Scheitelwert der Generatorspannung, geschlossen werden. Ist andererseits $L = 0$, ergibt sich der Phasenwinkel zu Null. Für diesen Grenzfall ist es allerdings ohne Bedeutung, wie groß φ_0 gewählt wird, weil der erste Summand in der eckigen Klammer stets Null ist. Man kann übrigens zeigen, dass der Ausdruck oben für $L = 0$ in die Form $i(t) = \frac{\hat{u}\cdot\sin\omega\cdot t}{R}$ übergeht. Das erfordert aber einige Umformungen und die Anwendung von Additionstheoremen. ◀

Beispiel 5

Ein rechteckförmiger Impuls der Dauer T wird auf einen passiven RC-Tiefpass zweiter Ordnung gegeben, Abb. 112.5a. Gesucht ist die Spannung am Ausgang als Funktion der Zeit.

Zunächst werden die Elemente direkt in die s-Ebene transformiert, Abb. 112.5b, und dann die Übertragungsfunktion $U_2(s)/U_1(s) = H(s)$ entweder über eine Netzwerkberechnung oder mit Hilfe der Zweitortheorie ($V_{uf}(s)$ für T-Zweitor mit $1/sC$ als Lastimpedanz, siehe Teil XIII „Nachrichtentechnik“, Kap. 93 „Zweitore“) ermittelt. Sie lautet:

$$H(s) = \frac{1}{R^2C^2s^2 + 3RCs + 1}$$

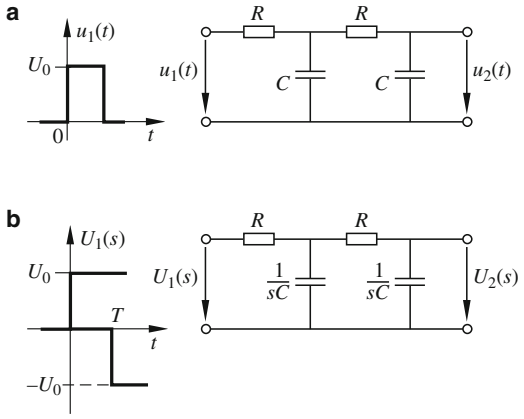


Abb. 112.5 Passiver RC-Tiefpass 2. Ordnung, mit Rechteckimpuls angesteuert. **a** Darstellung in der Zeitebene, **b** Darstellung in der Bildebene

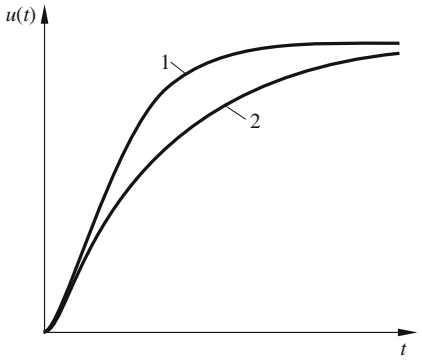


Abb. 112.6 Vergleich der Ausgangsspannung von Tiefpässen zweiter Ordnung bei einer Sprungfunktion am Eingang: 1 Tiefpass mit Operationsverstärker aufgebaut, 2 passiver Tiefpass nach Abb. 112.5

Für $0_+ \le t < T$ liegt die zum Zeitpunkt $t = 0_+$ eingeschaltete Spannung U_0 als U_0/s am Eingang. Dann ergibt sich:

$$U_2(s) = U_1(s) \cdot H(s) = \frac{U_0}{s} \cdot \frac{1}{R^2C^2} \cdot \frac{1}{s^2 + \frac{3}{RC}s + \frac{1}{R^2C^2}}$$

Die Nullstellen des Nenners erhält man durch Lösen der quadratischen Gleichung: $s^2 + \frac{3}{RC}s + \frac{1}{R^2C^2} = 0$.

Es ergibt sich:

$$s_{01,02} = -\frac{1}{RC}\left\{\frac{3}{2} \pm \sqrt{\frac{5}{4}}\right\}$$

$$\Rightarrow \quad s_{01} \approx -\frac{2{,}62}{RC}, \quad s_{02} \approx -\frac{0{,}38}{RC}$$

(numerische Näherung, soll die Übersichtlichkeit erhöhen). Die Rücktransformation mit Korrespondenz Nr. 15 (Tab. 112.3) liefert:

$$u_2(t) \approx \frac{U_0}{R^2C^2} \cdot \frac{1}{\frac{1}{R^2C^2}} \left\{1 + \frac{0{,}38 \cdot e^{-2{,}62 \cdot \frac{t}{RC}} - 2{,}62 \cdot e^{-0{,}38 \cdot \frac{t}{RC}}}{2{,}24}\right\}$$

$$u_2(t) \approx U_0 \cdot \left\{1 + \frac{0{,}38 \cdot e^{-2{,}62 \cdot \frac{t}{RC}} - 2{,}62 \cdot e^{-0{,}38 \cdot \frac{t}{RC}}}{2{,}24}\right\}$$

für $0 \le t \le T$.

Für $t > T$ bleibt die zum Zeitpunkt $t = 0_+$ eingeschaltete Spannung U_0 weiter eingeschaltet, und die Eingangsspannung Null wird für $t > T$ durch Addition der Spannung $-U_0$ zum bestehenden Wert $+U_0$ erzwungen. In der Berechnung ergibt das die Größe $\frac{-U_0}{s} \cdot e^{-Ts}$ (Verschiebungssatz). Das hat den Vorteil, dass zu dem Ergebnis aus der Berechnung $u_2(t)$ für $0_+ \le t \le T$ ein zweiter Term addiert wird, bei dem lediglich t durch $t - T$ ersetzt wird. Dieses Ergebnis ist dann gültig für $t \ge T$.

$$U_2(s) = U_1(s) \cdot H(s)$$

$$= \frac{U_0}{s}(1 - e^{-Ts}) \cdot \frac{1}{R^2C^2} \frac{1}{s^2 + \frac{3}{RC}s + \frac{1}{R^2C^2}},$$

$t \ge T$.

Zurücktransformiert erhält man:

$$u_2(t) \approx U_0 \cdot \left\{1 + \frac{0{,}38 \cdot e^{-2{,}62 \cdot \frac{t}{RC}} - 2{,}62 \cdot e^{-0{,}38 \cdot \frac{t}{RC}}}{2{,}24}\right\}$$

$$- U_0 \cdot \left\{1 + \frac{0{,}38 \cdot e^{-2{,}62 \cdot \frac{t-T}{RC}} - 2{,}62 \cdot e^{-0{,}38 \cdot \frac{t-T}{RC}}}{2{,}24}\right\},$$

$t \ge T$.

Auch hier wurde der Verschiebungssatz angewendet. Das Ergebnis stellt den tatsächlichen Verlauf der Spannung am Ausgang des Tiefpasses dar, ohne dass Idealisierungen oder Näherungen (numerische Näherung ausgenommen, siehe oben) verwendet wurden. Es treten auch keine Kausalitätsprobleme auf. ◀

Beispiel 6

Der mit passiven Elementen realisierte Tiefpass aus Beispiel 5 wird durch einen Tiefpass zweiter Ordnung mit der Übertragungsfunktion $\frac{1}{(1+sRC)^2}$ ersetzt. Eine entsprechende Schaltung kann z. B. mit Operationsverstärkern aufgebaut werden. Die Eingangsspannung hat wiederum rechteckförmigen Verlauf. Für $U_2(s)$ folgt damit:

$$U_2(s) = U_1(s) \cdot H(s) = \frac{U_0}{s} \cdot \frac{1}{R^2C^2} \frac{1}{(s + \frac{1}{RC})^2}$$

Zurücktransformiert mit der Korrespondenz Nr. 13 ergibt sich für $0 \le t \le T$:

$$u_2(t) = \frac{U_0}{R^2C^2} \cdot \frac{1 - e^{-t/RC} - \frac{t}{RC} \cdot e^{-t/RC}}{\frac{1}{R^2C^2}}$$

$$= U_0 \cdot \{1 - e^{-t/RC} - \frac{t}{RC} \cdot e^{-t/RC}\}, \quad t < T$$

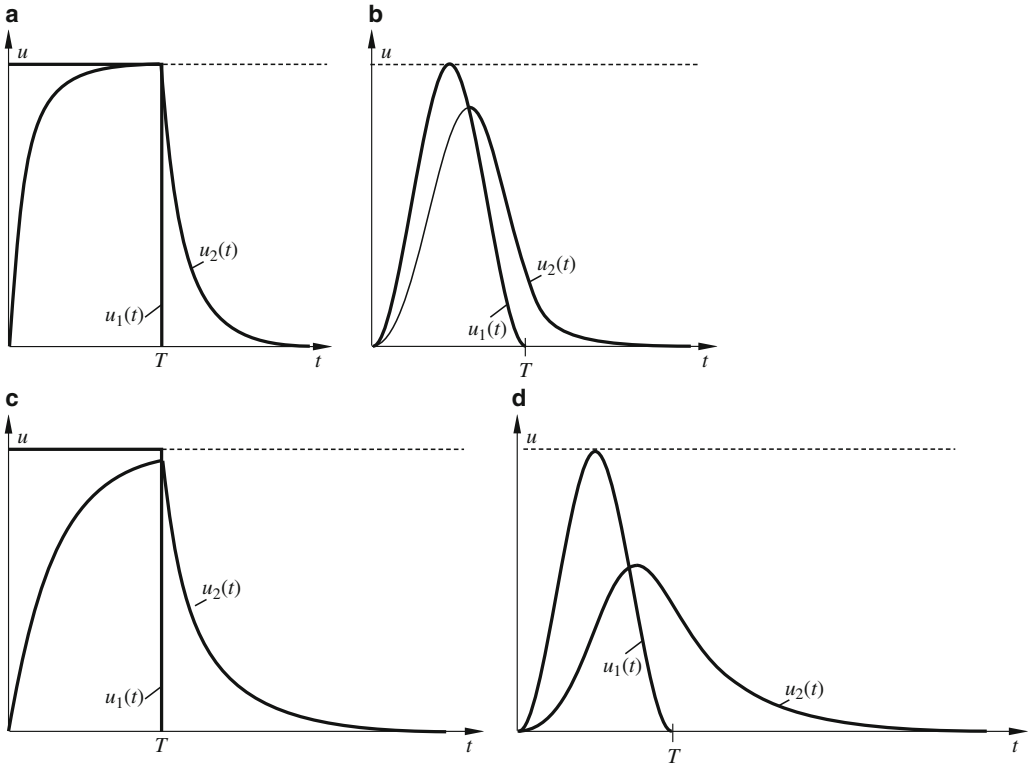


Abb. 112.7 Eingangsspannung $u_1(t)$ und Ausgangsspannung $u_2(t)$ an einem RC-Tiefpass erster Ordnung (Schaltung siehe Abb. 112.8b) mit der Grenzfrequenz f_g. **a** Rechteckimpuls der Dauer T: $T = 1/f_g$, **b** $\sin^2 t\pi/T$-Impuls der Dauer T: $T = 1/f_g$, **c** Rechteckimpuls der Dauer T: $T = 1/(2f_g)$, **d** $\sin^2 t\pi/T$-Impuls der Dauer T: $T = 1/(2f_g)$

Bei dem hier verwendeten Tiefpass steigt die Ausgangsspannung schneller an als beim Tiefpass nach Abb. 112.5, siehe Abb. 112.6.

Der Verlauf für $t > T$ kann entsprechend Beispiel 5 berechnet werden. ◀

Beispiel 7

Ein Rechteckimpuls der Dauer T und ein $\sin^2(\pi \cdot t/T)$-Impuls mit der Amplitude U werden auf einen passiven RC-Tiefpass erster Ordnung gegeben. Die Übertragungsfunktion lautet: $H(s) = \frac{1}{RC} \cdot \frac{1}{s+1/RC}$. Die Eingangsspannung für den Zeitbereich $0_+ \leq t < T$ ergibt sich zu: $U_1(s) = \frac{U}{s}$.

a) Rechteckimpuls: Erster Zeitbereich: $0 \leq t \leq T$:

$$U_2(s) = \frac{U}{s} \cdot \frac{1}{RC} \cdot \frac{1}{s+1/RC}$$

$$\Rightarrow \quad u_2(t \leq T) = \frac{U}{RC} \cdot RC \cdot (1 - e^{-t/RC})$$

$$= U \cdot (1 - e^{-t/RC})$$

Zweiter Zeitbereich: Für $t \geq T$ ergibt sich mit dem Ersatzschaltbild nach Abb. 112.2a und $u_C(t = T) = U(1 - e^{-T/RC})$:

$$U_C(s) = \frac{u_C(t=T)}{s} - I(s) \cdot \frac{1}{sC}$$

$$= \frac{u_C(t=T)}{s} - \frac{1}{sC} \cdot \frac{u_C(t=T)}{s \cdot (R + 1/sC)}$$

$$= \frac{u_C(t=T)}{s} - \frac{u_C(t=T)}{s \cdot (1 + RsC)}$$

$$u_2(t \geq T) = u_C(t=T) - u_C(t=T)\{1 - e^{-(t-T)/RC}\}$$

$$= u_C(t=T) \cdot e^{-(t-T)/RC}$$

Das gleiche Ergebnis erhält man bei der Anwendung des Verschiebungssatzes.

b) $\sin^2(\pi \cdot t/T)$-Impuls: 1. Zeitbereich: $0 \leq t \leq T$

$$u_1(t) = U \cdot \sin^2 \omega \cdot t$$

$$= \frac{U}{2} \cdot (1 - \cos 2\omega \cdot t)$$

$$\Rightarrow\ U_1(s) = \frac{U}{2}\cdot\left\{\frac{1}{s} - \frac{s}{s^2+4\omega^2}\right\}$$

$$U_2(s) = \frac{U}{2}\cdot\left\{\frac{1}{s} - \frac{s}{s^2+4\omega^2}\right\}\cdot\frac{1}{RC}\cdot\frac{1}{s+1/RC}$$

$$= \frac{U}{2}\cdot\frac{1}{RC}\cdot\frac{1}{s}\cdot\frac{1}{s+1/RC} - \frac{U}{2}\cdot\frac{s}{s^2+4\omega^2}\cdot\frac{1}{RC}\cdot\frac{1}{s+1/RC}$$

Zurücktransformiert ergibt sich:

$$u_2(t \le T) = \frac{U}{2}\left\{1 - e^{-t/RC} - \frac{e^{-t/RC}}{1+4\omega^2R^2C^2} + \sqrt{\frac{1}{1+4\omega^2R^2C^2}}\,\sin(2\omega\cdot t + \arctan 1/\omega RC)\right\}$$

Für $t \ge T$ kann der zeitliche Verlauf mit Hilfe des Verschiebungssatzes oder des Ersatzschaltbildes nach Abb. 112.2a berechnet werden. Abb. 112.7 stellt den Verlauf der Eingangsspannung und die zugehörige Ausgangsspannung für die zwei folgenden Dimensionierungen dar: 1. Die „Frequenz“ $f = 1/T$ des Impulses ist gleich der Grenzfrequenz $f_g = 1/(2\pi\cdot RC)$ des Tiefpasses, d. h. $f = f_g$; 2. $f = 2\cdot f_g$.
Der $\sin^2(tp/T)$-Impuls lässt sich, verglichen mit dem Rechteckimpuls, bei der Dimensionierung $f = f_g$ ebenfalls einsetzen, weil die Zustände 1 und 0 im Empfänger ausreichend sicher erkannt werden können. Gilt $f = 2\cdot f_g$, kann er u. U. auch noch verwendet werden, um die im Beispiel 2, Abschn. 112.1.4, aufgeführten Vorteile auszunutzen. ◀

Beispiel 8

Auch periodische Funktionsverläufe lassen sich mit Hilfe des Verschiebungssatzes einfach berechnen. Abb. 112.8a stellt einen rechteckförmigen Verlauf $u(t)$ dar, der zum Zeitpunkt $t = 0$ auf einen RC-Tiefpass erster Ordnung gegeben wird (Abb. 112.8b). Gesucht ist die Spannung $u_C(t)$ zu einem beliebigen Zeitpunkt $t \ge 0$. Die Übertragung in die s-Ebene geschieht nach Abb. 112.8a und Abb. 112.8c: Für $t = 0$ wird die Spannung U_0 eingeschaltet, für $t = aT$ folgt $-U_0$, für $t = T$ wiederum U_0 usw., wobei die einmal eingeschalteten Spannungen dauernd eingeschaltet bleiben. Der Tiefpass hat die Übertragungsfunktion

$$H(s) = \frac{1}{RC}\cdot\frac{1}{s+1/RC}.$$

Für die Funktion $U_1(s)$ folgt mit dem Verschiebungssatz (112.34):

$$U_1(s) = \frac{U_0}{s}\left(1 - e^{-aTs} + e^{-Ts} - e^{-(T+aT)s} + e^{-2Ts} \mp \ldots\right)$$

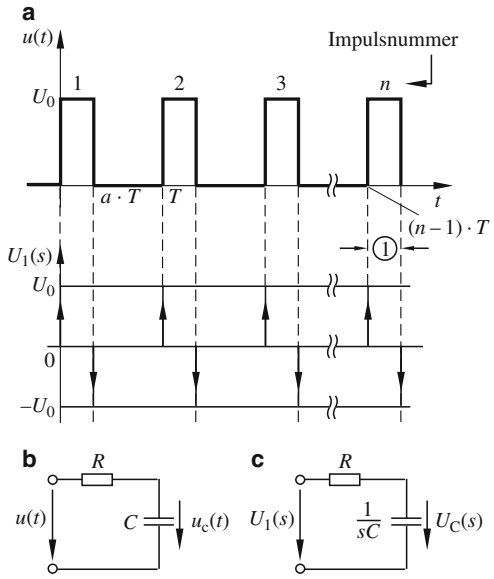


Abb. 112.8 Zu Beispiel 8: Speisung eines RC-Tiefpasses (**b**) mit einer bei $t = 0$ eingeschalteten Rechteckspannung $u(t)$ (**a**, oberer Teil), Darstellung in der s-Ebene (**a**, unterer Teil und **c**)

Daraus ergibt sich $U_C(s)$:

$$U_C(s) = U_1(s)\cdot H(s) = \frac{1}{RC}\cdot\frac{1}{s+1/RC}\cdot\frac{U_0}{s}\cdot\Big(\underbrace{1 - e^{-aTs}}_{\text{1. Impuls}} + \underbrace{e^{-Ts} - e^{-(T+aT)s}}_{\text{2. Impuls}} + e^{-2Ts}\ldots + \ldots + \underbrace{e^{-(n-1)Ts} - e^{-\{(n-1)T+aT\}s}}_{n\text{-ter Impuls}}\Big)$$

Zurücktransformiert in die Zeitebene erhält man mit dem Verschiebungssatz zunächst rein formal:

$$u_C(t) = \underbrace{U_0(1 - e^{-\frac{t}{RC}}) - U_0(1 - e^{-\frac{-aT+t}{RC}})}_{\text{1. Impuls}} + U_0(1 - e^{-\frac{-T+t}{RC}}) - U_0(1 - e^{-\frac{-T-aT+t}{RC}}) \pm \ldots + \underbrace{U_0(1 - e^{-\frac{-(n-1)T+t}{RC}}) - U_0(1 - e^{-\frac{-(n-1)T-aT+t}{RC}})}_{n\text{-ter Impuls}}$$

mit der Bedingung, dass der Ausdruck im Exponenten stets < 0 sein muss.

Soll der Spannungsverlauf des n-ten Impulses im Bereich $(n-1)T \le t \le (n-1)T + aT$ (Bereich ① in Abb. 112.8a) berechnet werden, lässt man zweckmäßigerweise im Punkt $(n-1)T$ die neue Zeit t^* mit dem Wert 0 beginnen. Dadurch erhält man nach einigen Umformun-

gen:

$$u_C(0 \le t^* \le aT) = U_0 + U_0 \cdot e^{-\frac{t^*}{RC}} \left\{ -1 + \left[e^{\frac{aT}{RC}} - 1 \right] \left[e^{-\frac{T}{RC}} + e^{-\frac{2T}{RC}} + \ldots + e^{-\frac{(n-1)T}{RC}} \right] \right\}$$

Für den Bereich $(n-1)T + aT \le t \le nT$ kann ein ähnlicher Ausdruck abgeleitet werden. Hinweis: Wird in obiger Formel zu Kontrollzwecken $a = 0$ gesetzt, heißt das, dass die Spannung erst ab $(n-1)T$ eingeschaltet wird. Damit ergibt sich:

$$u_C(t) = u_C(t^*) = U_0(1 - e^{t^*/RC}).$$

Für $a = 1$ heben sich alle Ausdrücke auf bis auf $u_C(t) = U_0(1 - e^{-\frac{(n-1)T+t}{RC}})$, d. h. es wurde für $t = 0$ die Spannung U_0 einmalig eingeschaltet. Weitere Spannungen werden nicht zugeschaltet, weil sich je eine positive und eine negative Spannung in den Zeitpunkten $T, 2T, 3T, \ldots$ aufheben. ◀

Spezielle Signale

113

Wilfried Plaßmann



113.1 Stoßfunktion, δ-Funktion, Dirac-Impuls 1356
113.2 Sprungfunktion $\sigma(t)$ 1356
113.3 Verknüpfung von σ- und δ-Funktion 1357
113.4 Harmonische Schwingungen 1357

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_113

In der Signal- und Systemtheorie werden spezielle Signalformen eingesetzt, die die Systemeigenschaften für den praktischen Gebrauch besonders vorteilhaft beschreiben bzw. die theoretische Erfassung und Berechnung erleichtern oder erst ermöglichen. Dazu gehören die im folgenden dargestellten Funktionen.

113.1 Stoßfunktion, δ-Funktion, Dirac-Impuls

Eine – ideale – harmonische Schwingung hat ein unendlich schmales Spektrum und eine unendliche Dauer. Ein komplementärer Vorgang weist eine unendlich kurze Dauer und ein unendlich ausgedehntes Spektrum auf. Man kann sich die Entstehung eines solchen Signals so vorstellen, dass z. B. bei einem Gauß'schen Glockenimpuls der Grenzwert $s(t) = \frac{k}{\sqrt{\pi}} e^{-k^2 t^2}\big|_{k\to\infty}$ gebildet wird. Die gesuchte Funktion $\delta(t)$ wird mit Stoßfunktion, δ-Funktion oder Dirac-Impuls bezeichnet und soll die zwei folgenden Bedingungen erfüllen:

1. $\delta(t) \to \infty$ für $t = 0$ und $\delta(t) = 0$ für $t \neq 0$. (113.1a)

2. $$\int_{-\infty}^{+\infty} \delta(t)dt = 1 \qquad (113.1b)$$

Aus der zweiten Bedingung folgt, dass die Fläche endlich ist. Mathematisch handelt es sich bei $\delta(t)$ nicht mehr um eine Funktion, sondern um eine Distribution, weil die oben angesprochenen Grenzwerte nicht existieren. Trotzdem ist die δ-Funktion in der Physik bzw. in der Technik eingeführt worden, weil sie die theoretische Behandlung der Systemanalyse und -synthese erleichtert: Sie tritt in der Regel nicht in singulärer Form auf, sondern in Verbindung mit anderen Funktionen, sodass die Grenzwertprobleme dann u. U. nicht zum Tragen kommen. Außerdem genügt es in der Praxis, den idealen Verlauf in physikalisch realisierbarer Form anzunähern, ohne dass sich die Ergebnisse merklich ändern. Im einzelnen liegt die Bedeutung von $\delta(t)$ in folgendem:

1. In der Beziehung

$$s(t) = \int_{-\infty}^{+\infty} s(\tau) \cdot \delta(t-\tau)d\tau \qquad (113.2)$$

kann $\delta(t-\tau)$ als Auswahlfunktion verstanden werden, die aus den Werten von $s(t)$ genau diejenigen Werte auswählt, für die das Argument Null ist. Dadurch wird die Signalabtastung auch formelmäßig beschreibbar.

2. Für die ideale δ-Funktion haben die Fouriertransformierte (Tab. 112.1) und die Laplacetransformierte den Wert 1 (Tab. 112.3). Wird $\delta(t)$ in Form der physikalisch möglichen Realisierbarkeit als Eingangssignal für ein System benutzt, so folgt mit (109.1a) und (109.1b): $S_2(f) = S_1(f) \cdot H(f) \approx 1 \cdot H(f)$ bzw. $F_2(s) = F_1(s) \cdot H(s) \approx 1 \cdot H(s)$. Man erhält also am Ausgang unmittelbar die Übertragungsfunktion $H(f)$ bzw. $H(s)$ und nicht deren Kombination mit dem Eingangssignal $S_1(f)$ bzw. $F_1(s)$. Abb. 113.1 zeigt die Ausgangsspannung $u_2(t)$ von Tiefpasssystemen erster Ordnung, $U_2/U_1 = 1/(s+a)$, und zweiter Ordnung, $U_2/U_1 = 1/(s+a)^2$, bei Anregung mit einer δ-Funktion.

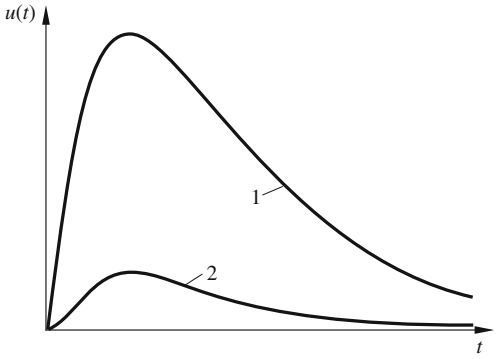


Abb. 113.1 Ausgangsspannung von Tiefpässen (TP) bei Anregung mit einer δ-Funktion: 1 TP erster Ordnung ($1/(s+a)$), 2 TP zweiter Ordnung ($1/(s+a)^2$)

113.2 Sprungfunktion $\sigma(t)$

Für die Sprungfunktion $\sigma(t)$ gilt: $\sigma(t) = 0$ für $t < 0$ und $\sigma(t) = 1$ für $t > 0$. Die δ-Funktion lässt sich als Differentialquotient der Sprungfunktion $\sigma(t)$ schreiben:

$$\delta(t) = \frac{d\sigma(t)}{dt} \qquad (113.3)$$

Duhamel hat gezeigt, dass der zeitliche Verlauf $s(t)$ aus unendlich kleinen Sprüngen $\sigma(t-\tau)ds(\tau)$ aufgebaut werden kann:

Duhamel'sches Integral

$$s(t) = s(0)\sigma(t) + \int_0^t \frac{ds(\tau)}{dt} \cdot \sigma(t-\tau)d\tau \qquad (113.4)$$

Auch wenn die Sprungfunktion für $t = 0$ nicht definiert ist, wird sie erfolgreich in der Technik angewendet:

1. Ein- und Ausschaltvorgänge können theoretisch erfasst und kalkuliert werden, auch wenn in der Praxis der Sprung $\sigma(t)$ nur näherungsweise realisiert werden kann bzw. nur angenähert auftritt.
2. Die Übertragung digital kodierter Signale kann mit Sprungfunktionen simuliert und berechnet werden (siehe Verschiebungssatz (112.34) und Beispiel 7). Für den Anwender lassen sich diese Ergebnisse auf das technisch ausgeführte System übertragen.
3. In der Schaltungs-, Steuerungs- und Regelungstechnik können Sprungfunktionen zur Untersuchung der Systemreaktion und zur Stabilitätskontrolle eingesetzt werden, weil die Änderung der Signalamplitude pro Zeiteinheit theoretisch unendlich groß und in der Praxis sehr groß ist. Damit wird

die für die meisten Systeme kritischste aller auftretenden Signalformen erfasst. Die Systemantwort in der Zeitebene wird auch mit Übergangsfunktion bezeichnet.

