



Technik-Trends 2024

Ausblick: Was bei Hardware und Software wichtig wird

Weitblick: Highlights aus Wissenschaft und Forschung

HARDWARE

- Display-Technik: Auf dem Weg zur Mikro-LED
- Die CPU-Pläne von AMD und Intel
- USB4: Schnelle Docks und SSDs
- Wi-Fi 7 beschleunigt das Heimnetz
- Wie Apples Vision Pro funktioniert
- Smart Home: Flexibel dank Matter



Gamechanger KI

Wirtschaftsmotor oder Stromfresser?

SOFTWARE

- Windows aus der Cloud
- Cloud frisst Open Source
- Passkeys lösen Passwörter ab
- Industrialisierung des Cybercrime
- EU-Bezahldienste: Günstig & datensparsam
- Warum Digitalisierung in Deutschland lahmt

Energiewende konkret

E-Autos: Schneller laden, weiter fahren • Batterie-Roadmap

Dynamische Stromtarife: Für wen sie sich lohnen

Energiespeicher gegen die Dunkelflaute



€ 5,90

AT € 6,50 | LUX, BEL € 6,90

NL € 7,20 | IT, ES € 7,40

CHF 9.90 | DKK 64,00

Für diesen heiklen Job braucht der Cyberdealer seine Crew!

Doppelter RAM am Orange Friday!

► 24.11.2023

Da gehen selbst dem Cyberdealer die Hände aus: Am Orange Friday zieht er bei Thomas-Krenn seinen nächsten großen Coup durch, und **schenkt allen Bestellern** ohne Aufpreis den **doppelten Arbeitsspeicher** für ihre gewählte Konfiguration. Und als wäre dieser verboten gute Deal nicht genug, ist außerdem ein mysteriöses **Glücksrad** mit wertvollen Sachpreisen in unserer Produktionshalle aufgetaucht...



Vorab Infos unter:
thomas-krenn.com/of23
+49 (0) 8551.9150-300

THOMAS
KRENN[®]
IT's people business



IT- und Technik-Trends: Vorsprung durch Wissen

Liebe Leserinnen und Leser,

die c't, die Sie gerade lesen, ist keine normale c't. Statt aktuelle Gadgets zu testen oder mit Hard- und Software herumzufricken, widmen wir uns in dieser Sonderausgabe ganz dem Thema Zukunft: Wir blicken auf die aus unserer Sicht wichtigsten IT- und Technik-Trends.

Dabei betreiben wir keine Spökenkiekerei. c't-typisch bleiben wir lieber bei den Fakten. So wollen wir Ihnen einen Wissensvorsprung verschaffen und Sie in die Lage versetzen, die aktuellen und auch kommende Hypes einzuschätzen.

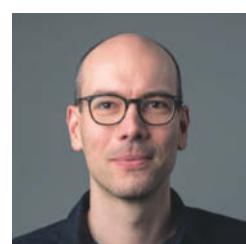
Welche Fortschritte machen die viel gerühmten Passkeys? Wie funktioniert eigentlich Windows in der Cloud? Welche Prozessoren bringen Intel und AMD 2024? Wie realistisch sind die Pläne der Bundesregierung zur Digitalisierung und zur Energiewende? Welche Technik steckt eigentlich in Apples Computerbrille Vision Pro? Woran forschen die Displayhersteller? Was passiert im Cyber-Untergrund? Das sind nur einige der vielen Fragen, die wir in diesem Heft beantworten.

Einige Beiträge blicken etwas weiter in die Zukunft. Klar, niemand weiß, ob wir künftig alle

einen Chip im Kopf tragen, mit 900 km/h durch Unterdruckröhren sausen und Elon Musk endlich auf den Mars schießen können. Trotzdem gibt es heute schon Handfestes zu berichten: Was Brain-Computer-Interfaces aktuell leisten, mit welchen Herausforderungen Hyperloop-Entwickler kämpfen und warum ein Nuklearantrieb die Flugzeit zum roten Planeten verkürzen würde.

Und manches Zukunftsszenario scheint plötzlich überraschend nah. Zum Beispiel entwerfen Mobilfunkbetreiber und Satellitenfirmen ein weltraumgestütztes 5G-Netz, das nicht für klobiges Sat-Equipment gedacht ist, sondern künftige Smartphones direkt versorgt. Das würde nervigen Funklöchern endlich den Garaus machen.

Wie gefällt Ihnen diese nichtnormale c't? Ich freue mich auf Ihr Feedback an cwo@ct.de.



Christian Wölbert

Christian Wölbert

Inhalt

Energiewende

- 8 Energie, Forschung, KI** Die Trends 2024
- 10 Erneuerbare Energien** Der Stand in Deutschland
- 16 Stromspeicher** Wege aus der Dunkelflaute
- 22 Dynamische Stromtarife** Für wen sie sich lohnen
- 28 E-Autos** Schneller laden, weiter fahren
- 32 Batterien** Energiedichte und Rohstoffe im Blick

Hardware

- 110 Hardware, Mobiles, Netzwerke** im Wandel
- 112 Hybrid-CPUs** Die Pläne von AMD und Intel
- 116 Kluge Chips** x86-Prozessoren mit KI-Einheiten
- 120 CPU-Mikroarchitektur** Was Intels x86-S bringt
- 124 Rechenbeschleuniger** für Supercomputer
- 126 Chip-Fertigung** soll umweltfreundlicher werden
- 132 USB4** Eine Schnittstelle für alles
- 136 Smart-TVs** Aktuelle und künftige Displaytechnik
- 140 Smart-Home-Standard** Matter kommt in Schwung

Forschung & KI

- 36 Forschungs-Highlights** Verkehr, Raumfahrt, Gehirn
- 44 Künstliche Intelligenz** Wirtschaft im Umbruch
- 50 Sprachmodelle** Es muss nicht immer riesig sein
- 58 KI-Energieverbrauch** Mücke oder Elefant?

Software & Security

- 64 Software, Sicherheit, Vernetzung** Alles as a Service
- 66 Windows** Zukunft in der Cloud
- 70 Open Source** Neue Lizizenzen gegen Cloudgiganten
- 76 Passkeys** Der Passwort-Killer wird erwachsen
- 80 Cyberkriminalität** Erpressen wie die Profis

Internet & Digitalisierung

- 86 Digitalgesetze** Was die neuen EU-Regeln bringen
- 92 Verbraucherschutz** Guter Service ist Glückssache
- 94 EU-Bezahldienste** Günstig und datensparsam
- 98 Social Media** Wie neue Plattformen X ablösen
- 104 Gesundheitssystem** Allheilmittel Digitalisierung?
- 108 Verwaltung** Wo die Digitalisierung stockt

Technik-Trends 2024



CPUs mit KI-Einheiten, Passkeys statt Passwörter, Windows aus der Cloud, Apples VR-Einstieg, Satelliten-5G-Netze für Smartphones ... was ist Hype, was echter Fortschritt? Wir stellen in diesem Heft die wichtigsten IT- und Technik-Trends für die nächsten Jahre vor und klären, was für Nutzer wirklich dabei herumkommt.

Mobiles & Netze

- 144 **Smartphones** werden teurer und langlebiger
- 148 **5G aus dem All** Vom Satelliten direkt ins Handy
- 154 **Mobile Gaming** Hype um Handheld-Konsolen
- 156 **3D-Brille** Wie Apples Vision Pro funktioniert
- 160 **Netzwerke** Mehr Speed für WLAN und Ethernet

Immer in ct

- 3 **Standpunkt** Vorsprung durch Wissen
- 7 **Schlagseite**
- 164 **Web-Tipps** Terminvorschau, virtuelles Tagebuch
- 166 **Buchkritik** Blockchains und Kryptowährungen
- 166 **Buchkritik** Digitale Freiheit trotz Social Media
- 168 **Story** Der neue Freund
- 175 **Stellenmarkt**
- 177 **Inserentenverzeichnis**
- 178 **Impressum**



Die unumgängliche Energiewende zwingt dazu, umzudenken und etablierte Pfade zu verlassen. Wir ergründen, wie es um die Erzeugung und Nutzung von klimaneutraler Energie in Deutschland steht, und liefern Perspektiven für eine zukunftssichere und umweltverträgliche Mobilität.



Hardcore

- 120 **x86-S** Intels kommende CPU-Mikroarchitektur

Hype

B U Z Z W O R D

B - I - N - G - O

| | | | | | |
|---|---------------|--------------------|-------------------|-------------------|---------------------|
| 1 | Disruption | Paradigmen-wechsel | Skalierbar-keit | Omni-channel | Neuronale Netzwerke |
| 2 | Voice Control | Benchmark | Data driven | Blockchain | Funnel-Optimierung |
| 3 | NFT | Deep Dive | Low-Hanging Fruit | Gamechanger | E-Sports |
| 4 | Agile | Im Loop halten | Proof of concept | Change-Management | Krypto |
| 5 | Pain Point | Trans-formation | Synergie | No Brainer | Mindset |



ct 3003 Newsletter



Kein Buzzword-Bingo, einfach nur Technik.
Erfahrt, was wirklich zählt!

Jetzt **KOSTENLOS** abonnieren:
ct.de/hype





Weitere Schlagseiten auf ct.de/schlagseite

Den Wandel gestalten

Trends bei Energiewende, Forschung und KI

Wir leben in spannenden Zeiten: Die Energiewende ist unumgänglich und zwingt uns, umzudenken und etablierte Pfade zu verlassen. Gleichzeitig krempeln immer mehr KI-Tools die Geschäfts- und Arbeitswelt um. Was steht uns da in nächster Zeit bevor und wie hilft die Forschung, die anstehenden Herausforderungen zu meistern?

Von Georg Schnurer



Einfach weitermachen wie bisher funktioniert immer seltener. Der Klimawandel zwingt uns zum Umsteuern: Die Erde wird immer wärmer, Extremwetter und Umweltkatastrophen häufen sich. Dekarbonisierung, also das Einsparen von CO₂-Emissionen, ist der einzige Weg, die Lebensbedingungen auf unserem Planeten menschenfreundlich zu erhalten.

Doch wie steht es um die Erzeugung und Nutzung von klimaneutraler Energie in Deutschland? Welche Perspektiven eröffnet die Energiewende, welche Hindernisse gilt es noch zu umschiffen? Wie speichern wir Strom von Windkraft- und Solaranlagen in Zukunft? Welche Möglichkeiten gibt es, den Energieverbrauch angebotsgesteuert zu lenken? Wie muss ein zukunftssicheres und umweltverträgliches Mobilitätskonzept aussehen? Antworten liefern die Beiträge ab Seite 10.

Künstliche Intelligenz, die zweite große Herausforderung für Gesellschaft und Wirtschaft, beschäftigt uns ab Seite 44. KI-Systeme erobern Wirtschaft und Gesellschaft: Chatbots übernehmen die Hotlines; Bildgeneratoren liefern auf Knopfdruck Illustrationen für Webseiten, Prospekte und Präsentationen; KI-Systeme analysieren Röntgenbilder, erkennen Hautkrebs und stellen Diagnosen für seltene Krankheitsbilder – schneller und mitunter zuverlässiger als der Mensch.

Vielen machen diese Neuerungen Angst. Dabei stehen KI-Tools erst am Anfang ihrer Entwicklung. Wie geht es also bei KI-Systemen und den dahinter stehenden Rechenmodellen weiter? Wir zeigen den Stand der Technik und beleuchten Perspektiven – damit Sie angstfrei und gut informiert in die Zukunft schauen können.

(gs@ct.de) **ct**





| | |
|--|----------|
| Trends bei Energiewende, Forschung und KI | Seite 8 |
| Sonne, Wind und Bürokratie: Energiewende in Deutschland | Seite 10 |
| Stromspeicher: Wege aus der Dunkelflaute | Seite 16 |
| Sparoption: Lohnen sich dynamische Stromtarife? | Seite 22 |
| E-Auto-Trends 2024: Größer, schwerer, vernetzter | Seite 28 |
| Roadmap: Batterien mit höchster Energiedichte | Seite 32 |
| Übermorgenland: Highlights aus Wissenschaft und Forschung | Seite 36 |
| Möglichkeiten: KI wirbelt die Wirtschaft durcheinander | Seite 44 |
| Kompakte Sprachmodelle: Herausforderung für die Giganten | Seite 50 |
| Mücke oder Elefant: KI-Energieverbrauch | Seite 58 |

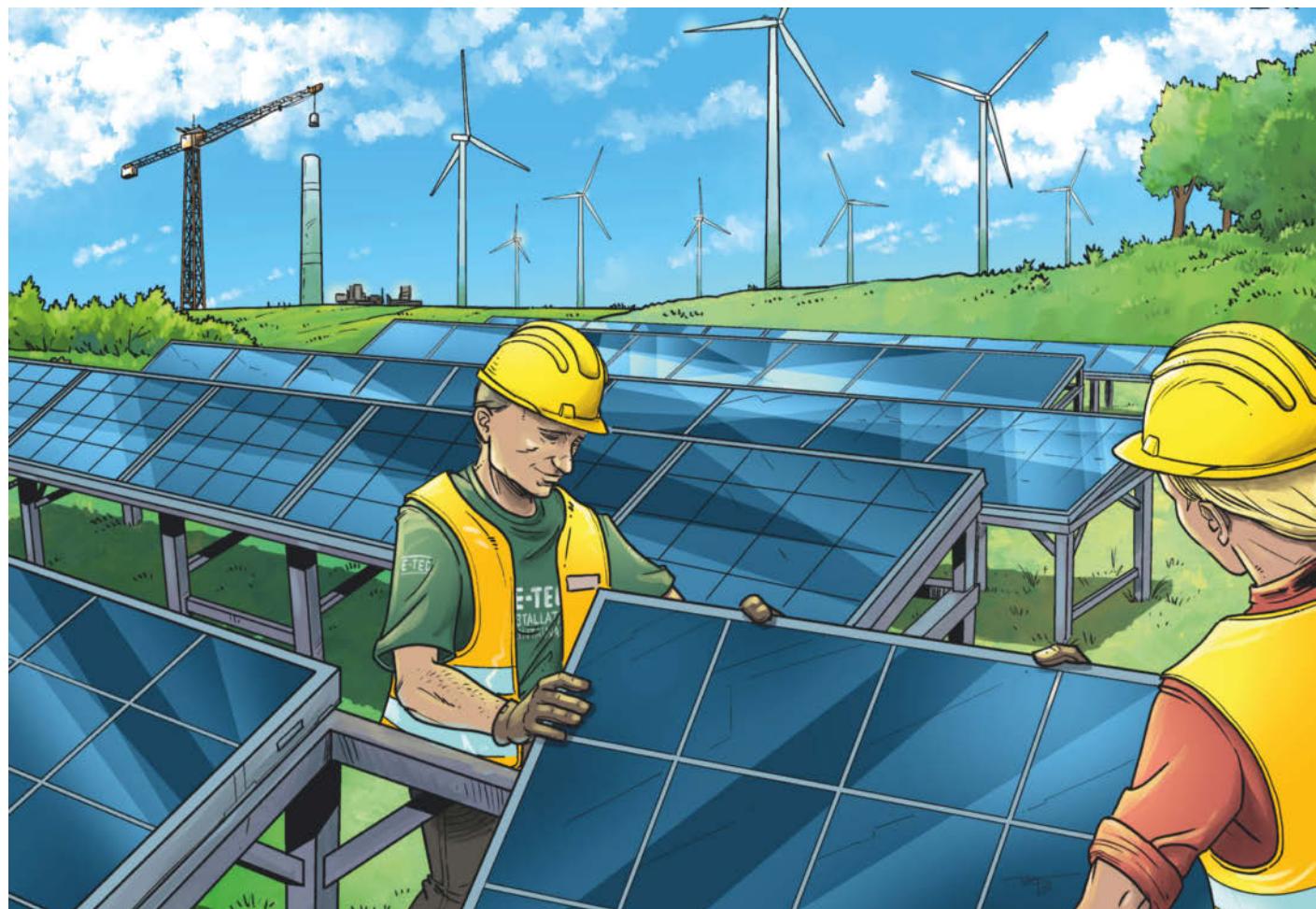


Bild: Michael Vogt

Große Ziele, große Baustellen

So schreitet der deutsche Ausbau erneuerbarer Energien voran

Neue Technik verspricht mehr Effizienz, Privat- und Industriedächer werden in ungekanntem Tempo mit Photovoltaik bedeckt. Aber reicht das, um die Ausbauziele zu erreichen und fossile Energieträger zu ersetzen? Ein Blick auf Zahlen, Technik und Gesetzesreformen.

Von Jan Mahn

Bis zum Jahr 2030 soll Deutschland seinen Stromverbrauch zu 80 Prozent aus erneuerbarer Energie decken – dieses legislativer periodenübergreifende Ziel hat sich die Bundesregierung gesetzt und sich gleichermaßen vorgenommen, bis zur nächsten Wahl die rechtlichen Voraussetzungen zu schaffen, damit es auch erreicht werden kann. Und das Ziel ist durchaus ambitioniert, lag der Anteil bei Amtsantritt von Scholz, Habbecke, Lindner & Co. nur bei rund 40 Prozent. Zwei Sektoren sollen bis 2030 massiv ausgebaut werden: Windkraft (an Land und auf See) und Photovoltaik.

Als die Regierung im Herbst 2021 antrat, war Energiepolitik nicht gerade das heiß diskutierte Topthema, in den Talkshows dominierte noch die Pandemie, und der Ausbau von Photovoltaik und Windkraft war nicht gerade Volkssport. Das goldene Photovoltaikjahr 2011 lag genau zehn Jahre zurück und wehmütig blickte man auf dieses Jahr, in dem dank hoher EEG-Einspeisevergütung über 7 Gigawatt Leistung installiert wurde. 2021 dagegen hatten viele nicht mal eine Vorstellung davon, was Strom pro Kilowattstunde kostet, geschweige denn, wie hoch der Jahres-

verbrauch in etwa ausfiel. Das änderte sich wenige Monate nach Amtsantritt der Regierung schlagartig mit dem russischen Angriffskrieg und der folgenden Diskussion über Versorgungssicherheit und Energiekosten. Die Energieversorger starteten mehrere Preiserhöhungsrunden, es kamen Debatten über die Strompreisbremse auf und spätestens da holten die allermeisten mal ihre Jahresrechnungen aus dem Schrank; es entstand ein neues Bewusstsein für Stromverbrauch und -kosten.

Neue Begeisterung

Nicht nur Energiesparen, Preise vergleichen und Anbieterwechsel wurden populär, in Deutschland stieg auch das Interesse an Photovoltaik – bei Privathaushalten und auch in Handel, Gewerbe und Industrie. Doch reicht die neue Begeisterung, um die hochgesteckten Ausbauziele zu erreichen? Welche Trends gibt es auf dem Markt, was hat der Gesetzgeber getan, um rechtliche und bürokratische Hürden abzubauen? Und wie geht es eigentlich mit dem Ausbau der Windkraft voran? Antworten auf diese Fragen haben wir aus öffentlichen Datenquellen und in Gesprächen mit Herstellern und Installateuren gesammelt.

Die besten Quellen für Informationen über in Deutschland installierte Erzeugungsanlagen, Speicher sowie über den Strommarkt liefert die Bundesnetzagentur in zwei Onlineportalen: dem Marktstammdatenregister (MaStR, www.marktstammdatenregister.de) und auf einer Seite mit dem simplen Namen Strommarktdaten (SMARD, smard.de). Die Daten des MaStR werden von den Betreibern der Anlagen eingestellt – teilweise also auch von Laien, die gerade eine PV-Anlage gekauft haben. Ein Problem in der Gesetzgebung. Während nur ein Elektrofachbetrieb die Anlage beim Netzbetreiber anmelden darf, ist die Eintragung im Marktstammdatenregister Aufgabe des Kunden. Der Netzbetreiber bekommt dann eine Nachricht und hat die Aufgabe, die Angaben mit denen des Installateurs abzulegen. Riecht für Sie nach vermeidbarer doppelter oder dreifacher Arbeit? Mehr dazu später.

Weil das System ist, wie es ist, sind einige Angaben grob falsch; immer mal wieder verwechseln Betreiber die Angaben Watt und Kilowatt und tragen versehentlich PV-Anlagen mit 10.000 Kilowatt ein, die eigentlich „nur“ 10 Kilowatt groß sind. Bei unserer Auswertung haben wir unplausible Einträge nach Möglichkeit ge-

c't kompakt

- Deutschland hat große Pläne beim Ausbau von Windkraft und Photovoltaik. Bis 2030 gibt es konkrete Ziele.
- Windkraft in Deutschland ist noch immer ein zähes Geschäft mit vielen bürokratischen Hürden – doch kleine Verbesserungen sind in Sicht.
- Photovoltaik erlebte 2023 ein Rekordjahr in Deutschland, beim Bürokratieabbau ist das Land aber noch am Anfang.

filtert. Falls Sie länger nicht mit großen Zahlen gerechnet haben: 1000 Watt sind ein Kilowatt (kW), 1000 Kilowatt ein Megawatt (MW) und 1000 Megawatt ein Gigawatt (GW). 1000 Gigawatt wiederum sind ein Terawatt (TW). Ein Verbraucher mit einer Leistung von 1000 Watt, der eine Stunde lang läuft, hat 1000 Wattstunden verbraucht, also eine Kilowattstunde, abgekürzt kWh. Während Watt eine Leistungsangabe ist, gibt die Kilowattstunde eine Energiemenge an. Ein Speicher kann 5 kWh haben, ein Einfamilienhaus 10 kWh pro Tag benötigen.

Balkonkraftwerke in Deutschland erreichen bereits 192 Megawatt Nettoleistung.

Dieser Artikel soll Zahlen einordnen, die teils hitzig geführte Debatte versachsen und wagt auch einen Blick in die Zukunft. Bei Diskussionen über Energiepolitik lauern aber diverse argumentative Fallen und Milchmädchenrechnungen. Marketing-Agenturen veröffentlichen gern Vergleiche wie „das entspricht der Leistung von zwei Atomkraftwerken“ oder „damit lassen sich 5000 Haushalte mit Strom versorgen.“ Beide Aussagen haben gewaltige Tücken: Wer die Spitzenleistung einer Photovoltaikanlage oder einer Windkraftanlage direkt mit der eines Atomkraftwerks vergleicht, der vergisst, dass alle drei Erzeugungsanlagen übers Jahr gesehen unterschiedlich lang auf voller Leistung arbeiten. Außerdem gibt es zu wenig Speicher, um den Strom dann nutzen zu können, wenn er gebraucht wird – mehr dazu auf Seite 16.

Eine PV-Anlage produziert je nach Ausrichtung und Ort zwischen 800 und 1000 Kilowattstunden pro installiertem Kilowatt Spitzleistung (kW Peak) und das logischerweise nur am Tage. Eine große Windkraftanlage auf See kommt auch mal auf 5500 kWh pro installiertem Kilowatt. An Land dagegen können es auch mal nur 2000 kWh sein, die Ertragsprognose für neue Windkraftanlagen ist eine eigene Disziplin. Neckarwestheim II schließlich, eines der letzten deutschen Atomkraftwerke mit einer Nennleistung von 1400 Megawatt, produzierte im Jahr 2022 ganze 7800 Kilowattstunden pro installiertem Kilowatt.

Auch die Vergleiche mit Haushalten sind schwierig. So gibt das Statistische Bundesamt den Verbrauch eines Einpersonenhaushalts im Jahr 2021 mit 2105 kWh an, bei drei und mehr Personen im Haushalt kommen die Statistiker auf 5411 kWh. Weil die meisten Vergleiche hinken, verzichten wir im Folgenden darauf.

Strom aus Wind

115 Gigawatt Windkraft an Land – das ist das Ausbauziel der Bundesregierung bis zum Jahr 2030. Der Weg bis dahin ist noch weit. Als die Regierung 2021 ihren Dienst antrat, waren knapp 55 Gigawatt Leistung installiert und seitdem ist kein Windkraft-Boom an Land zu erkennen. Gerade mal

5,2 Gigawatt wurden in den 24 Monaten seitdem installiert, im laufenden Jahr 2023 waren es bisher nur mickrige 2,3 GW an Land, dazu 0,2 GW auf See. Um das 115-Gigawatt-Ziel noch zu erreichen, müssten ab jetzt jedes Jahr 10 Gigawatt zugebaut werden.

Immerhin: Knapp 2600 Anlagen mit zusammen 12,7 Gigawatt Windkraftleistung stehen aktuell im Marktstammdatenregister als „In Planung“. Doch das heißt noch lange nicht, dass sie in den nächsten Wochen oder Monaten ans Netz gehen. Die ältesten eingetragenen Projekte verharren in diesem Zustand seit Februar 2019. Dass diese Zeiträume zu lang sind, hat auch die Bundesregierung als Problem ausgemacht und wollte reagieren.

Richten soll es die „Windenergie-an-Land-Strategie“, die Robert Habeck im Mai 2023 vorgestellt hat. Darin werden



Bild: EnBW / Paul Langrock

Den größten Solarpark Deutschlands hat EnBW in Brandenburg errichtet. 144 Megawatt Wechselrichterleistung und bifaziale Module hat das Unternehmen installiert.

Probleme aus verschiedenen Bereichen ausgemacht und Lösungen versprochen. Den Status quo fassen die Autoren der Regierungsstrategie drastisch zusammen: „Um Genehmigungsverfahren zu beschleunigen, muss die richtige Balance zwischen den widerstreitenden Belangen gefunden und das Verfahrensrecht deutlich gestrafft werden. Genehmigungsverfahren dauerten im Jahr 2021 im Bundesdurchschnitt über zwei Jahre ab Einreichung der Erstunterlagen und über zehn Monate ab Feststellung der Vollständigkeit der Antragsunterlagen.“ Ein paar Maßnahmen zur Beschleunigung hat die Regierung bereits umgesetzt, in ausgewiesenen Windenergiegebieten entfällt bis Juni 2024 die Umweltverträglichkeits- und -artenschutzrechtliche Prüfung. Unter anderem soll auch das Genehmigen von Windenergieanlagen in Industrie- und Gewerbegebieten einfacher werden. Das „Wind-an-Land-Gesetz“ (WaLG) soll bis 2027 die für Windenergie ausgewiesene Fläche in Deutschland von 0,8 auf 1,7 Prozent erhöhen. Weniger Bürokratie und mehr Fläche also – ob das einen Windrad-Boom entfacht, müssen die nächsten Jahre zeigen.

Einige nicht so intensiv thematisierte Hürden müssen Unbeteiligten erschreckend banal erscheinen, in Zeiten, in denen alle politischen Ebenen den Ausbau der Erneuerbaren zum Ziel erklärt haben. Da ist zum Beispiel der Transport von Windkraftanlagen – insbesondere die Ro-

toren sind lang und wurden über die Jahre immer länger, wie eine detailliertere Auswertung der Daten aus dem MaStR zeigt: Lag der Median der Rotordurchmesser aller im Jahr 2010 gebauten Anlagen noch bei 82 Metern, sind es im Jahr 2023 knapp 140 Meter. Logistikunternehmen, die sich auf Großraum- und Schwertransporte (GST) spezialisiert haben, müssen die immer größeren Teile zum Aufstellort transportieren. Doch schon seit Jahren klagt diese Branche, dass die logistischen Herausforderungen der nächtlichen Fahrten längst nicht mehr die meiste Arbeitszeit verschlingen, sondern die Genehmigungsverfahren und das Warten auf bearbeitete Anträge.

Die Folge sind Rotoren, die wochenlang auf Autobahnrasen geparkt sind. Außerdem fehle es, so die Branchenverbände, an einem zentralen System, in dem alle Baustellen und Hindernisse von der Autobahn bis zur Kreisstraße erfasst sind – denn eine unerwartete Fahrbahnverengung durch Bauarbeiten kann eine Schwerlastkarawane zum Anhalten zwingen. Technische Verbesserungen sind am Horizont bereits erkennbar. Die regionalen Niederlassungen der Autobahn GmbH stellen seit dem Sommer auf die gemeinsame Software GST.Autobahn um und versprechen, Anträge künftig innerhalb von wenigen Werktagen, nicht innerhalb von Wochen zu bearbeiten. Auch das gemeinsame Kartensystem für alle Straßentypen ist zumindest im Planung.

Fusionsenergiekollektoren

Das Ausbauziel von 115 Gigawatt bei Windkraft an Land liegt wegen großer und kleiner Hürden noch in weiter Ferne. Ganz anders sieht die Lage bei der Photovoltaik aus, die in den meisten Fällen ohne Genehmigung und ohne Gutachten auskommt. Solange eine PV-Anlage auf dem Dach von genehmigten Gebäuden errichtet wird, braucht sie keine separate Baugenehmigung. Und Photovoltaik gehört zu den günstigsten Verfahren, um elektrische Energie zu erzeugen.

Dafür reicht eine kleine Beispielrechnung: Eine Dachanlage mit 10 kW Peak auf dem Dach eines Einfamilienhauses bekommt man ab etwa 15.000 Euro (ab 1,50 Euro pro Watt). Im Jahr erzeugt sie bei optimaler Ausrichtung etwa 10.000 Kilowattstunden und das in vielen Fällen über 20 Jahre lang. Insgesamt kommen in ihrer Lebenszeit also 200.000 Kilowattstunden vom Dach. Die Kilowattstunde kostet bei dieser Rechnung 8 Cent in der Erzeugung. Schrottreif ist die Anlage nach diesem Zeitraum nicht automatisch, zumindest wenn man die Erfahrungen mit Komponenten heranzieht, die vor 20 Jahren gebaut wurden und heute noch laufen. Die Module degradieren über die Zeit, schaffen im Alter also ein paar Watt weniger und manch Wechselrichter muss zwischen 15 und 20 Jahren defekt getauscht werden, die gesamte Anlage arbeitet aber meist länger. Wer sich heute eine PV-Anlage installieren lässt, kauft einen großen Anteil des eigenen Energiebedarfs für die nächsten Jahrzehnte. Die Zeiten, in denen die Technik staatliches Fördergeld braucht, sind vorbei – dass Verkehrsminister Wissing im September dennoch ein Förderprogramm auflegte, steht auf einem anderen Blatt [1].

Je größer die Anlage, desto günstiger wird jedes installierte Kilowatt. Bei großen Freiflächenanlagen können die Betreiber den Preis auch unter 1 Euro pro Watt drücken und kommen inklusive Kosten für Wartung und Instandhaltung über die 20 Jahre auf bis zu 4 Cent pro Kilowattstunde, wie verschiedene Studien gezeigt haben. Deutschlands bisher größten Solarpark, der als einzelne Anlage im Marktstammdatenregister eingetragen ist, hat EnBW im Jahr 2022 in der Gemeinde Neuhardenberg in Brandenburg unter dem Namen „Solarpark Gottesgabe“ errichtet. Knapp 350.000 bifaziale Module – also solche ohne lichtundurchlässige Rückseitenfolie, die auch diffuses Licht von der Rückseite

nutzen – hat der Energieversorger an Wechselrichter mit einer Leistung von 144 Megawatt angeklemmt und dem Park noch einen Batteriespeicher für 3,5 Megawattstunden verpasst. Ein weiteres EnBW-Projekt gleicher Dimension steht in der Gemeinde Neutrebbin. EnBW betont, dass der Strom unabhängig vom Erneuerbare-Energien-Gesetz und garantierten Einspeisevergütungen zu Marktpreisen an der Strombörse und über Power Purchase Agreements (PPA) an Großabnehmer verkauft wird.

Import, Export

Verglichen mit Meldungen aus China wirken jedoch selbst diese Solarparks klein. Der Mythos, dass China seine Solarfabriken nur für den Export nach Europa errichtet, wackelt. Im Rahmen eines staatlichen Programms, um Energie in trockenen Regionen wie der Wüste Gobi zu erzeugen, entstehen dort Parks im Gigawattbereich. Im Juni verkündete die chinesische Energiebehörde NEA, dass im Land 380 Gigawatt PV-Leistung installiert sind, fast 50 davon wurden bis Ende April errichtet, 15 GW allein im April – das Jahresziel für 2023 soll bei über 100 GW liegen.

Da passt es gut, dass auch die größten Hersteller der Branche aus China kommen. Die Riesen wie Longi (110 Gigawatt Produktion im Jahr) Tongwei (80 bis 100 GW), JinkoSolar (75 GW), dominieren den Weltmarkt klar und die deutschen, österreichischen und schweizerischen Hersteller spielen in einer anderen Liga; sie spre-

chen vor allem das Publikum in ihrem Heimatmarkt an, das Wert auf made in Europe legt. Meyer Burger zum Beispiel mit Firmensitz in der Schweiz und Produktion in Freiberg in Sachsen arbeitet daran, seine Produktion auf 1,4 GW auszubauen und 2027 einmal 7 GW zu erreichen.

Die Standards im Geschäft bestimmen derweil andere: Im Juli 2023 verkündeten neun der großen Hersteller aus China, dass sie sich auf eine neue standardisierte Modulgröße geeinigt hätten und künftig alle Module in den Maßen 2382 x 1134 mm herstellen wollen. Diese Meldung sorgte in Deutschland teilweise für Panik – wurde sie zunächst so interpretiert, dass es bald gar keine kleineren Module mit etwa 1,8 m Länge mehr geben würde. Dazu wird es aber nicht kommen. Die kleineren Module für private Dächer wird es weiter geben, die großen Brecher (meist mit nicht so hübschem silbernen Rahmen) sind für Industriedächer, Freiflächenanlagen und einen der weiteren Trends des jüngsten Solarbooms gedacht: Agri-PV.

Zunehmend entdecken Landwirte, dass sich die Flächennutzung als Agrarfläche und für Photovoltaik nicht ausschließt – im Gegenteil. Gemüsepflanzen, Hühner und Weidetiere fühlen sich unter dem Teilschatten von halb-transparenten Modulen sehr wohl. Das Markstammdatenregister kennt noch kein Attribut für solche Anlagen, sodass ihre Zahl nicht zu



Bild: Sebastian Müller

Ein Beispiel für Agri-PV: Halbtransparente Module über Weinreben, dazu entschied sich Winzer Edgar Gimbel im Freiburger Ortsteil Munzingen. Das PV-Dach wurde 2023 fertiggestellt und schützt nicht nur vor zu viel Sonne und Hagel, sondern gewinnt nebenbei noch Strom.

just
DOCK IT.
Free your hands



**DREHBARER STÄNDER
FÜR TABLET/PHONE MIT
8-in-1 DOCKINGSTATION**

✓ **KLEIN, STABIL, ROBUST**
Für Smartphone/Tablet bis 11",
faltbare Aluminiumkonstruktion

✓ **VOLLE FLEXIBILITÄT**
360° Rotation und zwei
Neigungswinkel für Ablage-
fläche und Auslegearm

✓ **USB-C DOCKINGSTATION**
HDMI® 2.0, SD/microSD 3.0,
USB 3.2 Gen 1 Type-C®/
Type-A Ports, Headset-Buchse
und USB-C Power Delivery



Hier
erhältlich:

cyberport **reichelt**
elektronik **computeruniverse**
NBB **CONRAD**



JETZT MEHR ERFAHREN:

www.icybox.de

 [icyboxofficial](https://www.instagram.com/icyboxofficial)  ICY BOX

 [RaidSonic Technology GmbH](https://www.linkedin.com/company/raidsonic-technology-gmbh/)

ermitteln ist, doch Beispiele tauchen im Jahr 2023 immer wieder in lokalen Nachrichten auf.

Für PV-Interessierte sind es keine schlechten Nachrichten, dass die Module in großen Mengen nachgefragt und hergestellt werden. Höheres Angebot und teilweise auch staatliche Subventionen in

Versorgungssicherheit bei Stromausfällen zu erhöhen. Seit öffentlich mehr über die Möglichkeit von Black- und Brownouts diskutiert wird, steigt das Interesse an PV-Anlagen mit Speicher und Notstromfunktion. Bis Anfang 2021 hatten sich die meisten Betreiber von PV-Anlagen damit abgefunden, dass ihre netzsynchrone Anlage

Elektriker besprechen, denn Nachrüsten ist nicht immer möglich. Auf der einen Seite muss der Batteriewechselrichter tauglich sein. Dann gibt es zwei unterschiedlich teure Optionen: Deutlich günstiger sind Systeme, bei denen am Wechselrichter oder Speichersystem eine Schuko-steckdose angebracht ist. Fällt das Netz aus, kann man dort Kühltruhe oder Heizungssteuerung provisorisch betreiben, muss aber Kabel selbst umstecken. Teurer sind Notstromsysteme, die entweder das gesamte Heim oder einzelne Stromkreise versorgen. Der Elektriker muss dann einen Umschalter in der Verteilung einbauen, der reagiert, wenn das öffentliche Netz ausfällt und das Hausnetz davon vollständig trennt.

Aber auch ohne Notstromfähigkeit gehören Batteriespeicher bei Privatkunden 2023 zum Standard. 500.000 PV-Anlagen wurden im Jahr 2023 von Privatpersonen angemeldet, dazu kommen 366.732 Speicher. Etwa 75 Prozent der Kunden entscheiden sich demnach dafür, den Eigenverbrauch durch Zwischenspeichern zu erhöhen. Im Jahr 2020 lag der Anteil noch bei 48 Prozent.

Die Begeisterung für Batterien wächst nicht nur in Privathaushalten: Auf der einen Seite betreiben große Investoren Batteriegroßspeicher, mit denen sie am Strommarkt handeln [1], auf der anderen Seite bauen auch Industrieunternehmen Speicher, um Lastspitzen damit abzufangen oder eigenen PV-Strom in der Nacht zu nutzen. In Summe kratzen alle deutschen Batteriespeicher zusammen im Oktober an der Marke von 10 Gigawattstunden (GWh) Kapazität, wobei die Heimspeicher mit 8,1 GWh den größten Anteil ausmachen. Bemerkenswert: Vor drei Jahren, im Oktober 2020, standen erst mickrige 2 GWh Speicher in Deutschland. Auch international zeigt sich das gleiche Bild: Die Heim- und Industriespeicherbranche steht am Anfang einer Entwicklung, die voraussichtlich exponentiell ist.

Balkonspielzeuge?

Unter den knapp 3,5 Millionen Photovoltaikanlagen, die Anfang Oktober im Marktstammdatenregister eingetragen und in Betrieb waren, hatten 332.146 Anlagen eine Nettonennleistung kleiner als 1 Kilowatt – es handelt sich bei diesen Anlagen zum kleinen Teil um Erweiterungen von größeren Anlagen, zum überwiegenden Teil um Balkonkraftwerke. Das Marktstammdatenregister kennt für diese zu-

75 Prozent der privaten PV-Kunden haben sich im Jahr 2023 für einen Batteriespeicher entschieden.

China führten dazu, dass im Laufe des Jahres 2023 auch die Preise für PV-Module ins Rutschen kamen. Glas-Glas-Module aus der 400-Watt-Klasse kosteten Anfang des Jahres oft noch 180 Euro, über die Monate fielen die Preise teilweise auf rund 100 Euro. In einer 10-kW-Anlage macht sich das mit einer Ersparnis von 2000 Euro bemerkbar.

Gefühl der Sicherheit

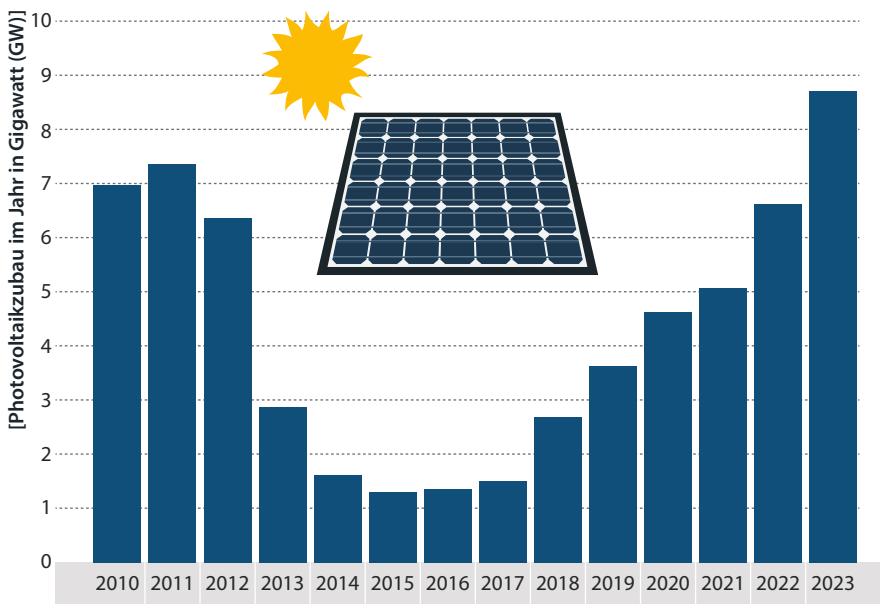
Die finanzielle Ersparnis ist für die meisten das wichtigste Argument für eine PV-Anlage auf dem Dach. Das gute Gefühl, damit Strombezug aus fossilen Energieträgern vermieden zu haben, ein weiteres. Beide Gründe gab es auch schon in der ersten PV-Welle, die im Jahr 2012 endete. Vergleichsweise neu ist die Motivation, die

über ein Relais innerhalb von Millisekunden abschaltet, wenn das öffentliche Netz mal ausfällt. Dass sich daran seit Beginn des Ukrainekriegs etwas geändert hat, berichten Installationsfirmen allerorten und die Daten im Marktstammdatenregister unterstützen diese subjektive Wahrnehmung: Angemeldet wird die Notstromfunktion zusammen mit dem Batteriespeicher. Vor dem 1. Januar 2021 gab es in Deutschland nur 23.080 Anlagen mit Notstromfunktion, die zusammen 0,25 Gigawattstunden speichern konnten. In der Zeit danach bis Oktober 2023 wurden gleich 91.181 neue Notstromanlagen mit einer Kapazität von zusammen rund 1 Gigawattstunde installiert.