113.3 Verknüpfung von σ- und δ-Funktion

Es kann gezeigt werden, dass für die Anwendung in der Elektrotechnik die folgende Beziehung zwischen beiden Signalen gilt:

$$\delta(t) = \frac{d\{\sigma(t)\}}{dt} \quad \text{bzw.} \quad \sigma(t) = \int_{-\infty}^{t} \delta(\tau)d\tau \qquad (113.5)$$

$\delta(t)$ Dirac-Impuls, $\sigma(t)$ Sprungfunktion.

Mit dem Duhamel'schen Integral nach (113.4) und dessen Interpretation lässt sich folgende Aussage ableiten:

Bekannt ist die Impulsantwort $s_2(t) = s_2(t)_\delta$ am Ausgang des Systems, verursacht durch einen Dirac-Impuls $\delta(t)$ am Eingang. Dann berechnet sich die Sprungantwort $s_2(t) = s_2(t)_\sigma$ am Ausgang durch Integration von $s_2(t)_\delta$. Andererseits folgt aus der bekannten Sprungantwort $s_2(t)_\sigma$ durch Differenzieren von $s_2(t)_\sigma$ die entsprechende Impulsantwort $s_2(t)_\delta$:

$$s_2(t)_\sigma = \int_{-\infty}^{t} s_2(\tau)_\delta d\tau \quad \text{bzw.} \quad s_2(t)_\delta = \frac{ds_2(t)_\sigma}{dt} \qquad (113.6)$$

Beispiel 1

Ein passiver RC-Hochpass erster Ordnung hat folgende Spannungs-Übertragungsfunktion in der Zeitebene:

$$h(t) = \delta(t) - \frac{1}{T} \cdot e^{-t/T}$$

Sie kann z. B. dadurch ermittelt werden, dass der Dirac-Impuls $\delta(t)$ als Eingangssignal verwendet wird. Damit gilt auch:

$$s_2(t)_\delta = \delta(t) - \frac{1}{T} \cdot e^{-t/T}$$

$s_2(t)_\delta$ Systemausgangsgröße bei einem Dirac-Impuls am Eingang (Impulsantwort).

Durch Integration ergibt sich:

$$\int_0^t s_2(\tau)_\delta d\tau = \int_0^t \{\delta(\tau) - \frac{1}{T} \cdot e^{-\tau/T}\}d\tau$$

$$= \left[1 + \frac{1}{T}T \cdot e^{-\tau/T}\right]_0^t = e^{-t/T} = s_2(t)_\sigma$$

$s_2(t)_\sigma$ Systemausgangsgröße bei einem Sprung am Eingang (Sprungantwort).

Bei der Berechnung wurde berücksichtigt, dass die Funktionen für $t < 0$ Null sind und dass das Integral des Dirac-Impulses den Wert 1 ergibt. ◄

113.4 Harmonische Schwingungen

Die harmonischen Schwingungen

$$s(t) = \hat{s} \cdot e^{jn\omega_0 \cdot t} = \hat{s} \cdot (\cos n\omega_0 \cdot t + j \cdot \sin n\omega_0 \cdot t),$$

$n = 1, 2, 3, \ldots$, haben mit $m = 1, 2, 3, \ldots$ und $n = 1, 2, 3, \ldots$ die Eigenschaft:

$$\int_0^T (\cos m \cdot \omega_0 \cdot t) \cdot (\cos n \cdot \omega_0 \cdot t)dt = 0 \quad \text{für } m \neq n;$$

$$\int_0^T (\sin m \cdot \omega_0 \cdot t) \cdot (\sin n \cdot \omega_0 \cdot t)dt = 0 \quad \text{für } m \neq n;$$

$$\int_0^T (\sin m \cdot \omega_0 \cdot t) \cdot (\cos n \cdot \omega_0 \cdot t)dt = 0 \quad \text{für alle } m, n. \qquad (113.7)$$

Es handelt sich damit um ein orthogonales Funktionensystem. In der Elektrotechnik gibt es noch weitere orthogonale Systeme, z. B. Walsh-Funktionen, bei denen beliebige periodische Signalverläufe durch Rechteckfunktionen geeigneter Dauer und Amplitude dargestellt werden. Dass die harmonischen Schwingungen eine besondere Bedeutung haben, liegt an ihrem Auftreten in der Physik:

- Projektion von Drehbewegungen auf eine Gerade und damit zusammenhängend die Form der Spannung im Wechsel- bzw. Drehstrom-Versorgungsnetz;
- Spannungs- und Stromverlauf eines verlustfreien elektrischen Schwingkreises aus L und C;
- Erfassung von Systemeigenschaften in der Frequenzebene für sinusförmige Spannungen und Ströme im eingeschwungenen Zustand (z. B. Netzwerke oder Zweitore aus den Elementen R, L, C). Ausgehend von diesen Eigenschaften sind gewisse Rückschlüsse auf das Verhalten bei nichtsinusförmigen Signalen und bei Einschwingvorgängen möglich. Die theoretische Behandlung wird erleichtert durch die Anwendung der komplexen Rechnung. Die sinusförmigen Größen gewinnen noch dadurch an Bedeutung, dass periodische nichtsinusförmige Funktionsverläufe durch die Fourierreihe und damit durch Sinusfunktionen dargestellt werden können. Siehe dazu auch Teil XIII „Nachrichtentechnik", Kap. 93, „Zweitore, Vierpole".

Leistung

114

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_114

Die Erfassung der am Ausgang eines Systems zur Verfügung stehenden Wirkleistung P ist nicht nur für die Energietechnik von Interesse. Je nach Art der Signalform von Strom und Spannung können verschiedene Berechnungsverfahren eingesetzt werden.

1. Spannung und Strom periodisch, sinusförmig

Für die Wirkleistung gilt:

$$P = U \cdot I \cdot \cos\varphi \tag{114.1}$$

U, I Effektivwerte von Spannung $\underline{U}$ und Strom $\underline{I}$ in V bzw. A, φ Phasenwinkel zwischen $\underline{U}$ und $\underline{I}$.

Liegen die Systemeigenschaften in Form von Zweitorgrößen vor, können letztere auch zur Bestimmung des Verhältnisses von Ausgangs- zu Eingangsleistung herangezogen werden. Nähere Einzelheiten zu den Zweitorgrößen sind im Teil XIII „Nachrichtentechnik“, Kap. 93 „Zweitore, Vierpole“ aufgeführt. Das Leistungsverhältnis v_P ist folgendermaßen definiert:

$$v_P = \frac{P_{\text{out}}}{P_{\text{in}}} \tag{114.2}$$

P_{out} Leistung am Ausgang, P_{in} Leistung am Eingang des Systems (Zweitors) in W.

Weiterhin gilt:

$$P_{\text{in}} = U_1^2 \cdot G_{\text{in}} = I_1^2 \cdot R_{\text{in}}; \quad P_{\text{out}} = U_2^2 \cdot G_a = I_2^2 \cdot R_a \tag{114.3}$$

U_1 Eingangsspannung und U_2 Ausgangsspannung des Systems (Zweitors) in V, I_1 Eingangsstrom und I_2 Ausgangsstrom des Systems (Zweitors) in A, G_{in} Realteil der Eingangsadmittanz $\underline{Y}_{\text{in}} = G_{\text{in}} + jB_{\text{in}}$ des Systems in $1/\Omega$, G_a Realteil der Lastadmittanz $\underline{Y}_a = G_a + jB_a$ in $1/\Omega$, R_{in} Realteil der Eingangsimpedanz $\underline{Z}_{\text{in}} = R_{\text{in}} + jX_{\text{in}}$ des Systems in Ω, R_a Realteil der Lastimpedanz $\underline{Z}_a = R_a + jX_a$ in Ω.

Daraus folgt das Leistungsverhältnis:

$$v_P = |\underline{v}_{\text{uf}}|^2 \cdot \frac{G_a}{G_{\text{in}}} = |\underline{v}_{\text{if}}|^2 \cdot \frac{R_a}{R_{\text{in}}} \tag{114.4}$$

$\underline{v}_{\text{uf}}$, $\underline{v}_{if}$ siehe Teil XIII „Nachrichtentechnik“, Kap. 93 „Zweitore, Vierpole“, Tab. 93.6 und 93.7; weitere Formelzeichen siehe (114.3).

Sind $\underline{Z}_{\text{in}}$ und $\underline{Z}_a$ reell, ergibt sich:

$$v_P = |\underline{v}_{\text{uf}}| \cdot |\underline{v}_{\text{if}}|$$

Stehen die $\underline{A}$-Parameter des Systems (Zweitors) zur Verfügung, gilt:

$$v_P = \frac{G_a}{\text{Re}\{(\underline{A}_{21} + \underline{A}_{22} \cdot \underline{Y}_a) \cdot (\underline{A}_{11}^* + \underline{A}_{12}^* \cdot \underline{Y}_a^*)\}} \tag{114.5}$$

Formelzeichen siehe (114.3), * konjugiert komplexe Größe, Re{...} Realteil des Klammerausdruckes.

2. Spannung und Strom periodisch, nicht sinusförmig

Es gibt zwei Möglichkeiten zur Leistungsberechnung.

2.1: Die gegebenen Signalverläufe von Spannung und Strom werden in eine Fourierreihe entwickelt (111.1):

$$u(t) = U_0 + \sum_{k=1}^{\infty} \{\hat{u}_{ka} \cdot \cos k\omega \cdot t + \hat{u}_{kb} \cdot \sin k\omega \cdot t\}$$

$$i(t) = I_0 + \sum_{k=1}^{\infty} \{\hat{i}_{ka} \cdot \cos k\omega \cdot t + \hat{i}_{kb} \cdot \sin k\omega \cdot t\}$$

Werden nach der Produktbildung aus beiden Reihen die Werte von $\hat{u}_{ka}$, $\hat{u}_{kb}$, $\hat{i}_{ka}$, $\hat{i}_{kb}$ bestimmt, ergeben alle diejenigen Integrale den Wert Null, die der Orthogonalitätsbeziehung nach (113.7) unterliegen. Damit folgt für jede Harmonische die Leistungsberechnung gemäß (114.1). Für die Gesamtwirkleistung müssen die Einzelleistungen aller Harmonischen einschließlich des Gleichanteiles addiert werden.

$$P = U_0 \cdot I_0 + U_1 \cdot I_1 \cdot \cos\varphi_1 + U_2 \cdot I_2 \cdot \cos\varphi_2 + \ldots$$

$$= U_0 \cdot I_0 + \sum_{k=1}^{\infty} U_k \cdot I_k \cdot \cos\varphi_k \tag{114.6}$$

Sind U und I über einen ohmschen Widerstand R verknüpft und haben damit den gleichen Kurvenformverlauf, kann die Beziehung $U_k = I_k \cdot R$ benutzt werden.

2.2: Zur Berechnung wird die folgende Formel verwendet:

$$P = \frac{1}{T} \cdot \int_0^T u(t) \cdot i(t) dt \tag{114.7}$$

Praktischer Hinweis zum Einsatz von (114.6) und (114.7): Die Erfahrung zeigt, dass die Anwendung von (114.6) nur dann zu empfehlen ist, wenn die Fourierreihenentwicklung für den gegebenen Kurvenformverlauf bereits vorliegt und damit die Effektivwerte U_k, I_k bekannt sind und die Fourierreihe nur aus wenigen Summanden besteht. Ein derartiger Verlauf ergibt sich beispielsweise, wenn ein rechteckförmiger Signalverlauf mit der Frequenz f ein ideales Tiefpassfilter mit der Grenzfrequenz $3f$ passiert und die Leistung am Ausgang des Filters bestimmt werden soll.

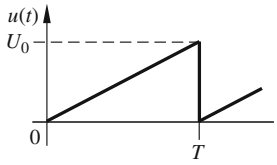


Abb. 114.1 Spannungsverlauf zu Beispiel 1

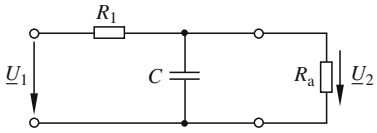


Abb. 114.2 System (Zweitor) zu Beispiel 2

Beispiel 1

Eine sägezahnförmige Spannung $u(t) = U_0 \cdot t/T$ speist einen Ohm'schen Widerstand R, Abb. 114.1. Gesucht ist die in R umgesetzte Wirkleistung.

Nach dem praktischen Hinweis zu (114.6) und (114.7) empfiehlt es sich, (114.7) zu verwenden. Für den Verlauf der Spannung erhält man: $u(t) = U_0 \cdot t/T$. Für $i(t)$ gilt: $i(t) = u(t)/R$. Damit ergibt sich

$$P = \frac{1}{T} \cdot \int_0^T \{u(t) \cdot \frac{u(t)}{R}\} dt = \frac{1}{T} \cdot \frac{1}{R} \cdot \int_0^T [u(t)]^2 dt$$

$$= \frac{1}{R \cdot T} \cdot U_0^2 \cdot \int_0^T \{\frac{t}{T}\}^2 dt = \frac{U_0^2}{R \cdot T^3} \cdot \frac{t^3}{3}\Big|_0^T = \frac{U_0^2}{3R}$$

Für den arithmetischen Mittelwert (Gleichanteil) folgt: $\bar{u} = U_0/2$. Es kann also nicht unmittelbar von $\bar{u}$ auf P geschlossen werden. ◀

Beispiel 2

Für das in Abb. 114.2 dargestellte System (Zweitor) soll das Leistungsverhältnis v_P berechnet werden. Mit Tab. 93.5 (Teil XIII „Nachrichtentechnik", Kap. 93 „Zweitore, Vierpole") ergeben sich die $\underline{A}$-Parameter wie folgt:

$$\underline{A}_{11} = 1 + j\omega \cdot R_1 C; \quad \underline{A}_{12} = R_1;$$
$$\underline{A}_{21} = j\omega \cdot C; \quad \underline{A}_{22} = 1$$

Daraus folgt mit (114.5):

$$v_P = \frac{G_a}{\mathrm{Re}\{(\underline{A}_{21} + \underline{A}_{22} \cdot \underline{Y}_a) \cdot (\underline{A}_{11}^* + \underline{A}_{12}^* \cdot \underline{Y}_a^*)\}}$$

$$= \frac{1/R_a}{\mathrm{Re}\{(j\omega \cdot C + \frac{1}{R_a}) \cdot (1 - j\omega \cdot R_1 C + \frac{R_1}{R_a})\}}$$

$$= \frac{1/R_a}{\mathrm{Re}\{j\omega \cdot C + \frac{1}{R_a} + R_1\omega^2 C^2 - \frac{R_1 j\omega \cdot C}{R_a} + \frac{R_1 j\omega \cdot C}{R_a} + \frac{R_1}{R_a^2}\}}$$

$$= \frac{1/R_a}{\frac{1}{R_a} + R_1\omega^2 C^2 + \frac{R_1}{R_a^2}}$$

$$= \frac{R_a}{R_a + R_1 + R_a^2 R_1 \omega^2 C^2}$$

Wird zur Kontrolle $C = 0$ gesetzt, ergibt sich die auch anschaulich einsichtige Lösung $v_P = \frac{R_a}{R_a + R_1}$. ◀

Faltungsintegral

115

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_115

Es gelten folgende Bezeichnungen:

$s_1(t)$, $(f_1(t))$ Zeitabhängige Eingangsgröße des Systems, bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_1(t)$ bezeichnet,

$h(t)$ Übertragungsfunktion des Systems in Abhängigkeit von der Zeit,

$s_2(t)$, $(f_2(t))$ Zeitabhängige Ausgangsgröße des Systems, bei der Laplacetransformation wegen Verwechslungsgefahr mit $f_2(t)$ bezeichnet,

$\underline{S}_1(f)$ Frequenzabhängige Eingangsgröße des Systems,

$\underline{H}(f)$ Übertragungsfunktion in Abhängigkeit von der Frequenz,

$\underline{S}_2(f)$ Frequenzabhängige Ausgangsgröße des Systems.

Nach (109.1a), (109.1b) und (109.1c) gilt:

$$\underline{S}_2(f) = \underline{H}(f) \cdot \underline{S}_1(f).$$

Für die zugeordneten zeitabhängigen Größen folgt die Beziehung

$$s_2(t) = \int_{-\infty}^{t} s_1(t-\tau) \cdot h(\tau) d\tau \tag{115.1}$$

oder, in Kurzschreibweise,

$$s_2(t) = s_1(t) * h(t) \tag{115.2}$$

Der allgemeine Zusammenhang zwischen Zeit- und Frequenzebene ist demnach durch die folgende Beziehung gegeben:

$$s_a(t) * s_b(t) \circ\!\!-\!\!\bullet\, S_a(f) \cdot S_b(f) \tag{115.3}$$

Gleichung (115.1) wird als Faltungsintegral bezeichnet, das Symbol „$*$“ bedeutet „gefaltet mit“. Das Faltungsintegral hat folgende Eigenschaft:

$$s_a(t) * s_b(t) = s_b(t) * s_a(t) \tag{115.4a}$$

Angewendet auf (115.1) kann man auch schreiben:

$$\int_{-\infty}^{t} s_1(t-\tau) \cdot h(\tau) d\tau = \int_{-\infty}^{t} s_1(\tau) \cdot h(t-\tau) d\tau \tag{115.4b}$$

bzw.

$$\int_{-\infty}^{t} s_1(t-\tau) \cdot h(\tau) d\tau = \int_{-\infty}^{t} h(\tau) \cdot s_1(t-\tau) d\tau \tag{115.4c}$$

Die untere Integrationsgrenze wird in den Gleichungen (115.1), (115.4b) und (115.4c) für kausale Systeme, die für $t < 0$ Null sind, zu Null gesetzt.

Die Faltung kann auch im Frequenzbereich angewendet werden. Es gilt:

$$s_a(t) \cdot s_b(t) \circ\!\!-\!\!\bullet\, S_a(f) * S_b(f). \tag{115.5}$$

Das Faltungsintegral wird u. a. in folgenden Fällen eingesetzt:

1. Die zeitabhängige Systemeingangsfunktion $s_1(t)$ und die zeitabhängige Übertragungsfunktion $h(t)$ sind gegeben. Dann kann die zeitabhängige Ausgangsgröße $s_2(t)$ mit (115.1) berechnet werden.
2. Gesucht ist die zeitabhängige Funktion $f(t)$. Die zugehörige Laplacetransformierte $F(s)$ kann in ein Produkt $F(s) = F_1(s) \cdot F_2(s)$ zerlegt werden, wobei die Rücktransformierten der beiden Faktoren bekannt sind. Mit $F(s) \bullet\!\!-\!\!\circ f(t)$, $F_1(s) \bullet\!\!-\!\!\circ f_1(t)$, $F_2(s) \bullet\!\!-\!\!\circ f_2(t)$ ergibt sich:

$$f(t) = \mathbf{L}^{-1}\{F_1(s) \cdot F_2(s)\} = \int_0^t f_1(\tau) \cdot f_2(t-\tau) d\tau \tag{115.6}$$

Beispiel 1

Ein passiver *RC*-Hochpass erster Ordnung hat die Übertragungsfunktion $\delta(t) - \frac{1}{T} \cdot e^{-t/T}$.

An den Eingang wird zum Zeitpunkt $t = 0$ eine Sprungfunktion mit dem Wert 1 für $t > 0$ angelegt. Für die Ausgangsgröße folgt dann mit (115.6):

$$\begin{aligned} s_2(t) &= \int_0^t \left\{\delta(t-\tau) - \frac{1}{T} \cdot e^{-(t-\tau)/T}\right\} \cdot 1 d\tau \\ &= \int_0^t \left\{\delta(\tau) - \frac{1}{T} \cdot e^{-\tau/T}\right\} \cdot 1 d\tau \\ &= \int_0^t \{\delta(\tau)\} d\tau - \int_0^t \left\{\frac{1}{T} \cdot e^{-\tau/T}\right\} d\tau \\ &= 1 + \frac{1}{T} T \cdot e^{-t/T} - 1 = e^{-t/T} \end{aligned}$$

Dabei wurde (115.4b) verwendet und berücksichtigt, dass die zeitabhängigen Funktionen für $t < 0$ Null sind. ◀

Beispiel 2

Gegeben ist die Laplacetransformierte

$$F(s) = \frac{1}{(s+a)\cdot(s+b)} = \underbrace{\frac{1}{s+a}}_{F_1(s)} \cdot \underbrace{\frac{1}{s+b}}_{F_2(s)}$$

mit $F(s) \bullet\!\!-\!\!\circ f(t)$.

Die Rücktransformierten der beiden Faktoren sind bekannt:

$$F_1(s) \bullet\!\!-\!\!\circ e^{-at}, \quad F_2(s) \bullet\!\!-\!\!\circ e^{-bt}$$

Mit (115.6) ergibt sich:

$$f(t) = \int_0^t f_1(\tau)\cdot f_2(t-\tau)d\tau = \int_0^t e^{-a\tau}\cdot e^{-b(t-\tau)}d\tau$$

$$= e^{-bt}\cdot\int_0^t e^{-(a-b)\tau}d\tau = e^{-bt}\cdot\left.\frac{e^{-(a-b)\tau}}{-a+b}\right|_0^t$$

$$= e^{-bt}\cdot\frac{e^{-(a-b)t}-1}{-a+b} = \frac{e^{-at}-e^{-bt}}{b-a}$$

Das Ergebnis ist in Übereinstimmung mit der Transformation Nr. 9 in Tab. 112.3. ◀

Abtasttheorem

116

Wilfried Plaßmann



© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_116

Die Übertragungskanäle der Nachrichten- und der Datentechnik sind in ihrer Frequenzbandbreite beschränkt. Häufig besitzen sie Tiefpassverhalten erster oder höherer Ordnung. Bei der theoretischen Behandlung wird in der Regel ein idealer Tiefpass angenommen, d. h. bis zur Grenzfrequenz f_g bleiben die Amplitudenverhältnisse aller Frequenzanteile untereinander konstant. Oberhalb der Grenzfrequenz werden keine Signalanteile mehr übertragen. *Shannon* hat gezeigt, dass es genügt, von einem frequenzbandbeschränkten Signal diskrete Probenwerte zu nehmen, weil aus diesen Probenwerten das Originalsignal wiederhergestellt werden kann. Die Ableitung wird hier in verkürzter Form dargestellt. Man macht für das bandbegrenzte Signal den Ansatz

$$s(t) = \sum_{n=-\infty}^{+\infty} a_n \cdot x_n(t)$$

und verwendet das orthogonale Funktionensystem

$$x_n(t) = \frac{1}{2B} \cdot \int_{-B}^{+B} e^{j\omega\cdot(t-\frac{n}{2B})} df = \frac{\sin 2\pi \cdot B \cdot (t-\frac{n}{2B})}{2\pi \cdot B \cdot (t-\frac{n}{2B})}$$

Darin ist B die Bandbreite, die bei einem Tiefpasssystem der Grenzfrequenz f_g entspricht. Eine Integration liefert:

$$\int_{-\infty}^{+\infty} s(t) \cdot x_n(t) dt = \int_{-\infty}^{+\infty} s(t) dt \cdot \frac{1}{2B} \int_{-B}^{+B} e^{j\omega\cdot(t-\frac{n}{2B})} df = \frac{a_n}{2B}$$

Die gesuchte Funktion lautet damit

$$s(t) = \sum_{n=-\infty}^{+\infty} s\left(\frac{n}{2B}\right) \cdot x_n(t) = \sum_{n=-\infty}^{+\infty} s\left(\frac{n}{2B}\right) \cdot \frac{\sin 2\pi \cdot B \cdot \left(t-\frac{n}{2B}\right)}{2\pi \cdot B \cdot \left(t-\frac{n}{2B}\right)} \qquad (116.1)$$

Die frequenzbandbegrenzte Funktion $s(t)$ ist also durch eine diskrete Folge von Abtastwerten eindeutig darstellbar. Man kann nämlich zeigen, dass sich durch die diskreten Abtastwerte nur *ein* möglicher Kurvenverlauf legen lässt. Nach (116.1) muss der zeitliche Abstand t_{ab} der Abtastwerte – theoretisch – halb so groß sein wie die Periodendauer T_g der höchsten im Signal vorkommenden sinusförmigen Frequenz f_g, oder, auf die Frequenzen bezogen,

$$\begin{aligned} f_{ab} &= 2 \cdot f_g && \text{(theoretisch)} \\ f_{ab} &\approx (2{,}2 \ldots 4) \cdot f_g && \text{(praktisch)} \end{aligned} \qquad (116.2)$$

In der Praxis wird $f_{ab} > 2 \cdot f_g$ gewählt, um u. a. den mehr oder weniger theoretischen Fall abzudecken, dass das abgetastete Signal sinusförmig mit f_g ist und in den Nulldurchgängen abgetastet wird. In der Praxis genügt erfahrungsgemäß $f_{ab} \approx (2{,}2 \ldots 3) \cdot f_g$. Wählt man dennoch $f_{ab} > 3 \cdot f_g$, hat das keinen Einfluss auf die originalgetreue Wiederherstellung des abgetasteten Signals, vielmehr spielen u. a. folgende Gesichtspunkte eine Rolle:

1. Verringerung des Quantisierungsgeräusches (Teil XIII „Nachrichtentechnik“, Abschn. 96.4.4.5).
2. Die für die Abtastung und damit für die originalgetreue Wiederherstellung des Signals erforderlichen idealen Tiefpassfilter (Antialiasingfilter) im oben geschilderten Sinne gibt es nicht. Frequenzanteile oberhalb f_g führen zum Aliasing-Effekt (Teil XIII „Nachrichtentechnik“, Abschn. 96.4.4 und Abb. 96.31b). Je höher aber die Abtastfrequenz f_{ab} im Vergleich zur Grenzfrequenz f_g ist, desto geringer sind auch die Amplituden der durch die nichtidealen Filtereigenschaften verursachten störenden Frequenzanteile oberhalb f_g.

Aus (116.1) folgt weiterhin, dass der *Abtastzeitpunkt* keine Rolle spielt, was anschaulich einsichtig ist und die Anwendung der Signalabtastung ganz wesentlich erleichtert. Schließlich ergibt sich aus dem Wertebereich von n in der Summe, dass das frequenzbandbeschränkte Signal zeitlich *nicht* beschränkt ist.