Wer sich für diese Funktion interessiert, muss das im Vorfeld mit seinem

PV-Ausbau in Deutschland

2012 war es mit der ersten PV-Welle schlagartig vorbei, es folgte ein längeres Tal. Bis Anfang Oktober 2023 wurden bereits 8,7 Gigawatt zugebaut – ein Rekord.



sätzlich den Filter „Lage der Einheit entspricht steckerfertige Solaranlage (sog. Balkonkraftwerk)“. Doch schon ein grober Blick auf die Daten und die Namen der Anlagen zeigt, dass längst nicht jeder Betreiber dieses Attribut gefunden und richtig ausgefüllt hat. Die ganze Wahrheit über Balkonkraftwerke kennt das Marktstammdatenregister ohnehin nicht, denn noch immer ist es unter Betreibern üblich, die Anlagen nicht zu melden, um sich das Formular des Netzbetreibers zu sparen oder schlicht, um den alten Ferraris-Stromzähler rückwärts drehen zu lassen.

Doch selbst die Anzahl der offiziell im MaStR eingetragenen Anlagen und die Summe der Erzeugungsleistung sind bemerkenswert. Balkonkraftwerke, in vielen Fällen auch Garagen-, Schuppen-, Dach- oder Gartenkraftwerke, sind längst keine Spielerei mehr. Auf 192 Megawatt Nettoleistung kommen die angemeldeten Erzeuger. Ebenfalls bemerkenswert: 134 Megawatt (230.000 Anlagen) wurden allein im Jahr 2023 in Betrieb genommen.

Staatliche Baustellen

In Summe sind 2023 in Deutschland Photovoltaikanlagen mit zusammen 8,7 Gigawatt ans Netz gegangen. Das liegt noch unter dem angepeilten Ziel von 10 Gigawatt, die bis 2030 jedes Jahr gebaut werden sollen, 2023 ist aber dennoch bereits ein Rekordjahr – und die Ziele erscheinen wesentlich realistischer als bei der Windkraft.

Einen Anteil hat auch die Bundesregierung geleistet, die in Form von Solar- und Osterpaketem eine Reihe kleiner und größerer Vereinfachungen umgesetzt hat – weitere sind in Planung. Da ist seit 1. Januar 2023 die Mehrwertsteuersenkung für privat installierte Anlagen unter 30 Kilowatt auf 0 Prozent, die Balkonkraftwerksbetreibern bares Geld spart und anderen Betreibern sowie dem Finanzamt die leidige Umsatzsteuervoranmeldung. Abgeschafft ist auch die Einkommensteuerpflicht für die Erträge aus verkauftem Strom – die Einnahmenüberschussrechnung können sich Betreiber kleiner Anlagen ebenfalls sparen.

Bereits durchs Kabinett und auf dem Weg durchs Parlament ist das Solarpaket, das unter anderem das Anmelden von Balkonkraftwerken erleichtert: Geht alles glatt, dürfen Netzbetreiber die Betreiber ab 1. Januar 2024 nicht mehr mit einem selbst ausgedachten Formular zur Anmeldung schikanieren – eine Anmeldung im Marktstammdatenregister soll reichen, die Grenze wird von 600 auf 800 Watt erhöht. Außerdem soll man den Wechselrichter direkt nach der Anmeldung einstöpseln dürfen; sofern dann ein alter Ferraris-Zähler rückwärts dreht, ist das ein Problem des Netzbetreibers, der dadurch einen Anreiz hat, sich mit dem Austausch durch einen digitalen zu beilegen. Letztlich passt der Gesetzgeber die Regeln damit an die Realität an.

Diese Änderungen bei Balkonkraftwerken könnten nach einer Testphase auch für größere Anlagen übernommen werden, wenn es nach dem Elektrohandwerk geht. Die Elektriker müssen aktuell bei jedem Netzbetreiber ein anderes Formular mit leicht anderen Feldern ausfüllen und immer wieder die gleichen Datenblätter von immer den gleichen Komponenten anhängen – Arbeiten, die schnell mehr Zeit in Anspruch nehmen als das Aufhängen und Anschließen eines Wechselrichters. Eine zentrale Anmeldung im Marktstammdatenregister könnte diese Kleinststaaterei der Netzbetreiber beenden und die Kosten für die Inbetriebnahme senken.

Ebenfalls angegangen ist der Gesetzgeber das Problem der ungenutzten Fläche auf Mehrfamilienhäusern. Das bisherige Mieterstrommodell, bei dem der Anlagenbetreiber (meist der Vermieter) seinen Mietern den gesamten Strom verkaufen und das, was die Anlage nicht erzeugen konnte, zukaufen musste, hatte sich als zu kompliziert und unattraktiv erwiesen und viele geeignete Flächen blieben leer. Lediglich 7670 solcher Anlagen gibt es aktuell, die zusammen auf 131 MW kommen. Das neue Modell heißt „Gemeinschaftliche Gebäudeversorgung“. Dabei darf jeder Mieter bei seinem Stromanbieter bleiben und kann gleichzeitig Strom von einer Anlage auf dem Dach ohne Netzentgelte kaufen. Verrechnet wird das über Zähler, die im Viertelstundenraster zählen können. Wie das System im Detail und die Abrechnung technisch funktioniert, lesen Sie in einer der nächsten Ausgaben der c't.

Fazit

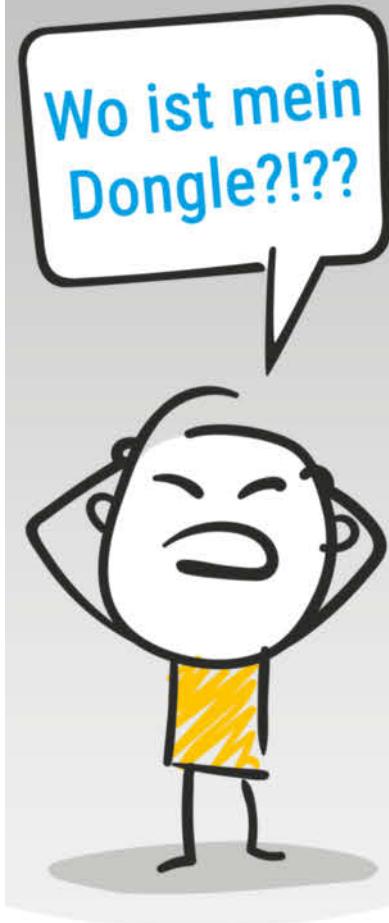
2023 war ein Rekordjahr für die Photovoltaik in Deutschland – vor allem durch eine neue Begeisterung der Privathaushalte und der Industrie. So rosig wie bei der Energie aus Sonnenlicht sieht es bei der Windkraft längst nicht aus. Die staatlichen Eingriffe konnten 2023 ihre Wirkungen noch nicht entfalten und die Ausbauziele liegen in weiter Ferne. Im Aufwind ist dagegen die Speicherbranche – Batteriespeicher, die langfristig auch dabei helfen, die Stromkosten zu senken, entstehen aktuell in allen Größenklassen von wenigen Kilowattstunden bis zu zig Megawatt.

(jam@ct.de) **ct**

Literatur

- [1] Jan Mahn, Bis zu 10.200 Euro für Photovoltaik von der KfW, c't 22/2023, S. 45
- [2] Jan Mahn, Stromlagerhaus, Wie Batteriespeicher das Stromnetz stabilisieren, c't 7/2023, S. 130

Strategiepapiere, Gesetzesänderungen, Datenquellen: ct.de/yctb



USB-Server hilft:

USB-Dongle & -Geräte an einem sicheren Ort platzieren & dank USB-Server bequem über das Netzwerk darauf zugreifen.



Nicht ärgern,
kostenlos testen:
www.wut.de/usb



W&T
www.WuT.de



Bild: KI Midjourney | Bearbeitung c't

Speichersysteme im Wandel

Pumpspeicherkraftwerke und ihre modernen Verwandten als Bausteine für die Energiewende

Wenn im Dunkeln kein Wind weht, gibt es keinen Strom aus erneuerbaren Energien und die Lichter in den Haushalten erlöschen – so die Befürchtung vieler Gegner der Energiewende. Doch eine solche Dunkelflaute ist gar kein Problem, wenn irgendwo ausreichend Strom gespeichert wurde. Ein Überblick über Möglichkeiten und Visionen.

Von Lutz Labs

W^{er} nördlich von Hamburg auf der A7 fährt, sieht eine ganze Reihe Windkrafträder neben der Autobahn stehen. Manchmal stehen sie still, obwohl der Wind durchaus weht. Warum? Es klingt absurd, aber einer der Gründe ist, dass der Betreiber für jede erzeugte Kilowattstunde bezahlen müsste, wenn der Strompreis an der Börse negativ ist. Auch könnte es in diesem Moment einfach keinen Abnehmer für den Strom geben.

Überschüssige Energie könnte man aber speichern und bei Bedarf nutzen, etwa bei einer Flaute oder bei Dunkelheit, wenn auch Solarzellen keine Energie mehr liefern. Passende Energiespeicher gibt es in vielen unterschiedlichen Größen, von der kleinen Knopfzelle bis zum riesigen Pumpspeicherkraftwerk. Letztere sind seit

Jahrzehnten die dominante Technik, um elektrische Energie in großem Maßstab zu speichern: Noch im vergangenen Jahr steckte 99 Prozent aller weltweit eingelagerten elektrischen Energie in Pumpspeicherkraftwerken.

Das Prinzip dieser Kraftwerke ist einfach: Überschüssiger Strom treibt eine Pumpe an, die Wasser von einem Becken in ein darüber liegendes pumpt; zur Stromerzeugung nutzt man die Schwerkraft, um das Wasser durch die gleichen Rohre wieder hinabfließen zu lassen – dabei arbeiten die Pumpen als Generatoren. Durch die Pumpen unterscheidet sich ein solches Kraftwerk von einem einfachen Laufwasserkraftwerk, welches Energie aus dem Gefälle eines Flussverlaufs oder einem Stausee erzeugt.

Die weltweit insgesamt installierte Leistung aller Pumpspeicherkraftwerke liegt bei etwa 130 Gigawatt (GW), in Deutschland waren es 2021 9,4 Gigawatt in 26 Kraftwerken – ein größerer Zubau ist hierzulande aufgrund der geografischen Gegebenheiten nicht zu erwarten (alle Quellen über ct.de/ybr6). Zwar baut EnBW das Speicherkraftwerk Forbach derzeit zu einem Pumpspeicherkraftwerk um, doch ist das eben nur ein mit veranschlagten Kosten von 280 Millionen Euro recht teurer Umbau: Das Kraftwerk als solches erzeugt bereits seit mehr als 100 Jahren zuverlässig Energie aus Wasserkraft, nun kommen Pumpen hinzu. Denkbar ist zudem der Umbau bereits vorhandener alter Steinkohlegruben oder Tagebauwerke, denn bei diesen sind bereits brauchbare Höhenunterschiede von mindestens 100 Metern vorhanden.

Das leistungsstärkste Pumpspeicherkraftwerk der Welt steht in Fengning in der chinesischen Provinz Hebei. Es hat eine Spitzenleistung von 3,6 GW und wurde erst Anfang 2022 in Betrieb genommen. Zum zweitstärksten Pumpspeicherkraftwerk im Norden des US-Bundesstaates Virginia liegen mehr Informationen vor: Die Bath County Pumped Storage Station nutzt Speicherbecken mit einem Höhenunterschied von 384 Metern und sechs Generatoren mit je 510 Megawatt Leistung. Bei Vollast fließen 852 Kubikmeter Wasser pro Sekunde durch die Generatoren. Deutschlands größtes Pumpspeicherkraftwerk Goldisthal im Thüringer Schiefergebirge ist mit einer Leistung von 1,06 Gigawatt gleichzeitig das größte in Europa.

Die Energiemenge aller deutschen Pumpspeicherkraftwerke liegt bei 6,7 GWh, sie wird durch die im Oberbecken gespeicherte Wassermenge und den Höhenunterschied zum Generator bestimmt. Die Kosten für die Speicherung einer Kilowattstunde Strom in einem Pumpspeicherkraftwerk liegen in Deutschland zwischen 3 und 5 Cent pro Tag. Zur Versorgungssicherheit in Deutschland leisten sie aber einen wichtigen Beitrag, denn sie sind schwarzstartfähig.

Weitere wichtige Kennziffern eines Energiespeichers sind die speicherbare Energiemenge und sein Gesamtwirkungsgrad. Dieser beträgt bei Pumpspeicherkraftwerken zwischen 75 und 80 Prozent. Verluste entstehen etwa durch Reibung und durch den Wirkungsgrad der Pumpen beziehungsweise Generatoren; dazu gesellen sich wie bei jedem anderen Energie-

speicher auch Übertragungsverluste durch die Umwandlung der lokal erzeugten Spannung in die vom Netz benötigte.

Schwarzstarter

Die Gasmangellage im vergangenen Winter hatte zu Furcht vor einem großflächigen Netzausfall geführt. Für das Wiederanlaufen des gesamten Netzes existieren Notfallpläne und vor allem sogenannte schwarzstartfähige Kraftwerke. Davon gibt es in Deutschland 174, eins der bekannten ist das Pumpspeicherkraftwerk Schluchsee – eine öffentliche Liste existiert jedoch nicht. Beim überwiegenden Teil der am Übertragungsnetz angeschlossenen und vertraglich zum Netzwiederaufbau gebundenen schwarzstartfähigen Kraftwerke handelt es sich um Wasserkraftwerke. In Ausnahmefällen kommen auch Erdgas oder Öl als Energieträger zum Einsatz.

Schwarzstartfähige Kraftwerke müssen unabhängig von externen Stromquellen starten können, auch der Nebenbedarf von Kontrollzentrale, Telefon- und Kommunikationssystemen et cetera muss vom Kraftwerk selbst erzeugt werden.

Zahlenspiele

Der Gesamtenergieverbrauch hat sich in den letzten 30 Jahren kaum verändert. In der Bundesrepublik werden jährlich rund 2500 Terawattstunden Energie benötigt. Ganz grob verbrauchen Industrie, Haushalte und Verkehr je etwa 700 Terawattstunden, dazu gesellen sich noch knapp 400 Terawattstunden für Gewerbe, Handel und Dienstleistungen.

Dabei liegt der Stromverbrauch im Sommer deutlich niedriger als im Winter. Bei steigender Verbreitung von Wärmepumpen dürfte sich dieses Ungleichgewicht noch verstärken.

ct kompakt

- Sonnen- und Windenergie sind nur unregelmäßig verfügbar.
- Speicherkraftwerke überbrücken Dunkelflauten.
- Neue Techniken zur Energiespeicherung sind in Entwicklung.

Bis zu ihrer Abschaltung liefen Atomkraftwerke durch, 24 Stunden am Tag. Mit Wasser- und Gaskraftwerken können die Betreiber den Strom bei Bedarf erzeugen. Während aber im Sommer die Sonnenenergie mehr zum Ertrag erneuerbarer Energien beiträgt, weht im Herbst und Winter der Wind stärker. Im Prinzip ergänzen Wind- und Sonnenenergie einander ideal, doch durch die ungleiche geografische Verteilung dieser beiden Energieformen und den ungleichmäßigen Bedarf kommen wir in Zukunft nicht ohne zusätzliche Energiespeicher aus.

Energie lässt sich nicht nur in Form von hochgepumptem Wasser speichern, sondern unter anderem auch, indem man überschüssigen Strom dazu nutzt, Methan oder Wasserstoff zu erzeugen, den man in Tanks speichert und später in Gaskraftwerken wieder verheizt.

Das Fraunhofer-Institut für Solare Energiesysteme hat in einer Studie über Wege zu einem klimaneutralen Energiesystem im Jahr 2045 zwei Szenarien untersucht. Sollte der Leistungsanteil der PV-Systeme dann doppelt so hoch sein wie der der Windkraft, sind für eine weitgehende Klimaneutralität insgesamt Speicherkapazitäten von 377 GWh notwendig; abzüglich der vorhandenen Pumpspeicherkraftwerke also rund 370 GWh zusätzlich. Ein Teil

Wind- und Sonnenenergie 2022

Windenergie trägt zum deutschen Strommix insgesamt wesentlich mehr bei als die Solarenergie. Diese aber hilft vor allem im windarmen Sommer, den Anteil erneuerbarer Energien einigermaßen konstant zu halten.

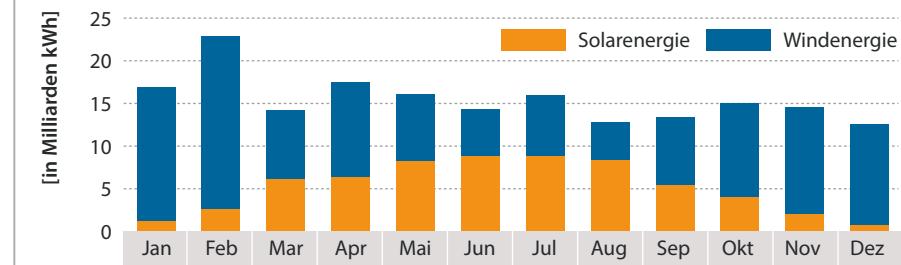




Bild: Tesla

Der Batteriespeicher von Tesla im belgischen Deux-Acren besteht aus 40 sogenannten Megapacks und erreicht damit eine Kapazität von 100 MWh.

der erzeugten Energie muss jedoch zusätzlich per Elektrolyse (Power-to-X) in Gas verwandelt werden, dazu braucht es Kraftwerke mit 40 GW Elektrolyseleistung sowie Gaskraftwerke mit einer Leistung von 130 GW.

Im zweiten berechneten Szenario beträgt der Leistungsanteil der PV-Anlagen das Siebenfache der Windkraft – und damit steigt der Speicherbedarf stark an: Stromspeicher für 900 GWh sind vonnöten, die Elektrolyseleistung steigt auf 160 GW und die Gaskraftwerke müssen 150 GW liefern. Dieses Szenario ist durch den starken Solarzellenausbau vor allem der privaten Verbraucher realistischer (siehe S. 10). Zudem geht die Studie von einem sinkenden Importanteil erneuerbarer Energie aus.

Gasspeicherkapazitäten spielen in diesen Szenarien keine Rolle, denn davon besitzt Deutschland genug. Was fehlt, sind die Kraftwerke für die Elektrolyse.

Guter und schlechter Wasserstoff

Beim Wasserstoff ist häufig die Rede von verschiedenen Farben, obwohl das Gas als solches natürlich farblos ist. Der Klassiker ist grauer oder schwarzer Wasserstoff, der durch Dampfreformierung von Erdgas, Kohle oder Öl entsteht, einem Prozess, bei dem Kohlenwasserstoffe unter Druck und hohen Temperaturen zu Methan, Kohlenmonoxid und Kohlendioxid umgewandelt und abschließend zu Wasserstoff katalysiert zu werden. Weil dabei CO₂ in die Atmosphäre entweicht, ist er nicht klimaneutral. Bei der Erzeugung blauen Wasserstoffs aus fossilen Brennstoffen speichert man das entstehende CO₂ zu etwa 95 Prozent unterirdisch mittels Kohlenstoffab-

scheidung und -speicherung (CCS-Technik, Carbon Capture and Storage).

Das bei der Methanpyrolyse, dem Prozess zur Erzeugung türkisen Wasserstoffs, entstehende Kohlendioxid kann ebenfalls abgespalten und unterirdisch gespeichert werden. Wenn die bei dem Prozess benötigte Energie aus erneuerbaren Energien stammt, spricht man ebenfalls von einer klimaneutralen Erzeugung. Da die Spätfolgen der Kohlenstoffspeicherung noch nicht absehbar sind, sind die klimaneutralen Einstufungen von blauem und türkisem Wasserstoff umstritten.

Grüner Wasserstoff entsteht durch Elektrolyse, der Aufspaltung von Wasser in seine Komponenten Sauerstoff und Wasserstoff durch Strom aus erneuerbaren Energien. Seine Herstellung ist klimaneutral, aber mit einem hohen Energieaufwand verbunden. Nutzt man für die Elektrolyse Strom aus Kernenergie, ist die Erzeugung zwar klimaneutral, hinterlässt aber radioaktiven Abfall. Der dabei entstehende Wasserstoff wird manchmal als pink oder gelb bezeichnet.

Blauer Wasserstoff lässt sich in den gleichen Anlagen produzieren wie grauer Wasserstoff, lediglich die CO₂-Abscheider kommen hinzu. Da die Abscheider ebenfalls Energie benötigen, sinkt der Wirkungsgrad des Prozesses. Zudem setzt der Produktionsprozess auch das besonders klimaschädliche Methan frei, wenn auch über die Mengen aufgrund des Pilotbetriebs der meisten Anlagen noch Uneinigkeit herrscht.

Der Wasserstoffbedarf vor allem der Stahl- und Chemieindustrien wird sich aller Voraussicht nach in naher Zukunft nicht mit CO₂-freiem grünem Wasserstoff decken lassen, auch wird grüner Wasserstoff als Brennstoff für Flugzeuge und Schiffe benötigt, da deren Betrieb mit Batterien in absehbarer Zeit kaum möglich erscheint.

In Deutschland gibt es aktuell sieben Anlagen zur Erzeugung von grünem Wasserstoff (sogenannte Elektrolyseure), zwei Anlagen produzieren grünes Methan; insgesamt erzeugen diese 27,8 MWh Gas pro Stunde. Der Wirkungsgrad der Elektrolyseure liegt derzeit bei maximal 70 Prozent, zudem kostet die Produktion grünen Wasserstoffs das Mehrfache von grauem Wasserstoff. Rechnet man den Wirkungsgrad eines Gaskraftwerks zur Wiederverstromung des Wasserstoffs hinzu, ergibt sich ein Gesamtwirkungsgrad von etwa 40 Prozent – deutlich weniger als bei einem Pumpspeicherkraftwerk.

Gasspeicher

Die vorhandenen Gasleitungen eignen sich Studien zufolge auch für Wasserstoff; bei den bestehenden Gasspeichern sind vor allem die großen Kavernenspeicher „besonders gut geeignet“. Unter Kavernen versteht man unterirdische Hohlräume oder Höhlen, die oft natürlich in Gesteins-

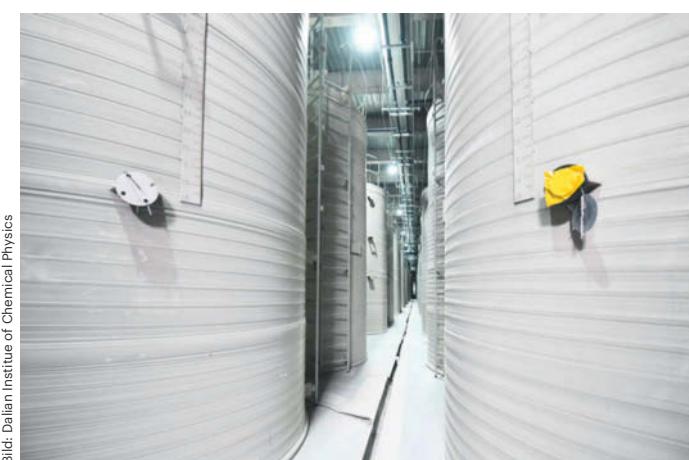


Bild: Dalian Institute of Chemical Physics

In der bislang größten Redox-Flow-Batterie im chinesischen Dalian lagert die Energie in großen Elektrolyt-Tanks. Die Kapazität der Anlage soll auf 800 MWh ausgebaut werden.

schichten oder Salzformationen vorkommen oder bergmännisch angelegt wurden. Es handelt sich dabei um Hohlräume mit einer Höhe von bis zu 500 Metern.

Weiterhin gibt es sogenannte Porespeicher, bei denen das Gas in poröses Gestein gepresst wird. Sie entstehen meistens durch ausgeförderte Erdgas- oder Erdöl-Reservoir. Porespeicher sind nach oben häufig durch Deckschichten aus Ton und Salz abgedichtet, unterhalb der gasdurchlässigen Gesteinsschichten begrenzt ein wasserführender Bereich das Reservoir.

Aktuell gibt es 16 Porespeicher in Deutschland, davon sind gemäß der Studie jedoch nur 4 für die Speicherung von Wasserstoff geeignet. Porespeicher sind im gesamten Bundesgebiet vorhanden, während Kavernenspeicher vor allem im norddeutschen Raum vorkommen.

Die Initiative Energien Speichern, ein Lobby- und Interessenverband von Betreibern deutscher Untergrund-Erdgasspeicher, hat im Mai eine Karte der Gasspeicher veröffentlicht. Dieser zufolge gibt es aktuell zwölf Betreiber von Wasserstoffspeichern in Deutschland, die insgesamt 22 Standorte befinden sich überwiegend im Norden, drei der vier Porespeicher jedoch im Süden Bayerns an der Grenze zu Österreich. Ob diese 22 Wasserstoffspeicher aktuell bereits in Betrieb sind, geht aus der Karte jedoch ebenso wenig hervor wie die Art der Erzeugung des dort eventuell gespeicherten Wasserstoffs.

Kapazität vorhandener Batteriespeicher

Viele Privathaushalte ergänzen ihre Solaranlage um einen Batteriespeicher, um den Autarkiegrad zu vergrößern. Das hat dazu geführt, dass die sogenannten Heimspeicher, deren Kapazität meist im Bereich einstelliger Kilowattstunden liegt, das Gros der Batterieleistung in Deutschland bereitstellen: Bei 8 GWh lag deren installierte Leistung im September 2023. Genau wie die Gewerbespeicher, deren Kapazität rund 380 MWh beträgt und die ähnliche Aufgaben bei Handwerks- und Industriebetrieben erledigen, tragen sie nur wenig zur Stabilisierung des Gesamtnetzes bei (mehr zu Batteriespeichern im Artikel auf S. 10).

Für die Netzregelung stehen vor allem die rund 1,3 GWh Kapazität der Großspeicher zur Verfügung, welche die Netzbetreiber zu genau diesem Zweck aufgebaut haben. Deren Kapazität hat sich in den



Bild: Institute Of Engineering Thermophysics (iet.cas.cn)

Der 100-Megawatt-Druckluftspeicher im chinesischen Zhangjiakou ist ein sogenannter adiabatischer Speicher, der die bei der Kompression entstehende Wärme nicht ungenutzt entweichen lässt.

vergangenen 15 Jahren verdreifacht, während die Kapazität der Heimspeicher im gleichen Zeitraum um mehr als das 15-Fache gestiegen ist.

Einer der größten Hersteller von Großspeichern ist das Kirchheimer Unternehmen Eco Stor, ein Tochterunternehmen des norwegischen Stromkonzerns A Energie. Ende 2025 will das Unternehmen in Sachsen-Anhalt einen Batteriespeicher mit einer Kapazität von 600 MWh in Betrieb nehmen, kurz darauf folgt ein 207-MWh-Speicher in Schleswig-Holstein. Der aktuell größte Speicher der Bundesrepublik fasst 72 MWh und steht im nordrhein-westfälischen Werne.

Alle aktuellen Großspeicher nutzen Lithium-Ionen-Akkus und viele stehen in der Nähe von großen Solarparks, um direkt vor Ort Spannungsschwankungen des Netzes auszugleichen. Neue Anlagen sind über das Marktstammdatenregister der Bundesnetzagentur oder die Battery-Charts der RWTH einsehbar, letzterer filtert bereits alle anderen Energiequellen heraus.

Der Autobauer Tesla ist mit seinen Megapack-Modulen ebenfalls im Hochkapazitätsgeschäft tätig. Ein Verbund von 40 solcher Akkupacks mit einer Kapazität von 100 MWh steht etwa in Belgien. Laut dem Installateur Corsica Sole betragen die Kosten für den Aufbau 33 Millionen Euro, die maximale Ausgangsleistung beträgt 50 MW.

Lithium-Ionen-Akkus sind zwar sehr weit verbreitet, es gibt jedoch mittlerweile fast ebenso leistungsfähige Alternativen. Der chinesische Akkuhersteller HiNa Battery etwa baut seit Herbst letzten Jahres Natrium-Ionen-Akkus, CATL und der Autobauer BYD wollen noch dieses Jahr damit starten – auch die letzteren Unternehmen sind in China beheimatet.

Neue Speichertechniken

Akkus verlieren mit der Zeit ihre Ladung. Damit sind sie nur eingeschränkt für die Speicherung elektrischer Energie über einen Zeitraum von mehr als ein paar Tagen geeignet. Speicherkraftwerke müssen zuverlässig und langlebig sein und dabei möglichst viel der eingespeicherten Energie wieder ausgeben – und das bei möglichst geringen Herstellungskosten. Im Folgenden stellen wir einige Möglichkeiten vor.

Unterwasser-Pumpspeicher

Unterwasser-Pumpspeicherkraftwerke bestehen nach einer Idee der Physiker Prof. Dr. Horst Schmidt-Böcking und Dr. Gerhard Luther aus etwa 30 Meter großen hohlen Kugeln, die in mindestens 700 Meter Tiefe auf dem Meeresgrund liegen. Bei Energieüberschuss pumpt man das Wasser aus der Kugel heraus, bei Energiebedarf lässt man Wasser hinein. Der dabei entstehende hohe Wasserdruk treibt einen Generator an, der Strom erzeugt. Der Wirkungsgrad soll bei 80 bis 85 Prozent liegen, ein Pilotprojekt hat die prinzipielle Funktion bereits 2016 gezeigt.

Das StEnSEA-Projekt (Stored Energy in the Sea, bisweilen auch als „Meer-Ei“ bezeichnet) wird von den beiden Physikern noch weiterentwickelt; 2020 etwa kam die Idee auf, den bis zu 400 Meter tiefen Hambacher Tagebau nach seiner Schließung als Grundlage zu verwenden. Die noch einmal vergrößerten Pumpkraftwerke sollten sieben Prozent des künftigen täglichen elektrischen Energiebedarfs Deutschlands zwischenspeichern können – seitdem ist es um das Projekt jedoch still geworden.

Redox-Flow-Batterien

Eine andere Art von Akkus sind Redox-Flow-Batterien. Die Besonderheit ist, dass



Bild: Energy Vault

In der ersten Version seines Schwerkraftspeichers nutzte Energy Vault noch Betonblöcke und stapelte sie übereinander. Im ersten kommerziellen Projekt bleibt nur das Prinzip gleich.

der Elektrolyt nicht nur in der Batteriezelle gelagert ist, sondern in externen Tanks. Pumpen tauschen ihn ständig aus, Energiespeicher und Energieumwandlung sind damit getrennt. Für einen größeren Speicher muss man lediglich das Tankvolumen vergrößern und mehr Elektrolyt einfüllen; damit ist das System sehr gut skalierbar. Zudem erwartet man eine hohe Lebensdauer, die Rede ist von bis zu 13.000 Ladezyklen. Der Wirkungsgrad einer Redox-Flow-Batterie mit in einem Lösungsmittel gelöstem Vanadium(V)-oxid als Elektrolyt liegt mit 70 bis 75 Prozent auf dem Niveau eines Lithium-Ionen-Akkus.

Als Mobilspeicher eignen sich Redox-Flow-Batterien weniger, weil sie aufgrund der geringeren Energiedichte größer und schwerer sind als Lithium-Ionen-Akkus. Prototypen stehen unter anderem im Fraunhofer-Institut für Chemische Technologie, im kommerziellen Einsatz ist ein 400-MWh-Speicher bereits im chinesischen Ort Dalian. Eine Zukunft haben Redox-Flow-Batterien jedoch wohl nur, wenn sich ein Ersatz für das teure und giftige Vanadium findet – daran forschen verschiedene Unternehmen.

Latentwärmespeicher

Auch das Deutsche Zentrum für Luft- und Raumfahrt (DLR) beschäftigt sich mit Energiespeichern: Es hat zusammen mit Partnern einen Latentwärmespeicher mit zwei Kubikmetern Nitratsalz aufgebaut. In der sogenannten Carnot-Batterie erhitzt

eine Hochtemperatur-Wärmepumpe das Salz, bis es bei etwa 150 °C schmilzt.

Die typische Speicherzeit liegt bei einigen Stunden bis Tagen. Zum Entladen des Speichers überträgt ein zweiter Kreislauf die Wärme zu einer Wärmekraftmaschine, die eine Turbine mit Generator antreibt. Zusätzlich zum damit erzeugten Strom kann die gespeicherte Wärme in Fernwärmennetze eingespeist oder für Industrieprozesse genutzt werden. Wenn die Wärmepumpe Unterstützung etwa durch eine Geothermie-Anlage erhält, ist ein Wirkungsgrad von bis zu 70 Prozent möglich. Angaben zur Skalierbarkeit liegen noch nicht vor, die DLR-Forscher erwarten in etwa zehn Jahren marktreife Systeme.

Druckluftspeicher

Ein Druckluftspeicher funktioniert auf den ersten Blick so einfach wie ein Pumpspeicher: Kompressoren drücken Luft in einen druckstabilen Behälter, bei Energiebedarf wird Luft daraus über einen Generator abgelassen.

Größtes Problem dieser Speicher ist das Wärmemanagement: Luft heizt sich unter Druck auf und kühlt beim Dekomprimieren wieder ab – das kann zum Einfrieren der Turbine führen. In früheren Druckluftspeichern erhitzte man die Luft beim Dekomprimieren durch das Verbrennen von Gas.

Die Betreiber des modernen Druckluftspeichers im nordchinesischen Zhangjiakou speichern die bei der Kompression

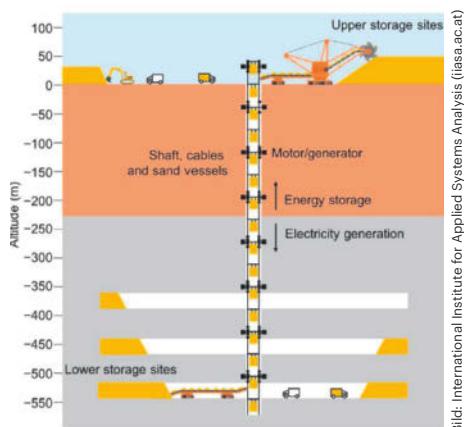
entstehende Wärme und erwärmen damit die Luft bei der Dekompression. Damit verhindern sie nicht nur das mögliche Einfrieren, es erhöht auch den Wirkungsgrad der Anlage: Nach Angaben der chinesischen Akademie der Wissenschaften liegt dieser mit 70,4 Prozent weit über dem herkömmlicher Druckluftspeicher, die maximal 54 Prozent erreichen. Das chinesische Speicherkraftwerk liefert maximal 100 MW und soll pro Jahr 132 GWh Strom speichern und wieder zur Verfügung stellen.

Gravitationsspeicher

Die Schwerkraft lässt sich nicht nur in Wasserkraftwerken, sondern noch in weiteren Energiespeichern einsetzen, beispielsweise in alten Bergwerksschächten. Heute braucht man diese Schächte kaum noch; zudem sind sie meistens bereits mit dem Stromnetz verbunden – das spart Anschlusskosten. Das Prinzip ist wieder einmal einfach: Man lässt Gewicht in Form von Sand über den vorhandenen Aufzug hinab, damit erzeugt ein Generator Strom. Hat man Energie übrig, zieht man das Gewicht wieder nach oben. Um die speicherbare Energiemenge zu erhöhen, könnte der Sand im Stollen und an der Oberfläche durch automatische Bagger oder Förderbänder be- und entladen werden – die Energie dafür erzeugt das System praktisch selbst.

Weltweit hätten solche Bergwerkspeicher eine Speicherkapazität von 70 TWh, der Wirkungsgrad solcher Anlagen könnte bei 70 bis 80 Prozent liegen. Bislang ist die Underground Gravity Energy Storage genannte Speichermethode nur eine Idee von Forschern des International Institute for Applied Systems Analysis (IIASA) in Österreich. Sie räumen allerdings bereits ein, dass die Speicherkosten beim Zwei- bis Vierfachen eines Wasserpumpkraftwerks liegen würden. Hinderlich ist zudem, dass viele ältere Gruben bereits mit Abraum befüllt wurden.

Eine andere Art von Gravitationsspeicher hat das Schweizer Start-up Energy Vault bereits 2018 ausprobiert, ein erster kommerzieller Einsatz folgt nun im chinesischen Rudon nördlich von Shanghai. Dort hat das Unternehmen ein Hochhaus ohne Wände errichtet, welches dem Rohbau eines Parkhauses ähnelt. Im Inneren sind einige Dutzend Kräne eingebaut, die große Gewichte nach oben ziehen beziehungsweise ablassen – während zum Hochziehen Energie benötigt wird, erzeu-



Gravitationsspeicher mal anders:
Speichert man Sand in der Nähe von alten Bergwerksschächten, kann man beim Herablassen Energie erzeugen.

gen die Generatoren beim Ablassen Strom.

Die Gewichte bestanden bei den Pilotversuchen des Unternehmens noch aus Beton; dessen Herstellung aber ist zum einen energieintensiv und zudem setzt der Prozess viel CO₂ frei. In China nutzt Energy Vault nun gepressten Erdaushub und andere Abfallstoffe. Der Wirkungsgrad der Anlage soll bei bis zu 80 Prozent liegen, doch es gibt Zweifel an der Wirtschaftlichkeit: Die Kosten für den Bau des Gebäudes sind hoch. Der Betreiber nennt eine maximale Leistung von 25 MW, die gespeicherte Energie soll 100 MWh betragen. Ende des Jahres soll das Speicherkraftwerk ans Netz gehen.

Ausblick

Der weitaus größte Teil der Speicherkraftwerke wird auf absehbare Zeit weiterhin mit Wasser betrieben werden. Der Aufbau alternativer Speicher kostet nicht nur Geld, sondern auch Zeit. Sicherlich wird nicht jede hier vorgestellte Technik auch im großen Umfang zum Einsatz kommen, aber das Versorgungsnetz der Zukunft dürfte mit verschiedenen Speichertechniken komplexer werden als das heutige.

Das größte Potenzial haben sicherlich Gasspeicher, auch wenn ihr Wirkungsgrad im Vergleich zu den anderen Speichern gering ist – der vorhandene Speicherplatz macht sie zu Favoriten und vor allem sind sie damit in der Lage, auch eine längere Dunkelflaute zu überbrücken.

(ll@ct.de) 

Links zu den verwendeten Quellen:
ct.de/ybr6

DER PARTNER FÜR SYSTEMHÄUSER!

EXPERTEN-TEAMS



- ✚ technische- und kaufmännische Beratung
- ✚ Planung von IT-Großprojekten
- ✚ Projektuntersuchung vor Ort



INDIVIDUALISIERUNG

- ✚ Co-Labeling / Branding von Gehäusen
- ✚ individuelles Bootlogo
- ✚ Masterimages RAID Konfiguration



SERVICE & SUPPORT

- ✚ Vor-Ort-Service
- ✚ Leih- und Teststellungen
- ✚ Pre-Sales Consulting kostenlos

JETZT PARTNER WERDEN!