Das Abtasttheorem lässt sich auch auf *zeitlich begrenzte* Signalverläufe anwenden. Entwickelt man das auf das Zeitintervall $-T \ldots +T$ beschränkte Signal in eine Fourierreihe, erhält man nach einigen Zwischenschritten:

$$S(f) = \sum_{n=-\infty}^{+\infty} S\left(\frac{n}{2T}\right) \cdot \frac{\sin 2\pi \cdot T \cdot \left(f-\frac{n}{2T}\right)}{2\pi \cdot T \cdot \left(f-\frac{n}{2T}\right)} \qquad (116.3)$$

Das Spektrum eines zeitlich beschränkten Signals ist frequenzmäßig nicht beschränkt und durch diskrete Abtastwerte der Spektraldichte im Abstand $1/2T$ vollständig bestimmt.

Die Rückgewinnung des Originalsignals geschieht nach (116.1) dadurch, dass die Abtastwerte auf ein Tiefpassfilter mit der Grenzfrequenz f_g gegeben werden.

Besondere Bedeutung hat das Abtasttheorem in der Form nach (116.1) für diejenigen Verfahren, bei denen mehrere analoge Signale quasi gleichzeitig über einen Kanal übertragen werden oder ein analoger Signalwert in einen digital kodierten Wert umgesetzt und übertragen oder gespeichert wird. Beispiele dafür sind:

- Pulsmodulationsverfahren, siehe Teil XIII „Nachrichtentechnik“, Abschn. 96.4.
- Speicherung von Sprache und Musik auf Compact-Disks (CDs). Die Abtastfrequenz beträgt 44,1 kHz.
- Fernsprechen: $f_g = 3{,}4\,\text{kHz}, f_{ab} = 8\,\text{kHz}, \Rightarrow f_{ab}/f_g \approx 2{,}35$.
- Anwendung des Zeitmultiplexverfahrens (Teil XIII „Nachrichtentechnik“, Abschn. 100.1).

Nichtkontinuierliche (zeitdiskrete) Signale

117

Wilfried Plaßmann



117.1 Diskrete Fouriertransformation (DFT) 1370
117.2 Schnelle Fouriertransformation (FFT) 1371
117.3 z-Transformation 1372

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_117

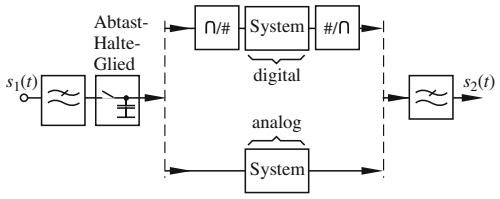


Abb. 117.1 Prinzip der zeitdiskreten Signalverarbeitung

Zeitdiskrete Signale werden häufig aus zeitkontinuierlichen Signalen durch Abtastung erzeugt. Dass beide Signale gleichwertig sind, zeigt das Abtasttheorem (Kap. 116) von *Shannon*, sofern die Bedingung nach (116.2), $f_{ab} \approx (2{,}2 \ldots 4) \cdot f_g$) eingehalten wird. Abb. 117.1 zeigt den prinzipiellen Aufbau einer zeitdiskreten Signalverarbeitung. Der abgetastete Wert kann analog (unterer Signalweg) oder digital codiert (oberer Signalweg) weiterverarbeitet werden. Der Tiefpass am Eingang ist ein Antialiasingfilter und sorgt für die Erfüllung von (116.2), der Tiefpass am Ausgang wandelt das zeitdiskrete Signal in ein zeitkontinuierliches Signal um. Der Einsatz dieser nichtkontinuierlichen Signale hat zwei Vorteile: 1. Es können Zeitmultiplexverfahren angewendet werden. 2. Liegen die Signale in digital codierter Form vor, können sie gespeichert werden (z. B. auf CDs) oder mit geringem Fehler übertragen werden (z. B. Datenfernübertragung). Deshalb hat die digitale Codierung der abgetasteten Werte die größte praktische Bedeutung. Für diesen Fall können die Signale am Ausgang des Systems (Abb. 117.1) auch mit der Laplacetransformation und Sprungfunktionen am Systemeingang erhalten werden.

117.1 Diskrete Fouriertransformation (DFT)

Es wird vorausgesetzt, dass die Bedingung nach (116.2), $f_{ab} \approx (2{,}2 \ldots 4) \cdot f_g$, erfüllt ist. Die Funktion $s(t)$ wird im Zeitbereich $0 \leq t < T$ betrachtet, wobei sie für $t > T$ nicht zwangsweise Null sein muss. Durch N-maliges Abtasten innerhalb des Zeitintervalls T erhält man die diskreten Funktionswerte $s(0)$, $s(T_{ab})$, $s(2T_{ab})$, $s(3T_{ab})$, …, $s(\{N-1\}T_{ab})$ mit $T_{ab} = T/N$, denen entsprechende Werte des Spektrums zugeordnet werden. Die Transformationsformeln lauten:

$$S\left(\frac{2\pi \cdot m}{NT_{ab}}\right) = \sum_{n=0}^{N-1} s(n \cdot T_{ab}) \cdot e^{-(j2\pi \cdot m \cdot n)/N} \qquad (117.1)$$

mit $m = 0, 1, \ldots, N-1$.

$$s(nT_{ab}) = \frac{1}{N} \sum_{m=0}^{N-1} S\left(\frac{2\pi \cdot m}{NT_{ab}}\right) \cdot e^{(j2\pi \cdot m \cdot n)/N} \qquad (117.2)$$

mit $n = 0, 1, \ldots, N-1$.

Auch bei der zeitdiskreten Signalverarbeitung haben die Sprungfunktion $\sigma(nT_{ab})$ und der Dirac-Impuls $\delta(nT_{ab})$ besondere Bedeutung. Impuls- und Sprungantwort am Ausgang des Systems hängen analog zu den Beziehungen bei kontinuierlichen Signalen nach Abschn. 113.3, (113.5) und (113.6), in folgender Weise zusammen:

$$s_2(nT_{ab})_\delta = s_2(nT_{ab})_\sigma - s_2(\{n-1\}T_{ab})_\sigma \qquad (117.3)$$

$$s_2(nT_{ab})_\sigma = \sum_{\mu=-\infty}^{n} s_2(\mu T_{ab})_\delta \qquad (117.4)$$

Die Indizes σ (Sprungfunktion) bzw. δ (Dirac-Impuls) weisen auf die Ursache des Ausgangssignals hin.

Ist die Systemantwort auf einen Dirac-Impuls am Eingang bekannt, kann das Ausgangssignal für beliebige Eingangssignale mit der Faltungssumme (entsprechend dem Faltungsintegral nach Gleichung (115.1) bzw. (115.4a)–(115.4c) bei zeitkontinuierlichen Signalen) berechnet werden:

$$s_2(nT_{ab}) = \sum_{\mu=-\infty}^{+\infty} s_1(\{n-\mu\}\, T_{ab}) \cdot h(\mu T_{ab}) \qquad (117.5)$$

Die Übertragungsfunktion $H(f)_{nT_{ab}}$ erhält man aus folgender Beziehung:

$$H(f)_{nT_{ab}} = \left.\frac{s_2(nT_{ab})}{s_1(nT_{ab})}\right|_{s_1(nT_{ab})=e^{jn\omega T_{ab}}} \qquad (117.6)$$

Ist die Impulsantwort $s_2(nT_{ab})_\delta$ am Systemausgang bekannt, gilt:

$$H(f)_{nT_{ab}} = \sum_{n=-\infty}^{+\infty} s_2(nT_{ab})_\delta \cdot e^{-jn\omega T_{ab}} \qquad (117.7)$$

Der Index δ weist auf den Dirac-Impuls als Eingangssignal hin.

Analog zur Fourierreihe aus Abschn. 111.1, (111.2a)–(111.2c), können die Koeffizienten der diskreten Fouriertransformation auch in der reellen Form dargestellt werden, wenn das Integral durch eine Summe ersetzt wird (wegen des Vergleiches mit (111.2a)–(111.2c) wurde k beibehalten):

$$a_0 = \frac{1}{N} \cdot \sum_{m=0}^{N-1} s_m \qquad (117.8a)$$

$$a_k = \frac{2}{N} \cdot \sum_{m=0}^{N-1} s_m \cdot \cos\left(k \cdot \frac{2\pi}{N} \cdot m\right) \qquad (117.8b)$$

$$b_k = \frac{2}{N} \cdot \sum_{m=0}^{N-1} s_m \cdot \sin\left(k \cdot \frac{2\pi}{N} \cdot m\right) \qquad (117.8c)$$

Man kann hier die Bedeutung der auftretenden Größen besonders anschaulich darstellen: N ist die Zahl der Abtastwerte im Intervall $0 \ldots T$, der zeitliche Abstand T_{ab} der Abtastwerte ist mit dem Intervall T über N verknüpft: $T_{ab} = T/N$ (wie bei (117.1) und (117.2)). Die Zahl der maximal zu berücksichtigenden Fourierkoeffizienten gibt k_{max} an. Für die höchste im Signal enthaltene Frequenz f_{max}, die über die Grenzfrequenz f_g des Tiefpasses (Antialiasingfilter) festgelegt wird, d. h. $f_{max} = f_g$, gilt: $f_{max} = k_{max}/T$. Dabei muss zwischen T_{ab} und f_{max} die Bedingung für die Abtastung nach (116.2) erfüllt sein: $f_{ab} \approx (2{,}2 \ldots 4) \cdot f_g$ oder $T_{ab} \approx 1/(\{2{,}2 \ldots 4\} \cdot f_{max})$.

Die Auswertung der Transformationsgleichungen erfordert wegen des erhöhten Rechenaufwandes den Einsatz von Rechnern. Je nachdem, ob die reelle oder komplexe Darstellung der Fourierreihe verwendet wird und ob das gegebene Signal Symmetriebedingungen erfüllt oder nicht, ergibt sich eine unterschiedliche Anzahl von Multiplikationen. Als Orientierungswert kann dienen, dass bei N Abtastwerten etwa N^2 Multiplikationen erforderlich sind. Dabei sind es weniger die Multiplikationen an sich, die zeitaufwendig sind, sondern vielmehr die Berechnungsroutinen zur Bestimmung der sin- und cos-Werte.

Für die Anwendung der diskreten Fouriertransformation werden die folgenden Formeln häufiger benötigt:

$$\sum_{\mu=0}^{\infty} a^{\mu} = \frac{1}{1-a} \qquad \text{für } |a| < 1 \tag{117.9}$$

$$\sum_{\mu=0}^{k} a^{\mu} = \frac{a^{k+1}-1}{a-1} \qquad \text{für } a \neq 1 \tag{117.10}$$

$$\sum_{\mu=k}^{\infty} a^{\mu} = \frac{a^{k}}{1-a} \qquad \text{für } |a| < 1 \tag{117.11}$$

Beispiel 1

Die Impulsantwort eines Systems ist gegeben zu:

$$s_2(nT_{ab})_{\delta} = \underbrace{A}_{1} \cdot \underbrace{\delta(0)}_{2} + \underbrace{\sigma(\{n-1\}T_{ab})}_{3} \cdot \underbrace{B}_{4} \cdot 0{,}5^{(n-1)} \tag{117.12}$$

Erläuterungen siehe Text, Funktionen sind Null für $t < 0$.

Erläuterungen zu den Ziffern: 1, 4: Die Konstanten A und B werden allgemein eingeführt, um die Zahlenwerte später geeignet zu wählen. 2: Die Konstante A erscheint zum Zeitpunkt $t = 0$ als Dirac-Impuls. 3: Die genannte Abhängigkeit tritt mit einer Verzögerung von T_{ab} auf.

Die Sprungantwort ergibt sich mit (117.4):

$$s_2(nT_{ab})_{\sigma} = A + B \cdot \sum_{\mu=1}^{n} 0{,}5^{\mu-1} = A + B \cdot \sum_{\mu=0}^{n-1} 0{,}5^{\mu} \tag{117.13}$$

Die Übertragungsfunktion $H(f)_{nT_{ab}}$ berechnet sich mit (117.7):

$$\begin{aligned} H(f)_{nT_{ab}} &= A + B \cdot \sum_{n=1}^{\infty} 0{,}5^{n-1} \cdot e^{-j\omega n T_{ab}} \\ &= A + B \cdot e^{-j\omega T_{ab}} \cdot \sum_{\mu=0}^{\infty} 0{,}5^{-j\mu\omega T_{ab}} \\ &= A + B \cdot e^{-j\omega T_{ab}} \cdot \frac{1}{1 - 0{,}5^{-j\omega T_{ab}}} \end{aligned} \tag{117.14}$$

Dabei wurde die Beziehung $\sum_{\mu=0}^{\infty} a^{\mu} = \frac{1}{1-a}$ (117.9), verwendet.

Bestimmt man das Quadrat des Betrages der Übertragungsfunktion (117.14), erhält man mit $|H(f)_{nT_{ab}}|^2 = H(f)_{nT_{ab}} \cdot H^*(f)_{nT_{ab}}$:

$$\begin{aligned} H(f)_{nT_{ab}} &= \left\{ A + e^{-j\omega T_{ab}} \cdot \frac{B}{1 - 0{,}5 \cdot e^{-j\omega T_{ab}}} \right\} \\ &= \frac{A - A \cdot 0{,}5 \cdot e^{-j\omega T_{ab}} + B \cdot e^{-j\omega T_{ab}}}{1 - 0{,}5 \cdot e^{-j\omega T_{ab}}} \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad & |H(f)_{nT_{ab}}|^2 \\ &= \underbrace{\frac{A + (B - 0{,}5 \cdot A) \cdot e^{-j\omega T_{ab}}}{1 - 0{,}5 \cdot e^{-j\omega T_{ab}}}}_{H(f)_{nT_{ab}}} \\ &\quad \cdot \underbrace{\frac{A + (B - 0{,}5 \cdot A) \cdot e^{+j\omega T_{ab}}}{1 - 0{,}5 \cdot e^{+j\omega T_{ab}}}}_{H^*(f)_{nT_{ab}}\ \{\text{konjug. kompl. zu } H(f)_{nT_{ab}}\}} \\ &= \frac{A^2 + A \cdot (B - 0{,}5 \cdot A) \cdot \{e^{-j\omega T_{ab}} + e^{+j\omega T_{ab}}\} + (B - 0{,}5 \cdot A)^2}{1 + \{-e^{-j\omega T_{ab}} - e^{j\omega T_{ab}}\} \cdot 0{,}5 + 0{,}25} \\ &= \frac{1{,}25 \cdot A^2 + B^2 - AB + (AB - 0{,}5 \cdot A^2) \cdot 2 \cdot \cos(\omega T_{ab})}{1{,}25 - \cos(\omega T_{ab})} \end{aligned}$$

Mit den Konstanten A und B lassen sich die Eigenschaften des Systems einstellen. So sind der Zähler und der Nenner gleich, wenn gewählt wird: $A = -0{,}5$, $B = 0{,}75$.

(Ansatz: $1{,}25 \cdot A^2 + B^2 - AB = 1{,}25$; $(AB - 0{,}5 \cdot A^2) \cdot 2 = -1$):

$$|H(f)_{nT_{ab}}|_{A=-0{,}5;\, B=0{,}75} = 1 \neq f(f)$$

Dann hat das System Allpassverhalten und kann z. B. zur Phaseneinstellung benutzt werden. ◀

117.2 Schnelle Fouriertransformation (FFT)

Die gesamte Rechenzeit bei der DFT kann nur dadurch wesentlich verringert werden, dass u. a. die Anzahl der Berechnungen für die sin- und cos-Werte reduziert wird. Die schnelle Fouriertransformation (*F*ast *F*ourier *T*ransform, FFT) nutzt zu diesem Zweck bestimmte Symmetrieeigenschaften der Sinus- und Cosinusfunktion aus, was besonders effektiv ist, wenn die Anzahl N der Abtastwerte auf einer Potenz der Zahlen 2, 4, 8 usw. basiert: $N = 2^M$ (Basis-2-FFT), $N = 4^M$ (Basis-4-FFT) usw. In [4] ist ausführlich dargestellt, wie z. B. eine Basis-2-FFT für $N = 1024$ entwickelt und für die Berechnung mit einem Rechner aufbereitet wird. Beispiel 2 zeigt die Symmetrieeigenschaften für $N = 2^4 = 16$.

Beispiel 2

Um zu zeigen, wie die Symmetrieeigenschaften bei der FFT ausgenutzt werden, wird $N = 2^4 = 16$ gewählt. Dann sind die sin- und cos-Werte von $n \cdot 22{,}5°$, $n = 0 \ldots 15$, erforderlich. Es genügt aber, *zwei* sin- und *einen* cos-Wert zu berechnen (jeweils unterstrichen) und je nach Quadrant das Vorzeichen entsprechend zu setzen, wenn die dargestellten Symmetriebedingungen beachtet werden:

$$\begin{aligned}
\underline{\sin 22{,}5°} &= \sin 157{,}5° = -\sin 202{,}5° = -\sin 337{,}5° \\
&= \cos 67{,}5° = -\cos 112{,}5° = -\cos 247{,}5° \\
&= \cos 292{,}5° \\
\underline{\cos 22{,}5°} &= -\cos 157{,}5° = -\cos 202{,}5° = \cos 337{,}5° \\
&= \sin 67{,}5° = \sin 112{,}5° = -\sin 247{,}5° \\
&= -\sin 292{,}5° \\
\underline{\sin 45°} &= \sin 135° = -\sin 225° = -\sin 315° \\
&= \cos 45° = -\cos 135° = -\cos 225° \\
&= \cos 315°
\end{aligned}$$

Die Werte für 0°, 90°, 180° und 270° brauchen als „Trivialwerte“ nicht berechnet zu werden. ◀

Die Erstellung eines entsprechenden Rechnerprogrammes ist zwar aufwendig, aber sinnvoll, weil der Anwender bei jedem Programmdurchlauf von der Zeitersparnis profitiert.

117.3 *z*-Transformation

Die z-Transformation dient ebenfalls zur Beschreibung von Abtastsystemen. Sie gibt eine geschlossene Darstellung der Folge der Abtastwerte ohne Rückgriff auf die Einhüllende. Die Regelungstechnik wendet sie häufig bei Abtastregelsystemen an, weil die dort auftretenden absichtlich eingefügten oder prinzipiell vorhandenen Zeitverzögerungen nT_{ab} einfach erfasst werden können. Hier sollen nur die Grundlagen dargestellt werden, weitere Einzelheiten sind z. B. in [5] enthalten.

Die Transformationsgleichungen sind wie folgt definiert:

Zweiseitige z-Transformation:

$$S(z) = \sum_{n=-\infty}^{+\infty} s(nT_{ab}) \cdot z^{-n} \tag{117.15}$$

oder, in symbolischer Darstellung:

$$S(z) \circ\!\!-\!\!\circ s(nT_{ab})$$

Einseitige z-Transformation:

$$S(z) = \sum_{n=0}^{+\infty} s(nT_{ab}) \cdot z^{-n}$$

d. h.

$$s(nT_{ab}) \equiv 0 \quad \text{für } n < 0 \tag{117.16}$$

Die Unterscheidung zwischen ein- und zweiseitiger z-Transformation wurde aus praktischen Erwägungen heraus eingeführt.

Die Rücktransformation geschieht nach folgender Beziehung:

$$s(nT_{ab}) = \frac{1}{2j\pi} \cdot \oint S(z) \cdot z^{n-1} dz \tag{117.17}$$

symbolische Darstellung: $s(nT_{ab}) \circ\!\!-\!\!\circ S(z)$.

Für die Anwendung sind folgende Eigenschaften von Bedeutung:

1. Multiplikation mit einer Konstanten K:

$$K \cdot s(nT_{ab}) \circ\!\!-\!\!\circ K \cdot S(z) \tag{117.18}$$

2. Addition zweier Zeitfunktionen:

$$s_1(nT_{ab}) + s_2(nT_{ab}) \circ\!\!-\!\!\circ S_1(z) + S_2(z) \tag{117.19}$$

3. Verschiebung:

$$s(nT_{ab}) \circ\!\!-\!\!\circ S(z) \quad \rightarrow \quad s(\{n-\mu\}T_{ab}) \circ\!\!-\!\!\circ z^{-\mu} \cdot S(z) \tag{117.20}$$

4. Faltung:

$$\begin{aligned}
s_1(nT_{ab}) * s_2(nT_{ab}) &= \sum_{\substack{n=-\infty \\ (n=0)}}^{+\infty} s_1(\tau T_{ab}) s_2(\{n-\tau\}T_{ab}) \\
&\circ\!\!-\!\!\circ S_1(z) \cdot S_2(z)
\end{aligned} \tag{117.21}$$

Tab. 117.1 enthält einige Korrespondenzen zur z-Transformation. Zum Vergleich sind teilweise die zugeordneten Laplacetransformierten mit aufgeführt.

Korrespondenz Nr. 11 in Tab. 117.1 zeigt, analog zur Laplacetransformation, eine allgemeine Lösung zur Rücktransformation. Sind die Null- und Polstellen bestimmt, kann eine Partialbruchzerlegung durchgeführt werden. Anschließend lassen sich die erhaltenen Summanden einzeln nach Tab. 117.1, Korrespondenzen 12 bis 14, zurücktransformieren. Der in der Transformation Nr. 13 auftretende Binomialkoeffizient wird wie folgt berechnet:

$$\begin{bmatrix} n-1 \\ \mu-1 \end{bmatrix} = \frac{(n-1)!}{\{(\mu-1)!\} \cdot \{[(n-1)-(\mu-1)]!\}} \tag{117.22}$$

Da die Koeffizienten $a_0 \ldots a_m$, $b_0 \ldots b_n$ der Korrespondenz Nr. 11 in Tab. 117.1 reell sind, ergeben sich reelle oder konjugiert komplexe Pol- und Nullstellen.

Für ein System mit der Übertragungsfunktion $H(z)$, der Eingangsgröße $S_1(z)$ und der Ausgangsgröße $S_2(z)$ gilt analog zur Fourier- und Laplacetransformation:

$$S_2(z) = H(z) \cdot S_1(z) \tag{117.23}$$

Tab. 117.1 Korrespondenzen der z-Transformation

Nr.	$f(t)$	$F(s)$	$F(z)$
1	$\sigma(t)$	$\frac{1}{s}$	$\frac{z}{z-1}$
2	t	$\frac{1}{s^2}$	$T \cdot \frac{z}{(z-1)^2}$
3	t^2	$\frac{2}{s^3}$	$T^2 \cdot z \cdot \frac{z+1}{(z-1)^3}$
4	e^{at}	$\frac{1}{s-a}$	$\frac{z}{z-e^{aT}}$
5	$t \cdot e^{at}$	$\frac{1}{(s-a)^2}$	$T \cdot e^{aT} \cdot \frac{z}{(z-e^{aT})^2}$
6	$\sin(\omega_0 \cdot t)$	$\frac{\omega_0}{s^2+\omega_0^2}$	$z \cdot \frac{\sin(\omega_0 \cdot T)}{z^2-2 \cdot z \cdot \cos(\omega_0 \cdot T)+1}$
7	$\cos(\omega_0 \cdot t)$	$\frac{s}{s^2+\omega_0^2}$	$z \cdot \frac{z-\cos(\omega_0 \cdot T)}{z^2-2 \cdot z \cdot \cos(\omega_0 \cdot T)+1}$
8	$e^{at} \cdot \sin(\omega_0 \cdot t)$	$\frac{\omega_0}{(s-a)^2+\omega_0^2}$	$z \cdot \frac{e^{aT} \cdot \sin(\omega_o \cdot T)}{z^2-2 \cdot z \cdot e^{aT} \cdot \cos(\omega_0 \cdot T)+e^{2aT}}$
9	$e^{at} \cdot \cos(\omega_0 \cdot t)$	$\frac{s-a}{(s-a)^2+\omega_0^2}$	$z \cdot \frac{z-e^{aT} \cdot \cos(\omega_0 \cdot T)}{z^2-2 \cdot z \cdot e^{aT} \cdot \cos(\omega_0 \cdot T)+e^{2aT}}$
10	$\delta(t-nT)$	e^{-nTs}	z^{-n}

Nr.	$F(s)$	$F(z)$
11	$\frac{b_0+b_1 \cdot e^{-Ts}+b_2 \cdot e^{-2Ts}+\ldots+b_n \cdot e^{-nTs}}{a_0+a_1 \cdot e^{-Ts}+a_2 \cdot e^{-2Ts}+\ldots+a_m \cdot e^{-mTs}}$	$\frac{b_0+b_1 \cdot z^{-1}+b_2 \cdot z^{-2}+\ldots+b_n \cdot z^{-n}}{a_0+a_1 \cdot z^{-1}+a_2 \cdot z^{-2}+\ldots+a_m \cdot z^{-m}}$

Nr.	$f(t)$	$F(z)$
12	$\sigma(n-1) \cdot a^{n-1}$ [a]	$\frac{1}{z-a}$, $\lvert z\rvert > \lvert a\rvert$
13	$\sigma(n-\mu) \cdot \binom{n-1}{\mu-1} \cdot a^{n-\mu}$ [a, b]	$\frac{1}{(z-a)^\mu}$, $\lvert z\rvert > \lvert a\rvert$
14	$\delta(n-\mu)$ [c]	$\frac{1}{z^\mu}$

[a] σ Sprungfunktion, [b] siehe (117.22), [c] δ Dirac-Impuls

Ob ein Abtastsystem stabil ist, lässt sich anhand der Übertragungsfunktion $H(z)$ feststellen. Mit

$$H(z) = \frac{P(z)}{Q(z)} \tag{117.24}$$

($P(z)$ Zählerpolynom, $Q(z)$ Nennerpolynom) gilt: Ein System ist stabil, wenn die Nullstellen des Nennerpolynoms innerhalb des Einheitskreises der z-Ebene liegen und der Zählergrad den Nennergrad nicht übersteigt.

Mit der z-Transformation lassen sich auch Differenzengleichungen erfassen. Abb. 117.2 zeigt eine Schaltung, die eine Differenzengleichung erster Ordnung ergibt. Es gilt für die Ausgangsgröße $s_2(nT_{ab})$:

$$\begin{aligned} s_2(nT_{ab}) = {} & b_0 \cdot s_1(nT_{ab}) + b_1 \cdot s_1(\{n-1\}T_{ab}) \\ & + a_1 \cdot s_2(\{n-1\}T_{ab}) \end{aligned}$$

bzw. umgeformt

$$\begin{aligned} & s_2(nT_{ab}) - a_1 \cdot s_2(\{n-1\}T_{ab}) \\ & \quad = b_0 \cdot s_1(nT_{ab}) + b_1 \cdot s_1(\{n-1\}T_{ab}) \end{aligned} \tag{117.25}$$

Für die Übertragungsfunktion folgt mit (117.7), (117.23) und (117.25):

$$\begin{aligned} & H(f)_{nT_{ab}} \cdot e^{jn\omega T_{ab}} - H(f)_{nT_{ab}} \cdot a_1 \cdot e^{j\{n-1\}\omega T_{ab}} \\ & \quad = b_0 \cdot e^{jn\omega T_{ab}} + b_1 \cdot e^{j\{n-1\}\omega T_{ab}} \end{aligned}$$

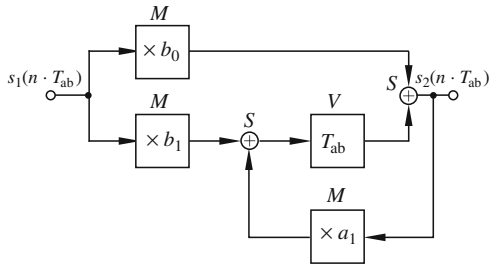


Abb. 117.2 Schaltung für eine Differenzengleichung erster Ordnung; M: Multiplizierer, V: Verzögerungsglied, S: Summierer

Umgeformt ergibt sich:

$$H(f)_{nT_{ab}} = \frac{b_0 + b_1 \cdot e^{-j\omega T_{ab}}}{1 - a_1 \cdot e^{-j\omega T_{ab}}} \tag{117.26}$$

Daraus erhält man die Transformierte $H(z)$ mit $e^{-j\omega T_{ab}} = z^{-1}$ zu:

$$H(z) = \frac{b_0 + b_1 \cdot z^{-1}}{1 - a_1 \cdot z^{-1}} \tag{117.27}$$

Der gemeinsame Faktor e^n kürzt sich in (117.26) heraus, was anschaulich auch einsichtig ist, denn $H(f)_{nT_{ab}}$ bzw. $H(z)$ dürfen sich nicht in Abhängigkeit von der Zahl der Abtastwerte ändern.

Durch geeignete Wahl von a_1, b_0 und b_1 kann das Verhalten der Übertragungsfunktion eingestellt werden.

Für eine Differenzengleichung zweiten Grades lässt sich in analoger Weise ableiten:

$$\begin{aligned} & a_2 \cdot s_2(nT_{ab}) + a_1 \cdot s_2(\{n-1\}T_{ab}) + a_0 \cdot s_2(\{n-2\}T_{ab}) \\ & \quad = b_2 \cdot s_1(nT_{ab}) + b_1 \cdot s_1(\{n-1\}T_{ab}) + b_0 \cdot s_1(\{n-2\}T_{ab}) \end{aligned}$$

Daraus ergibt sich $H(z)$:

$$H(z) = \frac{b_2 \cdot z^2 + b_1 \cdot z + b_0}{a_2 \cdot z^2 + a_1 \cdot z + a_0} = \frac{b_2 + b_1 \cdot z^{-1} + b_0 \cdot z^{-2}}{a_2 + a_1 \cdot z^{-1} + a_0 \cdot z^{-2}} \tag{117.28}$$

Häufig ist $b_2 = 1$ oder $a_2 = 1$, was durch entsprechendes Multiplizieren der gesamten Formel mit $1/b_2$ bzw. a_2 erreicht werden kann.

Beispiel 3

Ausgangspunkt sind (117.25) und Abb. 117.2. Das System ist stabil, wenn nach (117.24) gilt: $|a_1| < 1$. Anschaulich nimmt $s_2(nT_{ab})$ für $|a_1| > 1$ mit jeder Abtastung zu und wächst mit $n \to \infty$ über alle Grenzen. Wird als Eingangsgröße ein Sprung mit der Amplitude U verwendet, und sind alle Größen für $n < 0$ Null, ergibt sich für die

Ausgangsgröße $s_2(nT_{ab})$ nach der n-ten Abtastung (Zählbeginn mit $n = 0$; einige Rechnung erforderlich):

$$s_2(nT_{ab}) = U \cdot b_0 \cdot \sum_{\mu=0}^{n} c^{\mu} + U \cdot (b_0 + b_1) \cdot \sum_{\mu=0}^{n} c^{\mu-1} \tag{117.29}$$

Gilt $0 < a_1 < 1$, nähert sich $s_2(nT_{ab})$ asymptotisch dem Wert $U \cdot \frac{2b_0+b_1}{1-a_1}$. Für $-1 < a_1 < 0$ dagegen ergibt sich für $n \to \infty$ der Endwert $U \cdot \frac{2b_0+b_1}{1+a_1}$, dem sich $s_2(nT_{ab})$ oszillierend nähert.

Die in Abb. 117.2 dargestellte Schaltung kann damit wegen der Sprungantwort nach (117.29) als Tiefpass interpretiert werden. Für die Auswertung wurden die Summenformeln

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{bzw.} \quad \sum_{n=0}^{\infty} x^n \cdot (-1)^n = \frac{1}{1+x}$$

verwendet. ◀

Beispiel 4

Gegeben ist die Übertragungsfunktion $H(z)$ eines Systems:

$$H(z) = \frac{z}{z^2 - z + 0{,}5} = \frac{z}{(z - 0{,}5 - j \cdot 0{,}5) \cdot (z - 0{,}5 + j \cdot 0{,}5)} \tag{117.30}$$

Gesucht sind die Impulsantwort und eine Schaltung, mit der sich $H(z)$ realisieren lässt, sofern das System stabil ist.

Das System ist stabil, weil die Abstände x der zwei konjugiert komplexen Nullstellen des Nenners vom Koordinatenursprung in der z-Ebene kleiner als 1 sind:

$$x = +\sqrt{0{,}5^2 + 0{,}5^2} = +\sqrt{0{,}5} < 1$$

Außerdem ist der Zählergrad nicht größer als der Nennergrad. Für die Rücktransformation soll die Korrespondenz Nr. 12 aus Tab. 117.1 benutzt werden. Eine Partialbruchzerlegung liefert:

$$H(z) = \frac{z}{(z - 0{,}5 + j \cdot 0{,}5) \cdot (z - 0{,}5 - j \cdot 0{,}5)} = \frac{0{,}5 + j \cdot 0{,}5}{(z - 0{,}5 + j \cdot 0{,}5)} + \frac{0{,}5 - j \cdot 0{,}5}{(z - 0{,}5 - j \cdot 0{,}5)}$$

Mit der genannten Korrespondenz folgt:

$$s_2(nT_{ab}) = \{\sigma[(n-1)T_{ab}]\} \cdot \{[0{,}5 + j \cdot 0{,}5] \cdot [0{,}5 - j \cdot 0{,}5]^{n-1} + [0{,}5 - j \cdot 0{,}5] \cdot [0{,}5 + j \cdot 0{,}5]^{n-1}\}$$

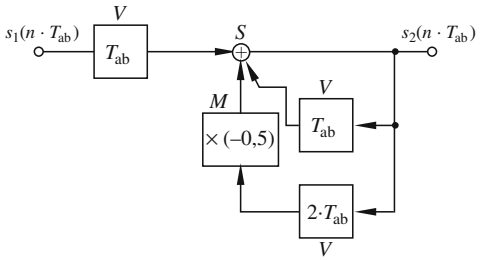


Abb. 117.3 Realisierung zu Beispiel 4; M: Multiplizierer, V: Verzögerungsglied, S: Summierer

Mit $[0{,}5 \pm j \cdot 0{,}5]^{n-1} = [\sqrt{0{,}5^2 + 0{,}5^2} \cdot e^{\pm j\varphi}]^{n-1}$ und $f = \arctan(1)$ ergibt sich:

$$\begin{aligned} s_2(nT_{ab}) &= \{\sigma[(n-1)T_{ab}]\} \cdot \{(\sqrt{0{,}5})^{n-1}\} \\ &\quad \cdot \{(0{,}5 + j \cdot 0{,}5) \cdot e^{-j\cdot(n-1)\varphi} \\ &\qquad + (0{,}5 - j \cdot 0{,}5) \cdot e^{j\cdot(n-1)\varphi}\} \\ &= \{\sigma[(n-1)T_{ab}]\} \cdot \{(\sqrt{0{,}5})^{n-1}\} \\ &\quad \cdot \{(0{,}5 \cdot e^{j\cdot(n-1)\varphi} + 0{,}5 \cdot e^{-j\cdot(n-1)\varphi}) \\ &\qquad + (-j \cdot 0{,}5 \cdot e^{j\cdot(n-1)\varphi} + j \cdot 0{,}5 \cdot e^{-j\cdot(n-1)\varphi})\} \\ &= \{\sigma[(n-1)T_{ab}]\} \cdot \{(\sqrt{0{,}5})^{n-1}\} \\ &\quad \cdot \{\cos[(n-1) \cdot \varphi] + \sin[(n-1) \cdot \varphi]\} \end{aligned}$$

Am Faktor $\{(\sqrt{0{,}5})^{n-1}\}$ wird auch anschaulich sichtbar, dass das System stabil ist. Das Ausgangssignal ist die Summe aus einer Sinus- und einer Cosinusschwingung mit abnehmender Amplitude. Mit den Ausführungen zu (117.28) kann die Differenzengleichung, hier zweiten Grades, aus $H(z)$ nach (117.30) aufgestellt werden. Besonders geeignet ist die Darstellung

$$H(z) = \frac{b_2 + b_1 \cdot z^{-1} + b_0 \cdot z^{-2}}{a_2 + a_1 \cdot z^{-1} + a_0 \cdot z^{-2}},$$

wenn man beachtet, dass ein Faktor $z^{-\mu}$ eine $(\mu \cdot T_{ab})$-fache Verzögerung gegenüber $(n \cdot T_{ab})$ bedeutet (siehe (117.20)). Dann folgt:

$$H(z) = \frac{z}{z^2 - z + 0{,}5} \cdot \frac{z^{-2}}{z^{-2}} = \frac{z^{-1}}{1 - z^{-1} + 0{,}5 \cdot z^{-2}}$$

und daraus

$$s_2(nT_{ab}) - s_2(\{n-1\}T_{ab}) + 0{,}5 \cdot s_2(\{n-2\}T_{ab}) = s_1(\{n-1\}T_{ab})$$

Abb. 117.3 zeigt eine Schaltung, die das berechnete Verhalten aufweist. Bei Bedarf lässt sich die Zahl der Verzögerungsglieder noch minimieren. ◀

Zufällige Signale

118

Wilfried Plaßmann



118.1 Grundbegriffe und Kenngrößen 1376
118.2 Verteilungen 1379
118.3 Rauschen 1381
118.4 Signalerkennung bei gestörter Übertragung 1382
Literatur 1383

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6_118

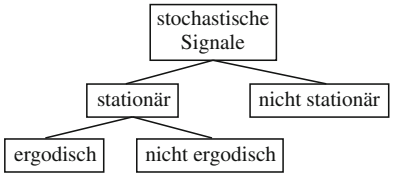


Abb. 118.1 Einteilung stochastischer Signale

Die hier betrachteten Signale werden auch als stochastische Signale bezeichnet. Sie sind zufällig, d. h. ihr Augenblickswert ist nicht vorhersagbar. Abb. 118.1 zeigt eine Unterteilung nach den Signaleigenschaften. Über die *nichtstationären Signale* lässt sich keinerlei Aussage machen, und sie werden hier deshalb auch nicht weiter betrachtet. Für die *stationären Signale* gilt, dass ihr Momentanwert zwar auch nicht vorhergesagt werden kann, dass sie aber trotzdem gewisse auswertbare Eigenschaften besitzen, die allerdings statistischer Art sind und deshalb nur mit den Gesetzen der Wahrscheinlichkeitstheorie erfasst werden können. Zwar sind – theoretisch – für die Anwendung der Wahrscheinlichkeitstheorie unendlich viele Probenwerte zu nehmen, in der Praxis hat sich aber gezeigt, dass man auch mit einer endlichen Anzahl dem gesuchten Wert schon ausreichend nahe kommen kann.

118.1 Grundbegriffe und Kenngrößen

Die nachfolgend dargestellten Grundbegriffe und Kenngrößen werden in den Beispielen 1 bis 7 anschaulich erläutert. Weiterhin haben einige der Begriffe eine zentrale Bedeutung in der Messtechnik und treten deshalb im Teil X „Messtechnik“, Kap. 80, ebenfalls auf. Es wird unterschieden zwischen einer Verteilung und einer Stichprobe. Die funktionale Abhängigkeit einer Verteilung ist in ihrem Verlauf durch häufiges Erfassen bzw. Messen relativ gesichert und damit auch ihre Kenngrößen, während es sich bei einer Stichprobe um einige wenige Werte handelt, von denen auf gewisse statistische Eigenschaften geschlossen werden soll.

- *Wahrscheinlichkeitsdichte(funktion)*, *Häufigkeit*, $h(x)$: Sie gibt die relative Häufigkeit an, mit der der Wert x vorhanden ist. Für einen bestimmten Wert $x = x_0$ ist ihr Wert gleich der Wahrscheinlichkeit $w(x = x_0)$, mit der dieser Wert auftritt. Sie kann aus diskreten Werten bestehen oder aus einer stetigen Funktion.
- *Summenfunktion*, *Summenhäufigkeitsfunktion*, $H(x)$: Sie ist die Summenfunktion der Wahrscheinlichkeitsdichte $h(x)$. Sie gibt die Wahrscheinlichkeit $w(x_0 \leq x)$ an, dass die Zufallsgröße $x = x_0$ kleiner oder höchstens gleich dem betrachteten Wert x ist:

$$w(x_0 \leq N) = H(N) = \sum_{x=0}^{N} h(x) \tag{118.1a}$$

für diskretes $h(x)$ bzw.

$$w(x_0 \leq x) = H(x) = \int_{y=-\infty}^{x} f(y)dy \tag{118.1b}$$

für stetiges $h(x)$.
Die Wahrscheinlichkeit dafür, dass die Zufallsgröße $x = x_0$ größer ist als der betrachtete Wert x, errechnet sich zu:

$$w(x_0 > N) = 1 - w(x_0 \leq N) = 1 - \sum_{x=0}^{N} h(x)$$

bzw.

$$w(x_0 > x) = 1 - \int_{y=-\infty}^{x} f(y)dy \tag{118.2}$$

Die Wahrscheinlichkeit, dass ein Wert zwischen einem unteren Wert $x_{\min}$ und einem oberen Wert $x_{\max}$ liegt, ist:

$$w(x_{\min} \leq X \leq x_{\max}) = \int_{x_{\min}}^{x_{\max}} h(x)dx = \int_{-\infty}^{x_{\max}} h(x)dx - \int_{-\infty}^{x_{\min}} h(x)dx \tag{118.3}$$

Die Wahrscheinlichkeitsdichte $h(x)$ einer stetigen Verteilung kann allerdings nicht mehr als die Wahrscheinlichkeit aufgefasst werden, dass genau der Wert x auftritt. Dieser Wert ergibt mit (118.3) Null.

- *Linearer Mittelwert* oder *Erwartungswert* $\bar{x}$ *einer Verteilung*:

$$\bar{x} = \frac{1}{N} \sum_{x=0}^{N} x \cdot h(x) \quad \text{bzw.} \quad \bar{x} = \int_{-\infty}^{+\infty} x \cdot h(x)dx \tag{118.4}$$

- *Linearer Mittelwert* oder *Erwartungswert* $\bar{x}$ *einer Stichprobe*: Die gemessenen Werte (z. B. Messwerte, Daten auf einer Datenleitung bei Mehrfachübertragung zur Fehlerreduktion) bilden eine Stichprobe. Wird daraus der Mittelwert gebildet, kann dieser häufig als Schätz- oder Näherungswert eingesetzt werden. In der Messtechnik wird in der Regel aus mehreren Einzelmessungen durch Mittelwertbildung auf den „wahren“ Wert geschlossen.

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{118.5}$$

(siehe dazu auch Teil. X „Messtechnik“, Abschn. 80.3 und Beispiel 3 in Kap. 109).

- *Geometrischer Mittelwert* $\bar{x}_g$: Er ist für diskrete Werte x_i, $1 \leq i \leq N$, definiert:

$$\bar{x}_g = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N} \tag{118.6}$$

- *Median* oder *Zentralwert* x_{50}: Die x-diskreten- Werte sind der Größe nach geordnet. Bei einer ungeraden Anzahl ist der in der Mitte liegende Wert der Median, bei einer geraden Anzahl wird aus den beiden mittleren der arithmetische Mittelwert gebildet. Auch dieser Wert wird in der Messtechnik als Annäherung an den „wahren“ Wert genommen, wenn Ausreißer bei den einzelnen Messwerten eine entsprechend geringe Wertigkeit haben sollen.

- *Varianz σ^2 einer Verteilung*: Mit dem linearen Mittelwert aus (118.4) folgt:

$$\sigma^2 = \sum_{x=0}^{N} (x - \bar{x})^2 \cdot h(x)$$

bzw.

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \cdot h(x)dx \qquad (118.7)$$

- *Varianz s^2 einer Stichprobe*: Sie ist ein Maß für die Streuung der Einzelwerte.

$$s^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2 \qquad (118.8)$$

(siehe dazu auch Teil. X „Messtechnik“, Abschn. 80.3 und Beispiel 4 in Kap. 109).
Der Faktor $N - 1$ im Nenner ergibt sich aus wahrscheinlichkeitstheoretischen Überlegungen und lässt sich praktisch so interpretieren, dass es nicht sinnvoll ist, aus *einem* Messwert die Varianz, d. h. die Streuung, zu bilden: $s^2(N = 1) \to \infty$.

- *Standardabweichung s*:

$$s = +\sqrt{s^2} \qquad (118.9)$$

Sie dient in der Messtechnik dazu, Messreihen zu beurteilen mit der Tendenz, dass um so eher vom Mittelwert der Stichprobe nach (118.5) auf den „wahren“ Wert geschlossen werden kann, je geringer die Standardabweichung ist.

- *Stationärer Prozess*: Die Wahrscheinlichkeitsdichte und damit die Bestimmung der Mittelwerte nach (118.4) bis (118.9) hängt nicht davon ab, zu welchem Zeitpunkt dies geschieht:

$$h(x)\big|_{t=t_1} = h(x)\big|_{t=t_2}$$

Daraus folgt z. B.

$$\bar{x}\big|_{t=t_1} = \bar{x}\big|_{t=t_2} \qquad (118.10)$$

- *Ergodischer Prozess*: Gegeben sind mehrere gleichzeitig ablaufende gleichartige Prozesse (Schar). Sind die Mittelwerte eines Prozesses, gebildet über die Zeit, gleich den Mittelwerten, gebildet über die Schar zum gleichen Zeitpunkt, ist der Prozess stationär und ergodisch. Bei Würfeln folgt die gleiche Aussage z. B. über den Mittelwert (Erwartungswert), unabhängig davon, ob mit einem Würfel N-mal gewürfelt wird oder mit N Würfeln einmal. Dazu ist ein Hinweis erforderlich: Es kann sein, dass man im Experiment nur eine endliche Anzahl von Proben durchführen kann und damit zwangsweise nicht genau zum gleichen Mittelwert (Erwartungswert) kommt.

- *Autokorrelationsfunktion Φ_{ss} (AKF)*: Sie gibt die Ähnlichkeit des Signales mit sich selber in Abhängigkeit von der gegenseitigen Verschiebung τ an:

$$\Phi_{ss}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{T} s(t) \cdot s(t+\tau)dt \qquad (118.11)$$

Sind $s(t)$ und $s(t + \tau)$ orthogonal zueinander, folgt:

$$\Phi_{ss}(\tau) = 0 \qquad (118.11a)$$

Für $\tau = 0$ erhält man:

$$\Phi_{ss}(\tau = 0) = \overline{s^2} = \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{+T} [s(t)]^2 dt \qquad (118.12)$$

Das ist das Quadrat des Effektivwertes und gibt die im Signal enthaltene mittlere Gesamtleistung an ($P \sim \overline{s^2}$). Es wird vorausgesetzt, dass der Wert des Integrals endlich ist.
Das Maximum hat die AKF für $\tau = 0$:

$$\Phi_{ss}(\tau = 0) \geq \Phi_{ss}(\tau \neq 0) \qquad (118.13)$$

Mit $\tau \to \infty$ folgt für ein beliebiges Signal:
Quadrat des zeitlichen Mittelwertes

$$\lim_{\tau\to\infty} \Phi_{ss}(\tau) = \overline{s^2} \qquad (118.14a)$$

Die Autokorrelationsfunktion eines stochastischen Signals mit verschwindendem Mittelwert geht mit zunehmender Verschiebung gegen Null:

$$\lim_{\tau\to\infty} \Phi_{ss}(\tau) = 0 \qquad (118.14b)$$

Mit der Beziehung
Parseval'sche Gleichung:

$$\int_{-\infty}^{+\infty} |s(t)|^2 dt = \int_{-\infty}^{+\infty} |S(f)|^2 df, \qquad (118.15)$$

(siehe auch Erläuterungen zu (112.15)), eingesetzt in (118.12), erhält man:

$$\overline{s^2} = \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{+T} [S(f)]^2 df \qquad (118.16)$$

- *Spektrale Leistungsdichte*, *Theorem von Wiener-Khintchine*: Der Ausdruck

$$\lim_{T\to\infty} \frac{1}{2T} \cdot |S(f)|^2 = S_{ss}(f) \qquad (118.17)$$

aus (118.16) wird als spektrale Leistungsdichte bezeichnet. Damit gilt auch:

$$\overline{s^2} = \int_{-\infty}^{+\infty} S_{ss}(f)df \qquad (118.18)$$

Man kann zeigen, dass die Fouriertransformierte der spektralen Leistungsdichte gleich ist der Autokorrelationsfunktion und umgekehrt (Theorem von Wiener-Khintchine):

$$\mathbf{F}\{S_{ss}(f)\} = \Phi_{ss}(\tau)$$

bzw.

$$\mathbf{F}^{-1}\{\Phi_{ss}(\tau)\} = S_{ss}(f) \tag{118.19}$$

Damit reicht das Vorhandensein der Autokorrelationsfunktion aus, um das spektrale Verhalten eines Vorganges vollständig zu beschreiben. Allerdings enthält die spektrale Leistungsdichte keine Phaseninformation mehr, denn sie ist durch die Betragsbildung verlorengegangen. Damit steht sie auch der Autokorrelationsfunktion nicht zur Verfügung.

Das Leistungsdichtespektrum von Signalen, auch stochastischen, ist durch Bandpassfilter bestimmbar:

$$S_{ss}(f = f_0) = \frac{1}{\Delta f} \cdot \overline{s_0^2} \tag{118.20}$$

Aus (118.16) folgt, dass es einen Vorgang, der bis zu unendlich hohen Frequenzen eine konstante Leistungsdichte hat, nicht geben kann, weil das Integral dann divergiert. Deshalb ist der Begriff „Weißes Rauschen“ so zu verstehen, dass das Spektrum stets frequenzbegrenzt ist. Stationäre stochastische Signale haben keine Fouriertransformierte. Korrelationsfunktionen gehen aber für größere Verschiebungen gegen Null, sodass sie transformierbar sind. Damit wird die Beziehung nach (118.18) möglich.

Allgemein gilt, dass Transformationen umkehrbar sind, Mittelungsoperationen dagegen nicht.

- *Kreuzkorrelationsfunktion* Φ_{s1s2} (*KKF*): Erfasst wird die Ähnlichkeit zweier unterschiedlicher Signale in Abhängigkeit von der Verschiebung t:

$$\Phi_{s1s2}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{T} s_1(t) \cdot s_2(t+\tau)dt \tag{118.21}$$

  s_1, s_2 beliebige zeitkontinuierliche Funktionen.
  Zwei Signale sind voneinander unabhängig (*nicht korreliert*), wenn gilt

$$\Phi_{s1s2}(\tau) = 0 \quad \text{für alle } \tau \tag{118.22}$$

Beispiel 2 zeigt, dass der Kreuzkorrelationsfunktion eine physikalische Bedeutung gegeben werden kann.

Die zwei Signale $s_1(t)$ und $s_2(t)$ haben die Energie

$$E_{s1} = \int_{-\infty}^{+\infty} \{s_1(t)\}^2 dt$$

bzw.

$$E_{s2} = \int_{-\infty}^{+\infty} \{s_2(t)\}^2 dt \tag{118.23}$$

Erfüllen die zwei Funktionen die Bedingung

$$s_1(t) \sim s_2(t), \tag{118.24}$$

hat die Kreuzkorrelationsfunktion ihren größten Wert:

$$\Phi_{s1s2}(0) = \sqrt{E_{s1} \cdot E_{s2}} \tag{118.25}$$

Dies gilt für jedes τ:

$$\int_{-\infty}^{+\infty} \{s_1(t)\}^2 dt = \int_{-\infty}^{+\infty} \{s_1(t+\tau)\}^2 dt \tag{118.26}$$

Danach liefert die Kreuzkorrelierte den Maximalwert, wenn für ein bestimmtes τ gilt:

$$s_1(t) \sim s_2(t+\tau) \tag{118.27}$$

Nach *N. Wiener* gilt auch:

$$\Phi_{ss}(\tau) = \int_{-\infty}^{+\infty} S_1(f) \cdot S_1^*(f) \cdot e^{j2\pi \cdot f} df \tag{118.28}$$

Verglichen mit

$$s_2(t) = \int_{-\infty}^{+\infty} S_1(f) \cdot H(f) \cdot e^{j2\pi \cdot f} df \tag{118.29}$$

folgt daraus:

$s_2\ (t = 0)$ ist maximal, wenn gilt:

$$H(f) = k \cdot S_1^*(f) \tag{118.30}$$

Darin ist k eine dimensionsbehaftete Konstante vom Betrag 1.