Erfahren Sie mehr auf
extracomputer.de/darum-wir



Unsere Eigenmarken:

exzone

calmo

Pokini

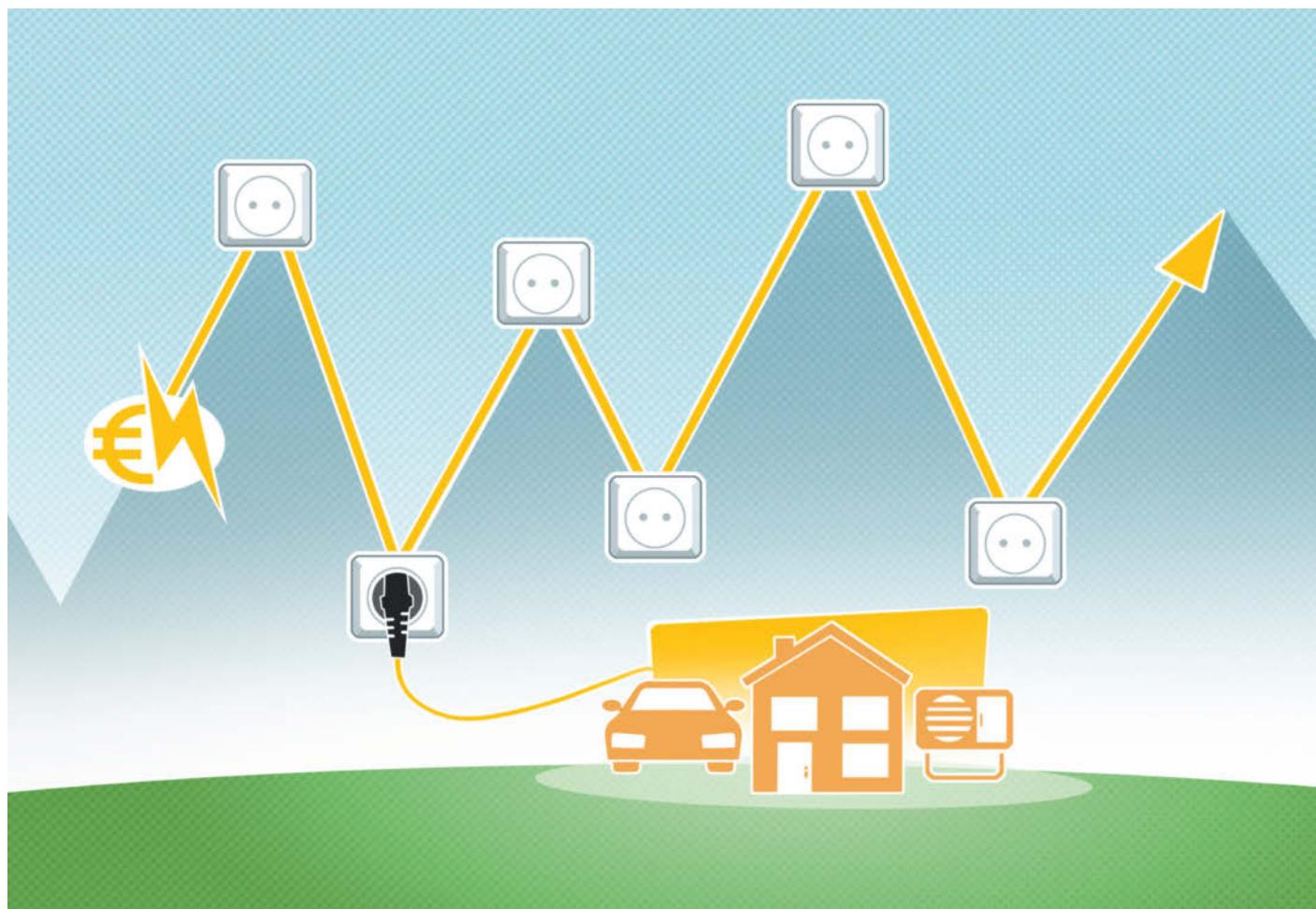


Bild: Albert Hüm

Sparen mit der Energiewende

Für wen sich dynamische Stromtarife lohnen

Dynamische Stromtarife können sich vor allem für Besitzer von E-Autos und Wärmepumpen richtig lohnen. Wir erklären die Funktionsweise, betrachten die Kosten und nennen Anbieter.

Von Clemens Gleich

Je höher die Stromproduktion bei viel Wind und Sonne ansteigt, umso niedriger fällt der Strompreis. Diese Preissignale nutzt die Industrie bereits. Schritt für

Schritt sollen nun auch die privaten Haushalte davon profitieren, die Gesetzgebung hierzu begann bereits im Jahr 2017.

Ab 2025 müssen die Netzbetreiber anfangen, größere Privathaushalte mit intelligenten, vernetzten Messeinrichtungen auszustatten. Diese Smart Meter und Smart-Meter-Gateways sind mit der Infrastruktur der Netzbetreiber verbunden. Die Pflicht beginnt ab 6000 Kilowattstunden jährlichem Verbrauch, betrifft also zunächst Kunden mit Elektroauto oder Wärmepumpe sowie (PV-)Stromerzeuger von 7 bis 100 Kilowatt maximaler Erzeugungsleistung. Kleinere Verbraucher können freiwillig Smart Meter beantragen.

Zur Umrüstpflicht für die Netzbetreiber kommt die Pflicht für die Energieversorger, dynamische Stromtarife anzubieten. Diese sollen auch den privaten Stromverbrauch tendenziell in Zeiten leiten, in denen die Erzeugungsleistung hoch und die Preise folglich niedrig liegen. Schon jetzt bieten viele Energieversorger solche Tarife an. Wir beleuchten Vorteile, Nachteile und Voraussetzungen.

Wie es funktioniert

Um einen dynamischen Stromtarif zu nutzen, brauchen Sie als Mindestausstattung eine Strommessung mit Zeitstempeln und Internetverbindung. Der Stromanbieter Tibber bietet seinen Kunden ein eigenes

Gerät namens „Pulse“ zum Kauf an, das jeden digitalen Stromzähler über die optische Schnittstelle ausliest und die Daten an Tibber zur Abrechnung schickt.

Die übliche Lösung sind allerdings Smart Meter mit Smart-Meter-Gateway. Die tun im Prinzip dasselbe, nur mit allen gesetzlichen und verbandlichen Vorgaben. Smart Meter melden den Momentanverbrauch beim Netzbetreiber, sodass dort a) stets aktuelle Stromnetzdaten ankommen und b) Sie den Strom theoretisch nach fünfzehnminütig genauen Tarifen bezahlen können. Üblich sind jedoch eher stündliche Tarife, weil die Anbieter den Strom am Vortag nach stündlichen Kontingenzen einkaufen (Day Ahead).

Wie viel der Strom wann kostet, erfahren Sie in der Regel über die Smartphone-App des Stromversorgers. Bei den üblichen Day-Ahead-Preisen wissen Sie also heute Abend, was der Strom Sie morgen stündlich kosten wird.

Beim Verbrauch gibt es zwei große Peaks, die den Strompreis hochtreiben: morgens, wenn die Leute sich zur Arbeit fertigmachen und abends, wenn sie heimkommen und kochen oder zocken. Dem Mittagsenergie-Peak beim Verbrauch steht vor allem im Sommer bei Sonnenschein ein großes Überangebot an Solarstrom gegenüber. Allein aufgrund von Angebot und Nachfrage ist Strom also morgens und abends teurer als mittags, siehe Grafiken. Dazu kommen günstige Preise in der Nacht ab circa 23 Uhr (jahreszeitabhängig), wenn

zwar wenig Strom erzeugt, aber eben noch weniger Strom nachgefragt wird.

Um die günstigen Strompreise nutzen zu können, müssen Sie nun zu den richtigen Zeiten Strom verbrauchen. Wer daheim arbeitet, kann leicht Spülmaschine, Waschmaschine und Trockner zu günstigen Zeiten betreiben. Wer auswärts arbeitet, lässt einen Energiemanager Steckdosen oder direkt verbraucherabhängig von den Strompreisen automatisch an- und ausschalten.

Dies führt aber nicht automatisch zu einer Ersparnis gegenüber einem Vertrag mit Festpreis: Wenn die Preise am Strommarkt explodieren (erinnern wir uns an die Energiepreiskrise ab 2021), dann explodiert auch Ihre Stromrechnung unmittelbar mit. Deshalb gibt es Tarife mit Preiskorridoren, bei denen Sie zwar weniger sparen, aber auch weniger verlieren können. Andere mitteln den Strompreis über einen Monat und setzen diesen Durchschnitt als Preis pro Kilowattstunde für Ihre Rechnung an. Solche Tarife berücksichtigen wir im Rahmen dieses Artikels nicht, weil sie nicht dynamisch sind und daher keine Möglichkeiten bieten, durch Verbrauchssteuerung die Stromrechnung direkt zu beeinflussen.

Attraktiv für Großverbraucher

Der größere Teil des Strompreises besteht aus Umlagen, Entgelten und Abgaben. Eine konsequente Zeitsteuerung von Trockner, Wasch- und Spülmaschine be-

c't kompakt

- Ab 2025 müssen die Stromnetzbetreiber größere Privatverbraucher mit smarten Messeinrichtungen ausstatten.
- Außerdem werden die Energieversorger verpflichtet, dynamische Stromtarife anzubieten.
- Von den schwankenden Preisen können vor allem Kunden mit E-Auto, Wärmepumpe oder Photovoltaikanlagen profitieren.

einflusst daher nur den Nettopreis der Stromlieferung, sodass die Ersparnis niedriger ausfällt, als man zunächst denkt. Damit sich ein dynamischer Tarif für Sie persönlich lohnt, muss Ihr jeweiliger Stromverbrauch möglichst hoch und möglichst flexibel sein. Am besten eignen sich daher Haushalte mit Großverbrauchern wie Elektroautos und Wärmepumpen.

Ein Elektroauto nutzt dynamische Stromtarife perfekt aus: Es braucht viel Strom, den es mit vergleichsweise hoher Leistung laden kann (zum Beispiel Wallbox mit bis zu 22 Kilowatt), benötigt diesen aber nicht zu festen Zeiten. Wenn Sie Ihr Auto nachts von 0 bis 4 Uhr laden, ist es morgens genauso voll, als hätten Sie es zum abendlichen Verbrauchs-Peak zwi-

Smart Meter für alle

Die gesetzliche Smart-Meter-Einführung beginnt schrittweise. Ab 2025 kommen verpflichtende Umrüstraten dazu. Flankierend müssen die Stromanbieter dynamische Tarife anbieten.

Freiwillige Einführung bis Ende 2024
für Verbraucher bis 100.000 kWh/Jahr einschl. § 14a EnWG und Erzeuger bis 25 kW installierter Leistung ab Inkrafttreten des Gesetzes (inkl. Möglichkeit für weitere Funktionen über Anwendungsupdates).

Optional:
Verbraucher unter 6.000 kWh/Jahr und Erzeuger ab 1 bis 7 kW installierter Leistung fallen unter den „optionalen Rollout“ - Smart Meter können auf Veranlassung des Messstellenbetreibers eingebaut werden, ab 2025 auch auf Wunsch des Kunden.

Pflicht zur Einführung
für Verbraucher ab 6.000 bis 100.000 kWh/Jahr einschl. § 14a EnWG und Erzeuger von 7 bis 100 kW installierter Leistung.

Pflicht zur Einführung
für Verbraucher über 100.000 kWh/Jahr und Erzeuger über 100 kW installierter Leistung.

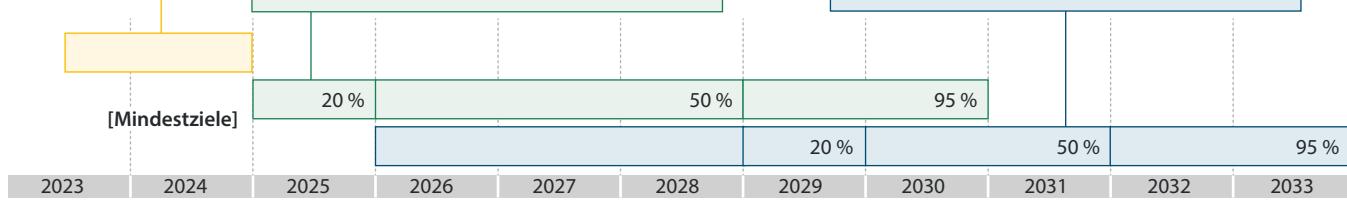




Bild: Tibber

Tibbers „Pulse“ liest über die optische Schnittstelle jeden digitalen Zähler aus.

schen 18 und 22 Uhr geladen. In einem fixen Tarif bezahlen Sie für beide Varianten den gleichen Betrag. In einem dynamischen Tarif gibt es meistens einen signifikanten Unterschied. Unabhängig von den Kosten wirkt sich die Entlastung der Spitzenverbrauchszeit auch positiv auf die Netzstabilität aus und das Auto trägt mit einer für Sie egal Verschiebung der Ladezeit zur Netzstabilität bei.

Etwas komplizierter ist die Lage bei der Wärmepumpe. Bei der Brauchwasserbereitung sind die Geräte noch relativ flexibel, wenn der Wärmepuffer groß genug ist. Ob die Legionellschutzfunktion um 12 Uhr oder um 15 Uhr das Wasser auf 70 °C heizt, ist für den Gesundheitsschutz egal. Dabei können Sie also gut auf günstigere Preise ausweichen. Beim Heizen der Räume dagegen bestimmen Leistung und Wärmespeicherkapazität der Anlage darüber, wie flexibel Sie agieren können.

Damit ein Unterschied auf der Rechnung herauskommt, muss die Leistung der Anlage so hoch sein, dass sie in billigen Zeiten genug Wärme auf Vorrat für die teuren Zeiten produzieren kann. Um diese Zeiten ohne Komforteinbußen zu überbrücken, muss das Gesamtsystem Haus-Heizung diese Wärme zudem auch tatsächlich speichern können (siehe c't 22/2023, S. 32).

Sie können sich hierzu beliebig komplexe Steuerungen vorstellen, in der Praxis fallen sie jedoch meistens eher einfach aus. Bei meinem Hersteller Nibe etwa stellt die Steuerung in Zeiten billigen Stroms einfach die Temperatur für das Heizwasser hoch und in teuren Zeiten herunter. Das bedeutet jedoch: Die Raum-

temperatur wird schwanken. Es ist nur die Frage, ob die Schwankungen auffallen oder gar stören.

Diese Einschränkungen haben einen Effekt: Als typische Einsparung für die automatische Strompreissteuerung gibt Nibe magere zehn Prozent an. Solange Ihr Wohlbefinden nicht leidet, lohnt sich das angesichts der hohen Energiemenge trotzdem. Sie müssen lediglich diese Funktion im Interface der Wärmepumpe aktivieren und eventuell erhöhte Verschleißkosten einkalkulieren.

Hausakkus und Geringverbraucher

Man könnte prinzipiell den Hausakku als Strompreispuffer verwenden: zu Minderzeiten laden, zu Peak-Zeiten den Haushalt mit Strom versorgen. Das bietet sich zum Beispiel im Winter an, wenn die eigene PV-Anlage kaum genug Strom für den Haushalt produziert, geschweige denn für die Speicherung im Akku. Dazu muss der Akku im Marktstammdatenregister als „sonstiger Stromspeicher“ registriert sein, weil Sie einen „EE-Stromspeicher“ nicht mit Netzstrom beladen dürfen (Begründung: Netzstrom speist sich nicht ausschließlich aus erneuerbaren Energien).

Es ist also möglich, aber selten lohnend.

Sie sollten unbedingt ausprobieren, wie viel Strom übrig bleibt, wenn er zweimal über den Umrichter und die Batterie fließt, weil hier selbst bei guten Markengeräten Verluste von 25 Prozent und mehr normal sind. Bedeutet: Wenn der Strompreis in Spitzenzeiten bei 40 Cent liegt und in Niederzeiten bei 30 Cent, machen Sie bei diesem Spread eine lokale Null und global ein Minusgeschäft durch den zusätzlichen Akkuverschleiß. Der Spread

muss also schon sehr groß sein, um sich zu rentieren.

Beim Hausakku kommt hinzu, dass die Ladeleistungen im Vergleich zu Wärmepumpe oder Auto meist niedrig liegen (zwei bis drei Kilowatt) und die Kapazitäten typischerweise bei um die zehn Kilowattstunden aufhören. Wenn der Strom für eine Stunde sehr günstig ist, kriegen Sie in dieser Zeit also nur zwei bis drei Kilowattstunden rein (und noch weniger raus). Alles in allem ist das Sparpotenzial per Akku eher zäh zu erschließen. Die Ladesteuerung nach Preisen müsste daher voll automatisiert und integriert werden. Der Speicherhersteller Fenecon etwa hat eine Akkusteuerung, die außer Wetterprognosen auch Day-Ahead-Strompreise berücksichtigt (App „Awattar Hourly“ nach dem gleichnamigen Tarif).

Wenn Sie weder Wärmepumpe noch E-Auto und auch sonst einen geringen Stromverbrauch haben, lohnen sich dynamische Stromtarife nach bisheriger recht einhelliger Expertensicht nicht, weil der Aufwand die Einsparpotenziale übersteigt. Sie können sich natürlich aus nicht-ökonomischen Gründen dennoch am System beteiligen. Vorteil für Geringverbraucher: Ihr Risiko liegt analog zum Verbrauch niedriger.

Kosten

Die Installationskosten für eine intelligente Messeinrichtung liegen meistens bei um die 100 Euro, können aber auch bis über 250 Euro gehen. Oder Sie können bei einem Jahresverbrauch über 6000 Kilowattstunden darauf warten, dass Sie ab 2025 per Pflichtprogramm umgerüstet werden, denn dann sind die Installationskosten auf 50 Euro begrenzt. Der Betrieb ist für Endverbraucher bis 10.000 Kilo-

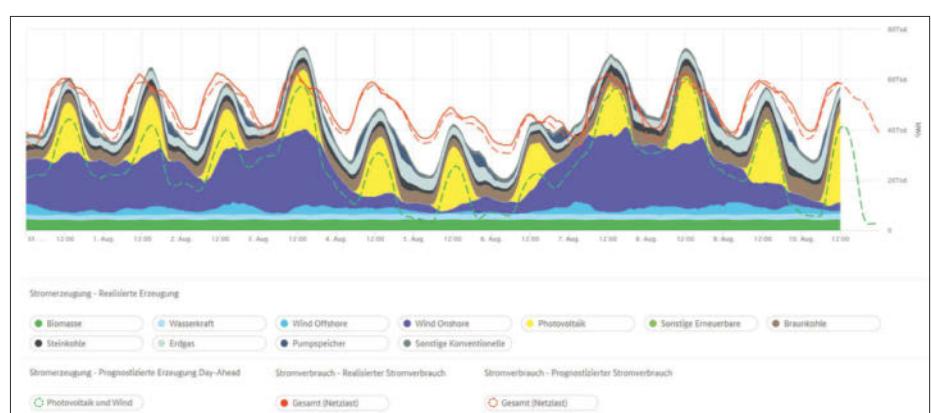


Bild: Bundesnetzagentur/SMARD

Wie viel Strom Wind und Sonne produzieren, bestimmt das Wetter, nicht der Bedarf. Dadurch schwanken die Strompreise immer stärker. Dynamische Stromtarife sollen helfen, Angebot und Nachfrage besser aufeinander einzustellen.

wattstunden auf 20 Euro jährlich gedeckelt. Bei höheren Verbräuchen zahlen sie maximal 50 Euro pro Jahr.

Zweirichtungszähler für PV kosten bis 15 Kilowatt-Peak 20 Euro, von 15 bis 25 Kilowatt-Peak 50 Euro und von 25 bis 100 Kilowatt-Peak 120 Euro. Dazu kommt die Grundgebühr des jeweiligen Anbieters. Günstige Anbieter fangen hier bei um die 5 Euro im Monat an (zum Beispiel Tibber mit 4,49 Euro). Die Netzbezugs- und Messstellengebühren hängen vom jeweiligen Verteilnetz ab, deshalb müssen Sie bei jeder Strompreisabfrage Ihre Postleitzahl angeben. Typisch sind um die zehn Euro im Monat, die obendrauf kommen. Diese Preise muss jeder Anbieter auflisten und aufschlüsseln.

Die variablen Kosten sind die für den verbrauchten Strom in Kilowattstunden. Der Börsenstrompreis macht aber nur einen Teil dieser Kosten aus, und noch dazu den kleineren. Alle Umlagen, Gebühren, Entgelte und Abgaben sowie je nach Anbieter außerdem Zusatzkosten und An-

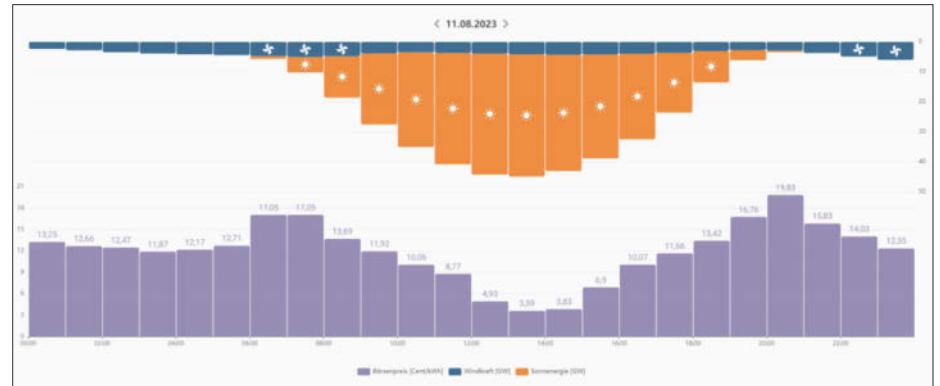


Bild: Awwatt

Einfluss der Solarstromproduktion auf den Strompreis: Der sommertägliche Peak bei der Solarleistung drückt kräftig auf den Strompreis. Wer sich darauf einstellen kann, kauft nachmittags vergleichsweise günstig Strom.

bietergebühren legen die Stromanbieter in einen Grundarbeitspreis pro Kilowattstunde. Der sollte Ihre größte Aufmerksamkeit finden, denn wenn der zu hoch liegt, wird es zäh mit dem Sparen. Dieser Grundarbeitspreis enthält außer fixen Zusatzkosten pro Kilowattstunde die variablen Steuern (Stromsteuer und Mehrwert-

steuer) abhängig vom Strom-Basispreis, deshalb schwankt er mit dem Bezugspreis.

Die Angabe des Anbieters bezieht sich also auf einen Bezugspreis, üblicherweise den durchschnittlichen lokalen Strompreis des letzten Monats. Günstige Anbieter liegen in meinem Verteilnetz aktuell bei um die 15 ct/kWh (plus Börsenstrompreis).

DOAG 2023
Konferenz + Ausstellung
21.-24. November
in Nürnberg

Development & Middleware

Data Analytics & KI

Strategie & Softskills

Datenbank & Infrastruktur

Deep Dives am Thementag

KI Navigator 2023
Konferenz zur Praxis der KI
in IT, Wirtschaft und Gesellschaft
ki-navigator.doag.org

anwenderkonferenz.doag.org

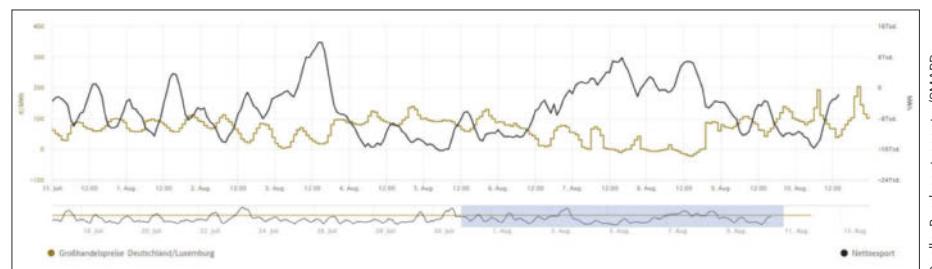
Blick auf die EPEX-Strombörse im Browser: Der Strom kostet zum Schreibzeitpunkt rund 90 Euro pro Megawattstunde, also 9 ct/kWh. Dazu kommen rund 17 Cent Stromnebenkosten, ergibt 26 ct/kWh für meinen Endkundenpreis. Heute Nachmittag ab 14 Uhr sind aber 4 Cent an der Börse angesagt, ergibt für mich rund 20 ct/kWh. Da könnte ich also ein E-Auto laden, bis etwa 17 Uhr wäre es günstig.

Gefahr der nackten Börse

Die Börse ist nicht für Privatkunden gemacht, sondern dort agieren Händler mit großen Strommengen und täglicher Börsenerfahrung. Dazu kommt: Die großen Energiefirmen sind Erzeuger, Händler und Börsenmitbesitzer in Personalunion. Als vom Stromanbieter durchvermittelter Privatkunde finden Sie dort also nicht dieselben Bedingungen vor wie, sagen wir, E.On.

Der Strompreis kann sogar negativ werden. Damit er aber für Sie netto negativ wird, muss er so stark ins Minus rutschen und solange dort verharren, dass er Ihre ganzen Zusatzkosten ausgleicht. Anfang Juli 2023 etwa durchbrach der Börsenstrompreis mittags die Grenze von -500 Euro/Megawattstunde. Da solche Situationen aber viel zu kurz andauern, können Sie selbst mit E-Auto nur wenig Kapital daraus schlagen.

Der Strompreis kann jedoch auch einfach extrem steigen. Im Sommer 2022 kostete eine Megawattstunde beispielsweise 563 Euro, heißt: Sie bezahlten 56,3 Cent plus alle Aufschläge, also brutto über 83 ct/kWh. Gegen solche Situationen haben Sie in einem rein dynamischen Stromtarif langfristig keinerlei Schutz. Kurzfristig bis April 2024 hilft die staatli-



Quelle: Bundesnetzagentur/SMARD

Die schwankende Stromerzeugung zeigt auch beim Im- und Export Effekte: Deutschland kauft Strom ein, wenn er teuer ist und verkauft Strom, wenn er billig ist. Der Preisunterschied wird sich künftig noch deutlich verschärfen. Gut auch für Sie als Inlandsstromkäufer, wenn Sie flexibel sein können.

che Strompreisbremse, die 80 Prozent des Verbrauchs auf 40 ct/kWh deckelt. Bei dynamischen Tarifen wird dabei der Monatsdurchschnitt hergenommen: Es gibt also keinen Schutz der Ersparnisse aus niedrigen Strompreisen, es gibt nur eine gewisse Kostensicherheit nach oben.

Nach der Strompreisbremse fällt die Sicherheit weg. Wenn Sie dann im Winter nicht frieren wollen, muss die Heizung Strom beziehen. Wird dieser wegen sinkendem Angebot und erhöhter Nachfrage teuer, blutet eben der Geldbeutel.

Mental dürften Preisachterbahnen den meisten Menschen ebenfalls zusetzen, wie wir es auch von den Finanzmärkten her kennen. Für Menschen, die nicht so ins Risiko gehen und sich nicht permanent mit der Strombörse beschäftigen möchten, könnten als Zwischending auch versicherte Tarife mit fixen Obergrenzen wie der „Hourly-Cap“ bei Awattar interessant sein. Der kostet allerdings bei mir rund 60 ct/kWh minus eventuelle Ersparnisse nach Börsenlage, wäre also signifikant teurer als mein fixer Stromtarif. Man

kann eben nicht alles haben. Wer fast alles haben möchte: Awattars „Hourly“-Tarif begrenzt den Preis pro kWh nach oben auf 80 Cent und nach unten auf -80 Cent. Da wären also die ärgsten Ausreißer versichert.

Empfehlenswerte Anbieter

Mittlerweile gibt es schon einige dynamische Stromtarife, zum Beispiel bei folgenden Anbietern: Awattar, E.On, Gasag, Oststrom, Rabot Charge, Stadtwerke Düsseldorf, Tibber, Voltego, Wuppertaler Stadtwerke. Einige andere versprechen in ihren Prospekten baldige Verfügbarkeit, darunter die großen Anbieter wie Vattenfall und EWE. Sie können also schon einmal bei Ihrem Stromanbieter vorfühlen, wenn Sie mit ihm zufrieden sind.

Die Tarife unterscheiden sich hauptsächlich bei der Höhe der sonstigen Kosten. Die Pioniere dynamischer Tarife (Awattar, Rabot, Tibber) haben immer noch mit die schlankesten Kostenstrukturen, und die sind für die Idee „nackte Börse“ am interessantesten.



Bild: SMA

Das Elektroauto ist als leistungsstarker, flexibler Großverbraucher der beste Freund von dynamischen Stromtarifen, PV-Überschussladen und Energie-management.



Clemens Gleich

Romantische Schneelandschaft, aber der Zähler rotiert dann im Bestandsbau. Wie viel Teuerzeit Sie überbrücken können, hängt an jenen Faktoren, an denen am liebsten gespart wird: Dämmung und Wärmespeicherkapazität von Gebäude und Heizanlage.

Tibber etwa gibt den Börsenpreis direkt weiter und macht sein Geschäft über die Grundgebühr, den Geräteverkauf (Wallbox, Pulse) und den Handel mit Regelenergie. Awattar schlägt drei Prozent des Börsenstrompreises als Kostendeckungsbeitrag auf, unter anderem um das versprochene Limit von maximal 80 ct/kWh refinanzieren zu können. Rabot Charge zwackt 20 Prozent der Ersparnis gegenüber dem Grundversorgungstarif als Provision ab, um sich mit den Kunden bei Ersparnissen ins selbe Boot zu setzen. Bei Mehrkosten gegenüber der Grundversorgung steigt Rabot allerdings aus dem Boot aus, die müssen Sie allein bezahlen. Immerhin fallen dann geringere Gebühren an.

Fazit

Menschen mit großem Sicherheitsbedürfnis sollten beim Grundversorger in einem fixen Tarif bleiben. Wer ein gewisses Risiko verkraften kann, sollte sich mit ihnen vergleichbare Nutzer solcher Tarife anschauen, um das Einsparpotenzial besser abschätzen zu können. Es liegt aufgrund der Stromkostenstrukturen meistens wesentlich niedriger, als man zunächst hofft, und ohne konsequente Automatisierung beziehungsweise Verhaltensänderung könnten Sie auch mehr zahlen als vorher.

Für die (Noch-)Minderheiten der E-Auto-Nutzer wird sich ein dynamischer Tarif mit etwas Automatisierung und ausreichend Fahrleistung meistens lohnen. Für die Wärmepumpenheizung müssen wir ausdifferenzieren: Je länger Haus(-technik) und Warmwasserspeicher die produzierte Wärme halten, umso mehr können die Bewohner sparen. Dabei hilft auch, wenn sie für ein paar Stunden Temperaturen unterhalb ihrer Komfortzone aushalten. Eine Anlage mit geringer Flexibilität, wie sie etwa in einem alten Haus mit hohen Wärmeverlusten und geringen Speicherkapazitäten öfter zu finden ist, kann jedoch auch sehr dynamische Mehrkosten generieren.

(anm@ct.de) 

Dieser Artikel erschien zuerst auf *heise+* unter heise.de/-9244963.

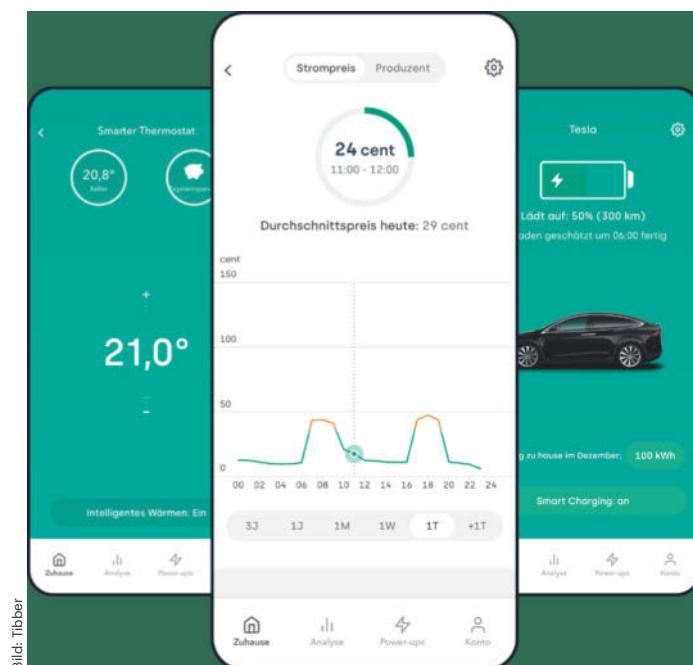


Bild: Tibber

Datenbetrachtung hat immer auch etwas von Gamification: Tibbers App weiß, welche Geräte im Haushalt gerade wie viel Strom zu welchem Preis verbrauchen – und berechnet und visualisiert die Kosten.

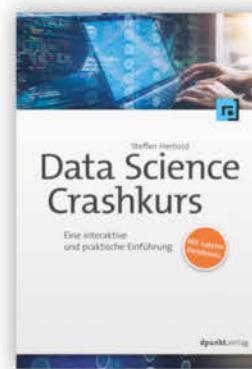
Deine Begleiter für Uni und Ausbildung!



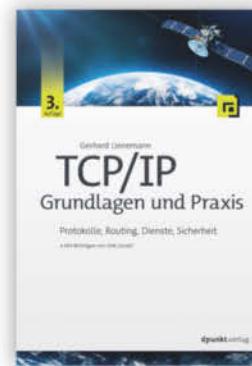
4. Auflage · 712 Seiten · 49,90 €
ISBN 978-3-86490-598-8



5. Auflage · 248 Seiten · 34,90 €
ISBN 978-3-86490-984-9



346 Seiten · 34,90 €
ISBN 978-3-86490-862-0

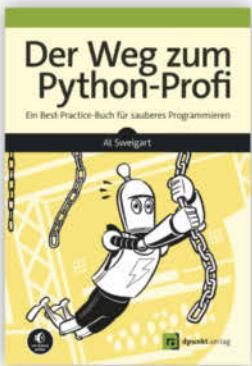


3. Auflage · 366 Seiten · 39,90 €
ISBN 978-3-86490-960-3

Entdecke weitere Lehrbücher unter
dpunkt.de/lehrbuecher



424 Seiten · 22,90 €
ISBN 978-3-86490-852-1



424 Seiten · 34,90 €
ISBN 978-3-86490-874-3



4. Auflage · 734 Seiten · 44,90 €
ISBN 978-3-86490-705-0

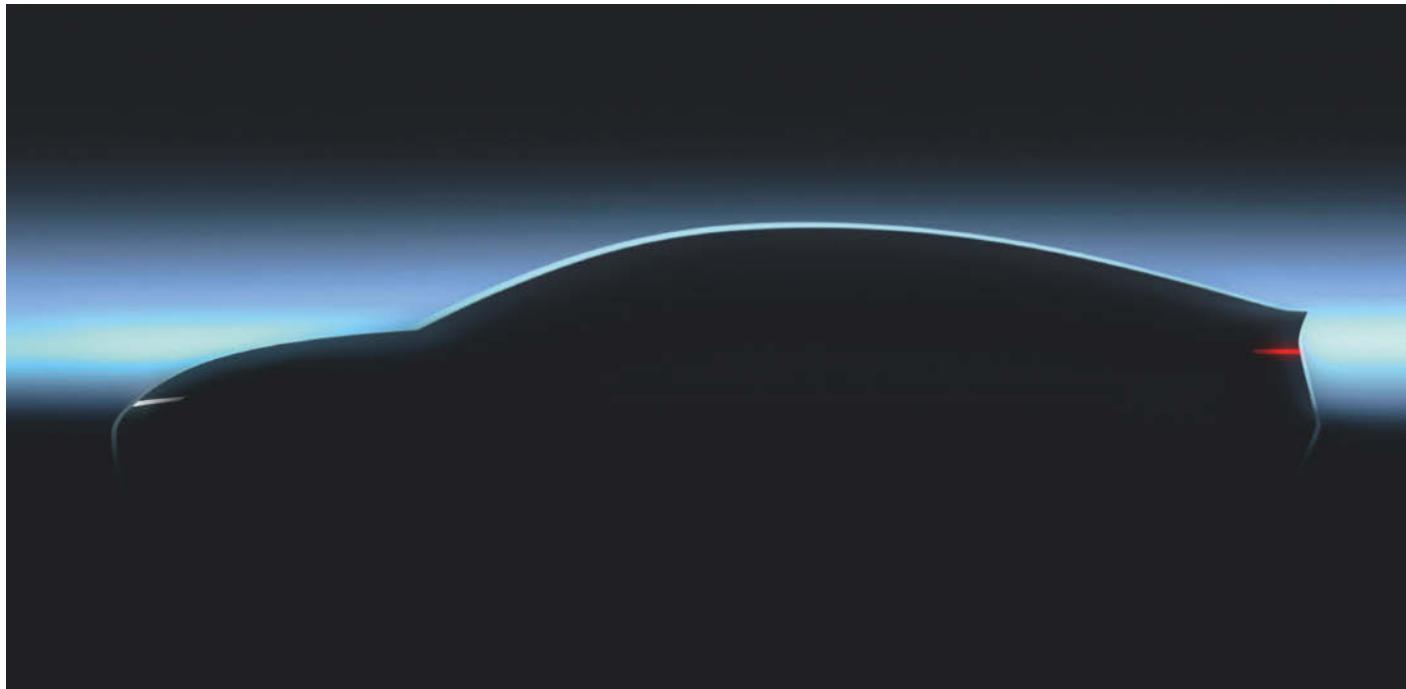


Bild Volkswagen AG

Größer, schwerer, Google

E-Auto-Trends 2024

Mit immer neuen Schnelllade-rekorden und dickeren Batterien wollen die Hersteller den Kunden das E-Auto schmackhaft machen. Im Hintergrund tobt ein Kampf um die Software-Hoheit, denn einfach „nur“ Auto ist nicht mehr genug.

Von Sven Hansen

Sie haben sich noch nicht an ein E-Auto herangetraut? Das könnte sich bald ändern. Lädt zu langsam und fährt nicht weit genug – das waren für die meisten Menschen die größten Hürden beim Wechsel zu den E-Modellen. Hier hat sich einiges bewegt: vor allem durch verbesserte Schnellladetechnik und größere Akkus.

Dabei benötigen die wenigsten Kunden im Alltag Kapazitäten jenseits der 100-kWh-Grenze, wie sie Mercedes beim EQS oder BMW im i7/iX anbieten. Denn die Batterie hat im Vergleich zum 85-Liter-Tank einen gewichtigen Nachteil: Auch

komplett entladen muss man die Last mit sich herumfahren, was die Reichweite im Vergleich zu einem leichteren Fahrzeug verringert. Mehr ist also nicht unbedingt auch besser.

Doch auch im mittleren Preissegment sind die Akkukapazitäten deutlich angestiegen oder die Hersteller bieten die Fahrzeuge zumindest mit unterschiedlichen Batterievarianten an, sodass man die zu den eigenen Bedürfnissen passende Reichweite selbst wählen kann.

Während die Akkugröße nach wie vor direkt mit dem Fahrzeugpreis verknüpft ist, lohnt beim Thema Ladegeschwindigkeit ein genauer Blick ins Datenblatt. Denn mit Hyundais Ioniq 5 findet sich auch ein erstes Fahrzeug im mittleren Preissegment, das in den aktuellen ADAC-Top-10 der schnellsten Schnelllader landet. Für den Alltag ist dies meist wichtiger als die Kapazität – denn was nutzt ein Riesenakku, den man nicht schnell befüllen kann?

Der ADAC hat rund 60 Modelle an die DC-Schnellladesäule gehängt: Modelle mit 800-Volt-Technik sind klar im Vorteil. Dazu zählen neben den elektrischen Rennmaschinen von Porsche und Audi

auch Fahrzeuge von Hyundai und der Schwestermarke Genesis; mit ihnen sind Peak-Leistungen jenseits der 200 Kilowatt durchaus möglich – und dann sind die um 100 kWh großen Akkus in rund 20 Minuten von 10 auf 80 Prozent geladen.

Viel spannender als die Peak-Ladeleistung sind allerdings die dazu gegenläufigen Bemühungen der Hersteller, die Lebenszeit des Akkus durch ein starkes Einbremsen des Ladevorgangs zu schonen. Der BMW i4 iDrive40 kommt so trotz 400-Volt-Technik zwar noch auf 209 kW Peak-Leistung beim Schnellladen, schafft bei einer Ladung von 10 auf 80 Prozent jedoch nur einen 133-kW-Durchschnitt – was etwa 30 Minuten an der Ladesäule bedeutet.

Etwas in den Hintergrund geraten scheint das AC-Laden – beim Rennen um den Schnellladerekord wird oft vergessen, dass längst nicht überall DC-Ladesäulen vorhanden sind. Ob am Urlaubsort an der Küste, in ländlichen Regionen oder daheim an der Wallbox mit Börsenstrom: Manchmal benötigt man das schnelle Laden mit Wechselstrom (AC).

Gerade die dynamischen Stromtarife dürften in Zukunft spannender werden.

Waren es bisher noch exotische Anbieter wie Tibber oder Awattar, bei denen Strom zum stundengenauen Börsenpreis zu haben war, steigen nun weitere wie EON ein; in ein paar Jahren muss jeder Stromanbieter einen variablen Tarif anbieten (siehe S. 22). Wenn der Strom dann mal für 3 statt 30 Cent pro Kilowattstunde zu haben ist, werden die Kunden auch daheim möglichst schnell laden wollen – mit der AC-Wallbox.

Das jüngst in nur binnen eines Tages verglühte Solarförderprogramm der KfW nahm diesen Trend auf. Statt wie beim alten Wallbox-Programm „bis zu“ 11 kW fördert man im Rahmen der Solarförderung Wallboxen „ab 11 kW“. Wer beim Kauf eines E-Autos im Konfigurator die Option „22 kW beim AC-Laden“ findet, sollte das Häkchen setzen.

Kampf um Kontrolle

Weiter zugespitzt hat sich der Kampf um die Vorherrschaft im Cockpit – Stichwort Software. VW hat sich von der konzernweiten Festlegung auf ein monolithisches VW-OS verabschiedet. Zu holprig lief der Start in den Fahrzeugen der ID-Reihe, mit hakenden OTA-Updates und viel Nachbesserungsbedarf, der mitunter Werkstattbesuche und den Austausch von Steuergeräten erforderte.

Ganz hat man das Projekt des haus-eigenen Betriebssystems nicht aufgegeben, will aber bei den hochpreisigen Marken im Konzern andere Wege beschreiten. VW-OS, vom ehemaligen Konzernchef Herbert Dies für 2025 angekündigt, hat sein Nachfolger Blume auf eine längere Bank geschoben: auf „Ende des Jahrzehnts“. Damit verzögert sich auch die Einführung der neuen Scalable Systems Platform (SSP), die dem eigentlich für 2026 angekündigten Modell Trinity zugrunde liegt (siehe Titelbild). Trinity soll den Weg für eine neue Generation von E-Fahrzeugen bereiten – die Zukunft scheint vage.