Beispiel 1

Es wird mit einem oder mehreren Würfeln mit den Augenzahlen 1 bis 6 gewürfelt. Die Wahrscheinlichkeitsdichtefunktion für das Auftreten der Augenzahlen $1, 2, \ldots, 6$ besteht aus diskreten Werten: $w(1) = w(2) = w(3) = w(4) = w(5) = w(6) = 1/6$. Wird die Augenzahl 4 betrachtet ($x = 4$), ist die Wahrscheinlichkeit für das Auftreten der Augenzahlen $x \leq 4$ nach (118.2): $w(x_0 \leq 4) = w(1) + w(2) + w(3) + w(4) = 2/3$. Für das Auftreten der Augenzahlen 5 und 6 gilt: $w(5) + w(6) = 1/3$, oder, mit (118.3): $w(x_0 > 4) = 1 - w(x_0 \leq 4) = 1 - 2/3 = 1/3$. Zählt man die Augenzahlen der einzelnen Würfe zusammen und teilt durch die Anzahl der Würfe, nähert man sich mit zunehmender Zahl der Würfe dem Erwartungswert $\frac{1+2+3+4+5+6}{6} = 3{,}5$. Der Median berechnet sich bei gerader Anzahl der Werte zu $(3+4)/2 = 3{,}5$. Für die Varianz s^2 erhält man mit (118.8): $s^2 = 3{,}5$. Weiterhin ist der Erwartungswert unabhängig davon, wie die einzelnen Würfe zeitlich liegen, der Prozess ist damit stationär. Er ist auch nicht davon abhängig, ob mit einem Würfel N Würfe nacheinander ausgeführt werden oder ob mit N Würfeln gleichzeitig gewürfelt wird: Der Prozess ist stationär und ergodisch. Hier zeigt sich anschaulich, dass die endliche Anzahl von Proben (Würfen) zu einem geringfügig voneinander abweichenden Erwartungswert für beide Experimente führt. ◀

Beispiel 2

Von den zwei Signalen $s_1(t) = \hat{u}\cdot\sin\omega_0 t$; $s_2(t) = \hat{i}\cdot\sin\omega_0 t$ soll die Kreuzkorrelationsfunktion Φ_{s1s2} nach (118.21) gebildet werden:

$$\Phi_{ui}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T}\hat{u}\cdot\sin\omega_0 t\cdot\hat{i}\cdot\sin\omega_0(t+\tau)dt$$

Unter Verwendung eines Additionstheorems und einiger Rechnung folgt mit $\varphi = \omega_0\cdot\tau$:

$$\Phi_{ui}(\varphi) = \frac{\hat{u}\cdot\hat{i}}{2}\cdot\cos\varphi = \frac{\hat{u}}{\sqrt{2}}\cdot\frac{\hat{i}}{\sqrt{2}}\cdot\cos\varphi$$

Das ist die Wirkleistung in Abhängigkeit von der Verschiebung t bzw. φ.

Wird nur s_1 verwendet und die Autokorrelationsfunktion nach (118.11) gebildet, erhält man:

$$\Phi_{uu}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T}\hat{u}^2\cdot\sin\omega_0 t\cdot\sin\omega_0(t+\tau)dt$$

und daraus $\Phi_{uu}(\varphi) = \frac{\hat{u}^2}{2}\cdot\cos\varphi$.

Die Ähnlichkeit ist maximal, wenn die Zeitverschiebung Null ist. Φ_{uu} stellt, bis auf den Maßstabsfaktor $1/R$ (Ohm'scher Widerstand), die Wirkleistung dar. Anschaulich einsichtig ist, dass sie am größten ist, wenn die Ähnlichkeit des Signals mit sich selbst am größten ist. Der für Φ_{uu} berechnete Ausdruck ist Null für $\varphi = \pm 90°$, und damit sind nach (118.11a) die Sinus- und die Cosinusfunktion orthogonale Funktionen. ◀

118.2 Verteilungen

Um die in der Praxis gegebenen Abhängigkeiten formelmäßig wiedergeben zu können, sind unterschiedliche Wahrscheinlichkeitsfunktionen entwickelt worden. Hier sollen drei betrachtet werden.

118.2.1 Binomialverteilung

Das Kennzeichen dieser Verteilung ist, dass nur zwei Ereignisse auftreten können, die sich aber gegenseitig ausschließen $(E, \overline{E})$. Die zugeordneten Wahrscheinlichkeiten w werden mit p und q bezeichnet:

$$w(E) = p;\quad w(\overline{E}) = 1-p = q \tag{118.31}$$

Das Experiment wird N-mal wiederholt. Es gilt:

$$w(x_0 = x) = h(x) = \binom{N}{x}\cdot p^x\cdot q^{(N-x)} \tag{118.32}$$

Die Funktion $h(x)$ gibt die Wahrscheinlichkeit an, dass bei N Proben (Versuchen) das Ereignis E mit der Wahrscheinlichkeit $w(E) = p$ genau x-mal eintritt und daraus folgend das Ereignis $\overline{E}$ mit der Wahrscheinlichkeit $w(\overline{E}) = 1-p = q$ genau $(N-x)$-mal.

Summenfunktion:

$$w(x_0 \le x) = H(x) = \sum_{m=0}^{N}\binom{N}{m}p^m\cdot q^{(N-m)} \tag{118.33}$$

Mittelwert:

$$\bar{x} = \sum_{x=0}^{N} x\cdot\binom{N}{x}\cdot p^x\cdot q^{(N-x)} = N\cdot p \tag{118.34}$$

Standardabweichung:

$$s = \sqrt{N\cdot p\cdot q} \tag{118.35}$$

Mit entsprechenden Programmen oder Tabellen können die Werte für $h(x)$ und $H(x)$ bei gegebenem p und x auf einfache Weise bestimmt werden.

Beispiel 3

Bei der Fertigung von Geräten sind 5 % fehlerhaft. Zur Qualitätskontrolle werden Stichproben durchgeführt, indem ein beliebiges Gerät entnommen, geprüft und wieder zurückgestellt wird. Anschließend wird das nächste Gerät entnommen usw. Das geschieht insgesamt 50mal. Gesucht ist die Wahrscheinlichkeit, unter den 50 untersuchten Teilen genau fünf fehlerhafte zu finden. Mit $N = 50$, $p = 0{,}05$, $q = 0{,}95$, $x_0 = 5$ ergibt sich:

$$w(x_0 = 5) = h(5) = \binom{50}{5}\cdot 0{,}05^5\cdot 0{,}95^{45} = 0{,}0658$$

Die Wahrscheinlichkeit beträgt 6,58 %. ◀

118.2.2 Poissonverteilung

Betrachtet werden Ereignisse, die zu zufälligen Zeitpunkten stattfinden. x ist die Anzahl der Ereignisse in einem vorgegebenen Zeitintervall, die nicht von der Lage des Zeitintervalls abhängt. Für unterschiedliche aber gleichlange Zeitintervalle ergeben sich auch unterschiedliche Werte für x. Mit steigender Zeitintervalllänge steigt auch die Anzahl der Ereignisse. Die Wahrscheinlichkeit dafür, dass in einem Zeitintervall genau x_0 Ereignisse stattfinden, ist durch folgende Formel gegeben:

$$w(x_0 = x) = h(x) = \frac{a^x}{x!}\cdot e^{-a}\quad \text{mit } a = \bar{x} \tag{118.36}$$

Sie erfüllt also die Nebenbedingung, dass der Mittelwert, gebildet nach (118.4), den Wert a ergibt. Für die Summenfunktion $H(x)$ gilt:

$$H(x) = \sum_{m=0}^{x} \frac{a^m}{m!} \cdot e^{-a} \tag{118.37}$$

Beispiel 4

Messungen in einer verkehrsberuhigten Zone haben folgendes ergeben: 1. Im Mittel fährt alle 10 Minuten ein Auto. 2. Es gilt die Poissonverteilung. Mit $a = 1$ können folgende Aussagen gemacht werden:

1. Wahrscheinlichkeit, dass in 10 Minuten kein Auto fährt:

$$w(x_0 = 0) = h(0) = \frac{1^0}{0!} \cdot e^{-1} = 0{,}368$$

2. Wahrscheinlichkeit, dass in 10 Minuten genau zwei Autos fahren:

$$w(x_0 = 2) = h(2) = \frac{1^2}{2!} \cdot e^{-1} = 0{,}184$$

3. Wahrscheinlichkeit, dass in 10 Minuten höchstens zwei Autos fahren:

$$w(x_0 \le 2) = H(2) = \frac{1^0}{0!} \cdot e^{-1} + \frac{1^1}{1!} \cdot e^{-1} + \frac{1^2}{2!} \cdot e^{-1} = 0{,}925$$

4. Wahrscheinlichkeit, dass in 10 Minuten weniger als zwei Autos fahren:

$$w(x_0 < 2) = w(x_0 \le 1) = H(1) = 0{,}74$$

5. Wahrscheinlichkeit, dass in 10 Minuten mindestens drei Autos fahren:

$$w(x_0 \ge 3) = 1 - w(x_0 \le 2) = 1 - H(2) = 1 - 0{,}925 = 0{,}075$$

6. Wahrscheinlichkeit, dass in 10 Minuten mehr als drei Autos fahren:

$$w(x_0 > 3) = 1 - w(x_0 \le 3) = 1 - H(3) = 1 - 0{,}98 = 0{,}02$$

◀

118.2.3 Normalverteilung, Gaußverteilung

Die Normal- oder Gaußverteilung wird sehr häufig eingesetzt. Sie beschreibt statistische Vorgänge immer dann, wenn viele unterschiedliche Einflussgrößen wirksam sind und diese sich untereinander nicht beeinflussen (siehe dazu „Zentraler Grenzwertsatz der Statistik“ unten). Dichte- und Summenfunktion sind wie folgt definiert:

$$h(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot (\frac{x-\bar{x}}{\sigma})^2} \tag{118.38}$$

$$w(x_0 \le x) = H(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{y=-\infty}^{x} e^{-\frac{1}{2} \cdot (\frac{y-\bar{x}}{\sigma})^2} dy \tag{118.39}$$

Diese Art der Darstellung wird mit $(\bar{x}, \sigma)$-Normalverteilung bezeichnet. Darin ist $\bar{x}$ der Erwartungswert und σ die Standardabweichung. Die Funktion $h(x)$ hat einen glockenförmigen Verlauf („Gauß’sche Glockenkurve“) und ist symmetrisch zum Mittelwert $\bar{x}$. Die Wendepunkte liegen bei $x_{1,2} = \bar{x} \pm \sigma$. Durch eine Koordinatentransformation kann der Mittelwert in den Koordinatenursprung gelegt werden:

$$h^*(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot y^2} \quad \text{mit } y = \frac{x - \bar{x}}{\sigma} \tag{118.40}$$

$$H^*(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-v^2/2} dv \tag{118.41}$$

Sie wird $(0, 1)$-Normalverteilung genannt, ihre Wendepunkte liegen bei $\pm y$. Folgende Angaben werden häufig benutzt:

Im Bereich $-1 \le y \le +1$ liegen 68,3 % der Werte (σ-Grenze),
im Bereich $-2 \le y \le +2$ liegen 95,5 % der Werte (2σ-Grenze),
im Bereich $-3 \le y \le +3$ liegen 99,7 % der Werte (3σ-Grenze). (118.42)

Außerhalb der 3σ-Grenze liegen demnach nur noch 0,3 % der Zufallsgrößen.

- *Zentraler Grenzwertsatz der Statistik*: Die Summe aus N voneinander unabhängigen stochastischen Vorgängen mit untereinander gleicher, aber beliebiger Wahrscheinlichkeitsdichte strebt für $N \to \infty$ gegen die Gaußverteilung.

Beispiel 5

Die Lebensdauer von Leuchtmitteln (mit Glühfaden) bei Nennspannung beträgt etwa 1200 Stunden und ist normalverteilt. Die Standardabweichung liegt bei etwa 100 Stunden. Die Wahrscheinlichkeit, dass das Leuchtmittel mindestens 1400 Stunden brennt, berechnet sich zu:

$$w(x_0 > 1400) = 1 - w(x_0 \le 1400) = 1 - H(1400) = 1 - 0{,}977 = 0{,}023 = 2{,}3\,\%.$$

Die Wahrscheinlichkeit für eine mindestens 1000stündige Brenndauer beträgt:

$$w(x_0 > 1000) = 1 - w(x_0 \le 1000) = 1 - 0{,}023 = 0{,}977 = 97{,}7\,\%.$$

Die Wahrscheinlichkeit für eine Brenndauer im Bereich (1000 ... 1400) h liegt bei

$$w(1000 \le x_0 < 1400) = H(1400) - H(1000) = 0{,}954 = 95{,}4\,\%. \quad \blacktriangleleft$$

118.3 Rauschen

Rauschen ist ein zufälliges Signal, das durch die ungeordnete Wärmebewegung von Ladungsträgern hervorgerufen wird. Im Teil XIII „Nachrichtentechnik“, Abschn. 92.4.11.1 „Rauschen“, sind die wichtigsten Rauschquellen und deren Rauschspannungen bzw. -ströme dargestellt. Hier sollen weitere Einzelheiten behandelt werden.

- *Zusammenschalten mehrerer rauschender Elemente*:
  Ein Widerstand liefert die folgenden Größen:
  Rauschspannung am Widerstand

$$\bar{u}_R = \sqrt{4kTBR} \tag{118.43}$$

  Rauschstrom im Widerstand

$$\bar{i}_R = \sqrt{4kTB/R} \tag{118.44}$$

  k Boltzmann-Konstante $1{,}38 \cdot 10^{-23}$ W s/K, T absolute (thermodynamische) Temperatur in K, B Bandbreite (ausgewerteter oder berücksichtigter Frequenzbereich) in Hz, R Widerstand in Ω.

Das Ersatzschaltbild eines rauschenden Widerstandes ist in Abb. 118.2 dargestellt. Zu dem als rauschfrei angenommenen Widerstand R wird eine Rauschspannungsquelle mit einer Spannung nach (118.43) in Reihe geschaltet bzw. eine Stromquelle mit einem Rauschstrom nach (118.44) parallelgeschaltet. Da es sich um statistische Größen handelt, wird der Mittelwert eingesetzt. Bei der Reihenschaltung zweier rauschender Widerstände werden die rauschfreien Anteile sowie die Quadrate der Rauschspannungen addiert. Das ist möglich und erforderlich, weil die einzelnen Rauschspannungen nicht korreliert sind (Abb. 118.2b):

$$\begin{aligned} R_g &= R_1 + R_2; \\ \overline{u_{Rg}^2} &= 4kT_1BR_1 + 4kT_2BR_2 \\ &= 4kB(T_1R_1 + T_2R_2) \end{aligned} \tag{118.45a}$$

Dabei können die Widerstände unterschiedliche Temperatur haben, die Bandbreite ist in der Regel als ausgewerteter Frequenzbereich für beide Widerstände gleich.

Sind die Widerstände parallelgeschaltet, wird das Stromquellenersatzschaltbild eingesetzt, und man erhält nach Abb. 118.2b:

$$\begin{aligned} R_g &= \frac{R_1R_2}{R_1 + R_2}; \\ \overline{i_{Rg}^2} &= 4kT_1B/R_1 + 4kT_2B/R_2 \\ &= 4kB(T_1/R_1 + T_2/R_2) \end{aligned} \tag{118.45b}$$

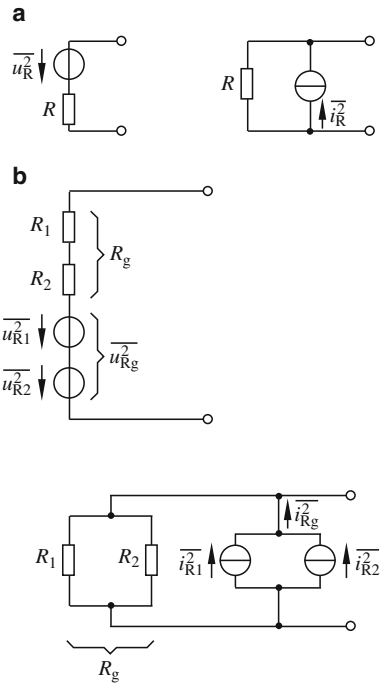


Abb. 118.2 Widerstände als Rauschquelle. **a** Ersatzschaltbild, **b** Zusammenschaltung von rauschenden Widerständen

Es ist möglich, für die Rauschquellen eine Spannungs- bzw. Stromrichtung anzugeben, obwohl das nach (118.43) und (118.44) nicht sinnvoll ist.

Enthält ein Signalkreis z. B. neben einem rauschenden Widerstand eine weitere Rauschquelle, die z. B. das Schrotrauschen erfasst (Teil XIII „Nachrichtentechnik“, Abschn. 92.4.11.1):

$$\overline{i_s^2} = 2 \cdot q_e \cdot I \cdot B \tag{118.46}$$

(q_e Elementarladung des Elektrons $1{,}6 \cdot 10^{-19}$ A s, I Anoden- bzw. Halbleiterstrom in A, B Bandbreite in Hz) folgt für das Quadrat des Gesamtrauschstromes $\overline{i_g^2}$:

$$\overline{i_g^2} = \overline{i_R^2} + \overline{i_s^2} = 4kTB/R + 2q_eIB \tag{118.47}$$

Formelzeichen siehe (118.44) und (118.46).

Für die Korrelationsfunktionen gilt:

$$\Phi_{NN}(\tau) = 0 \quad \text{für } \tau \neq 0 \quad \text{(theoretisch)} \tag{118.48a}$$

$$\Phi_{s1N} = 0 \quad \text{für alle } \tau \quad \text{(theoretisch)} \tag{118.48b}$$

Indizes: N für Rauschen (Noise), $s1$ beliebiges Signal außer Rauschen.

Diese theoretischen Werte werden in der Praxis nur angenähert erreicht.

- *Kennzeichnung von stochastischen Signalen*:
  Da die Signale in ihrem Verlauf nicht vorhersagbar sind, können nur statistische Größen zur Kennzeichnung verwendet

werden. Ist die Entstehung des stochastischen Signals bekannt, lassen sich Leistungsdichte und relative Häufigkeiten im voraus berechnen. Sie beschreiben dann das Signal.

Beispiel 6

Für den Eingangswiderstand eines Oszilloskopes gelten folgende Werte: $R = 1\,\mathrm{M\Omega}$, Temperatur $T = 300\,\mathrm{K}$ ($\hat{=} \approx 27\,^\circ\mathrm{C}$), berücksichtigte Bandbreite $B = 10\,\mathrm{MHz}$. Dann berechnet sich die Rauschspannung nach (118.43) zu:

$$\begin{aligned}\bar{u}_R &= \sqrt{4kTBR} \\ &= \sqrt{4 \cdot 1{,}38 \cdot 10^{-23} \frac{\mathrm{V\,A\,s}}{\mathrm{K}} \cdot 3 \cdot 10^2\,\mathrm{K} \cdot 10^7 \frac{1}{\mathrm{s}} \cdot 10^6 \frac{\mathrm{V}}{\mathrm{A}}} \\ &= 407\,\mu\mathrm{V}\end{aligned}$$

Eine Rauschspannung dieser Größenordnung müsste am Bildschirm sichtbar sein, wenn weitere Rauschquellen vernachlässigt werden und wenn nicht zwei Effekte den Wert verringerten: 1. Am Oszilloskopeingang liegt ein Kondensator von etwa 50 pF parallel zum Widerstand, sodass die höherfrequenten Rauschanteile in ihrer Amplitude abnehmen. 2. Der Innenwiderstand der den Oszilloskopeingang speisenden Quelle ist häufig niedriger als der Eingangswiderstand des Oszilloskopes. ◀

118.4 Signalerkennung bei gestörter Übertragung

An dieser Stelle können nur die Grundüberlegungen vorgestellt werden, weitere Einzelheiten sind z. B. in [6] enthalten. Einige zusätzliche Informationen können im Teil XIII „Nachrichtentechnik“, Kap. 108, nachgesehen werden.

118.4.1 Erkennen versteckter Periodizitäten

Ein Signal $s_g(t)$ besteht aus der Summe von Nutzsignal $s(t)$ und Störsignal $n(t)$: $s_g(t) = s(t) + n(t)$. Wird von $s_g(t)$ die Autokorrelationsfunktion gebildet, entstehen die folgenden Kombinationen:

$$\Phi_{s_g s_g}(\tau) = \Phi_{ss}(\tau) + \Phi_{nn}(\tau) + \Phi_{sn}(\tau) + \Phi_{ns}(\tau).$$

Darin sind $\Phi_{ss}(\tau)$ und $\Phi_{nn}(\tau)$ Autokorrelationsfunktionen, $\Phi_{sn}(\tau)$ und $\Phi_{ns}(\tau)$ Kreuzkorrelationsfunktionen. Sind Nutz- und Störsignal unkorreliert, was in der Regel der Fall ist, ergeben die zwei Kreuzkorrelationsfunktionen Null. Gilt für das Nutzsignal z. B. $s(t) = \hat{u} \cdot \sin(\omega t + \varphi)$, berechnet sich Φ_{ss} mit (118.11) zu:

$$\Phi_{ss}(\tau) = \frac{\hat{u}^2}{2} \cdot \cos \omega\tau$$

Der Wert der Autokorrelationsfunktion des Störsignals, das sein Entstehen dem Zufall verdankt, nimmt für wachsendes τ stetig ab. Dagegen ist $\Phi_{ss}(\tau)$ periodisch mit konstantem Scheitelwert, sodass das Verhältnis von Nutz- zu Störsignal mit wachsendem τ immer größer wird. In der Praxis können nur endliche Zeiten verwendet werden, und damit sind die „Kurzzeitkorrelierten“ $\Phi_{nn}(\tau)$, $\Phi_{sn}(\tau)$ und $\Phi_{ns}(\tau)$ noch nicht Null und erschweren die Signalerkennung.

118.4.2 Signalerkennung allgemein

Die Grundlage bildet (118.27). Das Signal $s_1(t)$ wird gespeichert und anschließend die Kreuzkorrelierte mit den nachfolgenden Signalen gebildet. Sind für ein bestimmtes τ das gespeicherte und das ankommende Signal einander proportional, ergibt sich ein maximaler Wert im Korrelator. Er ist gleich dem Maximalwert der Autokorrelationsfunktion Φ_{ss}.

118.4.3 Signalangepasste Filter (matched filter)

Nach (118.30) muss für ein optimales Filter gelten: $H(f) = k \cdot S_1^*(f)$. Das setzt voraus, dass das zu empfangende Signal in seinem Verlauf bekannt ist, was aber häufig der Fall ist. Im Teil XIII „Nachrichtentechnik“, Abschn. 108.2, sind weitere Einzelheiten aufgeführt.

Es kann noch eine Aussage über den Einfluss eines signalangepassten Filters auf das Verhältnis von Nutz- zu Störsignal gemacht werden. Die Spektraldichte von weißem Rauschen sei N_0, dann folgt daraus mit (118.12) und (118.15):

$$\lim_{T\to\infty} \frac{1}{2T} \cdot |S_N(f)|^2 = \frac{N_0}{2} \tag{118.49}$$

$S_N(f)$ im Intervall $-T \ldots +T$ enthaltener Rauschanteil.

Die im Signal $s_1(t)$ enthaltene Energie ist nach einem entsprechenden Ansatz

$$\int_{-\infty}^{+\infty} |S_1(f)|^2 df = E_{S1} \tag{118.50}$$

Das quadratische Verhältnis von Nutz- zu Störsignal beträgt:

$$a_{\mathrm{SN}}^{(2)}\big|_{\mathrm{Korr}} = \frac{2E_{S1}}{N_0} \tag{118.51}$$

Bei einem Bandfilter berechnet sich das entsprechende quadratische Verhältnis (hier ohne Ableitung) zu:

$$a_{\mathrm{SN}}^{(2)}\big|_{\mathrm{BaFi}} = \frac{E_{S1}}{N_0 BT} \tag{118.52}$$

B Bandbreite des Filters in Hz, T Signaldauer in s.

Darin ist für das – ideale – Filter angenommen worden: $|H(f)| = 1$ im Bereich $-B \ldots +B$, 0 sonst.

Bildet man den Quotienten

$$\frac{a_{\mathrm{SN}}^{(2)}\big|_{\mathrm{Korr}}}{a_{\mathrm{SN}}^{(2)}\big|_{\mathrm{BaFi}}} = \frac{\frac{2E_{\mathrm{S1}}}{N_0}}{\frac{E_{\mathrm{S1}}}{N_0 BT}} = 2BT \qquad (118.53)$$

so folgt daraus: Das quadratische Verhältnis von Nutz- zu Störsignal ist bei einem signalangepassten Filter gegenüber einem – idealen – Bandfilter um so größer, je größer die Bandbreite und die Signaldauer sind.

Beispiel 7

Die praktische Anwendung von Rauschsignalen und der Korrelation soll an einigen Beispielen gezeigt werden [6].

1. In der Automobiltechnik werden zum Testen der Federungseinrichtungen stochastische Signale eingesetzt, weil sie alle Möglichkeiten der in der Praxis vorkommenden Belastungsfälle erfassen.
2. Zum Messen der Schalldämmung zwischen nebeneinanderliegenden Räumen werden in einem Raum Rauschsignale über Lautsprecher ausgesendet und im Nachbarraum über Mikrofon empfangen. Dieses Verfahren hat gegenüber der Aussendung von Sinussignalen veränderbarer Frequenz den Vorteil, dass im Senderraum keine stehenden Wellen auftreten und das Ergebnis verfälschen.
3. Zum Testen der Störsicherheit von elektronischen Geräten (z. B. Router) werden Rauschsignale verwendet, weil bei genügend langer Testzeit so gut wie alle Möglichkeiten der in der Praxis auftretenden Störsignale vorhanden waren.
4. Im Teil X „Messtechnik“, Abschn. 85.4.2.3, ist ein Verfahren zur Abstandsmessung unter Verwendung der Korrelation beschrieben. ◄

Literatur

Literatur allgemein

1. *Holbrook, J. G.:* Laplace-Transformationen. Verlag Vieweg + Sohn Braunschweig. 1970.
2. *Küpfmüller, K.:* Die Systemtheorie der elektrischen Nachrichtenübertragung. Verlag S. Hirzel Stuttgart. 3. Auflage 1968.
3. *Weißgerber, W.:* Elektrotechnik für Ingenieure 3. Verlag Vieweg-Teubner Wiesbaden. 7. Auflage 2009.

Literatur zu Kap. 117

4. *Lechner, W. und Lohl, N.:* Analyse digitaler Signale. Verlag Vieweg Braunschweig/Wiesbaden. 1990.
5. *Doetsch, G.:* Anleitung zum praktischen Gebrauch der Laplace- und der Z-Transformation. Verlag Oldenbourg München Wien. 4. Auflage 1981.

Literatur zu Kap. 118

6. *Ehrenstrasser, G.:* Stochastische Signale und ihre Anwendung. Verlag Hüthig Heidelberg. 1974.

Sachverzeichnis

default
(Elektro-)Dynamisches Mikrofon, 1287
(Elektro-)Magnetisches Mikrofon, 1287
(Elektro-)Magnetisches System, 1289
(Feld-)Wellenwiderstand, 1181
(Kreis-)Frequenzhub FM, 1212
(Kreis-)Frequenzhub PM, 1212
$\lambda/2$-Dipol, 1195, 1199
$\lambda/2$-Faltdipol, 1199
$\lambda/4$-Transformation, 1187
$1/f$-Rauschen, 1139
1-aus-*n*-Decoder, 615
3/2-Wegeventil, 798
3-Excess-Code, 607
3-state, 640
4B/5B-Code, 759
4B3T-Code, 1302
5/2-Wegeventil, 798
5/3-Wegeventil, 798
7-Segment-Anzeigen, 534
80C515, 717
8051, 710
8085, 697, 718, 722, 726
8155, 700, 701
8255, 693
Überlagerungsempfänger, 1245

A
AB-Betrieb, 478
Abbildungsgleichung, 242
Abbildungsmaßstab, 242
A-Betrieb, 478
Abfallzeit, 500
Abgasbehandlung, 1017
Abgleichverfahren, 906
abhängige Variablen, 1185
Abhängigkeitsnotation, 572
Ablauf-Funktionsplan, 808
Ablaufkette, 808
Ablaufsteuerung, 808
Ablauftabelle, 570
Abschirmung, 944
absolute Dielektrizitätskonstante, 294
absolute Messabweichung, 877
absoluter elektroakustischer Übertragungsfaktor, 1284
absoluter Leistungspegel, 1144
absoluter Spannungspegel, 1144
absoluter Strompegel, 1144
Absorptionsgesetz, 594
Abstandsmessung, 960
Abtastfrequenz, 1138
Abtast-Halte-Glied, 953, 955, 1225
Abtasttheorem, 1138
Abtasttheorem, im Frequenzbereich, 1138
Abtasttheorem, im Zeitbereich, 1138
Abtastzeit T_A, 824, 835
Abweichungsfortpflanzung, 875, 879
Abweichungsgrenzen, 877, 959
active-stall-Regelung, 1024
Adaptive Deltamodulationsverfahren, 1228
Addierer, 611–614
Addierers, 491
Adressbus, 667
Adressdecoder, 690, 693
Adressdecodierer, 690
Adressen, 666
Adressenzwischenspeicher, 666
Adressierung, 723
Adressierungsarten, 723
Adress-Latch, 699
ADSL, 1304
Aiken-Code, 607
Akkumulator, 666
Aktionsblock, 809
aktiven Hochpass-Filter, 496
akustischer Kurzschluss, 1290
Akzeptanzwinkel, 1270
ALGOL, 743
Aliasing, 953, 955
Aliasing-Effekt, 1224
Allpass, 858
Allpassfilter, 1244
Alternativ-Verzweigung, 809
ALU, 663, 666, 709
Amplitudenmodulation, 1205
Amplitudenrand, 864
Amplitudenreserve, 864
Amplitudenspektrum, 1133, 1340
Amplitudenspektrum, komplexes, 1335
Amplitudenumtastung (ASK), 1217
Amplitudenverstärkung, 492
Analog-Digital-Wandler, 505
Analoge Signale, 1130
analoge Signale, 538
Analogverstärker, 450
Analyse, 599, 600
Ankerrückwirkungen, 992
Anlagenschutz, 1073
Anlassverfahren, 979
Anlaufmoment, 993
Anomalie des Wassers, 215
Anpassung, 1165
Anpassungsfaktor, 1171
Anregelzeit, 867
Anreicherungstyp, 434
Anschlusskennzeichnung, 566
Anschlussplan, 570
Anstiegsgeschwindigkeit, 507
Anstiegszeit, 500
Antennengewinn, 1196
Antennengewinnfaktor, 1196
Antennengewinnmaß, 1196, 1262
Antennenverlängerung, 1195
Antialiasingfilter, 1368
Antidiffusionsspannung, 396
Antiferromagnetismus, 348
$\underline{A}$-Parameter, 1150
Applikationsberichte, 1086, 1115
Approximation, 958
Äquipotentiallinien, 326
Arbeit, 203, 320
Arbeit im Wechselstromkreis, 361, 379
Arbeitsgerade, 450
Arbeitspunkt, 450, 849
Arbeitspunkt-Stabilisierung, 455
Arithmetischer Mittelwert, 875, 878, 882
arithmetischer Mittelwert, 408, 958, 1140
Aron, 913
Array, 634
ASCII-Code, 608
ASI, 947, 949
AS-i-Bus, 814
ASIC, 647
Assembler, 719, 726, 727, 742
Assoziativgesetz, 593
Astabile Kippstufen, 631
Asynchronzähler, 627
Ätzen, 574
aufgelöster Triebstrang, 1024
Auflösung, 739
Aufruf von Funktionen in AWL, 803, 805
Aufruf von Funktionsbausteinen in AWL, 803, 804
Aufruf von Funktionsbausteinen in FBS, 803, 804
Aufstellen von Transformatoren, 973
Auftrieb, 211
Auftriebskraft, 211
Auge, 244
Ausbreitungsgeschwindigkeit, 228
Ausbreitungskoeffizient, 1164, 1180
Ausgabeeinheit, 710
Ausgabeport, 692
Ausgangsreflexionsfaktor, 1184
Ausgleichsgerade, 959
Ausgleichsleitungen, 928
Ausgleichsstrom, 316, 971
Ausgleichszeit, 828
Auslösecharakteristik, 1074
Ausregelzeit, 838, 867
Ausschaltzeit, 500
Ausschlagverfahren, 906
Außenpolmaschine, 985
Austrittsarbeit, 252

© Springer Fachmedien Wiesbaden GmbH 2016
W. Plaßmann, D. Schulz (Hrsg.), *Handbuch Elektrotechnik*, DOI 10.1007/978-3-658-07049-6

Auswahl von Motoren, 963, 1001
Autokorrelation, 960
Autokorrelationsfunktion, 1140, 1377
Automatische Verstärkungsregelung, 1246
Automatische Verstärkungsregelung, 1245
Automatisierung, 774
Automatisierung von Messabläufen, 960
Automatisierung von Messabläufen, 957
Autonegotiation, 761
Avogadro-Konstante, 216

B
B2B, 748
Bahngeschwindigkeit, 198
Bandbreite, 377, 1205, 1230
Bändchenmikrofon, 1287
Bandfilter, 1234
Bandpass, 496
Bandpass aus *R*-*C*-Hochfilter, 1232
Bandpass aus Tiefpassfilter, 1232
Bandpassfilter, 1230
Bandpassschaltung, 377
Bandsperre, 496, 1230
barometrische Höhenformel, 211
BASIC, 743
Basisband, 1204
Basisband-Durchschaltung, 1267
Basisbandsignal, 1204
Basisbandübertragung, 751
Basiseinheiten, 190, 876
Basisschaltung, 454
Basisspannungsteiler, 455
Basisvorspannung, 455
Batch-Betrieb, 748
Baud, 758
Bauform, 1001
Bauformen, 555
Baugrößen, 1001
Baumnetz, 753
Bausteine, 654
BBAE, 1304
B-Betrieb, 478
BCD-Code, 606, 607
BCD-Zähler, 629
Bedienfeld, 812
Bedingte Wahrscheinlichkeit, 1310
Befehl, 742
Befehle, 720
Befehlsaufbau, 719
Befehlsdarstellung, 719
Befehlssatz, 662, 718
Befehlsvorrat, 661, 726
Befehlszähler, 663, 664
Befehlszeit, 724
Befehlszeiten, 723
Befehlszyklus, 723, 725
Begrenzungspegel, 408
Beharrungszustand, 826, 828
belasteter Generator, 1185
Beleuchtungsanlagen, 1075, 1077
Beleuchtungsstärke, 526, 1078
Beleuchtungswirkungsgrad, 1080
Benutzungsstundenzahlen, 1052
Berechnung von Beleuchtungsanlagen, 1080
Bernoulli-Gleichung, 212
Beschleunigung, 194
Beschleunigungsaufnehmer, 941
Bessel-Filter, 1241
Besselfunktionen, 1213
Bestimmungszeichen für Aktionen, 810
Betragskennlinie, 855
Betragsoptimum, 871
Betriebsarteneteil, 811
Betriebsgrößen, 453
Betriebsmittel, 549, 564
Betriebsmittelkennzeichnung, 566
Beugung, 231, 246
Beugung am Gitter, 247
Beugung am Spalt, 246
Bezeichnung von Elektroblechen, 287
Bezeichnung von Kabeln, 1067
BFOC(ST)-Verbindung, 1275
bibliotheksfähige Programme, 782
BIBO-Stabilität, 860
BICMOS, 645
BiCMOS, 634
Bildebene, 1343
BIMOS-Schaltstufe, 1101
Binär-Code, 606
binäre System, 590
Binomialverteilung, 1379
Biomassekraftwerk, 1033
Biot-Savart'sches Gesetz, 339
Bi-Phase-Verfahren, 1228
BISI, 1031
bistabile Kippstufe, 617
Bit, 590
bleibende Regeldifferenz, 833
Blindleistung, 380, 911
Blindleistungsmessung, 914
Blindleistungsstellung, 1043
Bloch-Wände, 346
Blockcodes, 1312
Blockheizkraftwerk (BHKW), 1019
Blockierbereich, 416
Blockierzustand, 1102
Blockkondensator, 328
Bluetooth, 1307
Bode-Diagramm, 492, 855
Bodenwelle, 1200
Bohr'sches Magneton, 345
Boltzmann-Konstante, 216, 250
Bottom-Up-Methode, 745, 746
Brechung, 230, 1270
Brechungsgesetz, 240
Brechungsindex, 240
Breitbanddipol, 1199
Breitband-Frequenzmodulation, 1213
Breitbandübertragung, 751
Brennstoffzelle, 1019
Brennstoffzellen-Stack, 1020
Brennstoffzellen-Typen, 1020
Brennweite, 242
Brewster'scher Winkel, 248
Brückenspannung, 318
Brummspannung, 409
Bürstenfeuer, 992
Buschbeck-Diagramm, 1186
Business-to-Business-System, 748
Busnetz, 752
Bustiming, 667
BUS-Treiber, 690
Butterworth-Filter, 1241
Byte, 590

C
C, 743
CAD (Computer Aided Design), 582
Candela, 250
CAS, 686
Cassegrain-Antenne, 1200
Cauer-Filter, 1241
CD-ROM, 735
CD-ROM-Laufwerk, 730
CELLFLEX-Kabel, 1178
Central-Process-Unit, 662
Chi-Quadrat-Test, 883, 958
CISC-Prozessoren, 662
Client, 749
Client/Server-Netz, 749
CMOS, 634, 642, 643
CMOS-Bausteine, 644
CMOS-Familie, 644
CMOS-Inverter, 642
CMOS-Schalter, 644
CMOS-Technik, 521
Code, 589, 606
Code, alphanumerischer, 606
Code, bewerteter, 606
Code, einschrittiger, 606
Code, gewichteter, 606
Code, mehrschrittiger, 606
Code, numerischer, 606
Code, tetradischer, 606
Code, unbewerteter, 606
Code, ungewichteter, 606
Codetabellen, 606
Codewandler, 614
Codierer, 614
Codierung, 606
COFDM, 1254
Compact Disc, 735
Compiler, 744
Complementary MOS, 642
Computer, 730
Coulomb'sches Gesetz, 324
C-Parameter, 1150
CPU, 664, 665, 667, 671, 672, 674, 709, 723
CSMA/CD, 761
CSMA/CD-Verfahren, 1300
Curie-Gesetz, 346
Curie-Weiß'sches Gesetz, 348
CW-Radar, 1278

D
Dahlander, 983
Dampfkraftwerk, 1015
Dämpfung, 496, 852
Dämpfungsfaktor, 1143
Dämpfungskoeffizient, 1164
Dämpfungskoeffizienten, 1181
Dämpfungsmaß, 1141, 1144, 1145, 1230
Dämpfungsmaß bei Bandfiltern, 1235
Dämpfungsverzerrungen, 1141
Darlington, 1095
Daten, 666, 782, 1130
Datenbus, 662
Datenreduktion, 1312
Datenreduzierung, 1253
Datenspeicher, 714
Datentransport-Befehlen, 720
Datentyp, 782
Dauergrenzstrom, 1102

Dauerkurzschlussstrom, 969
De Morgan'sche Gesetz, 594
Decoder, 614, 667, 691
Decodierer, 692
Decodierung, 606, 687
Decodierung, unvollständige, 687
Decodierung, vollständige, 687
Deemphase, 1214
Dehnung, 210
Dehnungsmessstreifen, 930
Deklaration einer Funktion, 781
Deklaration eines Funktionsbausteins, 781
Delay-Flipflop, 619
Delon-Schaltung, 412
Deltamodulation (DM), 1228
Delta-Sigma-Modulation, 1228
Demodulation von AM, 1207
Demodulation von Frequenzmodulation, 1216
Demodulation von Phasenmodulation, 1216
Demultiplexer, 615, 617
Denial-of-Service-Angriff, 748
Dense Wavelength Division Multiplexing, 751
D-Flipflop, 619
d-Funktion, 1355, 1356
D-Glied, 848, 851
DHCP, 769
DIAC, 1106
Diagramme, 564
Diamagnetismus, 345
Dielektrika, 327
dielektrische Polarisation, 293, 294
dielektrische Verlustzahl, 296
Differentiators, 494
Differenzierglied, 848
Differenz-Pulscodemodulation, 1253
Differenz-Pulscodemodulation (DPCM), 1228
Differenzverstärker, 486
Differenzverstärkung, 486
Diffusion, 396
Diffusionsverfahren, 402
Digitalbausteine, 639
digitale Abtastregelung, 831, 835
Digitale Filter, 960, 1236
Digitale Signale, 1130
digitale Signale, 538
digitaler Schrittregler mit PI-Verhalten, 831, 837
Digitalisierung, 953, 954
Digitalisierungsfehler, 954
Digitalmultimeter, 897, 898
Digitalschaltungen, 634
Digitaltechnik, 590
Digitalvoltmeter, 897, 898, 902
Dimmer, 1107
Dioptrie, 242
Dipol, 325, 1194
Dirac-Impuls, 1355, 1356
direkt gekoppelter Asynchrongenerator, 1026
direkt gekoppelter Synchrongenerator, 1028
Disjunktive Normalform, 594, 794
Diskette, 730
Diskettenlaufwerk, 732
Dispersion, 240, 241, 1166
Distributivgesetz, 593
DMA, 669
DM-Verfahren, 1228
DNS, 768
Domain Name Service, 768
Doppel-Kreisdiagramm, 1189
Doppelschlussmotor, 997
doppelt gespeister Asynchrongenerator, 1027
Doppel-T-Filter, 1159
Dopplereffekt, 227, 231
Dot Pitch, 739
Dotierung, 395
Drahtbruchsicherheit, 793, 794
Drahtpotentiometer, 923
DRAM, 685
Dreheisen-Messwerk, 885, 887
Drehfeld, 976
Drehfrequenz, 934
Drehimpuls, 207
Drehmomente, 932
Drehspul-Messwerk, 885, 886
Drehstromasynchronmaschine, 976
Drehstrombrücken, 1000
Drehstrommaschinen, 963, 976
Drehstromnetz, 1034
Drehstromsynchronmaschinen, 985
Drehstrom-Transformatoren, 970
Drehzahl, 198, 934
Drehzahlsteuerung, 981, 993
Dreieckschaltung, 385, 1035
Drei-Finger-Regel, 341
dreiphasiger Wechselrichter, 1028
Driftgeschwindigkeit, 306
Drosselspule, 1071
Drosselspulen, 974
Druck, 210
Drucker, 740
Druckluftspeicher, 1050
Druckwasserreaktor, 1021
DSL, 1303
DSQ-Kodierung, 760
DSR – Digitales Satellitenradio, 1267
Dual-Code, 606
Dual-Gray-Codierer, 601
Dualsystem, 602
Dualzahl, 603
Dualzahlen, 603
Duhamel'sches Integral, 1356
Durchbruch 2. Art, 1094
Durchflussmessung, 935
Durchflutung, 338
Durchlassbereich, 1230
Durchlassrichtung, 1088
Durchlassstrom, 1088
Durchlassverzögerungszeit, 1088
Durchtrittsfrequenz, 863
DVBT, 1251
DWDM, 751
Dynamik, 1132
Dynamische Spannungsregelung, 1041
dynamischen Viskosität, 212
dynamischer Durchlasswiderstand, 400
dynamischer Haltestrom, 1103
Dynamisches System, 1289

E

E/A-Bausteinen, 699
E_{01}-Welle im Rundhohlleiter, 1180
E_{11}-Welle im Rechteckhohlleiter, 1180
E-Banking, 748
Echtzeitübertragung, 1133
ECL, 634, 641
Editor, 744
EEPLD, 648
EEPROM, 682, 683
Effektivwert, 363, 875, 882, 1140
Eichbehörde, 876
Eichen, 876
Eigenleitfähigkeit, 274
Eigenleitung, 394
Eigensicherheit, 1110
Eindiodenmodell, 1031
Eindringtiefe, 1177, 1178
Einer-Komplement, 613
Einerkomplement, 603
Eingabeeinheit, 710
Eingabeport, 691
Eingangsimpedanz, 1167
Eingangsreflexionsfaktor, 1184
Eingangsruheströme, 489
Eingangsspannungsteiler, 902
Eingangswechselspannung, 450
Eingangswiderstand, 1197
Einheiten, 875, 876
Einleiterstromwandler, 975
Einmoden-Stufenindex, 1272
Einphasen-Asynchronmotoren, 963, 988
einphasiger Wechselrichter, 1028
Einquadrantenantrieb, 999
Einrichtungstreiber, 690
Einschaltzeit, 500
Einseitenbandmodulation (ESB, SSB), 1209
Einstellregel, 871
Eintakt-Betrieb, 478
Einzelelement-Variable, 782
Einzelkompensation, 1076
EIRP, equivalent isotropically radiated power, 1268
Eisenkerne, 964
Eisenverlustleistung, 966
elastischer Stoß, 205
Elastizitätsmodul, 210
Elektret-Kondensatormikrofon, 1288
Elektrisch kurze Leitung, 1168
elektrisch kurze Leitung, 1170
elektrisch lange Leitung, 1170
elektrische Feldkonstante, 294, 324
elektrische Feldstärke, 324
Elektrischer Wirkungsgrad, 1016
Elektrizitätswirtschaft, 1010
elektroakustischer Übertragungsfaktor, 1284
elektroakustisches Übertragungsmaß, 1284
Elektrochemie, 257, 264
Elektrochemische Spannungsreihe, 264
Elektrodynamisches Messwerk, 885, 887
elektromagnetischen Schwingungen, 526
Elektronegativitätsskala, 264
Elektronenvolt, 251, 326
elektronischer Schalter, 1110
Elektrostatisches System, 1289
elektrotechnische Normung, 562
Elementarladung, 306
Elementarvierpole, 1150
Elementarwelle, 230
elliptischer Wellenrohrhohlleiter, 1181
Emitter Coupled Logic, 641
Emitterschaltung, 453
Empfindlichkeit, 876, 925
EMV-Richtlinie, 818
Endgerät, 1296
Endgerät, 1293
Endstufe, 470

Endtransistoren, 482
Energie, 204
Energie eines Kondensators, 331
Energie in einer Spule, 358
Energieerhaltungssatz, 204
Energieerzeugung, 1009, 1013
Energiehandel, 1012
Energiespeicherung, 1050
Energieübertragung, 1034
Energieverbrauch, 1013
Energieversorgung, 1010
Energieverteilung, 1009, 1034
Energieverwischung, 1254
Energiewirtschaftsgesetz (EnWG), 1010
Energieübertragung, 1009
Entfernungsauflösung, 1278
Enthalpie, 1015
Entladen eines Kondensators, 332
Entladungslampen, 1079
Entropie, 1131
E-Nummer, 1182
Epitaxie-Verfahren, 403
EPROM, 681
Erdanziehungskraft, 209
Erdatmosphäre, 1201
Erdschlussgefahr, 795
E-Reihen, 554
ergodischer Prozess, 1377
Ersatzradius, 1055
Ersatzschaltbild, 571, 967
Ersatzschaltbild der Asynchronmaschine, 976
Ersatzschaltbild von Freileitungen, 1056
Ersatzschaltplan, 570
Ersatzspannungsquelle, 311, 315
Ersatzstromquelle, 311, 316
Erwartungswert, 878
Erzeugung von Frequenzmodulation, 1215
Erzeugung von Phasenmodulation, 1215
Erzwungene Schwingungen, 221, 223
E-Shopping, 748
Ethernet, 757

F

Fadenpendel, 222
Fall
  aperiodischer, 852
  periodischer, 852
Fallgeschwindigkeit, 195
Faltungscodes, 1312
Faltungscodierer, 1314
Fan Out, 635
Fano, 1312
Farbbildröhre, 739
Farbmonitor, 738
FAST, 638
Fast-Ethernet, 756
Faxgerät, 1303
FDM, 751
Federpendel, 222
Fehler, 877
Fehlererkennende Codes, 1312
Fehlerfortpflanzung, 875, 879
Fehlerkorrigierende Codes, 1312
Fehlerstrom-(FI)-Schutzschalter, 1073
Feld (Array), 782
Feld in einer Ringspule, 338
Feld um einen Leiter, 338
Feldbussystem, 814
Feldlinien, 325
Feldplatten, 919
Feldverläufe in einem Rechteck-Hohlleiter, 1179
Fensterung, 1309, 1321
Fernglas, 245
Fernsprechapparat, 1296
Ferrimagnetismus, 348
Ferritantenne, 1198
Ferromagnetismus, 346
Festplatte, 730, 733
Festplatten, 734
Festpunktzahl, 603
Festwertregelung, 844
Festwertspeicher, 680
Feuerungsarten, 1017
FIFO-Prinzip, 750
Flächenwiderstand, 1177
Flag, 721
Flagregister, 663, 721
Flags, 727
Flankenauswertung, 797
Flankensteilheit, 506
Flash-EPROM, 683
flexible Erzeugung, 1049
flexibler Verbrauch, 1050
FLEXWELL-Kabel, 1179
Flipflop, 617–619
Floppy-Disk, 730
Floppy-Disk-Controller, 733
Flügelradzähler, 935
fluktuierende Energiebereitstellung, 1046
Flussdichte, 340
Flüssigkeitskühlung, 1115
Flüssigkristalle, 533
Folgeregelung, 844
Foliendicke, Füllstandsmessung, 933
Foliendicke, Messung, 933
Footprint, 1268
Formatierung, 730
Formfaktor, 887, 905
FORTRAN, 743
Fotodiode, 925, 926
Fotoelement, 531
Foto-Thyristor, 1108
Fotovervielfacher, 926
Fotowiderstand, 926
Fourierreihe, 1134
Fourierreihe, komplexe, 1135
Fouriertransformation, 959, 1134, 1332, 1339, 1340
Fouriertransformation, diskrete, 1369, 1370
Fouriertransformation, diskrete (DFT), 1136
Fouriertransformation, schnelle, 1369, 1371
Fouriertransformation, schnelle (Fast Fourier Transform, FFT), 1136
Fourierzerlegung, 959
FPGA, 648
f-P-Regelung, 1036
Freilaufdiode, 502, 1096
Freileitung, 1053
Freiraumdämpfung, 1197, 1262
Freiraumdämpfungsmaß, 1197
Fremdkühlung, 1115
Frequency Division Multiplexing, 751
Frequenz, 198, 222
Frequenz, kritische, 1180
Frequenzabstand, 1138
Frequenzdiversity, 1262
Frequenzen, 897, 899
Frequenzgang, 492
Frequenzgangfunktion, 854
Frequenzhaltung, 1042
Frequenzkennlinie, 855
Frequenzkennlinien, 1037
Frequenzkompensation, 489
Frequenzmessung, 937
Frequenzmodulation (FM), 1211
Frequenzmultiplexverfahren, 1257, 1259
Frequenzregelung, 1036
Frequenzteiler, 625
Frequenzthyristor, 1103
Frequenzumtastung (FSK), 1218
Frequenzverhältnis, 1181
Führungsgröße, 822, 844
Führungsgröße w, 822
Funkelrauschen, 1139
Funkfelddämpfung, 1197
Funktion, 780
Funktionsbaustein, 780
Fusionsreaktor ITER, 1022
Fußpunktwiderstand, 1197

G

GAL, 648, 654
galvanische Trennung, 1110
galvanischen Kopplung, 471
GAN, Global Area Network, 750
Gangunterschied, 229
Gas- und Dampf-Kraftwerk, 1019
Gasableiter, 1121
Gasdruckkabel, 1066
Gaskonzentrationen, 929
Gasturbine, 1018
Gatespannungsteiler, 465
GATT, 1108
Gatter, 590
Gatterlaufzeit, 636
Gauß-Filter, 496
Gaußverteilung, 1380
Gedämpfte Schwingungen, 223
Gedruckte Schaltungen, 574
gedruckte Schaltungen, 582
Gedämpfte Schwingungen, 221
Gegeninduktion, 357
Gegeninduktivität, 357
Gegenkopplung, 456
Gegenstrombremsen, 994
Gegenstrombremsung, 981, 996
Gegentakt-Betrieb, 478
Genauigkeit
  stationäre, 866
Genauigkeitsklassen, 881
Geometrische Optik, 237, 238
geometrischer Mittelwert, 1376
geostationäre Umlaufbahn, 1266
geothermisches Kraftwerk, 1033
Geradeausempfänger, 1246
Geradeausempfänger, 1245
Geräusch, 1284
Geräusch-abstand, 954
Geräuschmessung, 960
Germanium, 275
Geräusch-abstand, 953
Gesamtrauschzahl, 1140
Geschwindigkeitssensor, 941
Gesetz von Boyle-Mariotte, 215

Gesetz von Gay-Lussac, 215
Gewicht eines Codewortes, 1312
Gewichtskraft, 199
Gezeitenkraftwerk, 1023
Gibbs'sches Phänomen, 1334
Gittermonochromatoren, 247
Gleichgewichtige Codes, 1312
Gleichstrombremsung, 980
Gleichstromgeneratoren, 997
Gleichstrommaschinen, 963, 991
Gleichstrommaschinen, Aufbau, 991
Gleichstrom-Messerrücken, 906
Gleichstromsteller, 1095
Gleichtaktunterdrückung, 487
Gleichtaktverstärkung, 486
Gleichtaktwiderstand, 488
Gleichung
  charakteristische, 860
Gleichwellennetz, 1255
Gleitpunktzahl, 603
Gleitpunktzahlen, 603
-Glied, 848
Glitches, 636
Global Area Network, 750
Glühlampen, 1078
GPS, 1268
Grad n eines Filters, 1230
Gradientenindexfaser, 757
GRAFCET, 808
Grafikkarten, 740
Graphische Darstellung, 957, 958
Gravitationskonstante, 209
Gravitationskraft, 209
Gray-Code, 607, 939
Grenzdaten, 634
Grenzfall
  aperiodischer, 852
Grenzfrequenz, 431, 1230
Grenzfrequenz (RC), 376, 377
Grenzfrequenz (RL), 376, 377
Grenzfrequenz, obere, 378
Grenzfrequenz, untere, 378
Grenzkosten-Einsatzreihenfolge, 1013
Grenzwert, 399
Größengleichungen, 876
Größenwert, 876
Großsignal-Stromverstärkung, 1092
Grundfunktionen, 590
Grundgrößen der Lichttechnik, 1078
Grundverknüpfungen, 590, 591
Gruppengeschwindigkeit, 1165, 1180
Gruppengeschwindigkeit im Plasma, 1201
Gruppenkompensation, 1076
Gruppenstrahler, 1199
GSM, 1305
GTO-Thyristor, 1106
Güte Q, 1230
Gütefaktor, Messung von L, C, 917
Gütefaktor, Messung von L, C, 901