Auch bei BMW steht man vor einem wichtigen Schritt weg vom proprietären Infotainment, hin zum offenen Android-System. Der neue 5er wird noch auf der alten Softwareplattform laufen – mit der „Neuen Klasse“ will man den Wechsel zu Android wuppen. Die Münchner betonen zwar, dass alle Fahrzeuge unabhängig von der Plattform mit Updates versorgt werden. Doch der Wechsel zu Android war nötig, um angesichts der ständig wachsenden Komplexität und der gesteigerten Er-

wartungshaltung der Kundschaft auf den deutlich größeren Pool von Android-Entwicklern und vorgefertigten Modulen und Apps zugreifen zu können. In der „Neuen Klasse“ dürfte in dieser Hinsicht deutlich mehr Bewegung sein als bei den bisherigen Fahrzeugen.

Was bei der Kernmarke erst noch kommt, ist bei der Tochter Mini schon heute zu sehen. Das Mini OS 9 setzt bereits auf einem offenen Android auf und zeigt, wie flexibel die Lösung ist. Statt der üblichen Rechteckdisplays kommt ein kreisrundes 24-cm-OLED von Samsung zum Einsatz, das an den mittlerweile ikonischen runden Tacho im Armaturenbrett erinnert. Der Displayspezialist hat das OLED extra für die Mini-Kooperation entwickelt. Das Infotainment bietet zahlreiche Möglichkeiten der Anpassung – voreingestellte Themen und Möglichkeiten zur Individualisierung mit eigenem Material.

Doch nicht nur für den technischen Unterbau, sondern auch für vordergründige Anwendungen auf dem Infotainment-Display bietet sich der Wechsel auf Android an. Hier haben die meisten Hersteller sichtbaren Nachholbedarf. Die Konkurrenz kommt – mit Ausnahme von Tesla vielleicht – dabei nicht aus der eigenen Branche, sondern aus dem Silicon Valley: Apple und Google haben mit CarPlay und Android Auto zwei starke Bedienoberflächen im Programm, die beliebte Smartphone-Apps via USB-Kabel oder drahtlos aufs Cockpit-Display beamen. Das Gros der Infotainment-Systeme der Autobauer hält weder optisch noch funk-

c't kompakt

- Der Reichweitenangst setzen die Hersteller größere Akkus und Schnelllader entgegen.
- Schon nächstes Jahr könnten rund die Hälfte der Neuzulassungen SUVs sein.
- Google, Apple oder Hersteller-Infotainment? Der Kampf ist in vollem Gange.

tional oder bei der Anzahl der unterstützten Apps mit.

Eine der im Auto meistgenutzten Apps dürfte zweifellos Google Maps sein. So wundert es kaum, dass Google auf der diesjährigen IAA verkündet hat, den Kartendienst optisch und technisch massiv aufzuwerten. Darüber hinaus will man den Fahrzeugherstellern den Einsatz von Google Maps erleichtern – sprich: künftig weniger Auflagen über zu nutzende Zusatzdienste machen. Eigentlich wollte der Konzern Maps nur im Paket mit Play Store und Google Assistant lizenziieren. In einem Brief an Google hatte das Bundeskartellamt diese Bündelung von Diensten unter Berufung auf den Anfang Mai des Jahres in Kraft getretenen Digital Markets Act (DMA) der EU kritisiert und ein Verfahren angedroht – offensichtlich mit Erfolg.



Bild: Smart

Trend zur Größe: Der Smart #1 ist fast doppelt so lang wie der Ur-Smart.

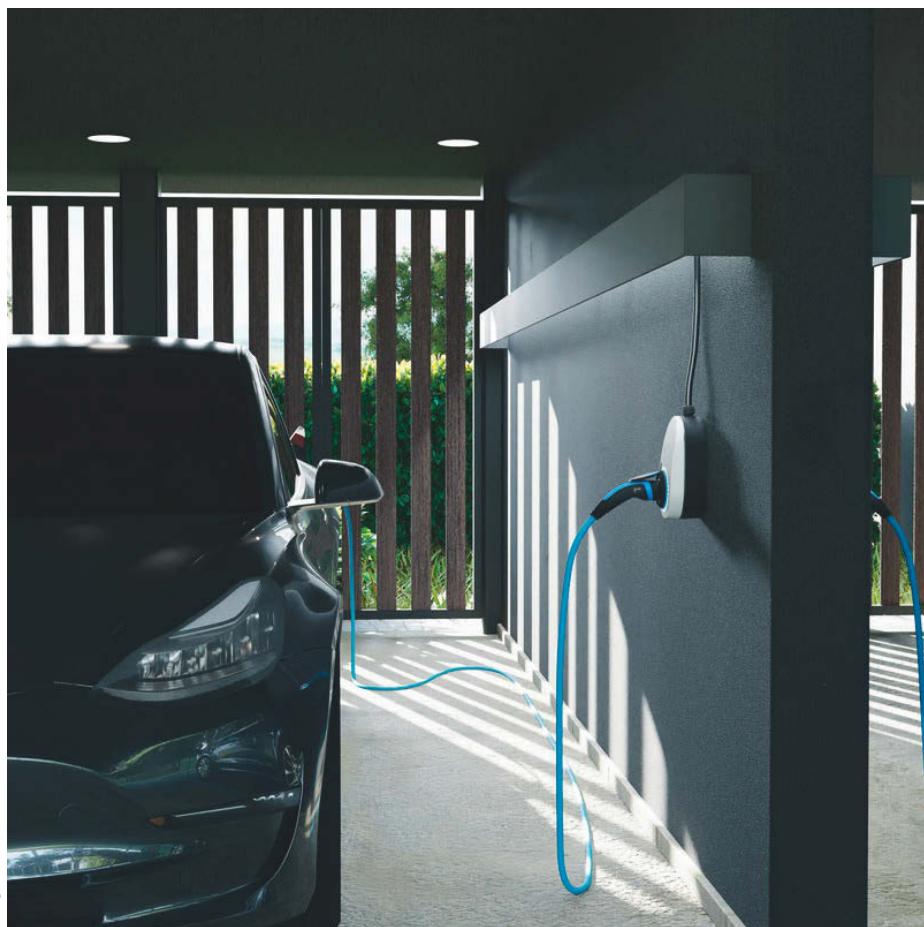


Bild: go-e

Der Go-e-Charger Gemini wurde im Rahmen des nach einem Tag vergriffenen KfW-Programms 442 in der 22-kW-Version gefördert.

Hochautomatisiert unterwegs

Beim autonomen Fahren geht es in kleinen Schritten voran, 2024 könnte den Durchbruch bringen. Mercedes hat den Staupilot auf die Straße gebracht und auch BMW steht nach eigenen Angaben kurz vor der Einführung erster Level-3-Funktionen. Momentan muss den Kunden noch Level 2+ reichen, wobei man die Geschichte bei der Deutschlandpremiere des bereits in den USA eingeführten Assistenzsystems ohne „+“ verkauft. Der Fahrer soll also ganz explizit **nicht** aus der Verantwortung entlassen werden.

Was aber geht: Hände vom Steuer, Augen auf die Straße – schon kann man mit 130 Sachen assistiert auf der Mittelpurfahren. Überholvorschläge des Assistenten lassen sich durch einen kurzen Blick in den Seitenspiegel annehmen. Ob das für mehr Sicherheit und Komfort sorgt, wird sich zeigen. Da die Option bei Neufahrzeugen nur auf Zeit buchbar ist, kann BMW sie zur Not auslaufen lassen.

Ansonsten dreht sich das Feld im Kreis. BMW setzt bei der Technik der Um-

gebungserkennung statt auf Mobileye auf Qualcomm, VW und Porsche arbeiten mit Mobileye, Mercedes hingegen mit Technik von Nvidia. Die auf der IAA bekannt gegebene Kooperation von Mobileye und

Valeo weist auf eine Verdichtung des Marktes hin. Valeo, Weltmarktführer bei Fahrerassistenzsystemen, wird mit Mobileye an Imaging-Radar-Lösungen für hochautomatisierte Fahren arbeiten. Das System soll Objekte auf 300 Meter im Umfeld des Fahrzeugs erkennen und sicher einordnen können. Dabei will es günstiger und robuster sein als Lidar-Lösungen mit statischen oder rotierenden Laserdioden – und unempfindlicher gegen schlechte Sichtverhältnisse.

SUVisierung schreitet voran

Schwerer, höher, breiter – kurz SUV. Schon im ersten Halbjahr kam der Fahrzeugtyp der Sport Utility Vehicles auf einen neuen Rekordwert von 41 Prozent bei den Neuzulassungen in Deutschland. Auf Deutschlands Straßen geht es enger zu – bald könnte jedes zweite Neufahrzeug ein SUV sein.

Der Trend zur Größe lässt sich wohl an kaum einem anderen Fahrzeug besser illustrieren als am neuen Smart #1: Als die Mutter Mercedes die komplette Elektrifizierung der Marke Smart verkündete, hatten die meisten Kunden wohl elektrifizierte Kleinwagen im Kopf. Doch der #1 steht auf derselben SUV-Plattform wie Volvos EX30 und bringt bei 4,72 Meter Länge und 1,82 Meter Breite (ohne Spiegel) satte 1,8 Tonnen auf die Waage. Zum Vergleich: Der „Ur-Smart“ wog in der Basisausstattung als Verbrenner 720 Kilogramm, war 30 Zentimeter schmäler und mit 2,50 Meter fast nur halb so lang wie der #1. Mit dem Kerngedanken des quer parkenden, smarten Stadtflitzers hat der



Bild: Mini

So schön geht Infotainment: Im Mini mit kreisrundem OLED tickt erstmals BMWs OS 9 auf Android-Basis.

Smart #1 so viel zu tun wie – ja, wie was eigentlich?

Wenn so die smarten Stadtautos der wilden 20er aussehen, wird man künftig häufiger die Spiegel wegklappen müssen, um sich in Schleichfahrt aneinander vorbeizukriegen. In den Parkhäusern finden sich immerhin häufiger „XXL“-Schilder und die Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) hat ihre Empfehlung für die Breite von Parkplätzen kürzlich von 2,50 auf 2,65 Meter erhöht. Mehr Auto, weniger Mensch.

Den Herstellern verschafft der Trend im wahrsten Sinne mehr Luft, den sie für zusätzliche Sicherheitstechnik und voluminöse Batteriespeicher nutzen können. Batterien sind nach wie vor in mehrfacher Hinsicht der Flaschenhals bei Entwicklung und Verkauf der E-Autos. Im oben erwähnten EX30 etwa verbaut Volvo erstmals LFP-Batterien (Lithium-Eisenphosphat). Sie sind robuster und zyklenfester, aber haben eine niedrigere Energiedichte und benötigen daher mehr Raum.

Auch durch steigende Sicherheitsanforderungen füllen sich die Innenräume, denn die Kriterienkataloge der Crash-Institutionen weltweit verlagern sich immer stärker hin zur Unfallvermeidung durch elektronische Systeme, statt nur die passive Sicherheit einzubeziehen. Wie steil die Anforderungskurve über die Jahre ansteigt, musste Renault beim Zoe erfahren: 2013 mit fünf von fünf Sternen im Euro-NCAP geadelt, holte sich der Kleinwagen zehn Jahre später eine Klatsche mit 0 Sternen.



Bild: Mobileye

Die israelische Firma Mobileye entwickelt in Kooperation mit Valeo und Wistron günstige Systeme zur sicheren Objekterkennung auch im urbanen Umfeld.

Mit dem in Kürze verfügbaren EX30 schwimmt Volvo als Premiumhersteller in zwei Punkten gegen den Strom: Während die Konkurrenz die Einstiegsmodelle eher tilgt, um den Gewinn zu optimieren, kommen die Schweden mit einem E-SUV zum Einstiegspreis von weniger als 37.000 Euro. Das klappt unter anderem deshalb, weil Volvo als Geely-Tochter direkten Zugriff auf die chinesische SEA-Plattform hat – auf selbiger steht der Smart #1. Den EX30 soll es allerdings nur mit 11-kW-AC-Lader geben – die Plattform beherrscht eigentlich mehr. Dennoch: Ein günstiges E-Auto aus

China mit dem europäischen Markenversprechen von Volvo – das könnte klappen.

Auch die Art, wie man sein Auto finanziert, ist im Wandel. Neben Barzahlung und Leasing ist eine Vielzahl zusätzlicher Möglichkeiten getreten. Der Trend zur Flatrate ist erkennbar. Ob Full-Service-Leasing oder Auto-Abo: Immer mehr Menschen wollen Rundum-sorglos-Pakete und sind auch bereit, dafür monatlich ein wenig mehr zu bezahlen. Ein Auto-Abo ist im direkten Vergleich zwar deutlich teurer als klassisches Leasing, dafür sind Versicherung, Steuer, Wartung oder Reparaturen schon inklusive. Gefällt einem das Auto nicht mehr, wechselt man zu einem neuen. Das klappt bei vielen Anbietern schon nach einem bis drei Monaten, während man sich beim Leasing üblicherweise zwei Jahre und länger bindet.

Fazit

Mehr Sicherheit, mehr Reichweite, mehr groß: Die E-volution ist in vollem Gange. Statt um elektrifizierte Zweitwagen geht es künftig um elektrische Familienkutschen, um Sport- und Dienstfahrzeuge. Mit Car-Play 2 in den Startlöchern und Google Automotive im Nacken ist der Innovationsdruck für die Hersteller gerade beim Thema Infotainment so groß wie nie. Sie werden alles daran setzen, sich im Kampf um die Vorrherrschaft im Cockpit gegen die Internetgiganten zu behaupten. (sha@ct.de) **ct**



Bild: Euro NCAP

Der „Fall“ des Zoe von 5 auf 0 Sterne zeigt: Für elektrische Kleinwagen wird es immer schwieriger, die Anforderungen an passive und aktive Sicherheit zu erfüllen. Auch das nährt den SUV-Trend.

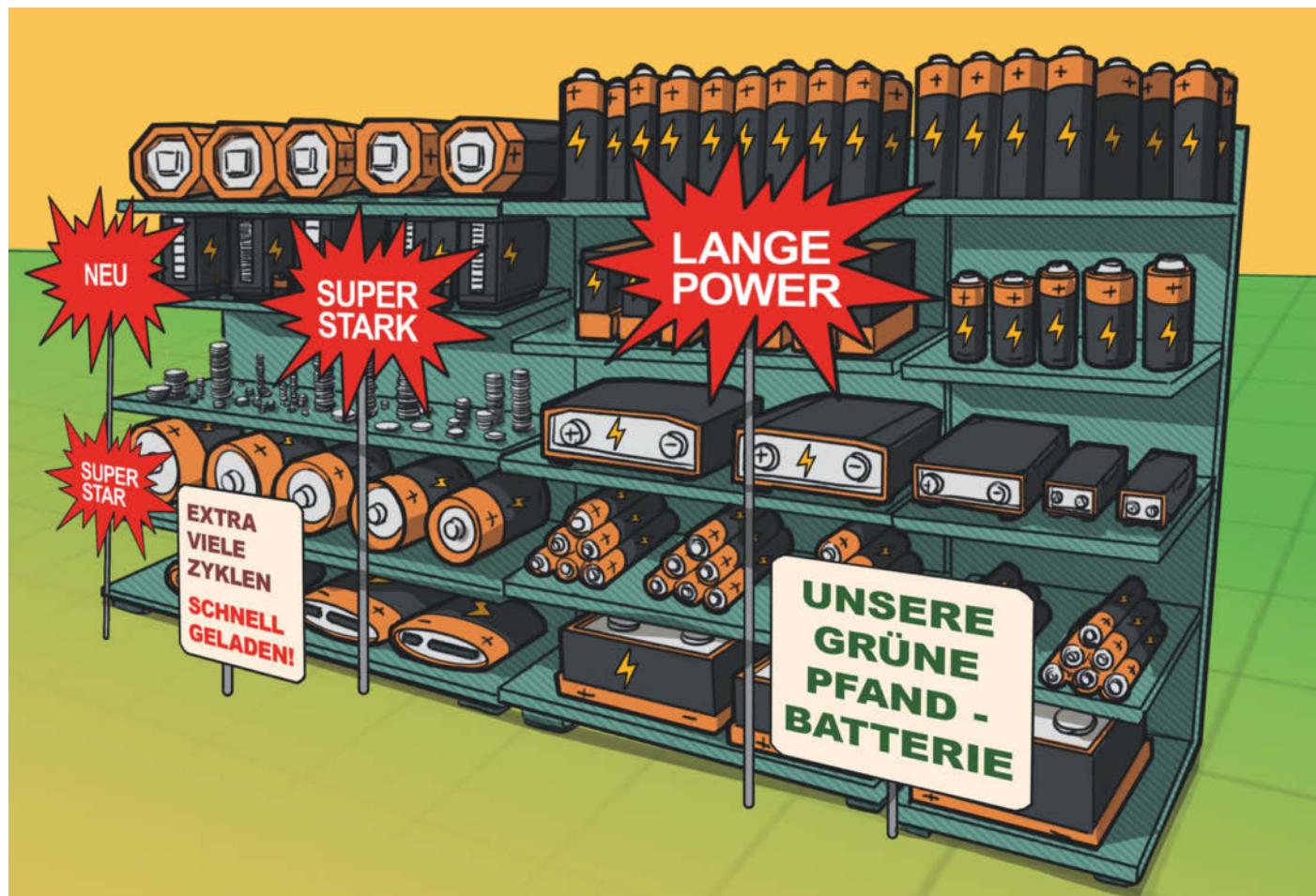


Bild: Thorsten Hübner

Muss es Lithium sein?

Roadmap: Batterien mit höchster Energiedichte oder ohne kritische Rohstoffe

Lithium-Ionen-Zellen sind der Standard bei der Akkutechnik. Das weist die jüngste Batterietechnik-Roadmap des Fraunhofer ISI aus. Aber auch Natrium-Ionen-Technik hat Vorteile und benötigt fast keine kritischen Rohstoffe.

Von Arne Grävemeyer

Batterietechnik macht mobil und der Bedarf wächst schnell. Smartphones und Mobilrechner steigern ihre Verbreitung noch, vor allem aber führen Elektroantriebe zu massiv wachsendem Bedarf an großen Akkus. Die derzeit dominierende Technik in kleinen Geräten und in großen Fahrzeugen ist die Lithium-Ionen-Batterie (LIB), die von den etablierten Techniken mit Abstand die höchste Energiedichte liefert. Eine im September 2023 veröffentlichte Roadmap-Studie des Fraunhofer-Instituts für System- und Innovationsforschung (ISI) (siehe ct.de/y377) untersucht die Alternativtechniken

und deren Einsatzchancen bis 2030 und darüber hinaus.

Die Lithium-Ionen-Technik hat ein großes Problem: Zur Herstellung der Akkumulatoren sind eine Reihe kritischer Rohstoffe erforderlich, gerade aus europäischer Sicht. „In Europa können nach heute angekündigten Projekten bis 2030 bis zu 20 Prozent des hier benötigten Lithiums gewonnen werden“, sagt Dr. Axel Thielmann, Leiter des Competence Centers Neue Technologien am Fraunhofer ISI. Andere Rohstoffe dieser Batterietechnik sind Nickel, das vorwiegend in Russland gewonnen wird, Grafit für den Minuspol,

das vornehmlich aus China stammt, und ein, wenn auch sinkender, Anteil an Kobalt, das man vor allem im Kongo abbaut. Schließlich steckt in LIB noch teures Kupfer.

Die nachhaltige Alternative

Im Vergleich dazu zeichnet sich bereits die Natrium-Ionen-Batterie (NIB) als eine nachhaltigere Speichertechnik ab. Nach Ansicht der Fraunhofer-Forscher stehen NIB für mobile Anwendungen schon heute kurz vor der flächendeckenden Einführung in elektrischen Zwei- und Dreirädern und sogar in Kleinwagen. Diese Technik erfordert weniger kritische Rohstoffe als das LIB-Konzept. Beispielsweise ersetzt die NIB Grafit durch Hartkohlenstoff, der sich aus natürlichen Rohstoffen erzeugen lässt, etwa aus Holz oder Nusschalen. Vor allem aber ist anstelle von Lithium lediglich Natrium erforderlich, ein in Form von Natriumcarbonat (Soda) leicht und günstig verfügbarer Rohstoff.

Der größte Nachteil der Natrium-Ionen-Technik ist ihre geringere Energiedichte. Während LIB laut Roadmap heute 200 bis 300 Wattstunden pro Kilogramm (Wh/kg) erreichen, liegen NIB bei 140 bis 160 Wh/kg. Auch für die Zukunft erwarten die Forscher einen deutlichen Leistungsabstand: LIB sollen bis 2035 Energiedichten von 320 bis 360 Wh/kg erreichen, NIB könnten bis dahin die 200 Wh/kg knacken. Im Volumen schneiden die NIB sogar noch schlechter ab: Sie erreichen 250 bis 300 Wattstunden pro Liter (Wh/l), LIB kommen derzeit auf 600 bis 750 Wh/l.

Auf der anderen Seite erfordern NIB zu ihrer Herstellung die günstigeren Ressourcen, verbunden mit einem geringeren CO₂-Fußabdruck dieser Rohstoffe. Darauf hinaus haben NIB auch technische Vorteile gegenüber LIB. NIB gelten als weniger entzündlich. Sie weisen sogar eine höhere C-Rate auf, also ein günstigeres Verhältnis zwischen Ladestrom und Kapazität, und erlauben damit eine kürzere Ladezeit. Und bei tiefen Temperaturen unter 0 °C verlieren sie wesentlich weniger Kapazität als LIB.

Trotz der günstigeren Rohstoffe dürfen die Kosten für NIB-Zellen zunächst höher liegen als die von LIB, wegen der LIB-Massenproduktion. Diese Rechnung geht sogar weiterhin auf, obwohl die Rohstoffpreise für Lithium die LIB-Kosten auf 140 bis 150 Euro pro Kilowattstunde getrieben haben. Die ISI-Forscher erwarten allerdings, dass die LIB-Preise in den kom-

ct kompakt

- In puncto Energiedichte setzt Lithium-Ionen-Technik bei vergleichsweise günstigen Herstellungskosten heute und in absehbarer Zukunft den Standard.
- Aus geopolitischer und ökologischer Sicht bietet die Natrium-Ionen-Batterie bereits heute eine technisch ausgereifte Alternative.
- Weitere Batteriekonzepte könnten in Zukunft Lithium-Ionen-Batterien überholen und beispielsweise Flugzeuge elektrifizieren.

menden zehn Jahren bis auf 50 bis 60 Euro fallen können.

Hochleistungskonkurrenz

Andere Konkurrenten der Lithium-Ionen-Technik sind heute noch nicht so weit wie LIB und NIB. Dazu zählen etwa Lithium-Schwefel-Batterien, denen die Forscher aufgrund ihrer hohen Energiedichte von über 300 Wh/kg im Labor zutrauen, ab 2035 Drohnen anzutreiben. Ab 2040 könnte sich diese Technik sogar in größeren elektrischen Fluggeräten durchsetzen. Dann müssten sich die Konstrukteure allerdings damit auseinandersetzen, dass die leichten Lithium-Schwefel-Batterien viel Platz brauchen. Sie erreichen derzeit nur bis zu 450 Wh/l.

Für stationäre Speichersysteme könnten Techniken mit geringerer Energiedichte interessant werden, darunter Redox-Flow-Batterien, Natrium-Schwefel-Hochtemperaturbatterien oder die schon genannten NIB. Bei diesen Anwendungen kommt es auf Zyklenstabilität und Tiefenentladefähigkeit an, nicht so sehr auf Gewicht oder Volumen der Batterie.

Industrielle Batteriezellformate beeinflussen die Batterieeigenschaften. Mit hoher Zellintegration lässt sich in großformatigen Zellen die höchste Energiedichte je Volumen erzielen.

Insbesondere bei Redox-Flow-Batterien sowie Lithium-Luft- und Aluminium-Ionen-Technik stehen EU-Forschungsstandorte aufgrund ihrer Patente besser da, während Forscher aus Japan und China in anderen Bereichen führend sind, insbesondere bei LIB.

Als einen Nachfolger der LIB mittelfristig betrachtet Thielmann die Feststoffbatterie, in der gegenüber der Lithium-Ionen-Technik der Flüssigelektrolyt durch einen Festelektrolyten ersetzt wird. Auch diese Batterietechnik hat im Prototypenstatus bereits an der Marke von 400 Wh/kg gekratzt. „Das wird eine echte evolutionäre Weiterentwicklung sein, die noch einmal höhere Energiedichten als LIB bringen kann“, schätzt Thielmann. Zusätzlich trauen Forscher dieser Technik eine sehr hohe Zyklenzahl zu, eine höhere Betriebssicherheit ohne giftige Flüssigkeiten und obendrein ein geringeres Entzündungsrisiko. Aber das ist derzeit nur ein gewagter Blick in die Zukunft.

Platzhirsch Lithium-Ionen-Technik

Im Moment führt an der verbreiteten LIB mit Flüssigelektrolyt bei Hochenergiebatterien für Pkw und in Zukunft sogar im Lkw kein Weg vorbei. Die ISI-Roadmap geht davon aus, dass die globale Nachfrage nach LIB von heute etwa 500 Gigawattstunden (GWh) auf mehr als das Zehnfache ansteigen wird und in den kommenden zehn bis zwanzig Jahren 6 bis 10 Terawattstunden (TWh) erreicht. Für die NIB bleibe in naher Zukunft nur etwa 10 bis 20 Prozent Marktpotenzial übrig.

„Angesichts dieser Entwicklung wird es darauf ankommen, die Umweltbilanz der Lithium-Ionen-Technik weiter zu verbessern“, unterstreicht Thielmann. Durch die Verbesserung der Lithium-Abbaumethoden, der Verarbeitung und des Recyclings lasse sich der CO₂-Fußabdruck dieses Rohstoffs senken. Zunehmende Men-



Bild: Fraunhofer ISI

gen von Primär- und Sekundärrohstoffen aus Europa helfen, die Abhängigkeit der hiesigen Industrie von anderen Regionen zu mildern. Zusammengerechnet mit alternativen Batterietechniken wie der NIB lasse sich die Lithium-Abhängigkeit Europas um bis zu 50 Prozent reduzieren.

Doch es gibt auch andere Stimmen, die den Verlauf der jüngsten Roadmap nicht in Stein gemeißelt sehen. „In der Natrium-Ionen-Batterie sind keine kritischen Rohstoffe aus politisch instabilen Regionen enthalten, außer vielleicht etwas Nickel“, betont Maximilian Fichtner. Er ist Direktor des Helmholtz-Instituts Ulm für elektrochemische Energiespeicherung und zugleich Sprecher des Exzellenzclusters „Post Lithium Storage“. Für den hohen Wirkungsgrad der NIB bei tiefen Temperaturen nennt er Zahlen: Bei -20 °C hat eine klassische LIB noch 70 Prozent ihrer Kapazität. Entsprechend sinkt bei einem E-Auto dann die Reichweite. Die NIB hat dann noch 90 Prozent und holt in der Reichweite auf.

Zudem gelten alle Metall-Ionen-Batterien als konzeptionell identisch. Damit gilt gerade die NIB als eine klassische Drop-in-Technik: Bestehende LIB-Produktionsanlagen lassen sich mit vergleichsweise geringem Aufwand auf eine NIB-Fertigung umstellen. Es ist demnach möglich, dass Produktionskapazitäten für LIB in Zukunft NIB herstellen.

Natrium-Ionen-Batterie im Kommen

Im Frühjahr 2023 hat der Akkuhersteller Hina mit dem Autohersteller JAC Sehol, einem Joint Venture der chinesischen JAC mit Volkswagen, den Kleinwagen E10X vorgestellt, der mit Strom aus NIB angetrieben wird. Es zeige sich, dass dieser Wagen sogar schneller geladen werden könne als vergleichbare Fahrzeuge mit LIB, sagt Fichtner.



Den Kleinwagen E10X haben der Akkuhersteller Hina und Autobauer JAC Sehol mit Natrium-Ionen-Batterien ausgestattet.

Bereits 2021 hat der Technikchef des größten chinesischen Akkuherstellers CATL bekannt gegeben, dass das Unternehmen in die Großproduktion von Natrium-Ionen-Zellen einsteigt. Außer stationären Speichersystemen entstehen dort bereits heute automobiltaugliche Zellen, die an Fahrzeughersteller ausgeliefert werden, wie Fichtner bei einem Besuch erfuhr. Man wolle mit NIB als Erstes die althergebrachten Bleibatterien im Automobil ersetzen und im zweiten Schritt auch Zweiräder sowie Stadtautos antreiben.

Hinzu kommen neueste Entwicklungssprünge im Batteriedesign. 2020 versprach Tesla mit der 4680-Zelle einen höheren Volumennutzungsgrad der LIB-Technik. Anscheinend haben der chinesischen Fertiger BYD mit seiner Blade Battery und CATL mit der Qilin Battery die Amerikaner 2022 überholt. Beispielsweise hat die Blade Battery mit 96 cm Länge, 9 cm Breite und

1,35 cm Dicke ein wesentlich größeres Format, etwa wie ein Dielenbrett. Mit ihren Konzepten gelang BYD und CATL, den Masseanteil des Speichermaterials deutlich zu erhöhen und dabei zugleich große Batterieeinheiten in die Großserie zu bringen. BYD hat damit bereits ein E-Auto, den Geely Zeekr 001, ausgestattet, der unter Testbedingungen 1000 Kilometer mit nur einer Batterieladung fuhr.

Fichtner sieht in dieser Entwicklung eine Chance für die NIB. Wenn diese neuen Designkonzepte auf die alternative Batterietechnik mit geringerer Energiedichte angewendet werden, dürften Autobatterien mit einer Reichweite von etwa 600 Kilometer dabei herauskommen. Der britische Hersteller Faradion hat bereits eine NIB-Zelle mit einer Energiedichte von 190 Wh/kg präsentiert. Indem sie ihre Technik testweise in eine Blade Battery von BYD verbauten, erzielten die Faradion-Ingenieure eine Energiedichte von 120 Wh/kg, rein rechnerisch ist das die Energiedichte, die auch in der Batterie eines Audi e-tron 55 steckt. Und das bei einer nachhaltigeren Bauweise und mit günstigeren Rohstoffen als in LIB-Technik. „Es ist paradox, aber die chinesischen Hersteller als Vorreiter zeigen den europäischen Unternehmen, wie sie sich von ihnen unabhängig machen können“, sagt Fichtner.

(agr@ct.de) **ct**

Die Natrium-Ionen-Technik im E10X ist aus zylindrischen Zellen aufgebaut, 32 Millimeter im Durchmesser, jeweils 14 Zentimeter lang.



ISI-Batterietechnik-Roadmap und Exzellenzcluster „Post Lithium Storage“:
ct.de/y377

MACH, WAS WIRKLICH ZÄHLT.

ZIVIL

FOLGE DEINER BERUFUNG.



Jetzt als IT-Projektmanager/in
(m/w/d) in Bonn bewerben!
bundeswehrkarriere.de



BUNDESWEHR

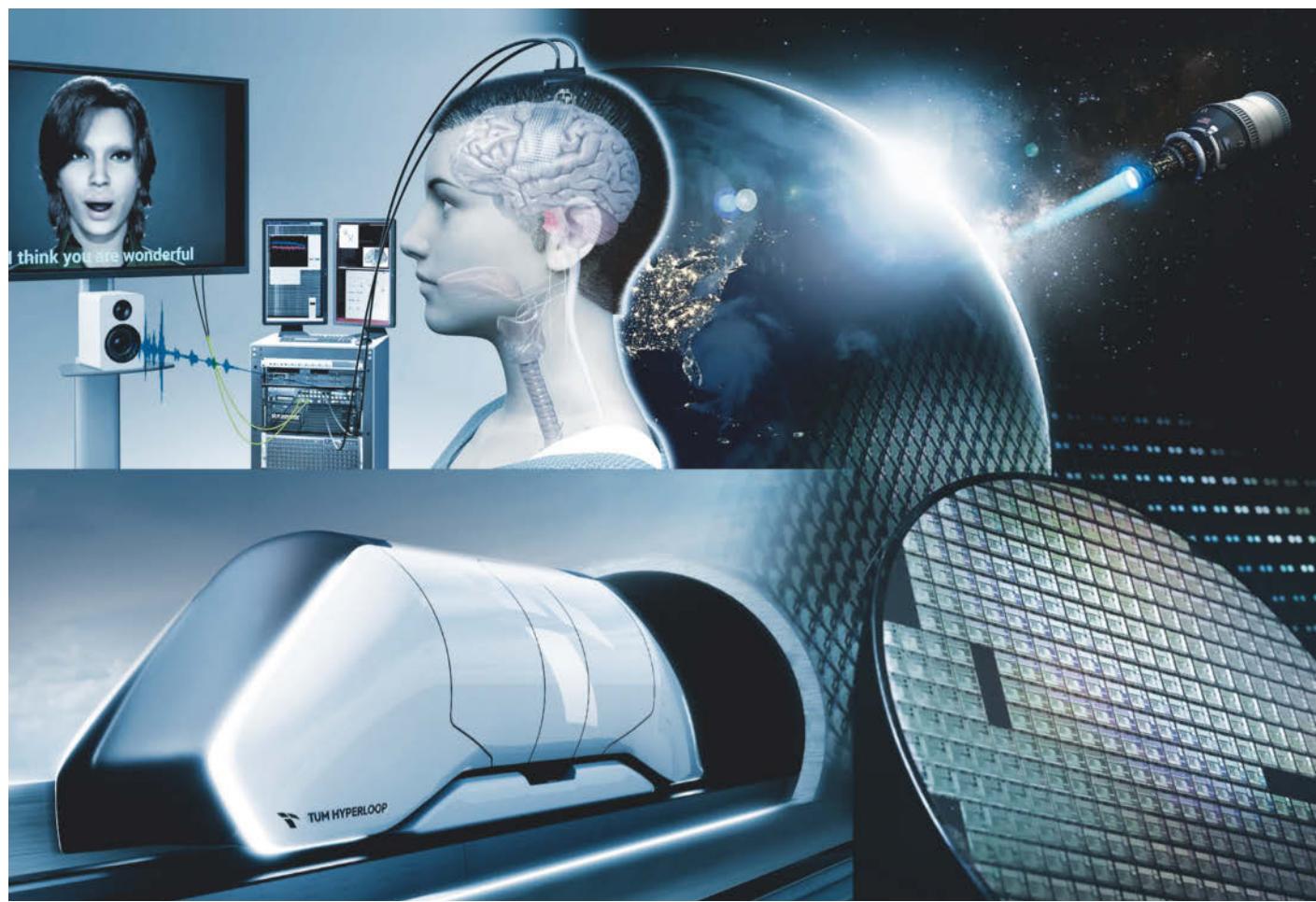


Bild: TU-München, NASA, University of California San Francisco, TSMC | Collage: c't

Die Welt von morgen

Highlights aus Wissenschaft und Forschung

Schon jetzt geben Forscher Einblicke in faszinierende Zukunftsvorprojekte. Nukleare Antriebe befördern Menschen zum Mars, Passagierkapseln sausen in Hyperloop-Röhren statt auf Gleisen dahin. Nicht zuletzt versprechen Hirn-Computer-Schnittstellen blitzschnelle digitale Kommunikation.

Von Arne Grävemeyer

Als das nächste große Fernziel der be- mannten Raumfahrt gilt der Mars, unser direkter äußerer Nachbar im Sonnensystem. Auf seiner Umlaufbahn nähert sich der rote Planet unserer Erde bis auf etwa 56 Millionen Kilometer Entfernung – kein Ver- gleich zu den knapp 400.000 Kilometern, die unsere Heimatwelt vom Mond trennt.

Nach Berechnungen der amerikani- schen Raumfahrtbehörde NASA benötigt eine Rakete mit herkömmlichem chemi- schem Antrieb zwischen 1000 und 4000 Tonnen Treibstoff, also neben Sauerstoff etwa Methangas oder Flüssigwasserstoff, um zum Mars und zurück zu fliegen. Sol- che Mengen an Treibstoff müssten in einem gigantischen und sehr schweren Raumschiff untergebracht werden.

Zudem müsste ein solches Raum- schiff mit seinem Treibstoff haushalten und könnte deshalb die Triebwerke nicht dauerhaft einsetzen. Bei der NASA rech- net man daher für ein chemisch ange- triebenes Raumschiff mit einer Reise- dauer von acht bis neun Monaten bis zum Mars – für die Besatzung bedeutet das eine schwere gesundheitliche und psychische Belastung. Allein der Dauer- beschuss durch kosmische Strahlung kann längerfristig dazu führen, dass die DNA der Raumfahrer permanent be- schädigt wird.

Uran im Tank

Einen Ausweg aus dem Dilemma ver- spricht die Nuklearenergie. Im Januar die-

ses Jahres gab die NASA mit der Forschungsagentur des US-Militärs DARPA (Defense Advanced Research Projects Agency) bekannt, gemeinsam für künftige Mond- und Marsmissionen einen nuklear-thermischen Antrieb zu entwickeln. Dafür kündigten die Amerikaner einen ersten Test bis zum Jahr 2027 unter der Bezeichnung DRACO (Demonstration Rocket for Cislunar Operations) an.

Ein nuklearer Raketenantrieb erzeugt große Hitze wie ein Atomreaktor in einem Kraftwerk: Neutronen werden auf das spaltbare Material gefeuert und stoßen einen Zerfallsprozess an, der wiederum Neutronen mit hoher Energie freisetzt. So entsteht eine Kettenreaktion.

Als Kernbrennstoff plant man im DARPA sogenanntes hochgradig schwach angereichertes Uran (high-assay, low-enriched uranium; HALEU) einzusetzen. Für den Einsatz im All sieht die NASA vor, den Anteil des spaltbaren Isotops Uran-235 auf etwa 20 Prozent anzureichern. Der aktivierte Kernbrennstoff erreicht Temperaturen bis zu 2700 °C und kann ein Triebwerk über Jahrzehnte befeuern.

Für einen kräftigen Schub soll im DRACO-Projekt die große Hitze der Brennstäbe einen hindurchgeleiteten Flüssigkraftstoff explosionsartig in ein sich ausdehnendes Gas umwandeln. Nach derzeitigen Plänen wird die NASA dafür (ähnlich wie bei klassischen chemischen Antrieben) Wasserstoff einsetzen, der dann über eine Düse am Korpus des Raumschiffs mit hoher Geschwindigkeit austritt. Zusätzlicher Sauerstoff ist dafür nicht erforderlich.

Der nuklearthermische Antrieb verbraucht nach ersten Berechnungen weniger als die Hälfte der Treibstoffmenge, die ein chemischer Antrieb benötigt. Die Planer haben überschlagen, dass eine Rakete damit den Mars drei- bis fünfmal so schnell erreichen kann wie ein chemisch angetriebenes Raumschiff. Die Flugzeit ließe sich nach ihren Berechnungen von sechs bis neun Monaten auf weniger als zwei bis höchstens drei Monate verkürzen.

Parallel zum DRACO-Projekt gibt es einen Ansatz von Ryan Gosse an der University of Florida in Gainsville. Sein Team will einen bimodularen Nuklearantrieb aufbauen, der die Vorteile von Plasma- und Ionenantrieb kombiniert. Bei diesem Ansatz nutzt ein Wellenmotor den beim Erwärmen entstehenden Druck, um den eingesetzten Treibstoff zu verdichten und gleichzeitig Strom zu erzeugen, der einen

ct kompakt

- Reisen im Vakuum: Mit über 900 km/h sollen Passagierkapseln derinst durch Hyperloop-Röhren flitzen; Raumschiffe mit Nuklearantrieb könnten in nur 45 Tagen den Mars erreichen.
- Erstmals konnten sprechunfähige Patienten mittels Brain-Computer-Interfaces die eigenen Worte in Dialoggeschwindigkeit hören lassen.
- Bochumer Forscher haben gezeigt, wie man Hardwaretrojaner auf manipulierten Chips sicher entlarven kann.

Ionenantrieb speist. Ionenantriebe gehen sehr sparsam mit ihrem Treibstoff um, da sie einzelne Ionen im elektrischen Feld nahezu beliebig hoch beschleunigen können. Das Team um Gosse hat für seinen Kombiantrieb einen im Vergleich zu chemischen Raketenantrieben doppelt so hohen Impuls errechnet. Demnach könnten Astronauten mit diesem Antrieb in gerade einmal 45 Tagen zum Mars fliegen.

Gefahren im All

Doch ein nuklearer Antrieb birgt selbst im Weltall Gefahren. Zusätzlich zum Risiko durch Strahlung wird den zuständigen Ingenieuren vor allem das Problem der immensen Hitze zu schaffen machen, welche die Kernspaltung freisetzt. Sollten mehrere Brennstäbe nötig sein, müssen diese

selbst bei Erschütterungen, wie sie etwa beim Raketenstart auftreten, sicher voneinander getrennt bleiben, um eine Kernschmelze zu vermeiden. Vor allem in der Startphase könnte eine Kernschmelze nicht nur den Verlust des Raumschiffs bedeuten, sondern überdies das angereicherte Uran über weite Landstriche auf der Erde verteilen.

NASA und DARPA wollen das Risiko eines nuklearen Unfalls durch das DRACO-Programm auch dadurch einräumen, dass der Reaktor erst im Orbit aktiviert wird – eine konventionelle Rakete müsste das nuklear bestückte Raumschiff zunächst dorthin transportieren. Draußen im All, weit außerhalb der Erdatmosphäre, müssen die Konstrukteure nur noch die Astronauten vor der radioaktiven Neutronenstrahlung des Antriebs beschützen. Schon deswegen ist in allen Entwürfen das Raumfahrerhabitat möglichst weit vom Nuklearantrieb entfernt platziert.

Passagiere als Rohrpost

Interplanetar oder terrestrisch: Überall kommt es auf möglichst hohe Reiseschwindigkeiten an. Wo heute schienengestützte Hochleistungszüge bereits als schnell gelten, dürfte das Konzept des Hyperloop völlig neue Maßstäbe setzen. In Röhren und bei starkem Unterdruck könnten Passagierkapseln derinst auf über 900 Kilometer pro Stunde beschleunigen und dabei gegenüber heutigen Bahnen noch Energie sparen.

Am 26. Juli dieses Jahres hat die TU München (TUM) in Ottobrunn eine erste für den Passagierbetrieb zertifizierte Test-

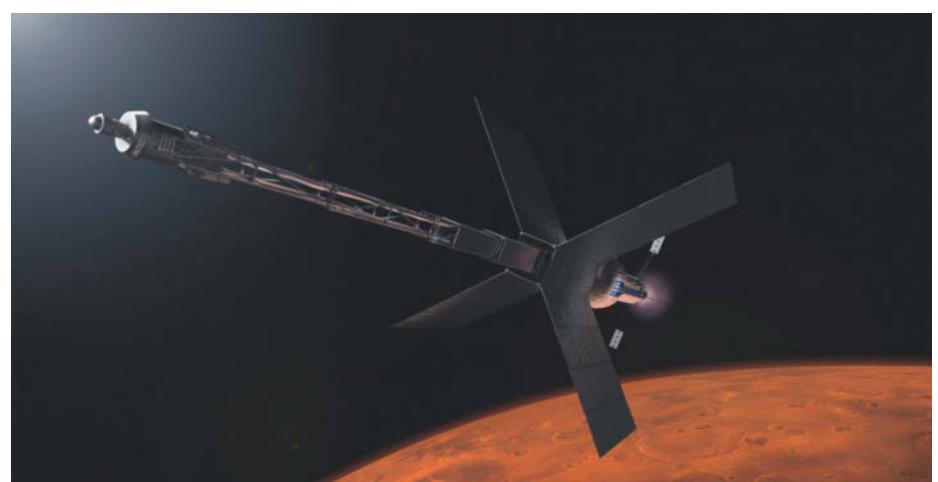


Bild: NASA

Nuklearantriebe gehen sparsam mit Treibstoff um. Ein derart angetriebenes Raumschiff – hier eine Designstudie – muss allerdings das Habitat der menschlichen Crew gut vor der Strahlung des Reaktorkerns am hinteren Ende abschirmen.



Die Hyperloop-Passagierkapsel wird in die Teströhre der TUM eingesetzt. Bereits im Juli sind damit zwei Passagiere im extremen Unterdruck der Röhre bewegt worden.

strecke eröffnet. Allerdings zeigt diese Röhre schon aufgrund ihrer Maße ganz klar, dass es sich noch um ein Forschungsprojekt handelt: Sie ist lediglich 24 Meter lang. Aber immerhin hat sie einen Durchmesser von vier Metern und erlaubt Tests mit einer Passagierkapsel in Originalgröße, die im Innenraum eine Höhe bis zu zwei Metern bietet.

Die Münchner entwickeln ein Transportsystem mit Magnetschwebetechnik. Bei dieser Technik steckt der eigentliche Antrieb in Magnetspulen in der Röhre. Beim Münchner System schwebt der Pod zwischen zwei Betonträgern mit Stahlschienen. Mit Elektromagneten zieht sich der Pod von unten an diese Schienen heran, hält aber einen festen Abstand zu ihnen ein. Von außen gesteuert zieht zusätzlich ein Linearmotor die Transportkapsel und beschleunigt sie in die eine oder die andere Richtung. Der Pod gleitet berührungslos durch die Röhre.

Der große Unterschied zur herkömmlichen Magnetschwebetechnik, die einst für den Transrapid entwickelt worden ist, besteht im sehr geringen Luftwiderstand. In der Demonstratorröhre testen die Münchner den Umgang mit Vakuumtechnik, ihr System ist auf einen Betrieb im Unterdruck bei ein bis zehn Millibar ausgelegt. Bereits im Juli haben die Forscher in einem solch extremen Unterdruck eine kurze Testfahrt mit zwei Passagieren durchgeführt.

Mit geringem Abstand folgt den Münchnern ein Forschungsteam um Walter Neu und Thomas Schüning an der

Hochschule Emden/Leer. Auf deren Campus soll bis Mitte 2024 ebenfalls eine Teströhre entstehen, in diesem Fall 26 Meter lang, aus verschweißtem Stahl mit einer Druckschleuse in der Mitte. Der Durchmesser dieses Demonstrators wird allerdings nur 1626 Millimeter betragen; er ist auf Frachtrouleaus etwa von Gitterboxen und Europaletten ausgelegt. Ein Forschungsschwerpunkt dieser Anlage ist das Ein- und Ausschleusen in den Unterdruckbereich.

„Der Personentransport lohnt sich auf lange Sicht mehr, aber bei der Entwicklung sollten wir zunächst weniger sicherheitskritische Cargosysteme beherrschen, bevor wir Passagiersysteme entwickeln“, sagt Schüning im Gespräch mit c't.

420-Meter-Hyperloop

Die Ostfriesen kooperieren mit der niederländischen Hardt Hyperloop, die im Juli mit dem Bau einer 420-Meter-Teströhre in der angrenzenden Provinz Groningen begonnen hat. Das holländische Magnetschwebesystem ist mit Elektromagneten an der Röhrendecke konzipiert, die Pods anziehen und wie Gondeln von oben halten, allerdings auch berührungslos. Dieses System vereinfacht den Bau von Verzweigungen und Weichen.

Über die technisch sinnvolle Höchstgeschwindigkeit eines Hyperloop-Systems herrscht noch keine Einigkeit. Klar ist, dass im Bereich der Schallgeschwindigkeit bei etwa 1200 Kilometer pro Stunde in einer geschlossenen Röhre problematische aero-

dynamische Effekte eintreten. Während man aber an der TU München mit einer Spitzenreisegeschwindigkeit von 900 Kilometer pro Stunde rechnet, was wesentlich größere Kurvenradien als bei heutigen Hochgeschwindigkeitsbahntrassen verlangt, konzipieren die Hardt-Entwickler ihr System mit Blick auf die Flächenverhältnisse in Europa für Geschwindigkeiten zwischen 600 und 800 Kilometer pro Stunde.

Bei den Beschleunigungen würde man sich an den Höchstwerten von Verkehrsflugzeugen orientieren, sagt Gabriele Semino, Projektleiter des Forschungsprogramms TUM Hyperloop. Wie in der Start- oder Landephase wären das also maximal 0,3 g. Wo Reisen ohne Anschnallpflicht möglich sein soll, müsste man sich wie im heutigen Bahnverkehr auf 0,1 g beschränken, also auf ein Zehntel der Erdbeschleunigung.

Das Hyperloop-System ist nicht nur schnell, sondern auch energieeffizient und Anwohner werden wenig dadurch gestört. Das berührungslose Antriebssystem erzeugt fast keinen Abrieb. Von Fahrgeräuschen kann kaum etwas nach außen dringen. Die in der Röhre eingeschlossenen Kapseln haben im annähernden Vakuum zudem nur wenig Luft zu verdrängen, was entstehende Geräusche ohnehin minimiert.

Bei anderen Hochgeschwindigkeitszügen wie dem ICE oder dem Shinkansen in Japan sind über 80 Prozent der aufgewendeten Energie nötig, um den Luftwiderstand zu überwinden. Eine Hyperloop-Kapsel hingegen häuft durch den drastisch verringerten Luftwiderstand in der Röhre nur sehr wenig Verlustenergie an. Beim Abbremsen aus hoher Geschwindigkeit kann die Technik sogar durch Rekuperation Strom zurückgewinnen. Schüning schätzt, dass der Hyperloop auf diese Weise lediglich 20 Prozent des Energieaufwands anderer Bahnsysteme erfordert. Dabei ist der Stromverbrauch der an der Strecke erforderlichen Vakumpumpen bereits eingerechnet.

Gedanken lesen

Einen Geschwindigkeitsrekord ganz anderer Art haben eigens entwickelte KI-Systeme im Zusammenspiel mit implantierten Gehirn-Computer-Schnittstellen (Brain-Computer-Interface, BCI) aufgestellt, wie zwei Veröffentlichungen im August 2023 belegen. Unabhängig voneinander lasen zwei Forscherteams aus



ICH WARTE NICHT AUF UPDATES. ICH PROGRAMMIERE SIE.

**40 %
Rabatt!**



C'T MINIBO PLUS AUF EINEN BLICK:

- 6 Ausgaben als Heft, digital in der App, im Browser und als PDF
 - Inklusive Geschenk nach Wahl
 - Zugriff auf das Artikel-Archiv
 - Im Abo weniger zahlen und mehr lesen

Jetzt bestellen:
ct.de/angebotplus



den Hirnströmen von gelähmten, sprechunfähigen Patientinnen das aus, was diese sagen wollten. Dabei erreichten sie Erkennungsraten von über 60 und sogar über 70 Wörtern pro Minute – ein gesunder Mensch kann etwa 150 Wörter in einer Minute sprechen.

Die eingesetzte Technik erfordert allerdings invasive Eingriffe unter dem geöffneten Schädelknochen. Dabei platzieren Chirurgen Elektrodenfelder direkt im Gehirn, und zwar in jenen Windungen, die bei Bewegungen der Gesichts- und Mundmuskulatur besonders aktiv sind. Dort erfassen die Elektroden wie bei einem herkömmlichen EEG (Elektroenzephalografie) die Verteilung der Hirnströme, nur eben direkt im Organ und dadurch sehr genau.

Eine Gruppe um Edward Chang an der University of California in San Francisco (UCSF) hatte ihrer Probandin, einer Schlaganfallpatientin, ein flexibles Silikonimplantat mit feinen Drähten und 253 Elektroden auf die Gehirnhaut gelegt. Eine vergleichbare Technik entwickelt auch das Unternehmen Neuralink, mit dem der Visionär, Tesla- und SpaceX-Gründer Elon Musk die Kommunikation zwischen Mensch und Maschine radikal verändern will.

Auch das Forscherteam um Francis Willet an der Stanford University ist mit Neuralink verflochten. Es hat einer Patientin mit der nicht heilbaren degenerativen Nervenerkrankung ALS eine feste Platine mit insgesamt 128 Nadelelektroden eingepflanzt, die in die Hirnrinde hineinstechen. Diese Technik ist ein Abkömmling des sogenannten Utah-Arrays: Jenes invasive BCI wurde bereits 1993 patentiert und wird seitdem bei vielen klinischen Studien eingesetzt. Auch diese Elektroden sind sehr nah an den Hirnströmen des Patienten dran, möglicherweise sogar noch etwas näher als die eines flexiblen Implantats auf der Hirnhaut.

KI liest in den Hirnströmen

Deep Learning hat nun große Leistungssprünge bei der Aufgabe ermöglicht, Wörter direkt aus Hirnaktivitäten herauszulesen. Beide Forschergruppen setzten eigens entwickelte Recurrent Neural Networks (RNN) ein, wie die jüngsten Veröffentlichungen in *Nature* im August 2023 belegen. Bei diesen neuronalen Netzen laufen die Eingangssignale nicht nur von einer Neuronenschicht zur nächsten, sondern es existieren auch Verschaltungen innerhalb einer Neuronenschicht und zu vorangegangenen Schichten. Durch diese

Architektur haben RNN die Möglichkeit, zeitliche Abläufe in Signalen zu berücksichtigen.

Um die neuronalen Netze auf die individuellen Hirnströme trainieren zu können, mussten die Patientinnen zunächst viele Stunden lang Vokabeln lesen, wobei die Wissenschaftler die Signatur der EEG-Signale aufnahmen. Diese EEGs repräsentierten also die Hirnaktivität, wenn eine Testperson an bestimmte Wörter dachte. Die EEG-Wort-Paare bilden dann das Trainingsmaterial, mit dem das RNN angeleert wird. Wenn sie schließlich versucht, Wörter zu bilden, erkennt das neuronale Netz die Muster wieder und kann die Wörter zuordnen. Der große Sprung nach vorn besteht in der Geschwindigkeit, mit der das nun möglich ist. Stanford-Forscher hatten 2021 einen Rekord mit 18 Wörtern pro Minute aufgestellt, die heutige Technik arbeitet bis zu viermal so schnell.

Der ALS-Patientin in Stanford gelang es, bis zu 62 Wörter in der Minute über ihr BCI an die KI zu übermitteln. Wenn sie sich dabei auf ein einfaches Vokabular mit 50 Wörtern beschränkte, wies die Erkennung laut Studie eine Fehlerquote unter zehn Prozent auf. Aber selbst in einem 125.000-Wörter-Vokabular erkannte das System mehr als drei von vier Wörtern korrekt: Die Fehlerquote lag bei 23,8 Prozent.

Das Team um Chang an der UCSF ging etwas anders an das Problem heran. Es trainierte sein RNN darauf, Wortlaute – die sogenannten Phoneme – zu erkennen und daraus Wörter zusammenzusetzen. Damit stand der Schlaganfallpatientin, mit der die Chang-Gruppe arbeitete, ein Wortschatz von 1024 Begriffen zur Verfügung. Wenn die Frau sich vorstellte, mit diesem Vokabular zu sprechen, gelang es der KI, bis zu 78 Wörter in der Minute zu erfassen. Die Fehlerrate bei der Erken-

nung der gedachten Wörter lag der Studie zufolge bei 25 Prozent.

Als zusätzlichen Clou begannen die Forscher an der UCSF, aus den Hirnströmen ihrer Patientin auch Gesichtsausdrücke auszulesen. Die erkannten Wörter und Sätze gaben sie nicht nur als Text aus. Stattdessen schufen sie einen Avatar, der die Sätze der Patientin möglichst mit der passenden Mimik ausspricht. In weiteren Studien wollen die Forscher den Wortschatz ihres Systems ausbauen und dem Avatar sogar die ursprüngliche Stimme ihrer Patientin verleihen.

Angesichts der schnellen Fortschritte, die die BCI-Forschung heute zeigt, kann man ins Grübeln kommen, wie lange die Menschheit wohl noch mit zwei Daumen Nachrichten in Handys eintippen wird. Wäre es nicht praktischer, mit digitalen Medien direkt im Gedankentempo zu kommunizieren? Heute ist Neuralink der stärkste Entwicklungstreiber, wenn es darum geht, Menschen durch Implantate einen direkteren Computerzugang zu eröffnen.

Diese Anwendungen sind durch medizinische Forschungen mit schwerwiegend beeinträchtigten Patienten seit Jahren etabliert. Neuralink setzt auf Implantationsmethoden, die angeblich verträglicher sind und bei denen man von außen keine großen Wunden mehr wahrnehmen soll. Zudem sieht das Unternehmen einen Elektroden-Patch mit über 1000 Kontakten vor. Die US-Zulassungsbehörde FDA hat in diesem Jahr klinische Tests mit Neuralink-Implantaten genehmigt.

Aber BCI sind entwickelt worden für Menschen, die sonst keine Kommunikationswege haben. Momentan sind diese Schnittstellen noch nicht so leistungsfähig, dass es sich lohnt, sie bei einem ge-

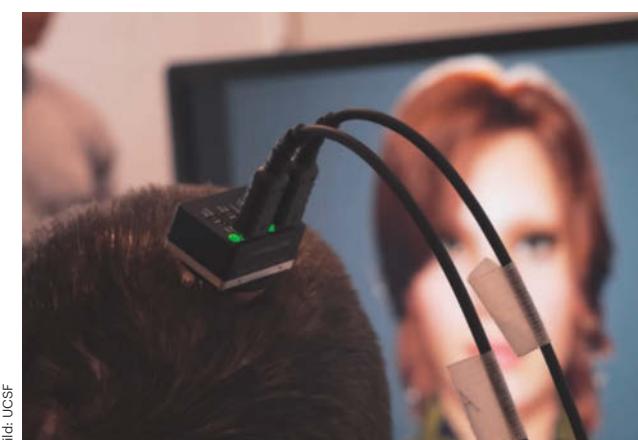


Bild: UCSF

Die unter der Schädeldecke eingepflanzte Gehirn-Computer-Schnittstelle einer sprechunfähigen Schlaganfallpatientin ist per Kabel mit dem Computer verbunden. Ein individuell trainiertes neuronales Netz liest Wörter, die nur gedacht sind, aus den Hirnströmen ab und legt sie einem Avatar in den Mund.



WIR TEILEN KEIN HALBWISSEN WIR SCHAFFEN FACHWISSEN

SECURITY
CHECK

20. - 21.11.



Dienste mit SELinux absichern

SELinux einfach abzuschalten, wenn es Probleme gibt, ist üblich, aber unklug. Der Workshop zeigt, wie man das System stattdessen so nutzt, dass alles besser abgesichert ist und trotzdem funktioniert.

21.11.



Fantastische Bildwelten mit KI

Das zweistündige Webinar bietet praktische Einblicke in die Nutzung des KI-Bildgenerators Midjourney und gibt Tipps zum richtigen Formulieren von Prompts.

23.11.



Einführung in GitLab

Der Workshop bietet einen Einstieg in den Betrieb einer eigenen GitLab-Instanz. Sie lernen GitLab initial aufzusetzen, sowie Ihre Instanz zu konfigurieren und an eigene Anforderungen anzupassen.

28. - 29.11.



Docker und Container in der Praxis

Der Workshop für Entwickler und Administrierende behandelt neben theoretischem Wissen über Container auch Herausforderungen im Alltag und eigene Container-Erfahrungen auf der Kommandozeile.

30.11. + 07.12



CI/CD mit GitLab

Der zweitägige Workshop bietet eine praktische Einführung in die GitLab-CI-Tools und zeigt, wie man damit Softwareprojekte baut, testet und veröffentlicht.

5.12.



ChatGPT, Midjourney & Co. – Rechtliche Aspekte beim Einsatz von KI-Generatoren im beruflichen Umfeld

Wir erklären die bestehende Rechtslage und ihre Auswirkungen auf den beruflichen Alltag.

Sichern Sie sich Ihren Frühbucher-Rabatt:
heise.de/ct/Events

sunden Menschen einzusetzen. Es gebe immer das Risiko, das Gehirn dabei zu verletzen, warnen Mediziner. Außerdem stellen die offenen Drähte am Kopf jederzeit ein Infektionsrisiko dar.

Es droht noch eine weitere Gefahr: Zusätzlich zur Hirnaktivität kann eine leistungsfähige BCI auch Biosignale aufzeichnen, beispielsweise Pupillenreaktionen, Herzschlag oder verschiedene Stresssymptome. Daraus könnte ein Computermodell in Form eines digitalen Zwillings entstehen, das über die Denkvorgänge auch Einblick in den Zustand einer Person gibt. Daraus könnte eine künstliche Intelligenz ablesen, was einem Menschen gefällt und was nicht. Ein Privatunternehmen, dass seinen Kunden BCI in die Köpfe pflanzt, würde so mehr über sie erfahren, als die Kunden wissentlich verraten wollen.

Öffentlicher Zugang zum Webwissen

Gegen zu viel Marktmacht und damit zu viel Wissen über jeden einzelnen Webnutzer wendet sich die Open Search Foundation, die zu ihrem fünfjährigen Gründungsjubiläum am 29. September erstmals den #FreeWebSearch-Day ausgerufen hat.

Was könnte man mit einem großen Web-Index alles machen, wenn dieser öffentlich frei verfügbar wäre? Man könnte alternative Suchmaschinen aufbauen oder fachlich tiefe Suchdienste etablieren, die auf ausgewählte Themen spezialisiert sind. Anwender hätten die freie Wahl und könnten ihre privaten Nutzerprofile besser schützen. Sprachforscher könnten in dem

Datenpool eines großen Web-Index verfolgen, wie sich unsere Sprache entwickelt, und Soziologen könnten beobachten, wie wir in den sozialen Medien miteinander umgehen. Webdienste könnten darin Hinweise auf beginnende Pandemien oder andere Katastrophenfälle suchen und damit ein Frühwarnsystem aufbauen.

Ein umfangreicher, auf europäische Quellen fokussierter Index würde zudem die Chance eröffnen, neue Sprachmodelle zu entwickeln. Davon könnten insbesondere wenig gesprochene Sprachen oder ausgewählte Sprachkombinationen wie etwa Tschechisch und Slowakisch profitieren, die bei den global ausgerichteten US-Diensten leicht durchs Raster fallen.

Das sieht auch die Europäische Kommission so, die zudem mehr digitale Souveränität anstrebt. Im September 2022 startete das EU-Projekt OpenWebSearch. Das Ziel besteht im Aufbau eines öffentlich zugänglichen Open Web Index (OWI) und in einer nachhaltigen, also auf Dauerbetrieb angelegten Infrastruktur. Bis 2025 soll ein Web-Index entstehen, der immerhin die Hälfte aller im Internet veröffentlichten Texte verzeichnet. Die beteiligten Partner rechnen dafür mit einem Speicherbedarf von etwa 5 Petabyte (5 Millionen Gigabyte). Im Vergleich zu den Indizes von Google oder Bing wäre das zunächst ein vergleichsweise kleiner Datenpool. Die etablierte Konkurrenz kommt mit Texten, Bilddateien, Multimedia, Nutzungsdaten und Logfiles aus dem Internet auf jeweils Hunderte von Petabyte.

„Wir sind kein europäisches Google“, sagt Michael Granitzer, Inhaber des Lehr-

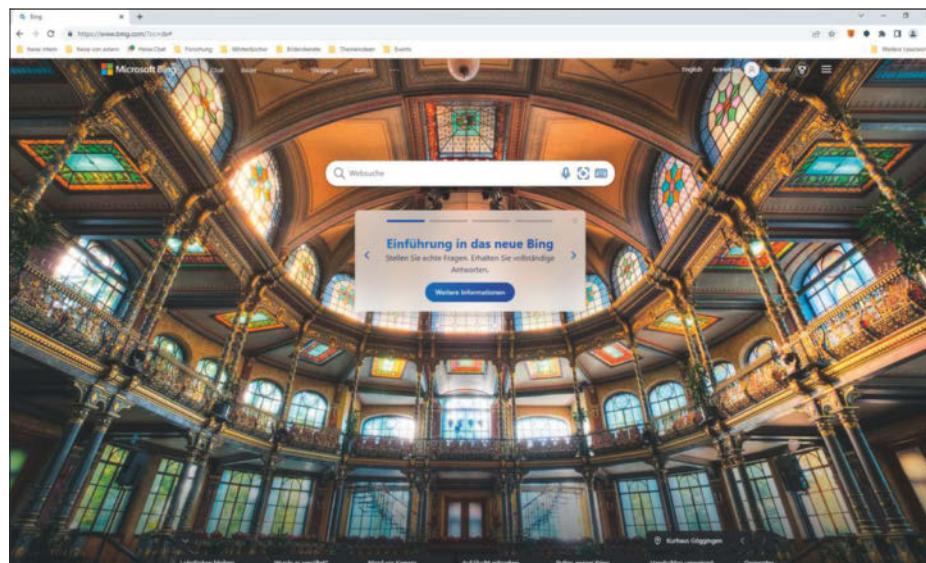
stuhls für Data Science an der Universität Passau, der das OpenWebSearch-Projekt koordiniert. Es gehe dabei nicht um den Aufbau einer großen Suchmaschine, sondern viel grundlegender darum, eine Infrastruktur zu etablieren, mit der später Suchmaschinen und andere Dienste arbeiten können. Googles Größe sei zunächst sicherlich unerreichbar. „Es wird eher wie bei Wikipedia sein, die im Vergleich zu großen Verlagen zunächst mit einem kleinen Kern startete und dann kontinuierlich wuchs.“

14 Projektpartner entwickeln Crawling-Techniken, wählen Metadaten aus, die der Index zusätzlich aufnehmen soll, und konzipieren eine dezentrale Aufteilung des OWI auf verschiedene Server und Standorte in Europa. So beteiligen sich etwa Infrastrukturpartner wie das Leibniz-Rechenzentrum in München, das CSC in Espoo, Finnland, das Europas größten Supercomputer betreibt, das tschechische National Supercomputing Center IT4Innovations sowie das CERN bei Genf.

Das Ende von SEO?

Langfristig könnte ein öffentlicher, transparent aufgebauter Web-Index die SEO-Landschaft komplett verändern. Heute versucht fast jeder Betreiber, seine Webseite für das Ranking in der Google-Suche zu optimieren. Wenn anstelle dieses Monopols eine Vielzahl von spezialisierten Suchdiensten treten sollte, würde sich die Situation grundlegend ändern. „Aus meiner Sicht sollten sich die Webseiten-Betreiber und die Web-User auf die Inhalts erstellung konzentrieren“, sagt Christian Gütl, Leiter des Cognitive and Digital Science Lab an der TU Graz. Wenn die Websuche nicht mehr durch einen Monopolisten beherrscht werde, komme es nur noch darauf an, Inhalte möglichst strukturiert aufzubereiten und mit aussagekräftigen Metadaten zu unterstützen.

Welche Suchfaktoren der neue Web-Index aufnehmen soll, ist noch nicht entschieden. Die Partner diskutieren zudem darüber, wie sie etwa Inhaltsqualität, Genres und Verlinkungen für das Ergebnisranking gewichten wollen. Forschungsprojekte untersuchen Möglichkeiten der Georeferenzierung, die Webdienste für einzelne Städte oder Gemeinden unterstützen soll. Allerdings erfordert die räumliche Einordnung von Weblinks zunächst eine semantische Analyse, denn schließlich ist Paris Hilton etwas anderes als das Hilton in Paris.



Suchmaschinen wie Bing oder Google beherrschen den Markt und behindern dadurch kleinere Dienste, die Websuche mit neuen Ideen weiterzuentwickeln.

Ein sehr wichtiger Unterschied zu den heutigen großen Indizes: User-Clicks und weitere Nutzeranalysen finden keinen Eingang in den OWI. Das können sie auch kaum, denn die Projektpartner planen nicht, selbst Suchmaschinen zu betreiben und damit Nutzerdaten einzusammeln.

Jagd auf Chip-Trojaner

Ob Netzsuche, Anlagensteuerung, Signalverarbeitung oder Kommunikation, all diese Schaltstellen des digitalen Alltagslebens sind auf Prozessoren und andere Mikrochips angewiesen. Die entstehen heute im arbeitsteiligen Prozess zwischen internationalen Partnern. Wie können Entwickler aber sicherstellen, dass nicht jemand in ihre Chipdesigns Trojaner oder Hintertüren einschleust? Ein Angreifer könnte das Design auf dem Weg zum Auftragsfertiger abfangen oder es direkt in der dortigen Produktion manipulieren.

In einem gemeinsamen Projekt hat eine Gruppe von Experten der Ruhr-Universität Bochum (RUB) und des ebenfalls in Bochum ansässigen Max-Planck-Instituts für Sicherheit und Privatsphäre (MPI-SP) frisch gefertigte Mikrochips mit dem Rasterelektronenmikroskop untersucht. Das Ziel bestand darin, Manipulationen im Vergleich mit den ursprünglichen Bauplänen aufzuspüren.

Die Forscher um Steffen Becker an der RUB zeigten damit, dass Hardware-Trojaner als Abweichung des gefertigten Chips vom ursprünglichen Layout nachweisbar sind.

Das Team aus Bochum untersucht Chips buchstäblich auf einer anderen Ebene. Um an die unterste Schicht der Transistoren, Dioden und Kondensatoren (Front End of Line, FEOL) direkt auf dem Siliziumchip heranzukommen, frästen sie zunächst den größten Teil der zugrundeliegenden Siliziumplatte weg und lösten den Rest chemisch auf. Anstatt also sämtliche höheren Schichten der Kontaktierung und metallischen Verbindungen des Chips nacheinander abzutragen, das sogenannte Delayering, arbeiteten sich die Wissenschaftler von unten durch das Substratmaterial.

Nach dem Präparieren der Unterseite konnten die Forscher die Schicht der aktiven Elemente auf dem Chip mit dem Rasterelektronenmikroskop (REM) aufnehmen. Für eine Chipfläche von etwa 1,5 Quadratmillimetern benötigten sie ungefähr 4000 Aufnahmen. Anschließend bearbeiteten sie die REM-Aufnahmen und

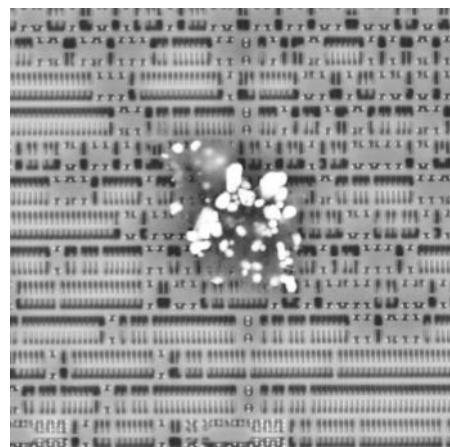


Bild: RUB

Ein Staubkorn auf einem Mikrochip unter dem Rasterelektronenmikroskop: Der Analyst erkennt schnell, dass er es hier nicht mit einem Hardware-Trojaner zu tun hat.

beseitigten zunächst störende Effekte wie etwa Rotationen oder sie glichen die Schieflagen nicht planarer Bereiche optisch aus. Ein Automatismus ordnete die Aufnahmen nebeneinander an und filterte die Überlappungen heraus. Danach konnte der Vergleich mit dem Chipdesign beginnen.

Der Bauplan verrät leider nicht, wie die unterschiedlichen Zelltypen auf dem Chip physikalisch aussehen. So muss also jemand für jeden Typ wie etwa Flipflop, NAND- und NOR-Gatter zunächst ein Referenzmodell auf den REM-Aufnahmen heraussuchen. Jedes dieser Beispiele stellt dann ein sogenanntes Golden Model dar, mit dem jedes weitere Auftreten des entsprechenden Zelltyps abgeglichen wird. Weist das Template Matching eine deutliche Abweichung aus, gilt das als Modifikation.

Echte Manipulationen versteckt

Dass es tatsächlich Manipulationen zu entdecken gab, dafür hatte Thorben Moos von der RUB als „Red Team“ gesorgt. Er veränderte nachträglich das Chipdesign leicht, fügte einzelne Zellen hinzu, tauschte da und dort ein NAND-gegen ein NOR-Gatter und Ähnliches. Auf vier Mikrochips mit insgesamt etwa 3,4 Millionen Zellen hatte er in den Bauplänen 40 unauffällige Manipulationen versteckt, eine Quote von wenig mehr als 0,001 Prozent. Chipdesign und gefertigte Chips stimmten also nicht überein.

Konkret lagen dem „Blue Team“ vier Mikrochips vor, gefertigt in verschiedenen

Technologien mit Strukturgrößen von 90, 65, 40 und 28 Nanometern. Bei den Chipgrößen von 90 bis 40 Nanometern entdeckte der Matching-Algorithmus zuverlässig alle Veränderungen. Im realen Fall hätte man jeden dieser Chips danach als manipuliert klassifiziert.

Gleichzeitig ermittelte der Algorithmus allerdings 500 False Positives. Dabei handelte es sich um Staubkörner oder Rückstände von der Präparation des Chips. Hinzu kamen Bildaufnahmeschwächen wie Verzerrungen, Rauschen oder zu wenig Kontrast. Ein Mensch kann diese Fälle relativ schnell in einer nachgeschalteten Analyse der Fehlermeldungen erkennen, das kostete die Beteiligten pro Chip etwa eine Stunde.

Bei sehr ähnlichen Zellen in der 28-Nanometer-Technik schwächelte das Verfahren allerdings: Drei subtile Veränderungen entdeckte der Algorithmus nicht. Bei diesen Strukturgrößen zeigten die Aufnahmen der Zellen in der verfügbaren Auflösung einfach zu wenige Details.

Die Forscher vermuten, dass verbesserte Analysealgorithmen und insbesondere Deep-Learning-Algorithmen noch bessere Ergebnisse liefern können, um auch die letzten False Negatives noch zu detektieren. In weiteren Versuchen planen sie, die Chipprüfung auf kleinere Strukturgrößen auszuweiten. Allerdings steht zu befürchten, dass REM-Aufnahmen mit 4,8 Nanometer pro Pixel dann einfach nicht mehr genug Auflösung für die Detektionsalgorithmen bieten können. Kommende REM-Generationen mit einer höheren Bildauflösung können hier weiterhelfen – letztlich genau die Technik, die auf der Seite der Lithografiesysteme für die Halbleiterindustrie die Entwicklungssprünge zu immer kleineren Strukturen ermöglicht.

(agr@ct.de) **ct**

Literatur

- [1] Arne Grävemeyer, Johannes Streeck, Uran im Tank, Für Marsmissionen tüftelt die NASA am nuklearthermischen Antrieb, c't 13/2023, S. 120
- [2] Arne Grävemeyer, Passagiere als Rohrpost, Hyperloop verspricht schnellste und nachhaltige Züge, c't 5/2023, S. 140
- [3] Arne Grävemeyer, Gedankensteuerung, Hirnströme können Prothesen bewegen und Texte diktieren, c't 22/2023, S. 122
- [4] Arne Grävemeyer, Basis für tausend Suchmaschinen, Die EU will bis 2025 einen öffentlichen Web-Index aufbauen, c't 9/2023, S. 122
- [5] Arne Grävemeyer, Strategien gegen Trojaner, Der Kampf gegen Chip-Manipulationen, c't 17/2023, S. 126

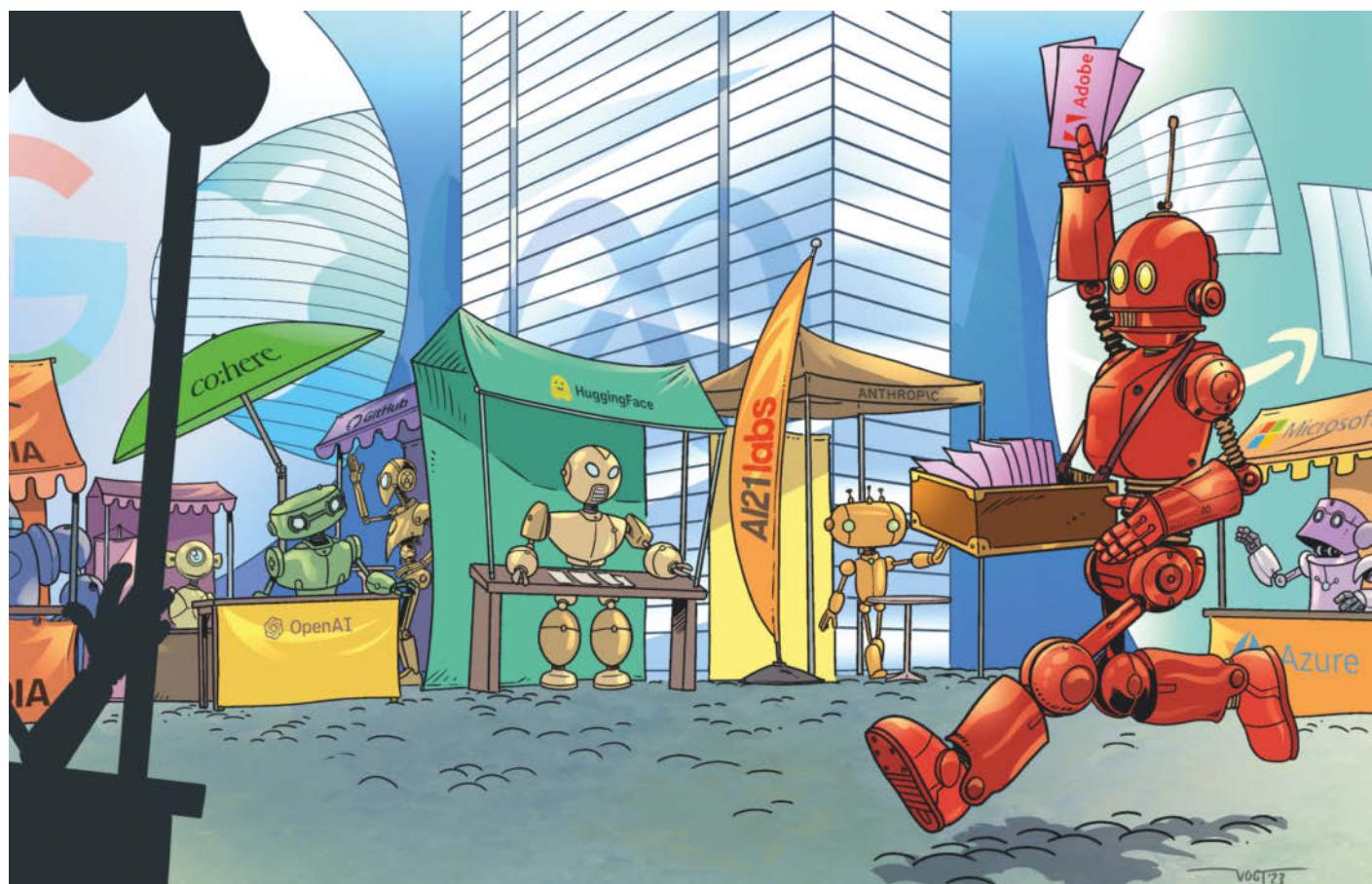


Bild: Michael Vogt

Die neuen Macher

Generative KI wirbelt den IT-Markt und die Wirtschaft durcheinander

ChatGPT, Midjourney & Co. verändern alles. Plötzlich stehen bislang unbekannte Unternehmen wie Hugging Face, OpenAI und Perplexity im Fokus der IT-Welt, und Branchenriesen wie Google und Amazon geraten unter Druck. Über eine Branche, die einen rasanten Umbruch durchläuft – und dabei die gesamte Wirtschaft umkrepelt.

Von Jo Bager

Microsoft stellt Word, Excel oder Outlook einen Copiloten an die Seite; ChatGPT kann hören, sehen und sprechen; Stability AI veröffentlicht zwei Sprachmodelle mit dem Namen FreeWilly auf Basis von Metas LLaMA-Modellen: In diesem Jahr wurde man von solchen Meldungen rund um KI förmlich überrollt. Die Masse solcher Einzelnachrichten verwirrt allerdings: Was bedeutet die Welle neuer generativer KI-Anwendungen ganz konkret für die IT-Welt und die Wirtschaft insgesamt?

Dieser Artikel beleuchtet am Beispiel der Internetsuche und des Journalismus, wie stark die IT-Welt mit anderen Branchen verzahnt ist, welche neuen Anwen-

dungen generative KI bereithält und wie sie Wertschöpfungsketten umwirft. Er gibt außerdem einen Überblick der IT-Welt in Zeiten generativer KI und wagt einen Blick in die Zukunft.

Google ist ein Riese in der IT-Branche und ein Vorreiter im Bereich KI-Forschung. Das „T“ in ChatGPT zum Beispiel steht für „Transformer“, eine bestimmte Architektur neuronaler Netze – die bei Google entwickelt wurde. Von Google sollte man also am wenigsten erwarten, dass es der Entwicklung bei generativer KI einmal hinterherläuft.

Dennoch: Am 7. Februar 2023 dürften die Alarmglocken bei dem Suchmaschinenkonzern geschrillt haben. An diesem Tag brachte Microsoft seinen KI-Copiloten für die Suchmaschine Bing und den Browser Edge heraus. Er setzt auf die Technik von OpenAI auf, das seinen eigenen Chatbot ChatGPT im November zuvor veröffentlicht hatte.

Suchmaschinen funktionieren seit über zwanzig Jahren nach dem gleichen Prinzip: Auf ein paar in den Suchschlitz eingetippte Begriffe hin stellen sie eine Liste von Links zusammen, von denen sie annehmen, dass diese für die Anfrage re-

levant sind. Rückfragen? Gibt es nicht. Wenn die Suchanfrage nicht die gewünschten Ergebnisse liefert, muss der Anwender neu ansetzen.

Das neue Bing jedoch schüttet den Anwender nicht mit Links zu, die er selbst abgrasen muss. Es sucht wie Google eine Reihe von Sites, die zur Anfrage passen. Diese ruft Bing dann auf, extrahiert die Inhalte und baut daraus eine Antwort zusammen. Auf diese Weise kann der Chatbot auch komplexe Fragen beantworten. Und es ermöglicht dem Anwender, nachzuhaken sowie Rück- oder ergänzende Fragen zu den Ergebnissen zu stellen.