H

H_{10}-Welle im Rechteckhohlleiter, 1180
H_{11}-Welle im Rundhohlleiter, 1180
Halbaddierer, 612
Halbbrücken, 908
Halbleiter, 267, 269
Halbleiterkühlung, 1112
Halbleiterkühlung, 1109
Halbleiter-Relais, 1111
Halbleiterspeicher, 661, 676
Halbwellenbetrieb, 419
Hall-Effekt, 343
Halleffekt, 267, 270
Hall-Koeffizient, 343
Hall-Sonde, 919
Hallsonde, 919
Hall-Sonden, 343
Hallspannung, 343
Haltestrom, 1102
Hamming-Code, 610
Hamming-Distanz, 1312
Handlungen im Notfall, 819
Hangabtriebskraft F_H, 200
Hardware, 582, 662
Harmonische Schwingungen, 1355, 1357
Harmonische Wellen, 227, 228
Hartmagnetika, 285, 290
Häufigkeit, 1376
Häufigkeitsverteilung, 882
Heizwerte, 1018
Helmholtz'sches Überlagerungsverfahren, 318
Hertzscher Elementardipol, 1194
Hexadezimalsystem, 603
Hexadezimalzahlen, 603
High-Speed-CMOS, 644
hinlaufende Welle, 1184
Histogramm, 882
Hochfrequenzdurchschaltung, 1267
Hochfrequenz-Koaxialkabel, 1177
Hochfrequenzleitungen, 1161, 1177
Hochfrequenzschaltung, 1288
hochohmige Leitungsabschnitte, 1183
Hochpassfilter, 1230
Hochpassschaltung mit RC-Glied, 376
Hochpassschaltung RL-Glied, 376
Hochspannungs-Gleichstromübertragung, 1034
Hochspannungskaskadenschaltung, 413
Hochtemperatur-Leiterseil, 1048, 1055
Hohlleiter, 1179
Hohlleiter, elliptische, 1180
Hohlleiter, Querschnittsformen, 1179
Hook'sche Gesetz, 210
Hörer, 1291
Hörer-Systeme, 1289
Horizontalfrequenz, 737
Horizontalsteuerung, 422
Hörschall, 1284
Hörschwelle, 1284
$\underline{H}$-Parameter, 1150
h-Parameter, 430
HUB, 752
Hubarbeit, 203
Huffman, 1312
Hurwitz-Kriterium, 861
Huygensches Prinzip, 227, 230
Huygensschen Prinzip, 246
Häufigkeitsverteilung, 875
Hybrid-Technik, 522
Hydrostatischer Druck, 211
Hyper Text Transfer Protocol, 754
Hysterese, 346
Hystereseschleife, 920

I

I-Regler, 834
I/O-mapped-Adressierung, 690
IEC-Bus, 947, 948
IGBT-Transistoren, 1087, 1100
I-Glied, 848, 851, 868
Impedanzen, 909
Impedanzwandler, 454
Impuls, 201
Impulsantwort, 850
Impulsbetrieb, 501
Impulserhaltungssatz, 202
Impulsverbreiterung, 1270
Impulswahlverfahren, 1297
Impulszeitplan, 590
Induktionsflussdichte, 344
Induktionsgesetz, 355
Induktive Aufnehmer, 925
induktiver Blindwiderstand, 366
Induktivität, 356
Induktivität von Leiterseilen, 1058
Induktivitätsmessbrücke, 918
Industrial Ethernet-TCP/IP, 816
Industrielle Kommunikation, 814
induzierte Spannung, 354
inelastischer Stoß, 205
Influenz, 327
Information, 1130
Informationseinheiten, 590
Informationsfluss, 1132
Informationsgehalt, 1130
Infraschall, 1284
Inklination, 337
Inkrementale Wegaufnehmer, 939
in-line-Röhre, 738
Innenpolmaschine, 985
Innenwiderstand, 316, 902
innere Reibung, 212
Installationsplan, 570
Instanziierung, 780
Instrumentenbeschriftungen, 885, 888
Integrationsdichte, 520
Integrator, 493
Integrierglied, 848, 868
integrierte Schaltungen, 633, 634
integrierte Standardbausteine, 592
integrierten Schaltungen, 520
Intelsat, 1268
INTERBUS-S, 814
Interface, 662, 690
Interface-Schaltungen, 535
Interfaceschaltungen, 633, 646
Interferenz, 229, 246
Internet, 749
Internet Protokoll, 762
Interpreter, 744
Interrupt, 668–671, 713, 722, 723
Interrupteingänge, 722
Interrupt-Programm, 669
Interruptroutine, 669
Intrinsicleitfähigkeit, 274
Inversion, 1189
IPv4, 762
IPv6, 765
I-Regelfunktion, 831, 834
I-Regler, 868
Irrelevanz, 1251, 1253, 1255
isentrope Zustandsänderung, 218
Isentropenexponent, 218
isobare Zustandsänderungen, 218
Isolator, 267, 268, 1054

Isolierscheiben, 1113
Isolierstoffe, 1063
Isolierstoffklassen, 1003
isotherme Zustandsänderungen, 217
I-Strecke, 829
IT-Netz, 1072

J
Jahreshöchstlast, 1014
Jahreswirkungsgrad, 970
JEDEC, 650
JEDEC-Datei, 651, 652
JK-Flipflop, 619
Justieren, 876

K
Kabel, 1063
Kabeldämpfung, 1177
Kabelkonzentrator, 753
Kalibrieren, 876
Kaltleiter, 1123
Kanalkapazität, 1132
Kanalkodierung, 1312
Kanalvermittlung, 750
Kapazität, 327
Kapazität einer Leitung, 328
Kapazität von Leiterseilen, 1058
Kapazitäten, 899
Kapazitätsmessbrücke, 918
Kapazitive Aufnehmer, 924
kapazitive Kopplung, 471
kapazitiver Blindwiderstand, 365
Kapazitäten, 897
Kenndaten, 399, 634
Kenngrößen, 959
Kenngrößen, 957
Kennimpedanz, 1284
Kennlinienfelder, 428
Kennlinienglied, 848
Kennwiderstand, 902
Kepler'sche Gesetze, 209
keramische Filter, 1235
Kernfusion, 1022
Kernkraftwerk, 1021
Kernspaltung, 1021
Kettenleiter, 1171
Keyboard, 736
K-Faktor, 932
kinetischen Gastheorie, 216
Kippmoment, 978
Kippschaltung, 503
Kippvorgang, 503
Klassierung, 883, 957, 959
Kleinsignal-Transistoren, 428
Kleinsignalverhalten, 430
KL-Glied, 848
Klirrdämpfung, 1143
Klirrfaktor, 1142, 1226
Knoten, 1184
Knotenregel, 312
Koaxialkabel, 756, 1177
Koaxialkabel, Feldlinienbild, 1179
Kodierung, 1227, 1309, 1310
Koerzitivfeldstärke, 346, 920
Kohärenzlänge, 246
Kohlemikrofon, 1286
Kollektorreststrom, 1092
Kollektorschaltung, 454
Kommutativgesetz, 593
Kompandierung, 1227
Komparator, 507, 614
Komparatoren, 614
Kompensationsanlagen, 1075, 1076
Kompensationswicklungen, 992
Kompensator, 903
Komplementärtransistoren, 479
komplexe Fourierreihe, 1333, 1334
komplexe Zahl, 364
komplexer Frequenzgang, 1141
komplexer Reflexionsfaktor, 1186
komplexer Übertragungsfaktor, 1141
komplexes Wellendämpfungsmaß, 1158
Kompressibilität, 210
Kondensatoren, 296, 323, 327
Kondensatormikrofon, 1287
Kondensatormotor, 988
konjugiert komplexe Anpassung, 1186
konjugiert komplexer Abschlußwiderstand, 1185
konstante Anpassungsfaktoren, 1186
Konstantstromquellen, 469, 481
Kontinuitätsgleichung, 212
konventionell richtigen Wert, 877
konzentrierter Triebstrang, 1024
Konzepte zur Netzkopplung, 1026
Kopfhörer, 1291
Koppelkondensatoren, 458
Kopplungsfaktor, 1234
Korpuskel, 238
Korpuskulartheorie, 238
Korrektion, 877
Korrekturglied, 868
Korrelation, 1309, 1317
Kosinus-Transformation, 1253
Kraft, 199
Kraftmessdose, 932
Kraftmessung, 930
Kraftstoß, 202
Kraftwerksarten, 1015
Kraftwerksleistung, 1014
Kreisdiagramm, 1161, 1186, 1187, 1189
Kreise konstanter Verstärkung im Kreisdiagramm, 1191
Kreisfrequenz, 198, 222, 362
Kreisgüte, 378, 379
Kreisprozesse, 218
Kreisradius für konstante Verstärkung, 1191
Kreisstrom, 345
Kreuzkorrelation, 960
Kreuzkorrelationsfunktion, 1140, 1378
Kreuzmodulation, 1209
Kreuzparitätsprüfung, 610
Kreuzschaltung, 1159
Kreuzsicherungsprüfung, 609
Kreuzvierpol, 1159
Kristallmikrofon, 1287
kritische Stromsteilheit, 1104
Kugelkondensator, 328
Kugelwelle, 228
Kühlart, 1002
Kühlkörper, 401, 1113
Kühlung, 964
künstlicher Sternpunkt, 912
Kunststoffkabel, 1065
Kurvenformfaktor, 887, 905
Kurzschlussläufer, 978
Kurzschlussstrom, 316
Kurzschlussversuch, 968
Kurzschlusswiderstand, 1183
KUSA-Schaltung, 980
KV-Tabelle, 598

L
Ladekondensator, 409
Laden eines Kondensators, 331
Lag-Glied, 869
Lambda-Sonde, 930
Lambertsches Gesetz, 249
Lambert-Strahler, 249
Lampen, 1078
LAN, 1304
LAN (local area network), 1300
LAN, Local Area Network, 750
Land, 736
Langdrahtantenne, 1200
Laplacetransformation, 1136, 1332, 1339, 1343
Laser-Dioden, 533
Laserdrucker, 740
Lasteinheit, 635
Late Collision, 761
Lautheit, 1285
Lautsprecher-Systeme, 1289
Lautstärkepegel, 1285
Lawineneffekt, 406
Lawinen-Fotodioden, 1274
Layout, 574
Layouts, 582
LCD-Bildschirm, 740
LC-Filter, 412
Lead-Glied, 868
Least Significant Bit, 590
Lebensdauererwartung, 296
Leerlauf, 965
leerlaufende Freileitung, 1060
Leerlaufleistung, 966
Leerlaufverstärkung, 487
Leerlaufversuch, 967
Leerlaufwiderstand, 1183
Leistung, 203, 320
Gleichgewicht, 1036
Leistungsanpassung, 320, 479
Leistungsaufnahme, 676
Leistungsbeiwert, 1024
Leistungsdämpfungsfaktor, 1143
Leistungsdämpfungsmaß, 1144
Leistungsfaktor, 381
Leistungsfaktormessung, 915
Leistungsfluss, 977
Leistungshyperbel, 406
Leistungsmessung, 901, 911
Leistungsschalter, 551
Leistungsspektralfunktion, 1332
Leistungsübertragungsfaktor, 1143
Leistungsübertragungsmaß, 1144
Leistungsverstärkung, 1186
Leiterformen, 1063
Leiterplatten, 522
Leiterseil, 1054
Leiterwerkstoffe, 1063
Leitfähigkeit, 307
Leitung als Transformator, 1170
Leitung als Vierpol, 1168
Leitung im Wechselstromkreis, 361, 379
Leitungsbeläge, 1162
Leitungsbeläge, 1161

Leitungsgleichungen, 1161, 1162
Leitungsmechanismus, 267, 268
Leitwert, 307
Leitwertdiagramm, 1188, 1189
Leitwerte, 1189
Lenz'sche Regel, 354
Leuchten, 1079
Leuchtstofflampen, 1079
Licht, 526
Lichtempfänger, 1274
Lichtempfindlichkeit, 526
Lichtempfänger, 1269
Lichtgeschwindigkeit, 238, 1180
Lichtleiter, 241
Lichtquant, 238
Lichtquellen, 1078
Lichtsender, 1269, 1274
Lichtstärke, 532
Lichtwellenleiter, 756
Lifo-Prinzip, 664
Likelihood-Bedingung, 1313
Likelihood-Viterbi-Decodierer, 1314
linearer Ausdehnungskoeffizient, 215
Linearisierung um den Arbeitspunkt, 850
Linearität, 1332
Linearmotor, 985
Linien-Netz, 754
Links-Rechts-Stereofonie, 1250
Linsen, 241
Lissajous-Figuren, 225
Listener, 948
Local Area Network, 750
Logik, 590
Logikschaltungen, 633, 634
Logische Grundschaltungen, 589, 590
logische Grundverknüpfung, 794
Logische Verknüpfungen, 591
Longitudinalwelle, 228
Lorentz-Kraft, 886
Lorentzkraft, 341, 919
Löschen von Thyristoren, 1105
Löschvorgang, 416, 1102
Low-Power-Schottky, 641
LS, 638
LSA-Verbindung, 1275
Lumen, 250
Lupe, 244

M

Machscher Kegel, 232
MAG, maximum available gain, 1186
Magnetbandgeräte, 735
Magnetfeld der Erde, 337
magnetische Feldkonstante, 341
Magnetische Feldstärke, 919
magnetische Feldstärke, 918
Magnetische Flussdichte, 919
magnetische Flussdichte, 918
magnetische Spannung, 349
magnetische Suszeptibilität, 344
magnetischer Fluss, 918
Magnetischer Widerstand, 349
magnetisches Moment, 342
Magnetostriktion, 348
MAN, Metropolitan Area Network, 750
Manchester-Kodierung, 759
Maschenregel, 312
Maschinencode, 718
Maschinenrichtlinie 89/392/EG, 818
Maschinensprache, 661, 718
Maschinenzyklen, 724
Maschinenzyklus, 667
Masken-ROM, 680
Massekabel, 1065
Massenspeicher, 730
Massenstrom, 212
Mast, 1053
Master, 621, 709
Master-Slave-Flipflop, 621
Master-Slave-JK-Flipflop, 621
Matrixdrucker, 740
Maus, 736
maximaler Amplitudenfehler, 1226
maximaler Quantisierungsfehler, 1226
Maximum-Likelihood-Verfahren, 1313
Maxterme, 594
Maxwellsche Geschwindigkeitsverteilung, 216
MC 8051, 709
Median, 883, 958, 1376
Mehrfrequenzwahlverfahren, 1297
Mehrmoden-Gradientenindex, 1271
Mehrmoden-Stufenindex, 1271
Memory-mapped-Adressierung, 690
Merit Order, 1012
Messabweichung, 875, 877, 902
Messaufnehmer, 921–923
Messbereich, 876
Messbereichserweiterung, 901, 902, 904
Messeinrichtung, 844
Messen, 876
Messergebnis, 876
Messfehler, 875, 877
Messgerät, 876, 886
Messgeräte, analog anzeigende, 881
Messgeräte, digital anzeigende, 881
Messgeräte, Fehlerangaben, 881
Messgeräte, Fehlerangaben, 875
Messglied, 844, 922
Messglieder, 922
Messkette, 922
Messkettenkoeffizient, 922
Messprinzip, 876
Messsignalanalyse, 957, 959
Messsysteme, Erdung, 945
Messumformer, 876
Messumsetzer, 876
Messung der Arbeit, 901, 916
Messung des Klirrfaktors, 960
Messverfahren, 876
Messverfahren, analoge, 876
Messverfahren, digitale, 876
Messverstärker, 945
Messverstärker, 943
Messwandler, 876
Messwerke, 886
Messwert, 876
Metropolitan Area Network, 750
M-Glied, 848
MIC-Verbindung, 1275
Mikrocomputer, 662
Mikrocontroller, 661, 708, 709, 714, 715
Mikrofone, 1285
Mikroprozessor, 663, 707
Mikroprozessoren, 662, 664
Mikroskop, 245
Mikrowellenfenster, 1262
Miller-Effekt, 1116
Mindestlast, 1014
Minterme, 594
Mischspannung, 450
Mischung, 1204
Mitkopplung, 456
Mittelwerte, 1140
Mitte-Seiten-Stereofonie, 1250
Mittlere Kodewortlänge, 1310
mittlerer Informationsgehalt, 1131
MLT3-Kodierung, 759
Mnemonik, 719
Modem, 1304
Moden, 1270
Modendispersion, 1270
Modulation an einer nichtlinearen, 1206
Modulation an einer quadratischen Kennlinie, 1206
Modulation durch Multiplikation, 1205
Modulation, additive, 1211
Modulation, multiplikative, 1211
Modulationsgrad, 1205
Modulationsindex h bei FM, 1212
Modulationstrapez, 1207
Modulator für Einseitenband-AM, 1210
Modulform, 1095
modulierendes Signal, 1204
Modulo-16-Zähler, 627
molare Wärmekapazität, 217
Monitor, 736, 738, 739
monochrome Licht, 533
Monoflop, 629, 631
Mono-Mode-Fasern, 757
MOS, 634
Most Significant Bit, 590
Motional Feedback-System, 1291
Motormoment, 995
Motorregel, 342
Motorschutz, 1003
MPEG, 1252
Multielement-Variable, 782
Multiplexer, 615
Multiplizierglied, 848
Multiscan-Monitor, 739

N

Nachricht, 1130
Nachrichtenquader, 1132
Nachrichtenübertragungssystem, 1128
Nachstellzeit, 834, 868
Nachstellzeit T_n, 834
Nahfeld–Fernfeld, 1194
NAND-Technik, 596
NAT, 764
Nebenschlussgenerator, 997
Nebenschlussmotor, 993
Nebensprechen, 1208
Network Address Translation, 764
Netzanbindung, 1024
Netzausbau, 1047
Netzformen, 1035, 1071
Netzimpedanz, 1048
Netzresonanz, 1049
Netzstatik, 1037
Neutralleiterstrom, 388
Newtonsche Axiome, 199
NF-Bootstrap-Verstärker, 494
Nibble, 590

nichtrekursive Filter, 1237
NICHT-Verknüpfung, 590
niederohmige Leitungsabschnitte, 1183
Niederspannungsrichtlinie 73/23/EWG, 818
Non Return to Zero/Invert, 759
Normalform, disjunktive, 594
Normalform, konjunktive, 594
Normalkraft, 200
Normalspannung, 210
Normalverteilung, 882, 958, 1380
normierte Parallelschaltung, 1187
normierte Serienschaltung, 1187
normierte Verstärkung, 1190
normierte Verstimmung, 1235
NOR-Technik, 596
n-Phasen-Umtastung, 1221
NRZI, 759
NRZ-Verfahren, 1228
NTBA, 1302
Nullkippspannung, 416, 1102
Nullstelle, 854
numerische Apertur, 1270
Nutzsignal, 1204
NVRAM, 685
Nyquistflanke, 1210
Nyquist-Kriterium, 862

O

Oberes Seitenband (OS), 1205
ODER-Verknüpfung, 591
Öffnungswinkel, 1196
Offset-Spannung, 489
Ohm'sche Aufnehmer, 923
ohmscher Wechselstromwiderstand, 365
Ohm'sches Gesetz, 305, 307
Öldruckkabel, 1066
OLMC, 654
OPC-Technologie, 816
open-collector-Ausgang, 490
Operationscode, 667, 726
Operationszyklen, 726
Operatoren, 364
Operatoren der Anweisungsliste (AWL), 784
Optimalcodes, 1312
Optimalfilter, 1309, 1316
Optimum
  symmetrisches, 872
Optische Aufnehmer, 925
Optische Schalter, 1269, 1275
Optokoppler, 535
Ortskurve, 855
Ortskurven für konstante Werte, 1190
OSI-Referenzmodell, 747, 754
Osmosekraftwerk, 1023
Ovalradzähler, 936
Oversampling, 955
Oversamplingrate OSR, 955

P

P-Regler, 832
Paketfilter-Firewall, 765
Paketvermittlung, 1294
PAL, 648–651
PAM-5-Kodierung, 760
Parabolantenne, 1200
Parabolspiegel, 239
Parallele Schnittstelle, 948
Parallelresonanz, 374, 379
Parallelresonanz an einer Streifenleitung, 1183
Parallelschalten von Transformatoren, 971
Parallelschaltung, 313
Parallelschaltung von
  Wechselstromwiderständen, 371
Parallel-Serienumsetzer, 624
Paramagnetismus, 345
Parität, 609
Paritäts-Bit, 608
Paritätsbit, 609
Parseval'sche Gleichung, 1341, 1377
PASCAL, 743
Passive R-C-Filter, 1231
Passive *R-C*-Hochpassfilter, 1231
Patch-Kabel, 756
PC, 730
PD-Regler, 868
  realer, 868
Peer-to-Peer, 948
Peer-to-Peer-Netz, 749
Pegel, 590, 635, 1144
Pegel, absoluter, 1144
Pegel, relativer, 1145
Pegeldiagramm, 1145
Pegelwerte, 635
Pendelzeiger, 1212
Periodendauer, 362
Peripheriebausteine, 661, 689
Permeabilität, 918, 920
Permittivitätszahl, 294, 1177
Personenschutz, 1073
P-Glied, 848, 850, 868
Phasenanschnittsteuerung, 419, 1107
Phasendifferenzkodierung (2-DPSK), 1220
Phasengeschwindigkeit, 229, 1165, 1180
Phasenhub, 1215
Phasenhub FM, 1212
Phasenhub PM, 1212
Phasenkennlinie, 855
Phasenkoeffizient, 1164
Phasenmaß, 1141
Phasenmodulation (PM), 1211
Phasenrand, 863
Phasenregelkreis, 1216
Phasenreserve, 863
Phasenschieberbetrieb, 987
Phasenschieberoszillator, 514
Phasenspektrum, 1340
Phasenspektrum, komplexes, 1335
Phasenumtastung (PSK), 1218
Phasenverschiebung, 363
Phasenverzerrungen, 1141
Photoapparat, 245
Photoeffekt, 251
photoelektrischer Effekt, 1030
Photometrie, 237, 248
Photonen, 251
Photovoltaikanlage, 1032
photovoltaisches Kraftwerk, 1030
pH-Wert, 929
Physikalisch-Technische Bundesanstalt, 876
PI-Regler, 834
PID-Regler, 834
PID-Abtastregler, 831, 836
PID-Regelfunktion, 831, 834
PID-Regler, 867
  realer, 868
piezoelektrische Konstante, 926
Piezoelektrischer Effekt, 926
Piezoelektrisches System, 1290
PIN-Fotodioden, 1274
PI-Regelfunktion, 831, 834
PI-Regler, 868
PI-Schrittregler, 838
Pit, 736
pitch-Regelung, 1024
Planarverfahren, 403
Planck'sche Strahlungsformel, 250
Plancksches Wirkungsquant, 250
Plasmafrequenz, 1201
Plattenkondensator, 329
PLD, 634, 647, 648
pLSI, 654
Poisson'sche Gleichungen, 218
Poissonverteilung, 1379
Poissonzahl, 210
Pol, 854
Polarisation, 247, 926, 927, 1194, 1267
Polarisation, elliptische, 1194
Polarisation, kreisförmige, 1194
Polarisationsfilter, 247
Polarisationswinkel, 248
Polling, 669
Polpaare, 976
Polradspannung, 986
Polradwinkel, 986
Popcorn-Rauschen, 1139
Portbaustein, 692, 702
Portbausteine, 690
Potential, 325
Potentialanschlüsse, 899, 907
POTS, 1302
Poynting-Vektor, 1194
Präzisionsmessgerät, 877
Preemphase, 1214
P-Regelfunktion, 831, 832
P-Regler, 868
Preisbildung, 1012
Primärradar, 1278
Prinzip der Pulsweitenmodulation, 1029
Prisma, 241
PROFIBUS-DP, 814
PROFINET, 814
PROFINET-Kommunikationskanal, 815
Program-Counter, 663
Programm
  gegliedertes, 787, 788
  lineares, 787, 788
  parametrierbares, 787, 788
Programm (P), 780
Programmiersprache, 742, 784
Programmiersprachen, 741, 742
Programmorganisationseinheit, 780
Programmspeicher, 714
Programm-Test, 746
Programmzähler, 723
PROM, 648, 680
Proportionalbeiwert, 868
Proportionalglied, 848, 868
Prozess, 844
Prozessor, 665
Pseudotetraden, 606
PSN-dotiert, 1089
PSRAM, 686
P-Strecke, 826
PT1-Strecke, 828

PT2-Strecke, 828
PT_1-Glied, 848, 851
PT_2-Glied, 848, 851
PTC-Widerstände, 1123
Pull-up-Widerstände, 690
Pulsamplitudenmodulation (PAM), 1221
Pulscodemodulation (PCM), 1224
Pulsdauermodulation (PDM), 1223
Pulsfrequenzmodulation (PFM), 1222
pulsierende Gleichspannung, 408
Pulsphasenmodulation (PFM)-, 1222
Pulsweitenmodulation, 1028
Pumpspeicherwerk, 1023, 1050
Punkt-zu-Punkt-Kanal, 750
Punkt-zu-Punkt-Verbindung, 1266
Pyroelektrischer Effekt, 927

Q
QAPSK, 1221
quadratischen Mittelwert, 408
quadratischer Mittelwert, 1140
Quadraturmodulation, 1210
Quantile, 883, 958
Quantisierung, 1226
Quantisierungsgeräusch, 1226
Quantisierungsrauschen, 954, 1226
Quantisierungsrauschspannung, 954
Quarzfilter, 1235
Quasi-Komplementär-Endstufe, 481
Quellenkodierung, 1310
Quellspannung, 316
Querkontraktionszahl, 932