Wir haben in mehreren Artikeln die Probleme mit Sprachmodellen im Allgemeinen und mit Suchdiensten im Besonderen behandelt: Sprachmodelle erfinden Dinge, die es nicht gibt. Der verlässlichste Suchdienst, den wir in [1] untersucht hatten, lieferte nur bei 80 Prozent der Testfragen eine richtige Antwort. Und auch die Microsoft-Dienste produzierten und produzieren bis heute viel Quatsch.

Angriff auf Googles Kern

Nichtsdestotrotz war die Veröffentlichung von Bing Chat ein Schock für Google. Denn die Suchmaschine ist der Kern von Googles Betriebssystem. Wie wichtig die Suche für Google ist, lässt sich schon daran ermessen, dass das Unternehmen Browserherstellern wie Apple und Mozilla viele Milliarden Dollar zahlt, damit diese Google als Standardsuche einrichten.

Diese Investition zahlt sich aus: Die Suchmaschine verzeichnet einen Jahresumsatz von 162 Milliarden US-Dollar. Mit einem Marktanteil von 80 bis 90 Prozent dominiert Google dieses Segment. Die Suche spielt aber nicht nur selbst Geld ein, sondern führt Nutzer auch anderen Google-Diensten zu.

Lange Jahre gab es neben Google und Bing keine ernstzunehmenden Player auf dem Feld der Suchmaschinen. Die nächstgrößeren Anbieter, Baidu und Yandex, spielen nur auf dem chinesischen und russischen Markt eine Rolle. Doch jetzt gibt es neue Unternehmen, die Google auf seinem Kernmarkt angreifen. Denn das neue Bing ist nur einer von vielen Diensten, die antreten, mit Sprachmodellen das Thema „Suche“ neu zu erfinden.

In den anderthalb Jahren vor dem Bing-Release haben sich, von der breiten Öffentlichkeit unbemerkt, weitere neue Anbieter auf den Markt getraut. Dazu zählen Kagi, Neeva, perplexity, Phind und

You.com. Dank Sprachmodellen bieten sie allesamt neue Bedienoberflächen und Suchfunktionen.

Einige dieser Start-ups vermeiden die direkte Auseinandersetzung mit Google, indem sie Nischen bedienen. Neeva hat sich von dem Cloud- und Analytics-Anbieter Snowflake kaufen lassen und kommt nur noch in dessen Produkten zum Einsatz. Phind wandelt sich derzeit zur Suchmaschine und zu einem Copiloten speziell für Programmierer. Kagi zielt eher auf Kunden, die bereit sind, die Werbefreiheit zu bezahlen. Auch You.com und Perplexity AI unterhalten abgestufte Preismodelle. In ihren kostenlosen Versionen lassen sie aber beliebig viele Suchen zu, weshalb man sie durchaus als Google-Konkurrenten ansehen kann.

Konnte Google die kleinen Konkurrenten noch ignorieren, so machte der Vorstoß des Branchenriesen Microsoft plötzlich die Öffentlichkeit auf die neuen Suchdienste aufmerksam. Google musste reagieren. Das hat einige Zeit gedauert, aber mittlerweile ist Googles Chatbot Bard auch hierzulande verfügbar.

Die Wertschöpfungskette bröckelt

Aber damit ist das Thema noch lange nicht durch, denn eine Chatbot-Bedienoberfläche ist wenig kompatibel mit dem Geschäftsmodell der Google-Suche. Google finanziert sich durch Werbelinks, die es seinen organischen Ergebnissen voran- oder hintanstellt und die von diesen kaum

c't kompakt

- Die plötzliche Verfügbarkeit von generativer KI, insbesondere von ChatGPT, hat selbst dominierende IT-Unternehmen wie Google unter Druck gesetzt.
- Unternehmen aus allen Branchen tun gut daran, sich mit dem Thema generative KI zu befassen.
- Insbesondere Open-Source-Modelle haben das Potenzial für Anwendungen, bei denen Unternehmen die Hoheit über die eigenen Daten und die Privatsphäre ihrer Kunden behalten.

zu unterscheiden sind. Mit ausformulierten Antworten funktioniert dieses Dark Pattern nicht mehr und die Einnahmequelle versiegt.

Derzeit machen die Ergebnisse, die Suchmaschinen mit Chatbots ausliefern, noch einen verschwindend kleinen Teil aus. Vielleicht wird das auch immer so bleiben, wenn Google, Microsoft & Co. es nicht hinkriegen, die Fehlerrate ihrer Bots zu drücken. Sollten sich die Chatbots bei Suchdiensten allerdings irgendwann durchsetzen, dürfte das Konsequenzen haben, die weit über die Suchdienste hinausreichen. Insbesondere wenn sich die Anwender mehr und mehr auf die fertig

What is Gen-1 and how does it work?

Sources

- Gen-1 by Runway (runwayml · 1)
- Generative AI - What is it and How Does it Work? (nvidia · 2)
- Generative artificial intelligence - Wikipedia (wikipedia · 3)
- Runway Researchers Unveil Gen-1: A New... (reddit · 4)
- GEN1 - New AI That Creates Animations Fro... (youtube · 5)
- Generative AI Models Explained - AltexSoft (altexsoft · 6)

Answer

Gen-1 is a new generative AI model developed by Runway that uses language and images to generate new videos out of existing ones. It is a type of generative AI model that uses foundation models to create custom models with features tailored to its use case. Specifically, Gen-1 realistically and consistently synthesizes new videos by applying the composition and style of an image or text prompt to the structure of the source video. The model uses neural networks to identify patterns and structures within existing data to generate new and original content. Gen-1 is an example of a generative AI model that uses multimodal inputs to generate new content, in this case, videos.

Ask follow-up...

Copilot

Sieht so die Zukunft der Suche aus? Perplexity liefert fertige Texte und die Links zu den Quellen.

The screenshot shows the Google Bard interface. On the left, there's a sidebar with a 'Neuer Chat' button and a list of 'Letzte Themen' including 'Ballettdirektor besch...', 'Vertrauenswürdigkeit ...', 'Werbung in Bard?', and 'Mehr anzeigen'. The main area has a search bar with the query 'Was hat es mit der Hundekotaffäre auf sich?'. Below the search bar, a summary box says: 'Die Hundekotaffäre war ein Skandal, der sich im Jahr 2023 in Deutschland ereignete. Der Ballettdirektor der Staatsoper Hannover, Marco Goecke, beschmutzte am 20. Juli 2023 der Kritikerin der Hannoverschen Allgemeinen Zeitung, Silke Bindseil, das Gesicht mit Hundekot.' Below this, a detailed text block continues: 'Goecke hatte Bindseil zuvor in einem Interview kritisiert, in dem sie seine Arbeit als "uninspiriert" bezeichnet hatte. Am Abend der Premiere seines neuen Balletts "Der Sandmann" in der Staatsoper Hannover kam es dann zum Eklat. Goecke traf Bindseil im Foyer der Oper und warf ihr eine Tüte mit Hundekot ins Gesicht.' At the bottom, there's a 'Prompt hier eingeben' input field and a microphone icon.

Diese Antwort von Googles Bard zeigt sehr anschaulich das Problem von Sprachmodellen mit den sogenannten Halluzinationen: Weder das Datum noch der Name der Kritikerin stimmen.

ausformulierten Ergebnisse verlassen und die Websites, von denen die Informationen stammen, nicht mehr aufsuchen, wird es auch für deren Betreiber eng.

Denn um wirtschaftlich arbeiten zu können, sind viele Websites abhängig von dem Strom an Besuchern, die Google & Co. ihnen zuführen. Die Befürchtung, dass Chat Bing und Konsorten auch deren Existenz gefährden, ist also nicht ganz abwegig: Zu einem ähnlichen Effekt haben die Infoboxen geführt, die Suchmaschinen bei einigen Anfragen schon seit ein paar Jahren einblenden. Webmaster und die Branche der Suchmaschinenoptimierer diskutieren derzeit, was sie tun müssen, damit ihre Inhalte auch in Zukunft berücksichtigt und verlinkt werden, wenn Chatbots mehr Gewicht in der Suche erhalten.

KI ändert alles

Wenn schon ein so großer Tech-Konzern wie Google derart massiv von der KI-Welle betroffen ist: Wie geht es dann kleineren Unternehmen? Besonders gut lässt sich das am Journalismus festmachen. Journalismus findet heute in vielen Fällen digital statt und wird auf den verschiedensten Ebenen von generativer KI berührt.

So speist sich manche KI aus redaktionellem Content. Man hat lange Zeit einfach alles aus dem Web zusammengezimmert und den Sprachmodellen zum Training vorgeworfen, was man finden konnte. Dabei sind auch die Onlineinhalte von Zeitungen und Zeitschriften in die Trainingskorpora von ChatGPT und anderen Sprachmodellen eingeflossen. Und niemand hat sich daran gestört.

Mit der zunehmenden Popularität von Sprachmodellen wollen Medien nun aber

nicht mehr, dass die KI-Entwickler ihre Inhalte derart ausschlachten – oder wollen zumindest daran mitverdienen. Die New York Times beispielsweise verhandelt mit OpenAI über eine Nutzungsgebühr für ihre Inhalte. Mit der Nachrichtenagentur AP hat OpenAI bereits einen Vertrag über die Verwendung des Textarchivs abgeschlossen.

Der ChatGPT-Betreiber hat mittlerweile technische Maßnahmen vorgesehen, mit denen Medien den OpenAI-Crawler von ihren Sites aussperren können. Und auch Googles Bard und Vertex AI, der Cloudplattform für alle Machine-Learning-Anwendungen Googles, können Webmaster den Zugriff auf ihre Inhalte verbieten.

Generative KI hilft bei der Produktion von journalistischen Inhalten. Manch ein Journalist setzt ChatGPT als Sparringspartner ein, der ihm bei der Ideenfindung und bei der Strukturierung von Texten hilft. Bei der c't nutzen einige Redakteure DeepL Write, um die Verständlichkeit ihrer Texte zu verbessern [2].

c't nutzt auch KI-Generatoren, um Bilder als Aufmacherillustrationen zu erzeugen. Pro Ausgabe entstehen mehrere Aufmacher mit Midjourney oder Stable Diffusion. Bei dem aufwendigen Transkribieren von Interviews und Podcasts hilft allen Redaktionen bei heise seit ein paar Monaten ein lokal installierter Whisper-Server.

Neue Produkte

Die Möglichkeiten gehen aber weit über das Optimieren bestehender Abläufe hinaus. KI ermöglicht es, ganz neue Produkte zu entwerfen oder bestehende Produkte

substanziell zu erweitern. So können Podcaster seit Kurzem ihre Beiträge von Spotify in andere Sprachen übersetzen lassen. Das soll fast vollständig automatisch funktionieren und die geklonte Stimme des Podcasters verwenden. So können sie völlig neue Zielgruppen erschließen.

Die Software des Unternehmens Hey-Gen macht so etwas auch mit Videos: Sie übersetzt nicht nur den gesprochenen Text in eine andere Sprache und klonst dabei die Stimme des Sprechers, sondern passt auch dessen Lippenbewegung an den Text an. Runway ML wiederum fertigt Videos anhand von Textprompts, tauscht Charaktere in Clips aus oder entfernt Objekte in Videos.

Es spricht grundsätzlich nichts dagegen, journalistische Inhalte von KI erzeugen zu lassen. Im Sport- oder Finanzjournalismus beispielsweise finden sich schon seit Längerem Angebote, die automatisiert Spiel- oder Finanzdaten zu Meldungen aufbereiten – die es ohne KI nicht gäbe. Der Einsatz von KI sollte aber für den Leser transparent sein.

In den letzten Monaten gab es allerdings bereits einige „Sündenfälle“, in denen Medien KI-generierte Inhalte nicht sauber gekennzeichnet haben. So hat der Burda-Verlag im Mai eine Extraausgabe des Rezeptemagazins Lisa Kochen & Bäckchen „99 Pasta-Rezepte“ größtenteils mit KI-Technik erstellt, ohne explizit darauf hinzuweisen. So etwas untergräbt die Glaubwürdigkeit des jeweiligen Mediums und des Journalismus im Allgemeinen.

Dass KI sich für Rationalisierungen einsetzen lässt, liegt auf der Hand. Warum soll man eigentlich eine Moderatorin für einen Radiosender bezahlen, wenn man die Moderationstexte auch von einer KI schreiben lassen und von einer anderen KI vorlesen lassen kann? Beim Sender bigGPT des Unternehmens Audiotainment Südwest „arbeitet“ die KI-Moderatorin bigLayla, die genau so funktioniert. Und in einer Mail der „Bild“-Geschäfts-führung anlässlich der Kündigung von 200 Mitarbeitern hieß es, man müsse sich „leider auch von Kollegen trennen, die Aufgaben haben, die in der digitalen Welt durch KI und/oder Prozesse ersetzt werden“. Laut Recherchen der Süddeutschen Zeitung musste die KI hier aber wohl eher als publicityträchtiger Vorwand herhalten: Die genannten Modernisierungsprozesse wie etwa das automatisierte Layout hat das Unternehmen der SZ zufolge noch gar nicht gestartet.

Immer einfacher zu bedienende KI-Werkzeuge eröffnen auch neue Möglichkeiten, Schmu damit zu treiben, etwa Fake-News-Inhalte zu erstellen. Das Foto vom Papst im weißen Designer-Daunenmantel konnte man noch als künstlich generiert erkennen. Aktuelle Versionen der Bildgeneratoren werfen aber derart realistische Bilder aus, dass sie auch ein geübtes Auge nicht mehr von einem Foto unterscheiden kann – eine neue Herausforderung nicht nur für Journalisten. Microsoft will derartigen Missbrauch des in seine Anwendungen integrierten Bildgenerators erschweren. Die erzeugten Bilder erhalten eine Art digitales Wasserzeichen, in dem die Herkunft kodiert ist.

Last, not least sehen sich Journalisten damit konfrontiert, dass Nutzer KI-Hilfsmittel einsetzen können, um ihre Inhalte anders als im ursprünglichen Format zu konsumieren. Der Newsreader Artifact (siehe auch S. 98) zum Beispiel erzeugt automatische Zusammenfassungen von Meldungen und formuliert sie zu reißerischen Schlagzeilen um. Und mit allerhand KI-Tools kann jeder sich YouTube-Videos zusammenfassen oder die Inhalte transkribieren lassen.

KI branchenübergreifend

Es werden wohl nur wenige Wirtschaftszweige so stark und auf so verschiedene Arten von generativer KI betroffen sein wie der Journalismus. Und es wird wohl auch nicht so extrem kommen, wie es Bundesminister Hubertus Heil voraussagt: dass es 2035 keinen Arbeitsplatz mehr geben werde, der nicht mit KI-Anwendungen zu tun hat [3].

Dennoch drängen Unternehmensberatungen Firmen, beim Einsatz von KI nicht abzuwarten, sondern sofort loszulegen. Hierzulande gehen Unternehmen allerdings noch recht verhalten auf die neue Technik zu. Nach einer im September durchgeführten Untersuchung des Branchenverbandes Bitkom haben gerade einmal 15 Prozent der Unternehmen KI bereits im Einsatz, 28 Prozent planen oder diskutieren den Einsatz und für mehr als die Hälfte der Unternehmen spielt KI derzeit keine Rolle.

Sich überhaupt nicht mit KI auseinanderzusetzen zu wollen, ist allerdings nicht ratsam, weil KI die Arbeitnehmer an jeder Ecke förmlich anspringt. In viele beliebte Anwendungen wie Notion, Adobes Creative Cloud sowie Googles Gmail und YouTube sind bereits KI-Funktionen eingezo-

gen. Die gesamte Softwareindustrie inklusive Hunderter Start-ups liefert sich ein Wettrennen, dem Anwender die heißesten, neuesten KI-Funktionen aufzudrängen.

Windows-Nutzer haben schon seit Längerem direkt aus ihrem Edge-Browser Zugriff auf Microsofts Chatbot. Sobald das „Moments 3“ genannte Windows-11-Update (23H2) herauskommt, erhält jeder Windows-Desktop einen KI-Copiloten. Die allgegenwärtige Verfügbarkeit von KI-Diensten könnte Mitarbeiter dazu verleiten, diese nützlichen Helfer einfach mal einzusetzen, etwa um sich beim Texten helfen zu lassen.

Risiko KI-Schatten-IT

Eine solche KI-Schatten-IT birgt aber das Risiko, dass vertrauliche Firmenunterlagen oder Kundendaten bei den KI-Diensten landen. Firmen sind daher gut beraten, KI-Richtlinien für ihre Mitarbeiter vorzugeben. Die soll festlegen, welche Mitarbeiter welche Dienste nutzen dürfen. Dabei sollten sie sehr kritisch auf die Nutzungs- und Datenschutzbedingungen der Anbieter schauen. Nicht jedes jetzt schnell aus dem Boden gestampfte KI-Start-up genügt europäischen Datenschutzstandards [4].

Unternehmen, die mit KI loslegen wollen, stehen ohnehin vor einer sehr unübersichtlichen Auswahl an Nutzungsszenarios und Diensten. Da gibt es zunächst die sogenannten Basismodelle (Foundation Models). Die bekanntesten Beispiele sind OpenAIs GPT-Modelle für Text sowie Midjourney für die Bildgenerierung.

Bei Basismodellen handelt sich um sehr große Modelle, die auf einem breiten Datensatz trainiert wurden und für eine Vielzahl von Anwendungen und nachge-

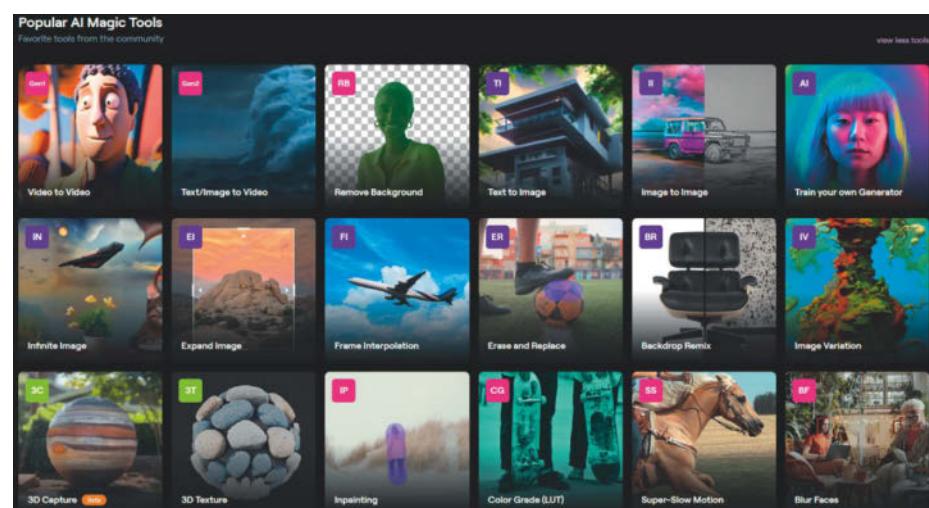
lagerte Aufgaben angepasst und fein abgestimmt werden können. Es gibt Basismodelle für die verschiedensten Modalitäten – Text, Bild, Video, Sprache sowie multimodal – und Anwendungen.

Der Pharmakonzern Moderna zum Beispiel nutzt IBMs Basismodell MoLMFormer-XL zur Entwicklung von mRNA-Medikamenten. Das Unternehmen Writer.ai betreibt eigene Sprachmodelle, unter anderem für Finanzdienstleister. Bei Runway läuft das Modell Gen-2, das zu Text-Prompts Videos generiert. Und der GitHub-Copilot stützt sich auf den OpenAI-Codex, eine modifizierte Version von GPT-3.

Bei den Modellen von OpenAI, Midjourney und Writer.ai handelt es sich um geschlossene Systeme. Weil viel Arbeit und Geld in die Modelle geflossen sind, stellen die Anbieter sie nicht als Open Source zur Verfügung. Sie laufen in der Cloud der Anbieter. Die beiden OpenAI-Modelle lassen sich allerdings sehr einfach per API-Call in bestehende Anwendungen integrieren oder noch einfacher mit einem No-Code-System wie Zapier aufrufen – beziehungsweise direkt über die Weboberfläche bei OpenAI.

Die Bitkom-Studie zum Einsatz von KI hat auch gezeigt, dass viele Unternehmen, die KI-Anwendungen bereits einsetzen oder künftig einsetzen wollen, auf das Herkunftsland ihres KI-Anbieters schauen. Außer den USA kommt für 81 Prozent der befragten Unternehmen nur Deutschland infrage. 74 Prozent können sich einen europäischen KI-Anbieter vorstellen.

Und es gibt durchaus deutsche und europäische Unternehmen, die leistungsfähige eigene KI-Lösungen anbieten. So ist das Kölner Unternehmen DeepL auf Über-



Runway ML bietet eine große Auswahl an neuartigen Bild- und Videofunktionen.

setzungen und Text spezialisiert, Aleph Alpha betreibt ein großes Sprachmodell. Das Onlinemagazin Sifted listet 150 europäische KI-Unternehmen (ct.de/yw4u).

Der weitaus größte Teil der Unternehmen, die unter dem Label „KI“ firmieren, betreibt aber gar keine eigene KI-Infrastruktur. Vielmehr zapfen sie die Modelle von OpenAI oder anderen Cloudanbietern an und verpacken sie in eine hübsche Oberfläche.

Unternehmen sollten also kritisch hinterfragen, welchen Mehrwert ihnen solche Anwendungen bieten. Es ist davon auszugehen, dass dort sehr bald eine Konsolidierung stattfindet – zumal allgemeine KI-Funktionen ohnehin bereits breit verfügbar sind und das Prompting demnächst zum Allgemeinwissen von Angestellten gehören sollte.

Zukunft offene Modelle

Zu den Anbietern geschlossener KI-Modelle und den Diensten, die darauf aufsetzen, ist im Laufe des Jahres noch eine dritte Kategorie hinzugekommen: Open-Source-Modelle. Stable Diffusion zum Beispiel ist ein solches als Open Source verfügbares Modell für die Bildgenerierung, das mit kommerziellen Systemen gut mithält.

Die zentrale Anlaufstelle für Entwickler, die sich mit offenen KI-Modellen auseinandersetzen wollen, ist Hugging Face. Das Repository dort listet mehr als 350.000 Modelle aller Art, darunter auch Tausende Sprachmodelle [5]. Derzeit reichen die Open-Source-Sprachmodelle aber nicht an die Leistungsfähigkeit großer geschlossener Systeme wie GPT-4 heran.

Bei der schnellen Entwicklung auf dem Gebiet sollten Unternehmen dieses

Feld allerdings im Blick behalten, denn der Einsatz offener Modelle bietet eine Reihe von Vorteilen. Wenn ein Unternehmen eine Open-Source-KI unter eigener Regie betreibt, verlassen keine Benutzereingaben und sonstige Daten das Unternehmen.

Ausblick

Nach dem ersten Schock durch die Veröffentlichung von ChatGPT und andere Angebote scheinen die großen IT-Unternehmen nach und nach die alte Hegemonie wiederherzustellen. Mittlerweile hat Microsoft KI-Funktionen nicht nur für Bing, Edge und Windows, sondern quasi für die gesamte Softwareproduktpalette im Repertoire.

Auch Meta durchsetzt seine gesamte Produktpalette derzeit mit KI-Funktionen: KI-generierte Sticker für WhatsApp, Chatbots in Instagram, Messenger und WhatsApp und Code Llama zum Beispiel, das auf Zuruf Programmcode schreibt.

Letztlich könnte die KI-Welle dazu führen, dass die großen IT-Unternehmen ihre Vormachtstellung sogar noch auf weitere Geschäftsfelder ausbauen, sprich: andere Unternehmen dort verdrängen. Googles Werbekunden etwa sollen in den kommenden Monaten den Chatbot Bard zur Erstellung von Werbekampagnen einspannen können. Die generative KI soll dabei „auf Agentur-Niveau“ arbeiten. Das klingt wie eine Kampfansage an eine ganze Branche.

Die Entwicklung beim Thema KI scheint zu belegen, dass das freie Spiel der Kräfte im Softwaremarkt nicht mehr funktioniert. Unternehmen wie Google, Microsoft oder Amazon sind einfach zu groß. Kommt ein neues Thema auf, bei dem einer dieser Riesen blank ist, kauft er sich

einfach die Expertise ein – wie Microsoft schon vor längerer Zeit bei OpenAI oder Amazon und Google nun bei Anthropic: Der Shopping- und Cloudkonzern hat erst kürzlich eine Investition von 4 Milliarden US-Dollar in das Start-up angekündigt. Anthropic entwickelt Claude, einen direkten Konkurrenten zu ChatGPT.

Vielleicht wirken ja Regelwerke wie der Digital Service Act und der Digital Market Act (siehe S. 86) ein Stück weit als Regulativ – oder Gerichte. So läuft gegen Google gerade ein Kartellverfahren: Das US-Justizministerium und fast alle US-Bundesstaaten erheben den Vorwurf, der Konzern habe seine Marktmacht missbraucht, um sein Geschäft mit Onlinewerbung, speziell rund um Suchergebnisse, illegal zu fördern.

Und dabei geht es auch um den Einsatz von KI: Wenn man eine Suchanfrage eingibt, schreibt Google schon seit einigen Jahren die Suchbegriffe mithilfe eines Sprachmodells um. Das ist eigentlich in Ordnung und verfeinert sehr allgemeine oder vieldeutige Anfragen. Der Vorwurf der Kläger: Google hat den Prozess genutzt, um die Suchanfragen so umzuschreiben, dass sie mehr Werbeerlöse generieren. Aus einer Suche nach „Kinderbekleidung“ soll so zum Beispiel „Kinderbekleidung der Marke NIKOLAI“ geworden sein. Das bringt dem Unternehmen mehr Geld ein, liefert aber Ergebnisse, nach denen gar nicht gesucht wurde.

In gewisser Weise lädt generative künstliche Intelligenz zum Missbrauch ein, weil viel von dieser „Intelligenz“ wie bei einer Blackbox verborgen von den Augen der Nutzer läuft. Bleibt zu hoffen, dass das Beispiel von Google nicht Schule macht – beziehungsweise adäquate Regeln und konsequente Ahndung von Verstößen dies verhindern.

(jo@ct.de) ct

Hugging Face ist mit seinen mehr als 350.000 Modellen die Anlaufstelle für KI-Interessierte.

Literatur

- [1] Jo Bager, Hartmut Gieselmann und Sylvester Tremmel, Gefährliches Halbwissen, Sieben Sudienste mit KI im Vergleich, ct' 10/2023, S. 18
- [2] Hartmut Gieselmann und Jo Bager, Die ct' in Zeiten der künstlichen Intelligenz, ct' 12/2023, S. 78
- [3] Dorothee Wiegand, Kollege KI, Wie sich generative KI auf den Arbeitsmarkt auswirken könnte, ct' 15/2023, S. 22
- [4] Jo Bager, Die KI-volution, Wirtschaft, Jobs, Gesellschaft – was generative KI verändert, ct' 15/2023, S. 16
- [5] Jo Bager und Hartmut Gieselmann, Die eigene Sprach-KI, Mit offenen Sprachmodellen experimentieren, ct' 23/ 2023, S. 16

Bitkom-Studie und Links: ct.de/yw4u



Horizons #23

by heise

Insights, Inspiration, Networking:
Strategien & Trends für Ihr Business

NEUE DIGITALE ARBEITSWELT

KÜNSTLICHE INTELLIGENZ

CYBER-SICHERHEIT

NEXT GENERATION & RECRUITING



27. NOVEMBER, **Horizons #23 Night**
Cumberlandsche, Hannover



28. NOVEMBER, **Horizons #23 Day**
Künstlerhaus, Hannover



Seien Sie mit dabei!

www.horizons-heise.de

eine Veranstaltung von

heise medien

Partner, Sponsoren

chromeOS

CISO
ALLIANZ

G DATA
academy

Digitalagentur
Niedersachsen

hannoverimpuls

HeNCKE

..htp
Internet Telefon Mobil

tn
digital
pioneers

UVN
Unternehmensverband
Netzwerkeisen e.V.

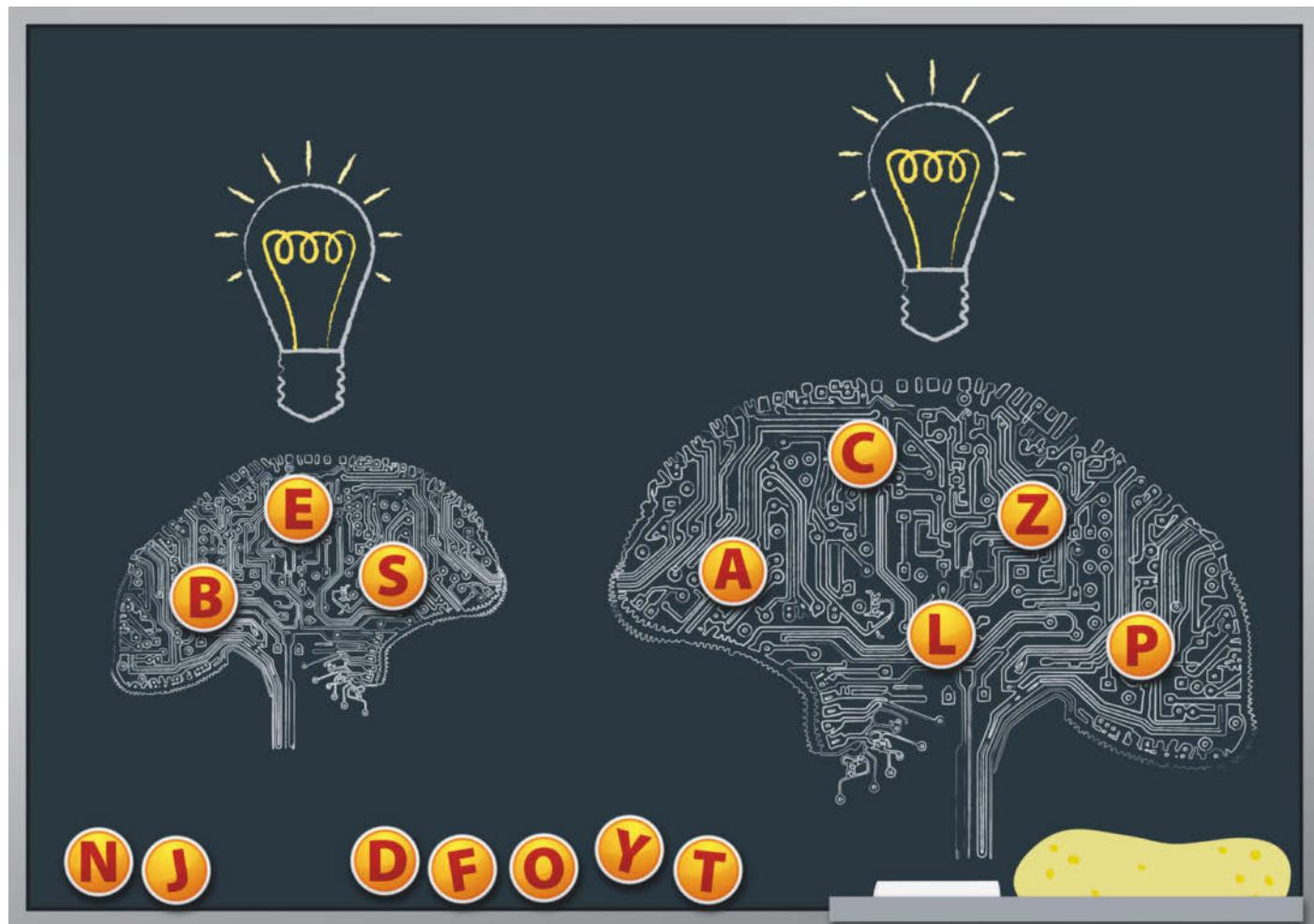


Bild: KI Midjourney | Collage c't

Klein, aber fein

**Weniger Parameter, solide Leistung:
Wie kompakte Sprachmodelle die Giganten herausfordern**

Eine Zeitlang kannte die Parameterzahl großer Sprachmodelle nur eine Richtung: steil nach oben. Mehr Parameter bedingen mehr und hochwertigere Fähigkeiten, so die Überzeugung. Doch 2023 schlug die Stunde der mittelgroßen Sprach-KIs: Sie sind genügsam – und erstaunlich konkurrenzfähig.

Von René Peinl

Mehr Parameter bedeuten für große Sprachmodelle mehr Anwendungsbereiche, mehr Aufmerksamkeit und damit mehr Investorenengeld. Dieses wiederum finanziert verlässlich mehr Rechenkapazität – womit sich die Parameterzahl noch weiter hochschrauben lässt. Die Spirale könnte sich ewig so weiterdrehen, wären da nicht die begrenzten Ressourcen (Technik und KI-Experten), der Vorwurf der Energieverschwendungen und die enormen Kosten, die selbst investorenengelvwöhnten Start-ups wie OpenAI und finanziestarken Tech-Giganten wie Google,

Facebook oder Amazon auf Dauer zu schaffen machen.

Und tatsächlich: Bereits jetzt, ein Jahr nach ChatGPT, scheint der Trend wegzugehen von aufgeblähten Large Language Models (LLMs) mit über 150 Milliarden Parametern hin zu mittelgroßen mit deutlich unter hundert Milliarden. Diese sind nicht nur vergleichsweise effizient, sondern auch überraschend gut. In mancher Disziplin rücken sie erstaunlich nahe an GPT-4 mit seinen kolportierten 1,8 Billionen Parametern heran. Damit tut sich ein riesiges Potenzial auf – auch für kleinere

und mittelgroße Unternehmen, die mit eigenen Anwendungen liebäugeln. Wir erklären, was die schlanken Verwandten der Giganten können, was sie so effizient macht und wie die Zukunft der Sprachmodelllandschaft aussehen könnte.

KI-Kunde in Kürze

Vorab etwas Grundlagenwissen: Gerade einmal fünf Jahre ist es her, dass Google im Oktober 2018 seinen Preprint des BERT-Aufsatzes (Devlin et al.) veröffentlichte und damit den Stein des aktuellen Sprachmodell-Hypes ins Rollen brachte. Was BERT (Bidirectional Encoder Representations from Transformers) so beeindruckend eloquent machte, war ein besonders umfangreiches sogenanntes Pretraining. Dabei wurde vor dem Training für die eigentliche Aufgabe zunächst einmal ein Training mit beliebigem Text absolviert. Die Texte dafür stammen aus dem Internet und müssen zuvor nicht aufwendig manuell annotiert werden. Sämtliche im Artikel referenzierten Forschungsarbeiten finden Sie unter ct.de/yzpx.

Für die größte Variante BERT-Large mit damals sagenhaften 340 Millionen Parametern wurden 256 GPU-Tage in dieses Pretraining investiert (16 Cloud-TPUs à 4 Chips für 4 Tage). Mittlerweile trainiert man LLMs mit rund 40.000 GPU-Tagen Rechenleistung. Das kürzlich vom Technology Innovation Institute in Abu Dhabi veröffentlichte Falcon-Modell hat 180 Milliarden Parameter. Es soll ein fast 300.000 GPU-Tage – also 7 Millionen Stunden – langes Basistraining absolviert haben, und das mit aktuellen Grafikkarten, die deutlich leistungsstärker sind als Googles damalige TPU-Chips. Der Ressourcenbedarf für das Sprach-KI-Training hat sich in dieser Zeitspanne somit mehr als vertausendfach. Der Bedarf explodiert, während die klassische Chiparchitektur an ihre physikalischen Grenzen stößt. Zum Vergleich: Gemäß Moore's Law, das ohnehin nicht mehr uneingeschränkt gilt, hat sich die Rechenleistung pro Fläche währenddessen nur um das 8- bis 16-Fache erhöht. Wie stark sich Bedarf und Produktionsfortschritt entkoppelt haben, zeigt die Grafik rechts.

Ausgelöst hat diesen Ressourcenhunger das Versprechen, dass Modelle mit mehr Parametern, die mit mehr Daten für längere Zeit (gemessen in investierter Rechenleistung) trainiert werden, auch bessere Ergebnisse liefern. OpenAI goss die-

ses Versprechen 2020 in eine griffige Faustformel und propagiert dieses Scaling Law seitdem erfolgreich (Kaplan et al. 2020). Forscher der Google-Tochter DeepMind verfeinerten und korrigierten die Formel 2022 in ihrem Chinchilla-Aufsatz (Hoffmann et al. 2022) [1]. Dementsprechend hat sich die Parameterzahl der großen Sprachmodelle seit 2018 stetig erhöht, ebenfalls grob um den Faktor 1000: Die größten aktuellen LLMs besitzen mehrere hundert Milliarden Parameter. Besonders spannend ist dabei das Auftreten von emergenten Fähigkeiten. Das bedeutet, dass LLMs plötzlich in Aufgaben gute Ergebnisse erzielen, für die sie gar nicht direkt trainiert wurden.

Talentierte Riesen

Über das Pretraining entwickeln LLMs bereits eine Art allgemeines Sprachgefühl, sie können also schon ordentliche Sätze bilden. Üblicherweise schließt sich an dieses Pretraining ein sogenannte Finetuning an, also das Verfeinern der Fähigkeiten für eine bestimmte Aufgabe. Durch das Finetuning lernen die Modelle dann typische Aufgaben wie Übersetzen, Fragen beantworten oder Kundenfeedback als positiv oder negativ zu klassifizieren.

Spätestens seit GPT-3 scheinen die Sprachmodelle aber immer mehr und immer nützlichere emergente Fähigkeiten

ct kompakt

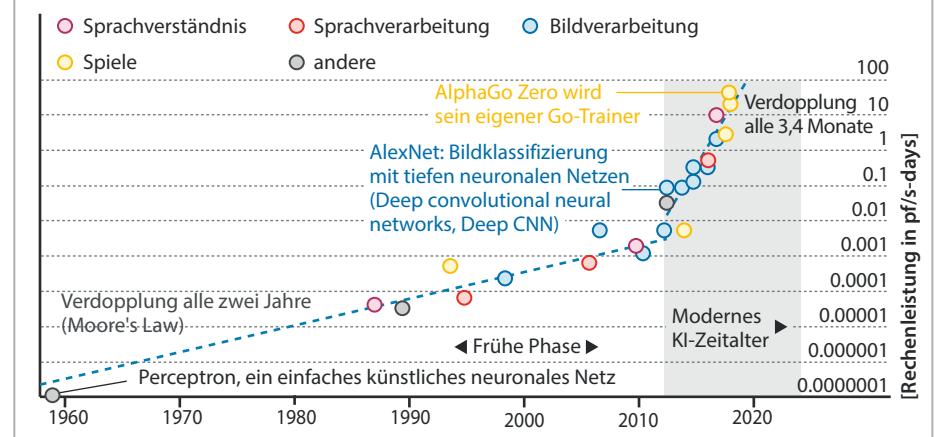
- Lange schien es so, als ob nur Sprachmodelle mit Hunderten Milliarden Parametern in der Lage sind, komplexere Denk-, Logik- und Wissensaufgaben zu lösen.
- Doch moderne, deutlich kleinere Sprach-KIs kommen erstaunlich nah an die riesigen Large-Language-Modelle heran oder übertreffen sie in einzelnen Disziplinen sogar.
- Das liegt an ausgefeilten Trainings- und Optimierungsmethoden. Es gibt aber auch Zweifel an den guten Resultaten.

zu entwickeln. Um diese zu ergründen und zu quantifizieren, testen Forscher und Entwickler sie mit sogenannten Zero-Shot-Aufgaben, also Anfragen, für die die KIs kein Finetuning bekommen haben (Wei et al. 2022). Dabei darf die Fragestellung auch keine Beispiele enthalten, aus denen die Modelle ableiten könnten, wie die Aufgabe korrekt gelöst wird (das wäre dann Few-Shot-Learning).

Visualisiert man für solche Aufgaben die Ergebnisse von Modellen unterschied-

Rechenleistung für das Training künstlicher neuronaler Netze

Bis zum Jahr 2012 folgten KI-Modelle brav Moore's Law. Die Rechenleistung pro Chipfläche verdoppelte sich ungefähr alle 2 Jahre, der Bedarf von Bild-, Sprach- und anderen KIs ebenfalls. Nachdem Google 2012 dem Deep Learning zum Durchbruch verholfen hatte, stieg der Aufwand fürs KI-Training überproportional an; er verdoppelte sich alle drei bis vier Monate (gemessen in Petaflop/s-days, pfs-day; logarithmische Skala). Ein pfs-day entspricht 10^{20} Berechnungen.



licher Größe in einer Grafik, so ergeben sich typischerweise Hockeystick- oder L-Kurven, siehe Grafik unten: Kleinere Modelle „raten nur rum“ und erzielen Ergebnisse im Bereich der Erwartungswerte nach Wahrscheinlichkeitsrechnung. Bei Multiple-Choice-Fragen mit vier Antwortmöglichkeiten und einer richtigen Antwort hätten sie also immerhin eine Trefferquote von 25 Prozent, wie im Bild unten bei Multi-task NLU zu sehen. Nach Überschreiten einer gewissen Größe können LLMs die Aufgaben aber plötzlich relativ gut lösen, obwohl sie nicht dafür trainiert wurden, und nähern sich schnell der menschlichen Genauigkeit.