R
Radarquerschnitt, 1278
Rahmenantenne, 1198
RAM, 662
RAS, 686
Ratiodetektor, 1217
Raumdiversity, 1262
Raumfaktor, 1080
Raummultiplexverfahren, 1260
Raumwelle, 1200
Raumwinkel, 249
Rauschabstand, 1140
Rauschen, 432, 1138, 1375, 1381
Rauschen, atmosphärisches, 1138
Rauschen, kosmisches, 1138
Rauschen, thermisches, 1138
Rauschleistung, 1139
Rauschmaß, 1139
Rauschzahl, 1139
Rayleigh-Streuung, 1270
RC-Filter, 412
Rechenverstärker, 486
Rechenwerk, 663
Rechteckgenerator, 509
Rechteckgeneratoren, 516
Rechteckhohlleiter, 1179
Rechteckhohlleitern, 1181
Redundanz, 608, 1131, 1251
Redundanz, relative, 1131
Reed-Solomon-Verfahren, 1254
reelle Fourierreihe, 1333, 1334
reelles Bild, 239
Referenzspannung, 490
Reflexion, 230, 1169, 1270
Reflexion des Lichtes, 238
Reflexionsfaktor, 1166
Reflexionsfaktoren, 1187
Reflexionsfaktoren, äußere, 1186
Reflexionsgesetz, 240
Regelabweichung, 844
regelbarer Ortsnetztransformator, 1041
Regeldifferenz, 844
Regeldifferenz e, 822
Regelfehler, 844
Regelglied, 844
Regelgröße, 822, 844
Regelgröße x, 822
Regelkreis, 822, 844
Regelstrecke, 822, 827, 844
Regelstrecke mit Ausgleich, 826
Regelstrecke mit Totzeit, 826
Regelstrecke ohne Ausgleich, 826
Regelung, 774, 822
Regelungsschema, 1039
Register, 622
Regler, 844
  digitaler, 824
Reglerausgangssignal y_R, 822
Regression, 883, 959
Reibungsarbeit, 203
Reichweite eines Radargerätes, 1278
Reihenfolgeverriegelung, 796
Reihenresonanz, 370, 378
Reihenschaltung von Wechselstromwiderständen, 367
Reihenschaltung von Widerständen, 313
Reihenschlussgenerator, 998
Reihenschlussmotor, 995
rekursive Filter, 1237
Relative Brechzahldifferenz, 1270
relative Dielektrizitätszahl, 294
relative Kurzschlussspannung, 969
relative Messabweichung, 877, 878
relative Permeabilität, 344
relativer Leistungspegel, 1145
relativer Spannungspegel, 1145
relativer Strompegel, 1145
relatives Leerlaufstromverhältnis, 968
Remanenz, 346
Remanenzflussdichte, 920
RESET, 712
Resonanz, 223
Resonanzfrequenz, 378, 379, 857, 1230
Resonanzkurve, 224
Resonanzüberhöhung, 857
Resonator, 514
Restseitenbandmodulation (RM, VSB), 1210
Restseitenbandübertragung, 1210
Restwelligkeit, 411
R-Glied, 854
Richtcharakteristik, 1285
Richtdiagramm, 1196
Ringmodulator, 1209
Ring-Netz, 754
Ringnetz, 753
Ringschieberegister, 624
Ringverstärkung, 514
RISC-Prozessoren, 662
RJ45-Stecker, 756
R-L-C-Bandpass, 1233
R-L-C-Bandsperre, 1233
RLL-Code, 760
R-Nummer, 1181
Robustheit, 844
Rohrschlitzstrahler, 1200
ROM, 662, 680
ROM-losen Versionen, 709
ROM-Versionen, 709
Rotationsenergie, 208
Router, 754, 765, 1304
RSA-Verfahren, 1318
RS-Flipflop, 618
Rückführgröße r, 822
Rückführung, 844
Rückkopplung, 456, 844
Rückkopplungsnetzwerk, 514
rücklaufender Welle, 1184
Rückübertragungsfaktor, 1184
Rückwärtsrichtung, 1088
Rückwärtszähler, 627
Rundhohlleiter, 1180, 1181
Rundsendekanal, 751
RZ-Verfahren, 1227

S
Sägezahngenerator, 417
Sägezahngeneratoren, 516
Sallen-Key-Filter, 1244
Sammellinsen, 242
Sättigung, 428
Satz von Shannon, 1310
Sauerstoffkonzentration, 929
Scatterverbindung, 1202, 1262
SC-Filter, 1237
Schadenfrüherkennung, 960
Schall, 1284
Schalldruck, 1284
Schalldruckpegel, 1284
Schalleistung, 1284
Schallempfänger, 1285
Schallempfänger, 1283
Schallgeschwindigkeit, 234, 1284
Schallintensität, 1284
Schallkennimpedanz, 1284
Schallmauer, 232
Schallpegel, 1284
Schallschnelle, 1284
Schallsender, 1283, 1289
Schaltalgebra, 589, 593
Schaltanlagen, 1068
Schalter, 1068, 1109, 1110
Schalterarten, 1069
Schaltgruppen, 971
Schalthysterese, 506, 838
Schaltnetze, 611
Schaltungsminimierung, 595
Schaltungsunterlagen, 563
Schaltwerke, 617
Schaltzeichen, 550, 564
Scheibenläufermotor, 990
Scheinleistung, 380, 911
Scheinleistungsmessung, 915
Scheitelwert, 362
Schenkelpol, 985
Schering-Messbrücke, 918
Schicht 1, 758
Schicht 2, 760
Schicht 3, 762
Schicht 4, 766
Schichten 5–7, 768
Schichtpotentiometer, 923

Schieberegister, 622
Schieberegisteranwendung, 624
Schiefer Wurf, 197
Schleifdraht-Messbrücke, 907
Schleife, 809
Schleifenbildung, 599
Schleifringläufer, 983
Schlitzleitung, 1182
Schlupf s, 977
Schmalband-Frequenzmodulation, 1213
Schmerzschwelle, 1285
Schmitt-Trigger, 505
Schnittstellenbausteine, 689
Schreib-Lesespeicher, 680, 683
Schritte, 808
Schrittmotor, 989
Schrotrauschen, 1139
Schubspannung, 210
Schutzart, 549, 1001
Schützen, 551
Schutzintervall, 1255
Schwarzer Strahler, 250
Schwebung, 225
Schwellenwertschalter, 505
Schwellspannung, 506
Schweredruck, 211
Schwerkraft, 199
Schwerpunkt, 206
Schwingkreise, 361, 378
Schwingungstypen, 1181
Schwingungszeit, 222
Schwund, 1262
Schwunderscheinungen, 1202
SC-Verbindung, 1275
SDM, 1260
Seildurchhang, 1056
Seitenbänder, 1205
Sekundärradar, 1278
selbstgeführter Wechselrichter, 1028
Selbstinduktion, 356
selektive Filter, 496
Sensor, 922
Sequentielle Logik, 617
Serielle Schnittstelle, 714, 948
Serienbetrieb, 500
Serien-Parallel-Umsetzer, 624
Server, 749
Servomotor, 990
SFN, 1255
S-Glied, 848
Shannon, 1368
SI-Basiseinheiten, 876
Sicherheitsbegriff, 817, 818
Sicherheitskategorie, 819
Siedewasserreaktor, 1021
Signal, 1130
  analoges, 790
  binäres, 790
  digitales, 789, 790
Signalangepasste Filter (matched filter), 1382
Signale, 1133
Signale, analoge, 1133
Signale, digitale, 1133
Signale, diskrete, 1133
Signale, kontinuierliche, 1133
Signale, nichtstationären, 1376
Signale, stationären, 1376
Signale, stochastische, 1138, 1376
Signale, stochastische, Kenngrößen, 1140
Signale, zufällige, 1138
Signalerkennung, bei gestörter Übertragung, 1382
Signalerkennung, bei gestörter Übertragung, 1375
Signalflussdiagramm, 1185
Signalflußdiagramm des beschalteten Vierpols, 1185
Signalflußdiagramme, 1184
Signalflüsse, 1185
Signalflüssen, 1186
Signalflussplan, 848
Signalleistung, 450
Signal-Quantisierungs-Geräuschabstand, 1226
Signal-Rausch-Verhältnis, 954
Signalverstärker, 1086
Silicium, 275
Simplex-Betrieb, 1300
Simultan-Verzweigung, 809
Sinus-Dreieck-Ansteuerung, 1029
Skin-Effekt, 1178
Slave, 621, 709
SMD-Technik, 523
SOAR-Diagramm, 1094
solarthermische Kraftwerksprinzipien, 1032
solarthermisches Kraftwerk, 1032
Solarzelle, 531, 1030
Soll-Istwert-Vergleich, 844
sone, 1285
Sonnenwind, 337
Spaltpolmotor, 989
Spannarbeit, 203
Spannenwert, 959
Spannung, 210, 306, 325
Spannungsdämpfungsfaktor, 1143
Spannungsdämpfungsmaß, 1144, 1158
Spannungsebenen, 1035
Spannungserhöhung, 1077
Spannungsfestigkeit, 1177
Spannungsinstabilität, 1062
Spannungsmesser, 902
Spannungsmessgeräte, analog, 902
Spannungsmessgeräte, digitale, 902
Spannungspfad, 888, 912
Spannungsregelung, 1038
Spannungsstabilisierung, 413
Spannungsstoß, 340
Spannungsteiler, frequenzkompensierter, 1160
Spannungstrichter, 1041
Spannungsübertragungsfaktor, 1143, 1166
Spannungsübertragungsmaß, 1144
Spannungswandler, 915, 975
Spannungszwischenkreisumrichter, 981
s-Parameter, 1161, 1184
s-Parameter im Kreisdiagramm, 1190
Spartransformatoren, 973
Speicher, 676–680
Speichererweiterung, 687
Speicherfunktion, 793, 795
Speicherkapazität, 662, 671, 676, 678
Speichermatrix, 677, 684
Speichern mit vorrangigem Rücksetzen, 796
Speichern mit vorrangigem Setzen, 796
Speicherorganisation, 676
Speicherprogrammierbare Steuerung SPS, 776
Speichertypen, 680
Speicherwerk, 663
Speicherzeit, 500, 1093
Speicherzelle, 685
Spektraldichte, 1332, 1340
Spektrale Leistungsdichte, 1141
spektrale Leistungsdichte, 1377
Spektralfarben, 526
Spektralfunktion, 1332
Spektrum, 1340
Sperrbereich, 1230
Sperrrichtung, 1088
Sperrschicht, 396
Sperrschichtkapazität, 400
Sperrschichttemperatur, 1112
Sperrstrom, 1088
Sperrverzögerungszeit, 400, 1088
spezielle Gaskonstante, 216
spezifische Wärmekapazität, 217
spezifischer Widerstand, 307, 1177
Spiegelgalvanometer, 887
Spitzensperrspannung, 1088
Spleißverbindung, 1275
Splines, 883, 959
Splitter, 1305
Sprung, 809
Sprungantwort, 850
Sprungfunktion $\sigma(t)$, 1355, 1356
Sprungstellen, 1334
SPS-Programmiernorm DIN EN 61131-3, 780
SPS-Regelstruktur, 837
SRAM, 683
Stabankermotoren, 990
Stabilität, 860, 1332
  asymptotische, 860
Stabilitätskriterium
  grundlegendes, 860
Stabilität, 859
Stack, 669, 721
Stackpointer, 663, 721
stall-Regelung, 1024
Standardabweichung, 878, 958, 1140, 1377
Standardregelkreis, 860
Stapelspeicher, 721
Start-Funktion, 819
State-Diagramm, 658
stationärer Prozess, 1377
statischer Durchlasswiderstand, 400
statischer Sperrwiderstand, 400
Statistischer Multiplex, 1255
Steckverbindung, 1275
Steckvorrichtungen, 552
Stefan-Boltzmann-Gesetz, 219
Stefan-Boltzmannsche Gesetz, 250
Stehende Wellen, 229
stehende Wellen, 1171
Stehwellenverhältnisse, 1186
Steigung des Amplitudenganges, 1230
Steiner'scher Satz, 209
Steinmetzschaltung, 989
Stellgröße, 822, 844
Stellgröße *y*, 822
Steradiant, 249
Stereoübertragung, 1250
Stereoübertragung, 1249
Stern-Bus-Netz, 753
Stern-Dreieck Umwandlung, 314
Stern-Dreieck-Anlauf, 979
Sternnetz, 752
Sternschaltung, 384, 1035

Stern-Stern-Netz, 753
steuerbare Last, 1050
Steuereinrichtung, 845
Steuergröße, 822
Steuerstrecke, 845
Steuerung, 774, 822, 844
Steuerwerke, 663
Steuerwortregister, 692
Stichleitungen, 1183
Stichprobe, Erwartungswert, 1376
Stichprobe, linearer Mittelwert, 1376
Stopp-Funktion, 819
Störabstand, 635, 1139
Störeinflüsse, 944
Störgröße, 822, 844
Störsicherheit, 635
Störungsbeseitigung, 1007
Stoßfunktion, 1356
Stoßstrombegrenzer, 1071
Stoßfunktion, 1355
Strahlcharakteristik, 249
Strahldichte, 249
Strahlstärke, 249
Strahlungsdichte, 1194
Strahlungsleistung, 249
Strahlungswiderstand, 1197
Streamer, 735
Strecke, 844
Streifenleitungen, 1182
Streuung, 1140
Strombörse, 1011
Stromdämpfungsfaktor, 1143
Stromdämpfungsmaß, 1144, 1158
Stromdichte, 307
Stromflusswinkel, 419
Stromlaufplan, 569
Strommesser, 904
Strommessgeräte, analoge, 903
Strommessgeräte, digitale, 903
Strompfad, 888, 912
Strompreis, 1012
Stromprodukte, 1011
Stromstärke, 307
Stromübertragungsfaktor, 1143, 1166
Stromübertragungsmaß, 1144
Strömungskraftwerk, 1023
Stromverdrängungsläufer, 979
Stromverstärkung, 427
Stromverstärkungen, 899
Stromverstärkungen, 897
Stromwandler, 915, 975
Strom-Zeit-Kennlinie, 1074
Stromzwischenkreisumrichter, 981
Struktur, 783
Struktur einer Ablaufsteuerung, 811
Strukturbild, 848
Stufenindexfaser, 757
Störeinflüsse, 943
Subtrahierer, 491, 613
Summenfunktion, 1376
Summenhäufigkeitsfunktion, 1376
Summenzeitkonstante, 869
Summierglied, 848
Superpositionsprinzip, 849
Suppressordioden, 1121
Supraleiter, 267, 270, 273, 276
Switch, 752, 1304
Symbol, 571
Symbole, 885, 888
Symmetriebedingungen, 1334
symmetrische Brückenschaltung, 999
synchrone Zählern, 628
Synthese, 600–602
System, 844, 1328
Systemanalyse, 1332
Systematische Abweichungen, 877
systematische Abweichungen, 877
Systematische Codes, 1312
Systematische Messabweichung, 903
Systemdienstleistungen, 1042
System-Kontrollsignale, 667
Systemsynthese, 1332

T

Taktfrequenz, 509
Talker, 948
Tastatur, 736
Tastgrad, 1089
Tastgrad beim Impulsradar, 1278
Tauchspulmikrofron, 1287
TCM, 760
TCP/IP, 766
TDM, 752
Teilklirrfaktoren, 1143
Temperaturbereich, 636
Temperaturen, 897, 899
Temperaturgradient, 1033
Temperaturkoeffizient, 308
Temperaturkoeffizienten, 406
Temperatur-Monitoring, 1047
TEM-Typ, 1162
Terminaladapter, 1303
Tetrade, 590
TE-Welle, Transversal-elektrische Wellen, 1179
T-Flipflop, 622
Theorem von Wiener-Khintchine, 1141, 1377
Thermischer Auslöser, 1003
Thermistor-Motorvollschutz, 1004
Thermoelemente, 927
Thermoempfindlichkeit, 927, 928
Thermoumformermesswerke, 905
Thin Client, 749
Thinwire Ethernet, 756
Thomson-Messbrücke, 907
Thyristoren, 1087, 1101
Tiefpassfilter, 1230
Tiefpassschaltung mit RC-, 377
Tiefpassschaltung RL-Glied, 377
Time Division Multiplexing, 752
TIMER, 696, 713
Timer, 510, 713
Time-Sharing-Betrieb, 749
Timing, 664
Tintenstrahlrucker, 740
TM-Welle, Transversal-magnetische-Wellen, 1179
TN-Netz, 1071
Toggle-Flipflop, 622
Token Ring, 753, 762
Token-(Ring-)Verfahren, 1300
Token-Verfahren, 948
Toleranzen, 554
Tonaderspeisung, 1288
Tonfrequenz-Rundsteuerung, 1071
Top-Down-Methode, 745, 746
Torsionsmoment, 933
Totalreflexion, 240
Totzeit, 828
Totzeitglied, 848
Träger, 1204
Trägerrückgewinnung, 1219
Trägerspeichereffekt, TSE, 1089
Trägheitsmoment, 207
Transformationen, 1189
Transformationsgleichung, 1186
Transformator, 964
Transformatorschutz, 972
transienter äußerer Wärmewiderstand, 1115
transienter innerer Wärmewiderstand, 1114
Transistor-Grundschaltungen, 451
Transistorschalter, 500
Transitfrequenz, 432
transitionaler Kopplungsfaktor, 1234
Transitionen, 808
Transponder, 1267
Transversalwelle, 228
Treiberstufen, 470
Trellis-Code-Modulation, 760
Trellis-Diagramm, 1315
TRIAC, 1107
Trigger-Flipflop, 622
Trinitron-Farbbildröhre, 738
Triplate-Leitung, 1182
Triple Play, 1305
Tristate, 640
Tristate-Treiber, 684
Tschebyscheff-Filter, 1241
TSE, 1088
TTL, 633, 634, 637
TT-Netz, 1072
Turboläufer, 985
Turn-Around-Time, 745
Twisted-Pair-Kabel, 756
Typenraddrucker, 740
TZ-Glied, 848, 851

U

Überkopfzünden, 1102
Überlagerungsempfänger, 1246
Überlagerungsprinzip, 849
Überlastung von Transformatoren, 972
Übernahmeverzerrungen, 479
Überreichweiten, 1202
Überschwingweite, 838
Übersetzungsverhältnis, 966
Übersichtsschaltplan, 569
Überspannungsschutz, 1073, 1120
Übersteuerung, 1092
Übersteuerungsfaktor, 500
Überstromschutz, 421, 1073, 1122
Übertragungsbeiwert K_{PS}, 826
Übertragungseigenschaften, 452
Übertragungsfaktor, 1143, 1166, 1197
Übertragungsfaktor, allgemein, 1332
Übertragungsfunktion, 854, 1135, 1141, 1328, 1332
Übertragungsglied, 848
- akausales, 849
- dynamisches, 849
- kausales, 849
- lineares, 849
- nichtlineares, 849
- statisches, 849
- zeitinvariantes, 849

zeitvariantes, 849
Übertragungsmaß, 1144, 1145
Übertragungsmaß Vierpole, passive
längssymmetrische, 1158
UDP, 767
Ultraschall, 233, 235, 1284
Ultraschall-Durchflussmessung, 936
Umfangsgeschwindigkeit, 198
Umkehrverstärker, 490
umrichtergekoppelter Asynchrongenerator, 1026
umrichtergekoppelter Synchrongenerator, 1027
UMTS, 1306
unabhängige Variablen, 1185
UND-Verknüpfung, 591
Unendlich lange Leitung, 1165
Ungleichung von Kraft, 1310
Universalmotor, 1001
Universalplatinen, 574
universelle Gaskonstante, 216
Unschärferelation, 1342
unsymmetrischen Einphasen-Brückenschaltung, 999
unsymmetrisches Dreileiternetz, 387
Unterbrechungssteuerung, 722
unteres Seitenband (US), 1205
Unterfrequenz-Lastabwurf, 1039
Unterprogramme, 721
untersynchrone Kaskade, 984
USART, 703
User Data Protocol, 767

V

V.24, 758
Valenzelektronen, 306
Variable, 782
direkte, 783
globale, 784
Varianz einer Stichprobe, 1377
Varianz einer Verteilung, 1377
Varistor, 1096
VDR-Widerstand, 1121
Verarmungstyp, 433
Verbinden von Lichtleitern, 1269, 1275
Verbindungshalbleiter, 276
Verbund-Leiterseil, 1054
Verdrahtungsplan, 569
Verdrängungsprinzip, 936
Vergleichsfunktion, 800
Vergleichsglied, 844
Vergleichsstelle, 927
Verify Check, 682
Verknüpfungssteuerung, 794
Verlustarme Leitung, 1169
Verlustfaktor, 1064, 1162
Verlustfaktor, Messung von L, C, 901, 917
Verlustleistung, 399, 637, 1089
Verlustleistungshyperbel, 1093
Verlustlose Leitung, 1164
verlustlose Leitung, 1161, 1169
Verlustziffer, 296
Vermaschung, 597
Verriegeln über die Rücksetzeingänge, 796
Verriegeln über die Setzeingänge, 796
Versorgungssicherheit, 1052
Verstärkerwirkung, 450
Verstärkungsmaß, 1144
Verstärkungsprinzip, 849
Verteilung, Erwartungswert, 1376
Verteilung, linearer Mittelwert, 1376
Verteilungen, 1375, 1379
Verteilungsgesetz, 593
Vertikalantenne, 1198
Vertragsbeziehung, 1012
Vertrauensbereich, 875, 882, 959
Vertrauensgrenzen, 882, 958
Vertrauensniveau, 882
Verzerrungen, 1141
Verzerrungen, lineare, 1142
Verzerrungen, nichtlineare, 1142
verzerrungsfreie Leitung, 1167
Verzögerungsglied erster Ordnung, 848
Verzögerungsglied zweiter Ordnung, 848
Verzögerungssystem, 870
Verzögerungszeit, 500, 1093
Verzugszeit, 828
Vielfachzugang, 1266
Vierphasenumtastung (4-PSK), 1220
Vierpole, aktive, 1148
Vierpole, Betriebskenngrößen, 1151
Vierpole, längssymmetrische, 1156
Vierpole, längssymmetrische passive, Wellenwiderstand, 1158
Vierpole, lineare, 1148
Vierpole, nichtlineare, 1148
Vierpole, passive, 1148
Vierpole, passive, Wellenparameter, 1147, 1157
Vierpole, rückwirkungsfreie, 1156
Vierpole, übertragungssymmetrische (reziproke), 1155
Vierpole, widerstandssymmetrische, 1156
Vierpole, Zusammenschaltung, 1148
Vierpol-Ersatzschaltbild, 452
Vierpolgleichungen, 452, 1148
Vierquadrantenbetrieb, 1000
Viertelbrücke, 908
Villard-Schaltung, 412
virtuelles Bild, 239
Viskosität, 212
Voice-over-IP, 748
VoIP, 748, 1305
Volladdierer, 612
Vollbrücken, 908
Vollgesteuerte Einphasen-Brückenschaltung, 999
vollständige Messschaltung, 912
Vollverbindung, 753
Volumenausdehnungskoeffizient, 215
Volumenstrom, 212
Vorhaltezeit, 868
Vorhaltzeit T_v, 835
Vorsteuerung, 845
Vorstufen, 470
Vorwärtsrichtung, 1102
Vorwärts-Sperrzustand, 1102
Vorwärtsübertragungsfaktor, 1184
Vorwärtszähler, 627
VZ_1-Glied, 848
VZ_2-Glied, 848

W

wahrer Wert, 877
Wahrscheinlichkeitsdichtefunktion, 1376
WAN, 1304
WAN, Wide Area Network, 750
Wanderwellen, 1169
Wärmekapazität, 217
Wärmeleitpaste, 1113
Wärmeleitung, 219
Wärmestrahlung, 219
Wärmeströmung, 219
Wärmeströmungsfeld, 1112
Wärmewiderstand, 401, 1112
Wartung von Maschinen, 1006
Wasser-Dampf-Kreisprozess, 1016
Wasserkraftwerk, 1022
Wasserstoff-Speicher, 1051
Wechselfeldpermeabilität, 283
Wechselrichter, 1028
Wechselspannungen, Messung, 901, 904
Wechselspannungsmessgeräte, 905
Wechselstrombrücken, 999
Wechselströme, Messung, 906
Wechselstromgegenkopplung, 459
Wechselstrom-Messbrücken, Impedanzmessung, 909
Wechselstromnetz, 1034
Wechselströme, Messung, 901
Wegaufnehmer, 941
Wegmessung, 938
Weißes Rauschen, 1139
Weiß'sche Bezirke, 346
Wellen, cosinuselliptische, 1181
Wellen, sinuselliptische, 1181
Wellenausbreitung, 1193, 1200, 1202
Wellenausbreitung im Plasma, 1201
Wellenfilter, 1171
Wellenkraftwerk, 1023
Wellenlänge, kritische, 1180
Wellenoptik, 237, 246
Wellentheorie, 238
Wellentheorie des Lichtes, 238
Wellentyp, 1180
Wellenwiderstand, 1164, 1177
Wellenwiderstand als Funktion von w/h, 1182
Wellenwiderstand des leeren Raumes, 1194
Wellenwiderstand des Plasmas, 1201
Wellenzahl, 228
Wellrohre, 1178
Wendepole, 992
Wertetabelle, 590
Wheatstone-Messbrücke, 906
Wheatstone-Messbrücke, Abgleichverfahren, 906
Wickelstromwandler, 975
Wicklungen, 964
Wide Area Network, 750
Widerstand, 307
Widerstandsänderung, 924
Widerstandsbremsen, 994
Widerstandsbremsung, 996
Widerstandsmessung, 906
Widerstandsrauschen, 1138
Widerstandsthermometer, 922
Wien-Brücke, 515
Wien-Robinson-Brücke, 516
Wien'sches Verschiebungsgesetz, 250
Windenergieanlage, 1023
Winkelbeschleunigung, 198
Winkelgeschwindigkeit, 198
Winkelmessung, 938
Winkelmodulation, 1211
Wired AND, 638
Wired OR, 638
Wirkfläche, 1197

Wirkleistung, 380, 911, 1185
Wirkleistungsmessung, 912
Wirkungsgrad, 204, 321, 969
Wirkwiderstand von Leiterseilen, 1057
WLAN, 1308
Woltmann-Zähler, 935

X
X-Y-Stereofonie, 1250

Y
Yagi-Antenne, 1199
Yellow Cable, 756
Y-Parameter, 1150
y-Parameter, 438

Z
Zahlensystem, 602
Zahlensysteme, 589, 602
Zahlenwertgleichungen, 876
Zähler, 625
Zähler in Automatisierungssystemen, 799
Zählerkonstante, 916
Zählpfeilsystem, 312
Zeichnungsnormen, 548
Zeiger, 224, 364
Zeigerdiagramm, 363, 1206
Zeilensprungverfahren, 738
zeitdiskreten Signalverarbeitung, 1370
Zeitgeber, 713
Zeitgeberbausteine, 696
Zeitglieder nach IEC 61131-3, 799
Zeitinvarianz, 1332
Zeitkonstante, 223, 332, 358, 871
Zeitmessung, 937
Zeitmultiplexverfahren, 615, 752, 1257, 1258
Zeitrelais, 551
Zener-Effekt, 405
Zentraleinheit, 662
Zentraler Grenzwertsatz der Statistik, 1380
Zentralkompensation, 1076
Zentralspeicher, 687, 689
Zentralwert, 958, 1376
Zentripetal-Beschleunigung, 201
Zentripetalkraft, 201
Zerstreuungslinsen, 242
Zielextraktor, 1279
Z-Parameter, 1150
z-Transformation, 1332, 1369, 1372
Zufällige Abweichungen, 877
Zugriffszeit, 676
Zünddiagramm, 1103
Zündimpulse, 423
Zündspannung, 418
Zündstrom, 418
Zündverzögerungszeit, 1103
Zündvorgang, 416, 1103
Zündzeit, 419
Zuverlässigkeit, 553
Zweierkomplement, 603, 604, 613
Zweig, 1185
Zweiphasenumtastung (2-PSK), 1219
Zweipunkt-Regelfunktion, 838
Zweiquadrantenantrieb, 999
Zweiquadrantenbetrieb, 1000
Zweirichtungstreiber, 690
Zweiseitenband-Amplitudenmodulation, 1205
Zwischenfrequenzdurchschaltung, 1267
Zyklische Codes, 1312
Zylinderkondensator, 328
Zylinderspule, 336