Das wird überwiegend als Hinweis auf vielversprechende Schritte hin zu einer allgemeinen künstlichen Intelligenz deutet: also auf ein System, das nicht nur Schach oder Go spielen kann und sonst nichts, sondern in möglichst vielen Aufgabenbereichen gut abschneidet. Wie stark und valide diese Indizien sind, darüber gehen die Meinungen allerdings auseinander; die unterschiedlichen Positionen haben wir in [2] zusammengefasst. Forscher der Universität Stanford haben kürzlich mit einem Aufsatz für Aufsehen gesorgt, der behauptet, dass die Art der Tests diese emergenten Fähigkeiten implizit erzeugen können (Schaeffer et al. 2023). Damit legen die Autoren den Finger grundsätzlich in die richtige Wunde: Vermutlich ist es korrekt, dass die Art der Messung in mancher Studie wohl beein-

flusst hat, dass neue Fähigkeiten eines LLM als emergent erschienen. Allerdings betonen die Forscher selbst, dass ihre Arbeit lediglich Schwächen aufzeigt. Die grundsätzliche Existenz des Emergent-Ability-Phänomens habe man damit nicht widerlegt.

Da der Bedarf an universell begabten LLMs derzeit riesig ist, aber nicht jeder gerade einen Supercomputer zur Hand hat oder Millionen Dollar in Cloud-Ressourcen investieren kann, kam schnell die Frage auf, ob nicht auch kleinere, deutlich billigere Modelle ähnlich gute Ergebnisse erzielen könnten. Das ist auch eine Frage, die OpenAI massiv umtreibt: Schön, wenn hundert Millionen Menschen ChatGPT nutzen wollen. Aber wie kann man die Unterhaltskosten dieses Dienstes so gering halten, dass die Firma mit einer Gebühr von 20 US-Dollar pro Benutzer und Monat Gewinne erzielt und nebenbei auch noch die vielen kostenlosen Accounts mitfinanziert? Mit einem 170-Milliarden-Parameter-Modell ist das schwierig.

Erste starke Indizien dafür, dass solche Dienste auch mit kleineren Modellen funktionieren könnten, lieferte DeepMind im Jahr 2022 mit seinem 70 Milliarden Parameter großen Sprachmodell Chinchilla. Es war zwar immer noch sehr aufwendig zu trainieren, lief aber dafür im späteren Einsatz – dem sogenannten Inferencing – deutlich kostengünstiger. Auch diverse Open-Source-Entwickler zeigten,

dass kleinere Modelle ein vergleichbares Niveau erreichen können. Man darf also getrost davon ausgehen, dass OpenAI mittlerweile – zumindest für ChatGPT 3.5 – ebenfalls kleinere Modelle einsetzt; zahlreiche Berichte über vermeintliche Qualitätsschwankungen legen diesen Schluss nahe.

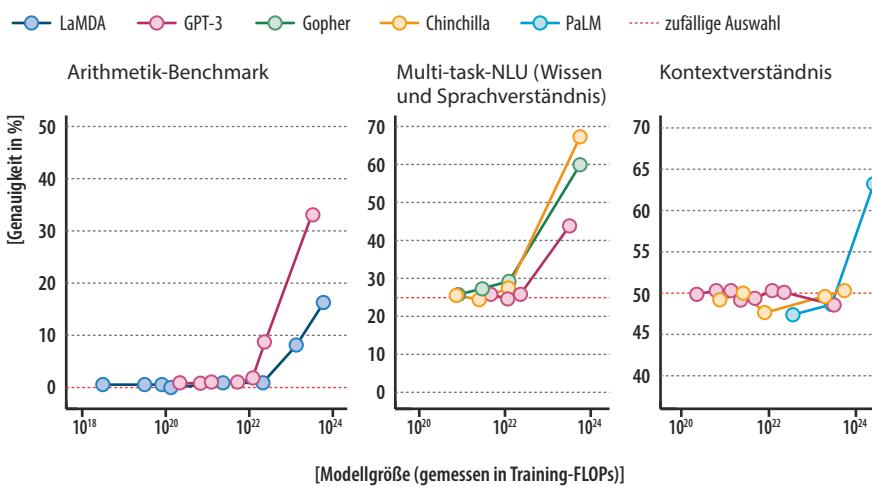
Ambitioniertes Nachwuchstraining

Wie aber gelingt das Kunststück, aus kleinen Modellen große Ergebnisse hervorzuzaubern? Ein wesentlicher Aspekt hierfür liegt im Pretraining mit mehr Daten und Rechenaufwand, wie DeepMind mit Chinchilla demonstrierte. Das 70-Milliarden-Parameter-Modell lieferte bei gleichem Trainingsaufwand genauso gute Resultate wie das aus einer früheren Generation stammende Gopher mit seinen 280 Milliarden Parametern. Auch bei Metas Open-Source-Modellen zeigt sich dieser Effekt, denn das 2023 veröffentlichte LLaMA schneidet mit 65 Milliarden Parametern in gängigen Benchmarks deutlich besser ab als das fast dreimal so große OPT (Open Pretrained Transformer, 175 Milliarden Parameter) aus dem Vorjahr. Die genauen GPU-Tage hat Meta in seinen Veröffentlichungen nicht dokumentiert. Für das aktuelle LLaMA v2 hat Meta nochmal mehr Daten und Rechenzeit spen- diert, weshalb auch dieses seinen Vorgänger wiederum übertrifft. Kein Wunder also, dass LLaMA v1 und auch schon v2 Grundlage für Tausende Open-Source-Modelle sind. Eine Übersicht sowie einen Praxisleitfaden zum Einsatz von Open-Source-LLMs finden Sie in unserer Artikelserie „Die eigene Sprach-KI“ in c't 23/2023 ab Seite 16, siehe ct.de/yzpx.

Der zweite Teil des Erfolgsrezepts ist das sogenannte Instruction Tuning. Dieser Finetuning-Ansatz erzieht das Modell dazu, eine ganze Reihe verschiedener Aufgaben zu erledigen (oft mehr als 1000), die man relativ einheitlich formulieren und gegebenenfalls mit ein paar Beispielen für die Eingabe und die erwartete Ausgabe anreichern kann. Ein Beispiel: „Ordne den folgenden RGB-Farbwerken in Hexadezimalschreibweise ein Farbwort zu! #16ade9 ergibt türkis, #fff600 ergibt gelb, was ergibt #fa05c2?“ Das oft zitierte Prompt Engineering ist also letztlich das Bemühen des Menschen, sein Anliegen möglichst so zu formulieren, dass die Aufgabenstellung den Trainingsdaten im Instruction Tuning nahekommt. Das erhöht die Chancen, dass

Wissen versus Raten

Kleinere Modelle „raten nur rum“: Sie erzielen zum Beispiel bei Multiple-Choice-Fragen mit vier Antwortmöglichkeiten (Mitte) lediglich eine Trefferquote von 25 Prozent – die auch bei zufälliger Auswahl der Antworten herauskommen würde. Je größer das Modell, umso besser löst es Aufgaben, auf die es nicht explizit trainiert wurde.



das Modell die Aufgabe versteht und trotz nur mittelmäßiger Generalisierungsfähigkeiten eine gute Antwort generiert.

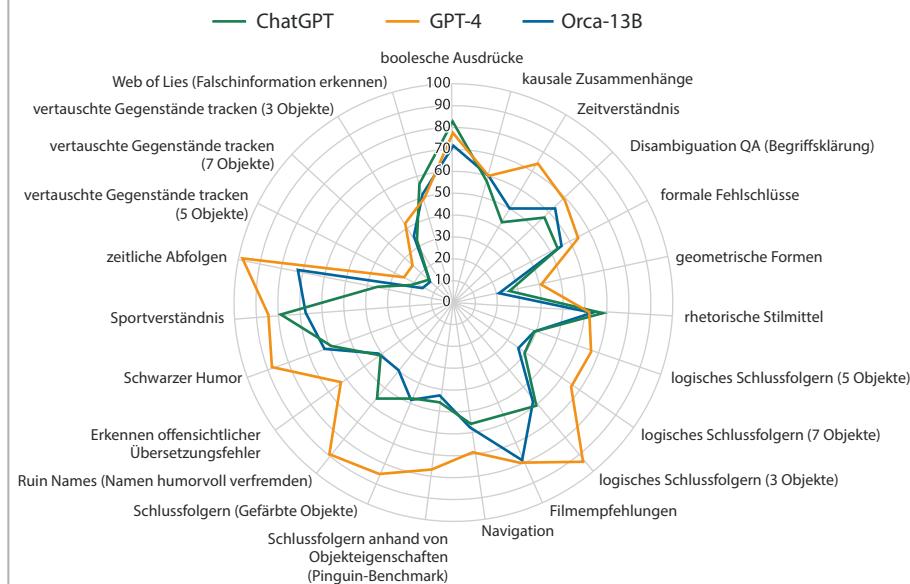
Da geeignete Trainingsdatensätze aufwendig zu erstellen sind, kamen Forscher der Stanford-Universität auf die Idee, ChatGPT dafür anzuwerben. Der mit diesem Datensatz verfeinerte LLaMA-Abkömmling ist unter dem Namen Alpaca bekannt geworden und hat eine veritable Nachahmerwelle ausgelöst. Der Zoo reicht von Guanaco über Camel bis Vicuna, und als die Lama-Verwandten ausgingen, gsellten sich Orca, Dolphin, Koala, Wombat und Platypus dazu. Die Modelle lassen sich in zwei Gruppen unterteilen: Die einen sind eher für einmalige Fragen ausgelegt (Instructions) und die anderen für Unterhaltungen mit mehreren Runden. Bei letzteren sind auch mal Nachfragen seitens des Modells erwünscht (Chat), wobei sich mittlerweile keine klare Grenze mehr ziehen lässt. Kritische Stimmen sagen aber, dass die Open-Source-Modelle nur den Antwortstil von ChatGPT oder GPT-4 imitieren, ohne wirklich die gleichen Fähigkeiten zu besitzen (Gudibande et al. 2023).

Als dritter Baustein, der LLMs deutlich verbessert, wird Reinforcement Learning from Human Feedback (RLHF) gehandelt. Bestärkendes Lernen kommt bisher hauptsächlich in der Robotik und bei Computerspielen zum Einsatz – etwa bei DeepMinds Agent57, der Atari-Spiele meistert. RLHF ist ein Verfahren, das menschliches Feedback simulieren soll. Denn beim KI-Training und -Benchmarking hat sich die menschliche Rückmeldung zwar als nützlich erwiesen – sie ist aber teuer und langsam. Um sie zu imitieren, wird ein eigenes neuronales Netz darauf trainiert, die menschlichen Präferenzen widerzuspiegeln.

Bei LLMs wie ChatGPT läuft das konkret so ab: Menschen geben Rückmeldung zu Antworten, die das LLM zu ein paar tausend Prompts generiert, indem sie diese etwa auf einer Skala von 0 bis 10 bewerten oder einfach nur entscheiden, welche von zwei Antworten besser ist. Anhand des so erstellten Trainingsmaterials aus Prompt, Antwort und menschlicher Bewertung lernt ein eigenes neuronales Netz, diese Rückmeldungen vorherzusagen. Das fertig trainierte Netz wird anschließend zum Finetuning des LLMs eingesetzt, um dort für eine Million Fragen und Antworten die Rückmeldung des Menschen zu ersetzen. Für ChatGPT war dies laut OpenAI der entscheidende

Orca – fast auf ChatGPT-Niveau

Das nur 13 Milliarden Parameter große Sprachmodell Orca liegt in diversen Benchmarks fast gleichauf mit ChatGPT und übertrifft es in einigen Disziplinen sogar. Hinter GPT-4 bleibt es aber noch deutlich zurück.



Schritt, insbesondere um das sogenannte Alignment zu verbessern, sprich: das Modell dazu zu erziehen, keine unangemessenen Antworten zu geben, die Kunden verärgern.

Im Open-Source-Umfeld hat sich diese automatisierte Feedback-Methode bisher nicht durchgesetzt. Der Grund dafür ist nicht ganz klar, vielleicht weil es aufwendiger ist als pures Instruction Tuning und sich nur in wenigen Benchmarks positiv auswirkt. Im Modell Wombat von Alibaba wurde eine Variante von RLHF eingesetzt. Das Modell schlägt sich im Vergleich gut, ohne jedoch an den besten Modellen mit Instruction Tuning vorbeiziehen.

ChatGPT hart auf den Fersen

In der Summe scheinen die diversen Tuning-Maßnahmen zu fruchten, denn einige kleinere Modelle schlagen sich in gängigen Benchmarks im Vergleich mit ChatGPT sehr beachtlich. Interessanterweise ist unter den ChatGPT-Jägern ausgerechnet Microsoft Research mit seinem 7-Milliarden-Parameter-Modell WizardLM ganz vorne dabei. Diese Konkurrenz verwundert angesichts der Milliardeninvestitionen von Microsoft in OpenAI und der exklusiven Partnerschaft, in deren Rahmen Microsoft die Modelle des Startups über seine Cloud vertreibt. Bei Wi-

zardLM (Xu et al. 2023) haben menschliche Tester 59 Prozent der Antworten als gleich oder besser beurteilt als diejenigen von ChatGPT und 76 Prozent als gleich oder besser als diejenigen von Alpaca. Bei besonders schwierigen Aufgaben des von den WizardLM-Autoren selbst erstellten Fragenkatalogs „Human Instruct Evaluation Set“ verschob sich das Verhältnis sogar noch stärker zugunsten von WizardLM.

Grundlage des ebenfalls von Microsoft entwickelten Orca (Mukherjee et al. 2023) ist ein 13-Milliarden-Parameter-Modell. Es wurde mit fünf Millionen ausgewählten Instruktionen inklusive Antworten aus fünf Open-Source-Datensätzen optimiert und durch GPT-4 noch weiter verbessert. Dadurch erreicht es 95 Prozent der Qualität von ChatGPT und 85 Prozent der von GPT-4 bei offenen Fragestellungen. Im AGIEval-Benchmark, der sich aus diversen Examensprüfungen zusammensetzt, fiel der Vergleich etwas schlechter aus, aber in Googles BIG-bench lag Orca in der Disziplin „hard“ mit besonders schweren Aufgaben sogar knapp vor ChatGPT. Die Grafik oben zeigt die konkrete Verteilung. Man sieht, dass Orca zum Beispiel deutlich besser mit zeitlichen Abfolgen klarkommt als ChatGPT und bei Filmempfehlungen sogar auf Augenhöhe mit GPT-4 liegt, aber bei booleschen Aus-

drücken und beim Sportverständnis nicht mithalten kann.

In eigenen Tests des Autors übertraf das Open-Source-Modell Airoboros des Entwicklers Jon Durbin mit 30 Milliarden Parametern nicht nur ChatGPT 3.5, sondern sogar die „Human Baseline“ (Peinl and Wirth 2023) [3], also den Durchschnittswert, den menschliche Probanden mit Abitur und zum Teil Hochschulabschluss im gleichen Test erreichten. Es ist jedoch unklar, ob die guten Ergebnisse durch „verunreinigte“ Trainingsdaten entstanden sind, in denen Teile der Testfragen enthalten waren. Insbesondere bei den Scherzfragen könnte das der Fall sein (zum Beispiel „Wie viele Tiere nahm Abraham mit in die Arche?“). Eine detaillierte Auseinandersetzung mit dem Benchmarking großer Sprachmodelle finden Sie in [4].

Energiesparmodus für LLMs

In allen Testszenarien zeigen sich die Modelle erstaunlich robust gegenüber einer Reduzierung der Genauigkeit der Modellparameter. Für das Training tiefer neuronaler Netze wählen Forscher und Entwickler normalerweise Gleitkommazahlen einfacher Genauigkeit als Datentyp für die Modellparameter (FP32). Neuerdings setzt sich zunehmend das Training mit „Mixed Precision“ durch, das dynamisch entscheidet, wo FP32 nötig ist und wo auch halbe Genauigkeit ausreicht (FP16). Beim

Bild: lightman et al., OpenAI

The denominator of a fraction is 7 less than 3 times the numerator. If the fraction is equivalent to $\frac{2}{5}$, what is the numerator of the fraction? (Answer: 14)

OpenAI

Let's call the numerator x.

So the denominator is $3x - 7$.

We know that $x/(3x-7) = 2/5$.

So $5x = 2(3x-7)$.

$5x = 6x - 14$.

So $x = 7$.

Differenziertes Feedback statt nur richtig oder falsch: Forscher plädieren dafür, den Sprachmodellen detailliertere Rückmeldung zu ihren Antworten und Lösungsvorschlägen zu geben. Hier demonstriert OpenAI, wie sich etwa Rechenschritte beurteilen lassen.

Inferencing, also in der späteren Anwendung, reicht in der Regel FP16, sodass man dort die Hälfte des GPU-Speichers sparen kann.

Sogenanntes Quantisieren [5] reduziert die Genauigkeit während des laufenden Betriebs noch weiter, etwa auf ganzzählige Werte mit einem Byte (int8) oder sogar auf nur vier Bits wie beim derzeit populären GPTQ-Verfahren zur Quantisierung von Generative-Pretrained-Transformer-Modellen. Die aktuellste Methode nennt sich „Activation-aware Weight Quantization“ (AWQ) und belässt ausgewählte Parameter in FP16-Genauigkeit, während sie den Rest auf vier Bit quantisiert. Viele Modelle auf Hugging Face bekommt man von zwei bis acht Bit in allen Stufen, bei vier Bit scheint es einen Sweet-spot aus Größe und Qualität zu geben.

Während das Betreiben eines Modells wie LLaMA v2 mit 70 Milliarden Parametern in voller Genauigkeit mindestens zwei Hochleistungsgrafikkarten vom Typ A100 erfordert, was mit entsprechendem Server mindestens 50.000 Euro kostet und einen erheblichen Stromverbrauch nach sich zieht, kann ein auf vier Bit quantisiertes Modell problemlos und schnell auf einer A6000-GPU betrieben werden. Das drückt den Preis inklusive Workstation auf unter 10.000 Euro und läuft erheblich energieeffizienter. Modelle mit 30 Milliarden Parametern, die immer noch sehr gute Leistung zeigen (etwa acht Prozent schlechter in typischen Benchmarks), lassen sich sogar auf einer Consumer-GPU wie der RTX 4090 betreiben, was den Hardwarepreis auf unter 3500 Euro senkt.

Vollends unübersichtlich wurde die Anzahl an offengelegten LLMs kürzlich durch ein Verfahren namens LoRA (Low Rank Adaptation, Hu et al. 2021), einem effizienten Finetuning-Verfahren. Dabei werden nicht alle zig Milliarden Parameter des Transformer-Netzes aufwendig nachjustiert, sondern nur ein paar wenige zusätzliche. LoRA ergänzt nun für jeden Layer eigene trainierbare Matrizen, wodurch sich der Rechenaufwand um Faktor 500 und mehr reduzieren lässt (Ding et al. 2023). Dieses Verfahren wurde ursprünglich für Bilderzeuger wie Stable Diffusion entwickelt, um das Modell effizient auf spezifische Konzepte wie eine bestimmte Person oder den eigenen Hund trainieren zu können, ohne die bisher erlernten Fähigkeiten zu beeinträchtigen. Künstliche neuronale Netze neigen nämlich zum „catastrophic forgetting“, also dem Verlust ihrer schon beherrschten Fähigkeiten, wenn sie etwas Neues dazulernen sollen.

Model Input (Multistep Arithmetic)

<Few shot chain-of-thought exemplars omitted for brevity>

Q: $((4 + 7 * 4 - 5) - (4 - 1 - 4 - 4))$
A: Let's think step by step.

Model Output

Let's recall that the order of operations in mathematics is as follows: (1) Parentheses, (2) exponents, (3) multiplication and division (from left to right), (4) addition and multiplication (from left to right). So, remember to always compute the expressions inside parentheses or brackets first.

This equation can be written as "A - B", where A = $(4 + 7 * 4 - 5)$ and B = $(4 - 1 - 4 - 4)$.

Let's calculate A = $(4 + 7 * 4 - 5) = (4 + (7 * 4) - 5) = (4 + (28) - 5) = (4 + 28 - 5) = (4 + 28 + 5) = 37$.

Let's calculate B = $(4 - 1 - 4 - 4) = ((4 - 1) - 4 - 4) = ((-5) - 4 - 4) = ((-5 - 4) - 4) = ((-5 + 4) - 4) = (-1 - 4) = -5$.

Then, the final equation is A - B = $37 - 5 = 37 + 5 = 42$. So the answer is 42. ✓

Model Input (Snarks)

Q: Which statement is sarcastic?
Options:
(A) Working at a restaurant has made me an amazing chef
(B) Working at a McDonald's has made me an amazing chef
A: Let's think step by step.

Model Output

If we look at (A), it says that working at a restaurant has made the speaker an amazing chef. This is a neutral statement.
If we look at (B), it says that working at a McDonald's has made the speaker an amazing chef. This is a neutral statement. Above the above, the sarcastic option is (A). So the answer is (A). X

Bild: Sargin et al., Google Research / Stanford University

Eine sehr nützliche emergente Fähigkeit ist das Anfertigen von Schritt-für-Schritt-Begründungen (links). Um einem Sprachmodell solche Antworten zu entlocken, wurde eine spezielle Prompting-Technik mit Beispielen (Chain of Thought) entwickelt. Manchmal verheddet sich die KI aber in ihren Begründungen, wie rechts bei der Sarkasmus-Erkennung.

Was bringt die Zukunft?

Während ChatGPT (3.5) in vielen Benchmarks von deutlich kleineren Modellen eingeholt wurde, bleibt GPT-4 bislang souverän an der Spitze. Nach Gerüchten zur nicht offengelegten Architektur besteht es aus 120 Schichten mit insgesamt rund 1,8 Billionen Parametern, also mehr als zehnmal so vielen wie GPT-3.5. Diversen ebenfalls nicht bestätigten Informationen zu folge verwendet es aber immer nur einen Teil davon zur gleichen Zeit, gemäß Googles Mixture-of-Expert-Architektur (MoE), so dass 16 „Experten“ mit je 111 Milliarden Parametern verbunden werden. Diese Modelle werden auch als „sparse“ bezeichnet, analog zu dünn besiedelten Matrizen.

Es war eigentlich zu erwarten, dass mit der Zeit fast alle LLMs auf diese Architektur umschwenken. Denn Google hatte mit GLaM (Generalist Language Model, Du et al. 2021) schon Ende 2021 gezeigt, dass MoE eine bessere Qualität als die damalige Referenz GPT-3 ermöglicht, während es nur ein Drittel des Rechenaufwands beim Training benötigt und im operativen Betrieb (Inferencing) lediglich die Hälfte der Rechenleistung beansprucht.

Über Googles Chatbot Bard ist ähnlich wenig bekannt wie über GPT-4. Er stützte sich wohl zunächst auf das haus-eigene LaMDA, stieg dann aber aus Genauigkeitsgründen auf PaLM um. Der auf PaLM v2 beziehungsweise dessen Instruct-Variante (Flan-PaLM 2) basierende

Bard reicht zwar immer noch nicht an die Qualität von GPT-4 heran, kann aber in vielen Benchmarks gut mit ChatGPT mithalten (wobei die Studienlage hier kein eindeutiges Bild ergibt). Diese Ergebnisse würde zu den kolportierten Daten mit 340 Milliarden Parametern und 3,6 Billionen Trainingstoken passen.

Man darf auch gespannt sein, ob sich die nächste Version von Metas LLaMA in die MoE-Reihe eingliedert oder nochmal versucht, mit der bisherigen Architektur besser zu werden.

Derweil verschiebt sich der Fokus wieder weg von den allgemeinen hin zu spezialisierten Sprach-KIs. Bloomberg hat es im April dieses Jahres schon für die Finanzindustrie vorgemacht und Fintech-Unternehmen werden folgen. Mit ChatLaw beschrieben Forscher im Juni ein auf das chinesische Rechtssystem spezialisiertes Modell. Selbst für die Lebensmittelbranche, den E-Commerce und sogar Hausrenovierungen wurden schon eigene Modelle entwickelt. Seit Mai 2023 erschienen zunehmend Modelle, die aufs Programmieren spezialisiert sind, zu den prominenteren gehören StarCoder, WizardCoder und Code Llama.

Schlussendlich dürfte sich die Wissenschaft zunehmend auch mit ausgefeilten „didaktischen“ Methoden beschäftigen. Ähnlich, wie menschliche Schüler profitieren, wenn sie statt einer bloßen Schulnote eine differenzierte Rückmeldung über die Stärken und Schwächen

eines Aufsatzes oder eines mathematischen Rechenwegs erhalten, könnte sich damit auch die Leistung von Sprachmodellen verbessern lassen. Gerade bei längeren Antworten und schrittweise erarbeiteten Ergebnissen genügt eine Gesamt-rückmeldung nicht, um hinreichend feine Impulse für das Training eines LLMs zu geben, weshalb Peinl und Wirth bereits im Mai dieses Jahres für ein differenzierteres Feedback plädierten. In dieselbe Kerbe schlägt ein kurz danach veröffentlichter Aufsatz von OpenAI-Forschern (Lightman et al.). Sie stellen darin ein auf mathematische Aufgaben trainiertes Feedback-Verfahren vor, das nicht nur die Lösung als richtig oder falsch bewertet, sondern auch die Teilschritte beurteilt. (at@ct.de) ct

Literatur

- [1] Jordan Hoffman et al., Training compute-optimal large language models, arxiv:220315556
- [2] Pina Merkert, Philipp Bongartz, Skalierungshypothesen vs. Neurosymbolik, Welche nächsten Schritte muss die KI-Forschung gehen?, c't 23/2022, S. 124
- [3] René Peinl, Johannes Wirth, Evaluation of medium-large Language Models at zero-shot closed-book generative question answering, arxiv:230511991
- [4] Hartmut Gieselmann, Die 80-Prozent-Maschinen, Warum KI-Sprachmodelle weiterhin Fehler machen und was das für den produktiven Einsatz bedeutet, c't 21/2023, S. 16
- [5] Pina Merkert, Byte-Intelligenz, Neuronale Netze für TensorFlow-Lite auf 8-Bit-Integer umrechnen, c't 22/2019, S. 182

Forschungsaufsätze, Studien, c't-Artikel zum Thema: ct.de/yzpx

Heft + PDF mit 29 % Rabatt

Online-Shopping ohne Probleme: c't hilft.

- Die wichtigsten Regeln für den Onlinekauf
- Schützen Sie sich vor Betrug
- Kaufprobleme lösen
- Käuferschutz richtig einsetzen
- Digital bezahlen
- Auch als Heft + digitale Ausgabe mit 29 % Rabatt

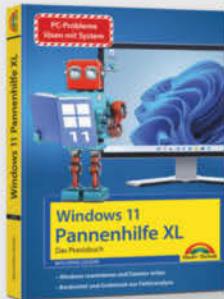
Heft für 14,90 € • PDF für 12,99 € • Bundle Heft + PDF 19,90 €



shop.heise.de/ct-sicher-einkaufen23

Für Wissenshungrige...

Ausgewählte Fachliteratur



Wolfram Gieseke

Windows 11 – Power-Tipps

Ob ein externes Gerät nicht erkannt wird, Programme nicht mehr wie gewohnt laufen oder ein Ihnen unbekannter Update-Fehler auftritt: Wenn Sie den unterschiedlichen Fehlermeldungen selbst auf den Grund gehen möchten, hilft Ihnen dieses Buch weiter.

19,95 €

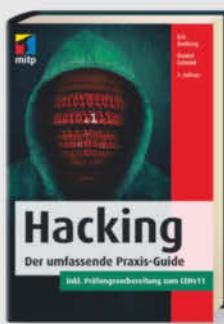


Brian Svidergol, Bob Clements, Charles Pluta

Microsoft 365 Mobilität und Sicherheit

Bereiten Sie sich auf die Microsoft-Prüfung MS-101 vor und zeigen Sie, dass Sie die erforderlichen Fähigkeiten und Kenntnisse für die Verwaltung von Mobilität und Sicherheit in Microsoft 365 sowie die damit verbundenen Verwaltungsaufgaben in der Praxis beherrschen. Dieses Prüfungstraining wurde für erfahrene IT-Profis entwickelt.

49,90 €



Eric Amberg, Daniel Schmid
Hacking – Der umfassende Praxis-Guide (2. Auflage)

Dies ist ein Leitfaden für angehende Hacker, Penetration Tester, IT-Systembeauftragte, Sicherheitsspezialisten und interessierte Poweruser. Mithilfe vieler Workshops sowie Tipps und Tricks lernen Sie die Vorgehensweise eines professionellen Hacking-Angriffs kennen.

49,99 €



Michael Weigend
Python 3 für Studium und Ausbildung

Alle wichtigen Grundlagen der Python-Programmierung werden erklärt. Es sind keine Vorkenntnisse notwendig und die Themen werden fachunabhängig erläutert.

19,99 €



Christian Immler
Haus und Wohnung smart vernetzt

Ob Sie Daten, Musik und Medien im ganzen Haus nutzen, Ihr WLAN optimieren oder per App aus der Ferne Ihre Heizung anstellen, diese und weitere relevante Themen rund um Ihr vernetztes Zuhause werden in diesem Buch ausführlich besprochen.

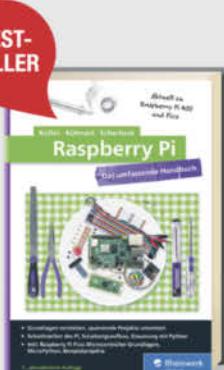
19,95 €



Thomas Kaffka
3D-Druck – Praxisbuch für Einsteiger (3. Auflage)

Entdecken Sie die nahezu unbegrenzten Möglichkeiten des 3D-Drucks in allen Varianten: vom Einsatz des eigenen 3D-Druckers zu Hause über die Verwendung von öffentlich zugänglichen Druckern bis hin zur Nutzung von 3D-Druckservices.

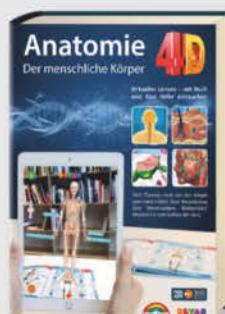
29,99 €



Michael Kofler, Charly Kühnast, Christoph Scherbeck
Raspberry Pi (7. Auflage)

Das Standardwerk in 7. Auflage, aktuell zum Raspberry Pi Pico. Die RasPi-Experten Michael Kofler, Charly Kühnast und Christoph Scherbeck bieten Ihnen auf über 1.000 Seiten das komplette Wissen, damit Sie mit dem Raspberry Pi richtig durchstarten.

44,90 €



Anatomie 4D – Der menschliche Körper

Mithilfe einer kostenlosen App und bahnbrechender Augmented Reality kann der Aufbau der Knochen, die Muskeln in Aktion, das Nerven- und Kreislaufsystem sowie das größte menschliche Organ, die Haut, beobachtet werden.

14,95 €



shop.heise.de/highlights2023

PORTOFREI AB 20 € BESTELLWERT INNERHALB DEUTSCHLANDS



Zubehör und Gadgets



Oxocard Artwork Creative Coding

Mit dem leistungsfähigen Dual-Core Chip ESP32 liefert die Oxocard genügend Power für Ihre Experimente. Lernen Sie in kurzer Zeit wie man beeindruckende visuelle Effekte erzeugt, wie wir sie aus Spielen und Filmen kennen.

69,90 €



musegear® finder Version 2

Finden Sie Schlüssel, Handtasche oder Geldbeutel bequem wieder statt ziellos zu suchen. Mit dem Finder können Sie z.B. das Smartphone klingeln lassen oder Wertgegenstände einfach tracken und noch mehr.

24,90 €



Joy-IT LCR-T7 Messgerät

Mit Hilfe des LCR Messgerätes können Sie die Induktivitäten (L) von Spulen, Kapazitäten (C) von Kondensatoren und deren Widerstände (R) als Verlust messen. Die automatische Bauteilerkennung von dem Messgerät kann elektronische Komponenten (Dioden, Z-Dioden, Dioden, Widerstände, Kondensatoren, Induktoren, Thyristoren, Triacs, Feldeffekttransistoren, Bipolartransistoren und Batterien) erkennen.

29,90 €



Nitrokey 3A NFC

Der Nitrokey 3 vereint die Funktionen vorheriger Nitrokey Modelle: FIDO2, Einmalpasswörter, OpenPGP Chipkarte, Curve25519, Passwort-Manager, Common Criteria EAL 6+ zertifiziertes Secure Element, Firmware-Updates. Damit werden Ihre Accounts zuverlässig gegen Phishing und Passwort-Diebstahl geschützt.

59,90 €



Joy-IT OR750i: Freifunk- & OpenWrt-Dual-Band-Router

Der Einstieg in die Freifunk- und OpenWrt-Welt kann oft schwierig sein. Deshalb hat Joy-IT in Zusammenarbeit mit Freifunk Hannover und c't den OR750i entwickelt. Dank Webinterface kann man beliebige Firmwares einfach hochladen – ohne komplizierte Kommandos oder inkompatible Hardware-Revisionen; ideal für OpenWrt-Einsteiger und solche, die Freifunk einfach nur nutzen wollen.

39,90 €



JOY-IT DSO-138 M mini Oszilloskop

Das Mini- Oszilloskop mit einer Bildschirmgröße von 2,4" kann per USB oder Akku betrieben werden. Eine Verbesserung ist der externe Triggereingang, welcher TTL- und LVTTL-Signale als Quelle akzeptiert und serielle Ausgabe von Wellenformdaten.

54,90 €



Die Reise mit dem micro:bit V2

Mit der Electronic Adventure Experimentier-Box ab 8 Jahren lernt man in aufeinander aufbauenden Lektionen wie sich auf Basis des BBC micro:bit spannende Experimente verwirklichen lassen.

49,90 €



REINER SCT Authenticator

Der REINER SCT Authenticator speichert die elektronischen Schlüssel für die Logins sicher in seiner Hardware und generiert die TOTP-Einmalpasswörter hochgenau alle 30 Sekunden. Er arbeitet ohne Internetverbindung und kann deshalb online nicht angegriffen werden. Zusätzlich kann seine Funktion noch mit einem PIN-Schutz abgesichert werden.

44,90 €



Bild: Nmedia, saran25, stock.adobe.com | Collage c't

Mücke oder Elefant?

Weshalb sich der KI-Stromverbrauch schlecht einschätzen lässt

Der Boom der generativen künstlichen Intelligenz (KI) beruht auf sehr leistungsstarken Servern, die viel Strom schlucken. Dadurch wiederum könnte der IT-Energiebedarf in Zukunft enorm steigen. Weil das aber sehr teuer würde, wollen Programmierer und Chiphersteller die Effizienz verbessern.

Von Christof Windeck

Derzeit tobt eine regelrechte Materialschlacht unter den großen Firmen, die generative KI-Modelle entwickeln, etwa große Sprachsysteme wie ChatGPT. Dieser KI-Hype lässt befürchten, dass der ohnehin schon hohe Stromverbrauch der internationalen Rechenzentren noch schneller anwächst.

Das sogenannte Training großer KI-Modelle frisst nämlich Rechenleistung in bisher beispielloser Menge. Davon profitiert vor allem die US-Firma Nvidia, die bis Ende 2024 mehr als 2 Millionen KI-Rechenbeschleuniger des jüngsten Typs H100 „Hopper“ verkaufen will. Wenn sie alle unter Voll-

last laufen, fressen sie mit über 1,6 Gigawatt mehr Strom, als einer der größten Atomreaktoren liefert. Und außer Nvidia werfen noch mehrere weitere Firmen KI-Beschleuniger auf den explodierenden Markt, zurzeit sind mehrere KI-Supercomputer zu Preisen über jeweils 1 Milliarde Euro im Bau. Allerdings gibt es auch belastbare Hinweise, dass sich die aktuelle Entwicklung nicht auf Dauer fortsetzen wird. Um das zu verstehen, müssen wir ein bisschen ausholen.

KI ≠ KI

Mit dem Begriff KI bezeichnet man höchst unterschiedliche Dinge, die im Grunde

kaum mehr Gemeinsamkeiten haben, als dass sie Daten verarbeiten. Folglich schlucken sie auch unterschiedlich viel Strom. KI-Algorithmen sind schon seit Jahrzehnten Bestandteil verbreiteter Onlinedienste, etwa der Google-Suche. Genau deshalb hat Google eigene KI-Beschleuniger für die hauseigenen Server entwickelt, die so genannten Tensor Processing Units (TPU). Dabei führt der Begriff „Beschleuniger“ ein wenig in die Irre. Denn die etwa für den PageRank-Algorithmus der Google-Suche verwendeten TPU-Karten erledigen die Aufgabe nicht bloß schneller als gewöhnliche Prozessoren von AMD oder Intel, sondern vor allem auch energetisch effizienter.

Ähnliches gilt für die eingebauten KI-Einheiten in vielen Smartphoneprozessoren etwa von Apple, Qualcomm und MediaTek, darunter Apples Neural Engine in den iPhones. Letztere lässt sich für die Spracherkennung einspannen und führt die KI-Algorithmen zur Verschönerung von Fotos und Videos aus. Dabei braucht sie nur Bruchteile eines Watts und leert den Akku nicht wesentlich schneller als andere Apps.

Die Befürchtung, der KI-Stromverbrauch gehe durch die Decke, bezieht sich also nicht auf KI-Algorithmen allgemein, sondern vor allem auf generative KI-Systeme (Generative AI), die zurzeit boomen.

ct kompakt

- Generative KI-Modelle schießen zurzeit wie Pilze aus dem Boden.
- Diese KI-Modelle benötigen sehr viel Rechenleistung und deshalb auch viel Strom.
- Die enormen Kosten für Energie und Hardware könnten den KI-Boom aber auch abwürgen.

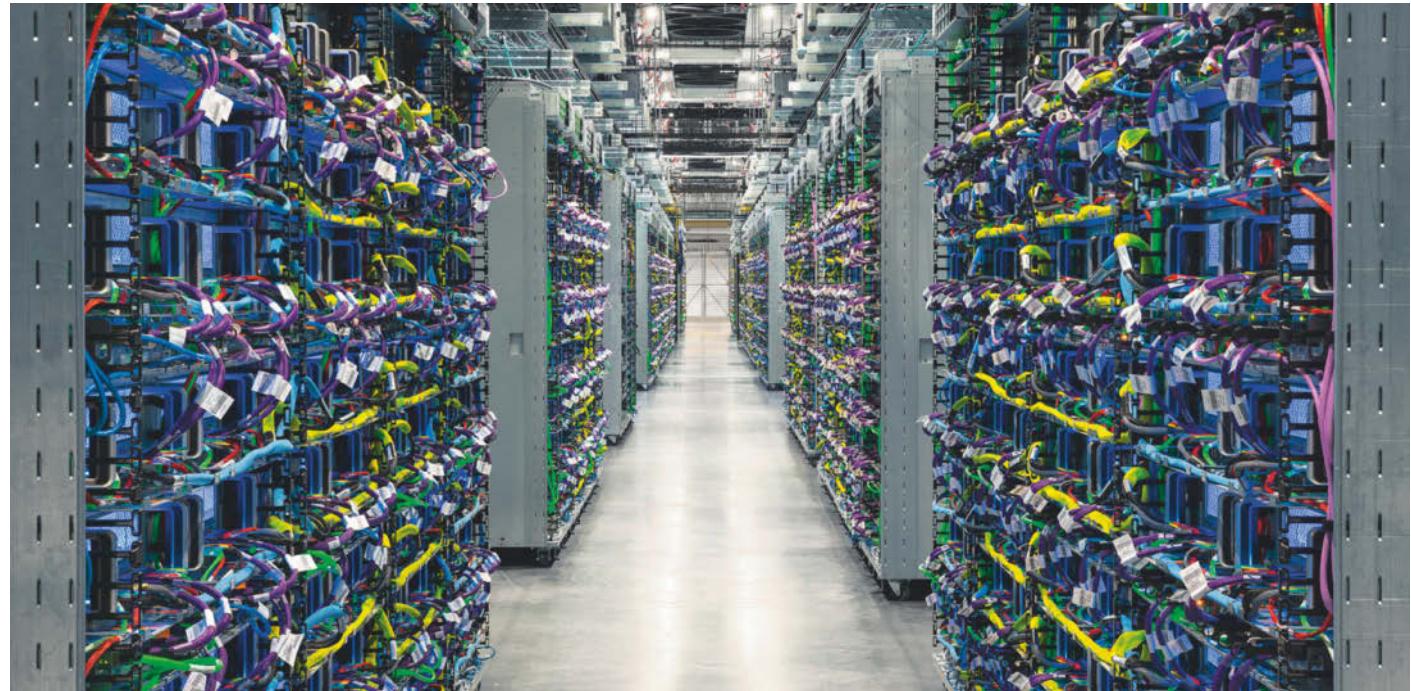
Dazu gehören Sprachmodelle wie ChatGPT und LLaMA sowie Bilderzeuger wie Midjourney und Dall-E. Andere generative Modelle komponieren etwa Musik oder produzieren Videos. Viele dieser generativen KI-Systeme benötigen besonders leistungsstarke Hardware. Der konkrete Energiebedarf hängt jedoch von vielen weiteren Faktoren ab, etwa von der Anzahl der jeweils genutzten Parameter oder generierten Tokens. Was mit letzteren gemeint ist, erklärt der Textkasten auf Seite 60.

Training oder Inferencing?

Die meisten KI-Dienste haben zwei wesentliche Lebensphasen, die völlig unterschiedliche Anforderungen an Rechen-

leistung, Arbeits- und Massenspeicher stellen, nämlich Training und Inferencing. Beim Training lernt das Modell anhand oft gigantischer Testdatensätze, was es später während der Nutzung genau tun soll. Diese Nutzungsphase nennt man Inferencing, gemeint ist die Anwendung eines fertig trainierten Modells auf Dateneingaben. Beispiele für Inferencing sind die erwähnten Bildoptimierer in Smartphones, das Benutzen der Google-Suchmaschine oder die Erzeugung von Bildern aus Texteingaben, den sogenannten Prompts.

Während der Trainingsphase kann man mit einem KI-Modell noch kein Geld verdienen. Es soll daher möglichst schnell gehen. Typischerweise nutzt man daher fürs KI-Training sehr viel stärkere und stromdurstigere Hardware als fürs Inferencing. Die Trainingsphase dauert ein paar Stunden bis Monate, während das Inferencing ein kontinuierlicher oder oft wiederholter Prozess ist. So greifen Tausende bis Millionen Anwender gleichzeitig auf ChatGPT-Dienste oder die Google-Suche zu, also auf Inferencing-Server. Beim Inferencing auf Endgeräten gibt es unterschiedliche Nutzungsarten. Der Bildverbesserer im Smartphone läuft etwa kurzzeitig nur dann, wenn die Kamera ein Foto schießt. Hübscht die Beauty-App hingegen das eigene Antlitz in einer Videokonferenz



Weltweit werden zurzeit riesige KI-Supercomputer aufgebaut, einige für jeweils mehr als 1 Milliarde Euro. Google rüstet die KI-Leistung seiner Cloud unter anderem mit dem selbst entwickelten Beschleuniger TPU v5e hoch.



Der Nvidia-Beschleuniger H100 (der obere, glänzende der beiden großen Chips) verwandelt unter Volllast 700 Watt in Abwärme.

auf, arbeiten die KI-Rechenwerke so lange durchgehend.

Viele generative KI-Modelle benötigen fürs Inferencing deutlich mehr Rechenleistung als andere KI-Modelle. Über den Energiebedarf, der für die Verarbeitung eines Prompts nötig ist, sagt das für sich genommen jedoch wenig aus. Vielmehr hängt der Energiebedarf davon ab, wie viele Rechenwerke wie lange arbeiten müssen, um die gewünschte Antwort zu liefern. Und dabei gibt es riesige Unterschiede.

Enorme Leistungsdichte

KI-Algorithmen laufen sowohl auf normalen x86- oder ARM-Prozessoren als auch auf speziellen Rechenbeschleunigern. Letztere werden oft verkürzt und leicht falsch auch „GPUs“ genannt, denn speziell die Nvidia-Beschleuniger wurden im Laufe der Jahre auf der Basis von PC-Grafikchips – also GPUs – entwickelt. Die bereits erwähnte H100 hat mit gewöhnlichen PC-Grafikkarten aber nicht mehr viel gemeinsam. Mit Stückpreisen von 25.000 bis rund 30.000 Euro und einer Leistungsaufnahme von bis zu 700 Watt ist sie wesentlich teurer und stromdurstiger als eine PC-GPU. Allerdings verarbeitet sie KI-Datenformate wie BF16 (BF16), TF32 oder Int8 um ein Mehrfaches schneller

und daher effizienter als Grafikkarten. Außerdem hat sie viel mehr und viel schnelleren Speicher. Während auf aktuellen PC-Grafikkarten der 600-Euro-Klasse 16 GByte GDDR6-RAM mit einer Datentransferrate von gut 500 GByte/s sitzt, kommt die H100 als SXM-Modul mit 80 GByte HBM3-Stapelspeicher, der mehr als 3,3 TByte/s schafft. Damit auch die Daten riesiger KI-Modelle in den schnellen Speicher passen, lassen sich H100-Module über superschnelle Verbindungen mit 900 GByte/s koppeln. Diese NVLink-Schnittstellen sind direkt in die KI-GPU eingebaut.

Weil Datentransport mit derartiger Geschwindigkeit viel Energie kostet, packen AMD, Intel und Nvidia einerseits immer mehr RAM auf jede Karte – bald mehr als 100 GByte – und quetschen immer mehr Rechenwerke in jeden Beschleuniger. Dadurch steigt die Leistungsaufnahme. Kam die 2020 vorgestellte Nvidia A100 noch mit 400 Watt aus, sind es bei der seit 2023 ausgelieferten H100 schon 700 Watt, also 75 Prozent mehr. Doch die H100 liefert wesentlich mehr Rechenleistung pro Watt als die A100, weil sie je nach Datenformat zwei- bis sechsmal so schnell rechnet und ihr Speicher doppelt so groß ist. 2024 kommt eine H100-Variante mit 141 GByte noch schnellerem HBM3e-RAM.

Nvidias hauseigener KI-Rechenknoten DGX H100 vereint acht H100-Module mit zwei Intel-Prozessoren, kostet mindestens 300.000 Euro und schluckt unter Volllast rund 10 Kilowatt. In einen Serverschrank (Rack) für Rechenzentren kann man nur vier DGX H100 mit insgesamt 32

H100 einbauen, weil die kumulierte Leistungsaufnahme von über 40 Kilowatt pro Rack Stromversorgung und Kühlung vieler Rechenzentren überfordert. Noch vor einigen Jahren waren pro Rack eher 3 bis 8 Kilowatt üblich, erst wenige Colocation-Rechenzentren bieten mehr als 20 Kilowatt an. Um noch mehr Abwärme zu bewältigen, baut man in die rückseitigen Schranktüren Luft-Wasser-Wärmetauscher ein. Sie bieten mehr Flexibilität bei der Hardware-Auswahl und sind oft billiger als die direkte Wasserkühlung der Chips, die aufwendige Konstruktionen aus Rohren und Schläuchen verlangt.

Die Effizienz der KI-Beschleuniger ist innerhalb der vergangenen Jahre um ein Mehrfaches schneller gewachsen, als es die Halbleiter-Strukturverkleinerungen gemäß „Moore’s Law“ bewirken könnten. Laut Nvidia-Chef Jensen Huang nahm die KI-Rechenleistung pro System in den vergangenen 10 Jahren um den Faktor 1 Million zu. Das liegt an neuartigen Rechenwerken, den erwähnten KI-Datenformaten wie BF16, an neuen Chiplet-Verfahren wie HBM-Stapelspeicher, an stark verbesserten Algorithmen sowie auch an Software-Optimierungen. Für die Zukunft erwartet Nvidia ähnliche Steigerungsmöglichkeiten – aber diese Einschätzung teilen nicht alle Experten.

Aus dem Blickwinkel des KI-Stromdursts gibt es außerdem ein Henne-Ei-Problem: Der Siegeszug der generativen KI wurde durch die enormen Steigerungen bei Rechenleistung und Effizienz überhaupt erst möglich. Der KI-Stromverbrauch der Zukunft kann daher nur schrumpfen, wenn die höhere Effizienz

KI-Token: Wortschnipsel für Maschinen

Ein KI-Sprachmodell versteht Sprache nicht wie ein Mensch, sondern berechnet aus eingespeisten Textschnipseln bloß neue. Dazu zerlegt das System die Anfrage (den Prompt) in sogenannte Token. Ein Wort aus der englischen Sprache belegt im Mittel rund 1,3 Token, französische oder spanische Wörter hingegen mehr als 2 Token. Die neu berechneten Token wandelt das KI-System vor der Ausgabe wieder in Wörter um.

Je nach Art und Größe des Sprachmodells benötigt die Verarbeitung jedes

Token unterschiedlich viel Rechenleistung; die dafür aufgewendete Energie hängt wiederum von der Effizienz der jeweils verwendeten Rechenwerke ab. Der ChatGPT-Entwickler OpenAI rechnet die Benutzung seiner KI-Modelle nach der Anzahl der verarbeiteten Token ab. Dabei addiert OpenAI die Token des Prompts mit denen der Antwort. Je nach KI-Modell verlangt OpenAI pro Token unterschiedliche Preise. Diese Unterschiede spiegeln unter anderem den pro Token nötigen Bedarf an Rechenleistung und Energie.

Geschätzter Energieaufwand fürs Training generativer KI-Sprachmodelle

| KI-Modell | Laufzeit auf 256 A100-GPUs [Tage] | GPU-Stunden | Energiebedarf [kWh] | Preis USA (0,08 US-\$/kWh) | Preis DE (0,27 €/kWh) |
|------------|-----------------------------------|-------------|---------------------|----------------------------|-----------------------|
| GPT3-1.3B | 0,14 | 860 | 344 | 28 US-\$ | 93 € |
| GPT3-6.7B | 2,32 | 14.254 | 5.702 | 456 US-\$ | 1.539 € |
| GPT3-30B | 35,98 | 221.061 | 88.424 | 7.074 US-\$ | 23.875 € |
| GPT3-70B | 176,55 | 1.084.723 | 433.889 | 34.711 US-\$ | 117.150 € |
| LLaMA-7B | k. A. | 82.432 | 32.973 | 2.638 US-\$ | 8.903 € |
| LLaMA-33B | k. A. | 530.432 | 212.173 | 16.974 US-\$ | 57.287 € |
| LLaMA-65B | k. A. | 1.022.362 | 408.945 | 32.716 US-\$ | 110.415 € |
| OPT-175B | k. A. | 809.472 | 323.789 | 25.903 US-\$ | 87.423 € |
| BLOOM-175B | k. A. | 1.082.880 | 433.152 | 34.652 US-\$ | 116.951 € |

Daten von MosaicML (GPT-3) und Meta (LLaMA, OPT, BLOOM); Industriestrompreise geschätzt

nicht wieder aufgezehrt wird, beispielsweise durch noch riesigere Modelle.

Energie und Arbeit

Die zur Lösung einer Rechenaufgabe verwendete Energie ist das Produkt aus der tatsächlichen Leistungsaufnahme und der Dauer der Berechnung. Nimmt ein Server beispielsweise kontinuierlich 1000 Watt (1 Kilowatt, kW) Leistung auf, dann hat er nach 1 Stunde (1 h) genau 1 Kilowattstunde (kWh) sprichwörtlich verbraucht. Kürzer ausgedrückt: Energie ist Leistung mal Zeit.

Das Training sehr großer generativer KI-Modelle kann Cluster mit mehreren Hundert Rechenbeschleunigern über Monate beschäftigen. Einige Studien beziffern daher den jeweils nötigen Aufwand in „Grafikchipstunden“ (GPU hours). Das ermöglicht eine grobe Einschätzung des Energiebedarfs für das Training, weil die jeweilige Leistungsaufnahme von Nvidia A100 und H100 bekannt ist und man bei typischen KI-Servern mit jeweils acht solcher GPU-Module den Strom für die restliche Hardware vernachlässigen kann.

Die US-Firmen MosaicML und Meta (Facebook) nennen Daten für Trainingsläufe der KI-Modelle GPT-3, LLaMa und BLOOM, siehe ct.de/yyzq und die Tabelle auf Seite 61 oben. Der Rechenaufwand und somit die Zahl der zum Training nötigen GPU-Stunden steigt überproportional mit der Anzahl der Parameter, die das Modell berücksichtigen soll. GPT-3 mit 70 Milliarden Parametern hat um den Faktor 54 mehr als GPT-3 mit 1,3 Milliarden, der Rechenaufwand steigt aber um den Faktor 1260. Im Extremfall verbrauten Tausende A100-Karten beim Training zusammen 434.000 kWh.

Der hohe Energieaufwand verursacht enorme Kosten. MosaicML nennt fürs

GPT-3-Training je nach Modell Preise von 2000 bis 2,5 Millionen US-Dollar. Alleine schon wegen der Laufzeit von bis zu mehreren Monaten je nach verfügbarer Hardware kann man ein großes KI-Modell nur in größeren Abständen neu trainieren.

Inferencing-Energiebedarf

Eine wirtschaftlich rentable, gehostete KI-Dienstleistung muss deutlich mehr Geld einbringen, als der Strom fürs Training und das Inferencing kostet. Energiekosten fürs Training fallen in größeren Abständen blockweise an. Doch der Strom fürs Inferencing fließt kontinuierlich durch die KI-Server, auf die die Nutzer zugreifen. Für den Energiebedarf vieler KI-Dienste ist daher langfristig gesehen das Inferencing wichtiger. Es gibt Schätzungen, dass 90 Prozent des Energiebedarfs großer Sprachmodelle aufs Inferencing entfallen, nur 10 Prozent aufs Training.

Weil es viele unterschiedliche Sprachmodelle gibt und diese auch noch in jeweils unterschiedlichen Varianten und Größen, variiert der Rechenaufwand fürs Inferencing enorm. Wie viel Energie die Berechnung verschlingt, hängt dann auch noch von der jeweils verwendeten Hardware ab. Die Tabelle auf Seite 61 unten zeigt anhand von Daten der Website ML

Leaderboard, wie weit der Energiebedarf je nach konkreter Konfiguration auseinanderliegt.

Die Tabelle nennt den durchschnittlichen Energiebedarf für die Berechnung von jeweils 1 Million Token für ein bestimmtes Modell. Wenn ein KI-Modell ein Prompt, das aus rund 50 englischen Wörtern besteht, mit 110 englischen Wörtern beantwortet, verarbeitet es dabei insgesamt etwa 120 Token. Mit einem A100-Beschleuniger würde folglich 1 kWh Strom reichen, um 8300 Anfragen an das Modell LLaMA-7B zu beantworten; bei LLaMA-13B reicht es nur für die Hälfte. Die älteren V100-Beschleuniger brauchen bei großen Modellen deutlich mehr Energie pro Token und die A40-Karten halten bei sehr großen Modellen nicht mehr mit.

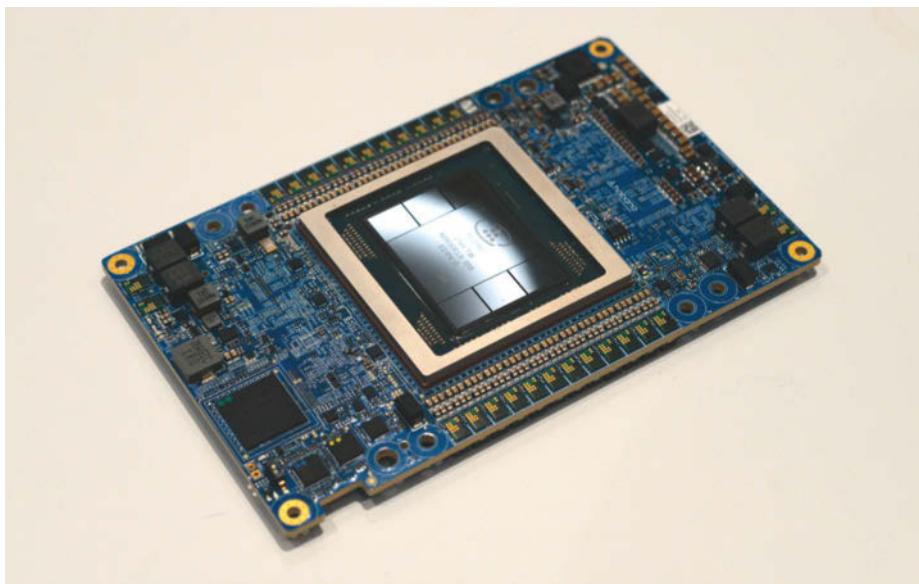
Der Energiebedarf bei der Verarbeitung von KI-Algorithmen hängt auch stark vom jeweils verwendeten Softwareunterbau ab. Viele KI-Anwendungen setzen auf sogenannte Frameworks wie TensorFlow oder Caffe auf, für die die Hardwarehersteller wiederum ihre jeweiligen Treiber optimieren und angepasste Bibliotheken liefern. Dabei gibt es riesige Unterschiede, wie gut ein bestimmter KI-Algorithmus das Rechenpotenzial der Hardware nutzen kann. Ein Beispiel dafür liefern unsere Experimente mit dem Open-Source-Bilderverzweiger Stable Diffusion, über die wir in [1] berichtet hatten. Um das Bild „Glas-Kugel“ mit 1680×1024 Pixeln zu berechnen, brauchte Stable Diffusion mit derselben Nvidia-Grafikkarte GeForce 3060 zwischen 96 und 331 Sekunden (Faktor 3,5) – je nachdem, welche Treiber und welche Optimierungen zum Einsatz kamen.

Wegen der zurzeit stürmischen Fortentwicklung der KI-Modelle und der dafür verwendeten Hardware – es gibt viele unterschiedliche KI-Beschleuniger [2] – ist es bisher also unmöglich, die für einen bestimmten KI-Dienst einigermaßen genau einzuschätzen. Dazu müssten die jeweili-

Energieaufwand fürs Inferencing der Text-KI LLaMA

| KI-Modell | Beschleuniger | Token/s | kWh/1 Million Token |
|-------------|---------------|---------|---------------------|
| LLaMA-7B | A100 | 27,5 | 3,4 |
| LLaMA-7B | V100 | 26,1 | 3,6 |
| LLaMA-7B | A40 | 25,8 | 5,7 |
| LLaMA-13B | A100 | 23,2 | 7,6 |
| LLaMA-13B | V100 | 17,3 | 12,4 |
| LLaMA-13B | A40 | 15,7 | 22,9 |
| LLaMA 2-13B | A40 | 17 | 70,9 |

Datenquelle: ML Leaderboard (ml.energy/leaderboard)



Nvidia verdient sich mit KI-Beschleunigern eine goldene Nase, also versuchen das auch andere Firmen wie Cerebras, Graphcore oder Groq. Die Intel-Tochter Habana schickt den abgebildeten Gaudi 2 ins Rennen.

gen Anbieter konkrete Daten aus der Praxis liefern. Die liegen derzeit kaum vor und wären jeweils innerhalb weniger Monate überholt, wenn neue KI-Modelle auf neuer Hardware laufen.

Womit vergleichen?

Selbst wenn Zahlen für den KI-Strombedarf vorlägen, stellt sich die Frage, womit man vergleichen möchte. Denn auch bei bisherigen Onlinediensten ist der spezifische Energiebedarf sehr unterschiedlich und meistens unbekannt. Der Zugriff auf einen simplen Zeitserver schluckt weniger Strom als eine komplizierte Datenbankabfrage oder gar Game-Streaming, bei dem das Spiel auf einem Server mit Grafikkarten läuft. Der weitaus größte Anteil des Internetdatenverkehrs entfällt seit Jahren auf Videodienste; und dabei wiederum schlucken die Transportnetze mehr Energie als die Rechenzentren, in denen Server die Videodaten speichern und ausliefern.

An vielen Alltagsanwendungen sind zahlreiche Server beteiligt. Das belegt ein Blick auf die große Anzahl der Trackingdienste, die etwa kommerzielle News-Websiten einbinden. Ein scheinbar simpler Zugriff auf eine solche Seite löst Aktivitäten auf mehreren Dutzend Servern aus: Im Hintergrund laufen dann in Echtzeit automatisierte Bieterverfahren um die Werbeplätze ab, Datenbanken werden durchforscht, Bilder und Videos geladen und ausgespielt. Es ist kaum zu beurteilen, ob der eine oder andere KI-Dienst dabei einen nen-

nenswerten Anteil am Energieaufwand hat.

Im Vergleich zu anderen Energieverbrauchern wirkt der KI-Bedarf eher bescheiden. Einerseits könnte man mit den für einen besonders aufwendigen KI-Trainingslauf nötigen 434.000 kWh zwar fast 30 Einfamilienhäuser ein Jahr lang heizen, die dazu je 15.000 kWh Gas, Öl oder Strom benötigen. Andererseits entsprechen

434.000 kWh etwa 36 Tonnen Kerosin, die ein Airbus A330-300 für einen einzigen Mittelstreckenflug von Frankfurt nach Dubai verbrennt.

Einige irreführende oder schlicht falsche Einschätzungen zum IT-Energiebedarf aus der Vergangenheit zeigen, wie schwierig solche Prognosen sind. Eine Zahl von Google aus dem Jahr 2009 zum Energiebedarf einer Suchanfrage etwa dient noch bis heute auch seriösen Medien als Anhaltspunkt – 14 Jahre und mehr als fünf Hardwaregenerationen später. Dabei war sie schon damals Quatsch, weil sie den Energiebedarf der Netzwerke vernachlässigt [3]. Und ein Forscher, der 2005 vor extrem stark steigendem Strombedarf von Rechenzentren warnte, konnte sich 15 Jahre später über eine deutlich günstigere Entwicklung freuen. Diese Hinweise sollen den immensen KI-Stromdurst nicht beschönigen, sondern liefern Indizien dafür, dass aktuelle Zahlen kaum Vorhersagen ermöglichen. Für 2022 melden Google und Microsoft jedenfalls stark gestiegenen Wasserverbrauch für die Kühlung ihrer Rechenzentren: Der von Microsoft steigt um 34 Prozent, der von Google um rund 20 Prozent.

Strombegrenzer

Mehrere Faktoren begrenzen den Energieverbrauch von KI-Diensten, an erster Stel-

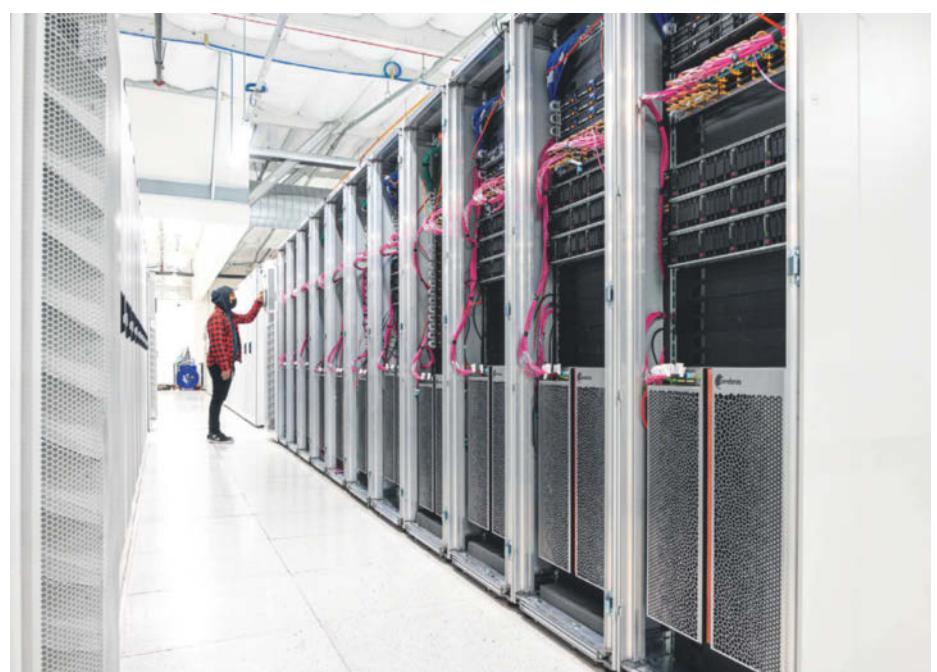


Bild: Cerebras

In jedem dieser Racks sitzt unten ein KI-System vom Typ Cerebras CS-2, das rund 20 Kilowatt aufnimmt. Zusammen mit den oben eingebauten x86-Servern muss jeder Schrank rund 30 Kilowatt Abwärme verdauen.

le die Kosten. Die erwähnten Nvidia-DGX-Systeme kann sich nicht jede Firma leisten und auch die Mietpreise für H100-Instanzen auf Cloud-Servern sind hoch. Die Stromkosten für den oben beschriebenen Trainingslauf mit 434.000 kWh betragen über 30.000 Euro, selbst wenn das System in den USA steht, wo man Industriestrom für 8 US-Cent/kWh bekommt; hierzulande bei 27 Euro-Cent/kWh wären es fast 120.000 Euro. Mit riesigen Sprachmodellen experimentieren daher zurzeit vor allem Großkonzerne wie Microsoft, Google und Meta sowie gut finanzierte KI-Start-ups, die Risikokapital verbrennen.

Momentan befinden wir uns in einer wilden Experimentierphase. Den dafür in Kauf genommen Aufwand kann man nicht einfach in die Zukunft extrapoliieren. Denn generative KI wird sich nur dann weit verbreiten, wenn es Geschäftsmodelle gibt, die die Kosten für Energie und Hardware tragen. Zurzeit treten die Firmen dafür meist noch in Vorleistung, um den Aufsprung nicht zu verpassen.

Die Effizienz der KI-Rechenwerke steigt rasch an, alleine schon durch Moore's Law, also durch die Fortschritte bei der Halbleitertechnik. Man kann grob damit rechnen, dass sich der Energiebedarf pro Rechenleistung alle zwei Jahre halbiert. Zurzeit geht es durch zusätzliche Optimierungen bei KI-Datentypen und Algorithmen sogar schneller. So zeigt sich, dass sich der Inferencing-Rechenbedarf bei manchen Modellen durch geschickte Quantisierung der Parameter enorm verringern lässt. Dann genügt es für ähnliche Qualität beispielsweise, mit Int8- statt mit FP16-Zahlen zu rechnen [4].

Allerdings gibt es einen starken Rebound-Effekt: Immer aufwendigere Modelle zehren den Zuwachs an Effizienz wieder auf. Schaut man auf die IT-Entwick-

lung der vergangenen Jahrzehnte zurück, ist aber zu erwarten, dass mit zunehmender Reifung der Technik irgendwann der spezifische Energiebedarf pro Prompt sinkt.

Kommerzielle KI-Dienste laufen praktisch ausschließlich in sehr effizienten Cloudrechenzentren, von denen viele bereits regenerativ erzeugten Strom beziehen. Das mindert die Umweltlasten durch KI. Viele große Cloudanbieter bauen eigene Wind- und Solarparks aber auch deshalb, um mit möglichst billig erzeugtem Strom ihren Profit zu steigern.

Außerdem müssen die fetten KI-Server erst einmal hergestellt und um den halben Globus transportiert werden. Die Fertigung der vergleichsweise großen Siliziumchips der Beschleuniger sowie der vielen RAM-Bausteine benötigt wiederum viel Strom sowie aufwendig produzierte Rohmaterialien.

Fazit

Angesichts der explodierenden Verkaufszahlen von Nvidia-Rechenbeschleunigern lässt sich zuverlässig vorhersagen, dass der Stromverbrauch für KI weiter stark steigen wird. Der Börsenwert von Nvidia schoss auf fast 1,2 Billionen US-Dollar; der KI-Rausch ist ein lukratives Geschäft, das man sich nicht von Bedenken wegen des Energiehunders vermiesen lassen will. Allerdings verlagern viele Cloudanbieter ihre Einkäufe weg von „normalen“ Servern hin zu KI-Maschinen. Der Energiebedarf von Cloudrechenzentren wächst deshalb zurzeit vor allem durch KI.

Das kann nicht ewig weitergehen, schon weil nur wirtschaftlich sinnvolle KI-Dienste langfristig überleben werden. Darum arbeiten Forscher und Entwickler nicht nur mit Hochdruck an immer neuen KI-Modellen, sondern auch an der Steigerung der Effizienz. Doch Energiekosten



Bild: Vertiv/Liebert

Um mehr als 30 Kilowatt Hitze aus einem Server-Rack abzuleiten, kann man in die hintere Schranktür Luft-Wasser-Wärmetauscher einbauen.

sollten nicht die einzige Leitplanke sein, die den KI-Ressourcenbedarf in der Spur hält. Für Verbraucher und Umwelt wäre es hilfreich, würden gesetzliche Regelungen die Betreiber solcher Dienste zur Transparenz beim Energiebedarf zwingen. Dann könnten die Kunden zum jeweils effizientesten oder sparsamsten Anbieter wechseln. Das wäre ein starker Anreiz, um den Effizienzwettlauf zu beschleunigen.

(ciw@ct.de) ct

Literatur

- [1] Carsten Spille, Giga-Chips, Die besten Grafikkarten für Stable Diffusion AI, ct' 9/2023, S. 64
- [2] Christof Windeck, Intelligenzbestien, So funktionieren Rechenwerke für KI-Algorithmen, ct' 11/2023, S. 124
- [3] Christof Windeck und Christian Wölbert, Klimakiller oder Klimaretter?, Der CO₂-Fußabdruck der IT, ct' 6/2020, S. 62
- [4] Pina Merkert, Byte-Intelligenz, Neuronale Netze für TensorFlow-Lite auf 8-Bit-Integer umrechnen, ct' 22/2019, S. 182

Studien zum KI-Energiebedarf:
ct.de/yyzq

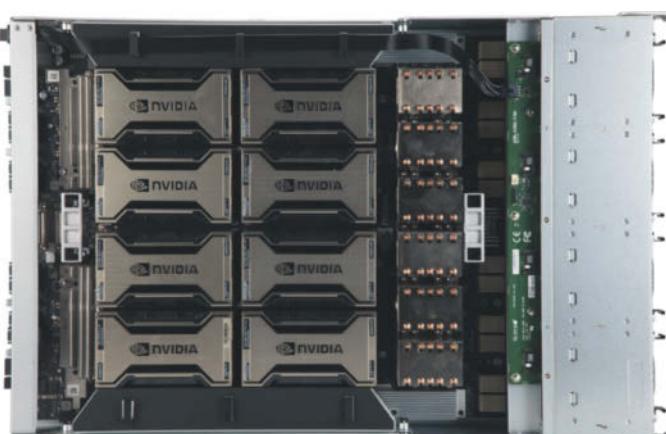


Bild: Supermicro

Auf unsicherem Terrain

Trends bei Software-Entwicklung, Sicherheit und Digitalisierung

Der Marktdominanz amerikanischer Internetgiganten versucht Europa mit rechtlichen Schutzschirmen und eigenen Bezahlsystemen zu begegnen. Die Digitalisierung kommt derweil nur schleppend voran – vielleicht ist das auch gut so, denn halbfertige Systeme sind eine allzu leichte Beute für die immer professioneller agierenden Cyberkriminellen.

Von Hartmut Gieselmann

Bei Software schreiben die Hersteller den Servicegedanken groß. Schließlich lässt sich mit SaaS (Software as a Service) nicht nur einmalig, sondern monatlich Geld verdienen. Was in vielen Anwendungsbereichen inzwischen Standard ist, hält auch bei Betriebssystemen immer mehr Einzug, wie Microsofts Cloud-Anbindung von Windows zeigt, die wir ab Seite 66 untersuchen. Auch kommerzielle Anbieter entdecken zunehmend die Vorteile von Open Source. Dabei entfernen sie sich immer weiter von der ursprünglichen Idee, dass Software für alle frei zugänglich sein sollte (Seite 70).

Mit all diesen Diensten steigt auch die Zahl der Accounts, die jeder heute verwalten muss. Hunderte von Passwörtern kann sich ohnehin niemand mehr merken, deshalb werden sie durch Passkeys ersetzt (Seite 76). Solche Umwälzungen bringen natürlich auch Unsicherheiten mit sich, die sich die immer professioneller werden den Cyberkriminellen zunutze machen, wie wir ab Seite 80 berichten.

In der EU sieht man derweil den freien Markt durch die Dominanz der US-Konzerne bedroht. Erste Gesetzesinitiativen scheinen bereits Wirkung zu zeigen (Seite 86). So will man bei den Bezahlsystemen unabhängig von Visa und Mastercard werden und bastelt an europäischen Alternativen (Seite 94).

Meta & Co. halten neue Nachrichtendienste wie Threads vorerst aus Europa fern, weil sie hier nicht so viele Daten sammeln dürfen, wie sie gerne

möchten. Das schafft Chancen für alternative Anbieter und bringt Bewegung bei den Social-Media-Plattformen (Seite 98). Beim Verbraucherschutz setzt sich der oben erwähnte Servicegedanke leider nicht durch. Das zeigt unser Blick in die Wüste auf Seite 92.

Bei der Digitalisierung des Gesundheitswesens (Seite 104) und der Verwaltung (Seite 108) steckt Deutschland weiter im Stau. Die Umstellungen verursachen enorme Kosten. Und die sind angesichts von Wirtschaftskrise und Aufrüstungsprogrammen nicht mehr so leicht zu stemmen. Langfristig erhofft man sich zwar Kosten- und Zeitersparnis durch Rationalisierungen. Doch auf absehbare Zeit werden Ärzte und Verwaltungsangestellte davon nicht profitieren. Hier sind neue Ansätze gefragt.





| | |
|---|------------------|
| Trends bei Software, Sicherheit und Digitalisierung | Seite 64 |
| Windows: Zukunft in der Cloud | Seite 66 |
| Open Source: Die Revolution frisst ihre Kinder | Seite 70 |
| Passkeys: Mehr Sicherheit statt Passwörter | Seite 76 |
| Cyberkriminalität: Erpressen wie die Profis | Seite 80 |
| EU: Neue Digitalgesetze sollen freien Markt sichern | Seite 86 |
| Verbraucherschutz: Servicewüsten wachsen an | Seite 92 |
| Bezahlsysteme: Europäische Alternativen | Seite 94 |
| Social Media: Völkerwanderungen auf neue Plattformen | Seite 98 |
| Gesundheitssystem: Apps statt Ärzte | Seite 104 |
| Digitalisierung: Baustellen der Verwaltung | Seite 108 |



Bild: KI Midjourney | Collage c't

Wie Windows in die Cloud wandert

Sortiert: Windows 365 Cloud-PC, Azure Virtual Desktop und die Zutaten

Im Sommer 2023 pfiffen die Newsticker der Welt von den Dächern, dass das Betriebssystem, auf dem der Erfolg Microsofts gründet, bald in die Cloud wandern soll. Was ist da dran?

Von Peter Siering

Allass für die Berichte im Sommer 2023 war eine interne Präsentation von Microsoft aus dem Jahr 2022, die offenbar erst mit Verzögerung den Weg in

die Öffentlichkeit gefunden hat. Das darin skizzierte Produkt ist für Firmenkunden längst Realität geworden: Als Windows 365 mieten Unternehmenskunden in der Microsoft-Cloud Azure ausgeführte Windows-Instanzen oder konfektionieren solche als Azure Virtual Desktop selbst.

In der Präsentation und in den kolportierten Berichten war die Rede davon, dass Windows „komplett“ in die Cloud wandern soll. Das Wort „komplett“ kann man technisch oder marktorientiert lesen. Die Aussage, dass Microsoft alle laufenden Windows-Instanzen zukünftig dort ausführen will, scheint eine überspitzte Darstellung, die der heutigen

Newsaufmerksamkeitsökonomie geschuldet ist.

Die wirkliche Neuigkeit war, dass Microsoft darüber nachdenkt, sein Cloud-Windows auch Endkunden zugänglich zu machen. Bisher steht dieses Angebot nämlich nur Unternehmen offen und die nutzen das weidlich. Das Folgende erklärt die Technik, wirft einen Blick auf den Markt und betrachtet die Kosten. Das hilft bei der Einschätzung, ob und was es mit Windows aus der Cloud auf sich hat.

Zwei Dinge sind entscheidend für ein Cloud-Windows: Zum einen muss es dort eine Art Computer geben, der eine Instanz des Betriebssystems ausführt. Zum anderen braucht es eine Möglichkeit, die Aus- und Eingaben dieses Computers lokal darzustellen. Das sind keine neuen Problemstellungen, sondern Dinge, für die es auch unter Windows seit langem schon bewährte technische Lösungen gibt, Virtualisierung und Fernzugriff – wir erlauben uns einen Rückblick auf beides:

Rückblick: Windows aus der Ferne

Schon 1998, also zu Zeiten von Windows NT, dem technischen Vorläufer aller heutigen Windows-Versionen, führte Micro-

soft eine spezielle Windows-Server-Variante namens „Terminal Server“ ein. Sie erlaubte es mehreren Nutzern gleichzeitig über das Netzwerk interaktive Sitzungen herzustellen. In diesen Sitzungen sehen Sie einen regulären Desktop und können auf dem Terminal Server installierte Anwendungen nutzen.

Hoffähig hatte diese Technik gar nicht Microsoft selbst gemacht, sondern die Firma Citrix. Die hatte schon für den Windows-Konkurrenten OS/2 ähnliche Erweiterungen gebaut und sich dafür gegen den Willen IBMs bei Microsoft die Quelltexte des Betriebssystems besorgt. Mit NT4 hatte Citrix Ähnliches vor, aber Microsoft kaufte der Firma kurzerhand die Rechte an der Idee ab und veröffentlichte sie als eigenes Produkt mit dem Terminal Server.

Das von Citrix erdachte Protokoll, mit dem die Clients auf den Server zugriffen (ICA) kaufte Microsoft nicht von Citrix, sondern leitete ein eigenes von Netmeeting ab; einer Entwicklung, die Microsoft ebenfalls aufgekauft hatte. So entstand schließlich das Remote Desktop Protocol (RDP), das noch heute für den Zugriff auf entfernte Windows-Sitzungen verwendet wird. Es hat seitdem allerdings diverse Optimierungen erfahren. In großen Installationen mit vielen Tausend Nutzern sitzt Citrix weiter mit am Tisch, weil sein ICA-Protokoll Bandbreite schont und so in größeren Installationen besser als RDP funktioniert.

Ein RDP-Server steckt als Remote-Desktop inzwischen in allen „professionellen“ Windows-Versionen, also ab Pro und in den Servern, dort allerdings nur für Admins. Clients gibt es von Microsoft auch für viele andere Plattformen, unter anderem für macOS und in einer Variante für HTML5-fähige Browser – auch andere bauen passende Software. Ein RDP-Server findet sich zum Beispiel in den grafischen

Versionen von Pi OS, also dem Raspi-Betriebssystem.

RDP erlaubt nicht nur die Darstellung typischer Windows-Grafikaufrufe, sondern verbindet auch Maus, Tastatur, Massenspeicher, Audio- und USB-Geräte sowie Drucker des lokalen Geräts mit der entfernten Windows-Instanz. In Grenzen ist der Zugriff auf entfernte GPU-Ressourcen möglich. Mit dem in Windows 10 eingeführten RDP 10.0 nutzt Microsoft den Videocodec H.264, um eine performantere Darstellung zu erreichen, seitdem ergibt es auch technisch Sinn, von Streaming zu sprechen.

Wie weit die Techniken rund um RDP in Windows Einzug gehalten haben, zeigt sich auch an anderer Stelle: Wenn Windows eine virtuelle Maschine (VM) unter Hyper-V ausführt und der Zugriff darauf als „Erweiterte Sitzung“ konfiguriert ist, dann nutzt es RDP für den Zugriff auf die Konsole der VM. Auch Oracles VirtualBox verwendet das in der kommerziellen Ausgabe, um den Zugriff auf die Konsole der VMs zu vereinfachen.

Rückblick: Windows virtuell

Virtualisierung hat Windows vergleichsweise spät erlernt. Seit Windows 8 und Server 2008 steckt ein optional nutzbarer Hypervisor in Windows. Die als Hyper-V eingeführte Funktion verfrachtet dabei Windows selbst in eine VM, die dann parallel zu anderen VMs läuft, aber eine privilegierte Rolle übernimmt, weil sie unter anderem die Treiber zum Ansteuern der Hardware ausführt. Zwischenzeitlich nutzt Microsoft Mini-VMs für viele Zwecke in Windows, etwa für die Linux- und Android-Subsysteme, für Sandboxen und, um Anmeldeinformationen zu schützen (Credential Guard).

Das Modell, das Microsoft gewählt hat, ähnelt der Technik, die die Xen-Entwickler erdacht hatten. Da läuft primär ein Linux in einer privilegierten VM, dort Dom0 genannt. Übrigens: Xen landete zwischenzeitlich im Portfolio von Citrix. Sei es drum. Microsoft nutzt wie Xen paravirtualisierte Treiber, um die eine Virtualisierung bremende Emulation bei Netzwerk- und Plattenzugriffen zu eliminieren.

Eine Zeit lang führten die Microsoft-Entwickler sogar die Liste der zum Linux Kernel Beitragenden an, als sie paravirtualisierte Treiber ergänzten, damit Linux auf Hyper-V möglichst rund lief. Das war unter anderem wichtig, um in der Microsoft-eigenen Cloud Azure den Kunden auch Linux

ct kompakt

- Windows als Betriebssystem ist schon lange in der Cloud präsent, sei es in VMs oder auf dem Server.
- Die dazu nötigen Fähigkeiten hat es bereits im letzten Jahrzehnt erlernt und in der Praxis bewiesen.
- Die momentanen Angebote sind auf Unternehmenskunden gerichtet.

nuPro AS-2500: mit Dolby Digital (AC3), DTS Digital Surround, Analog- und Digitalanschlüssen, 200 Watt Musikleistung, Mehrschichtlack seidenmatt Weiß oder Schwarz, Stoßfeste in Schwarz, 495,- Euro (inkl. 19% MwSt., ggf. zzgl. Versand) Nuert electronic GmbH, Nuertstr.1, 73529 Schwäbisch Gmünd, 30 Tage Rückgaberecht · Webshop: nuert.de · Beratung 07171 8712-0

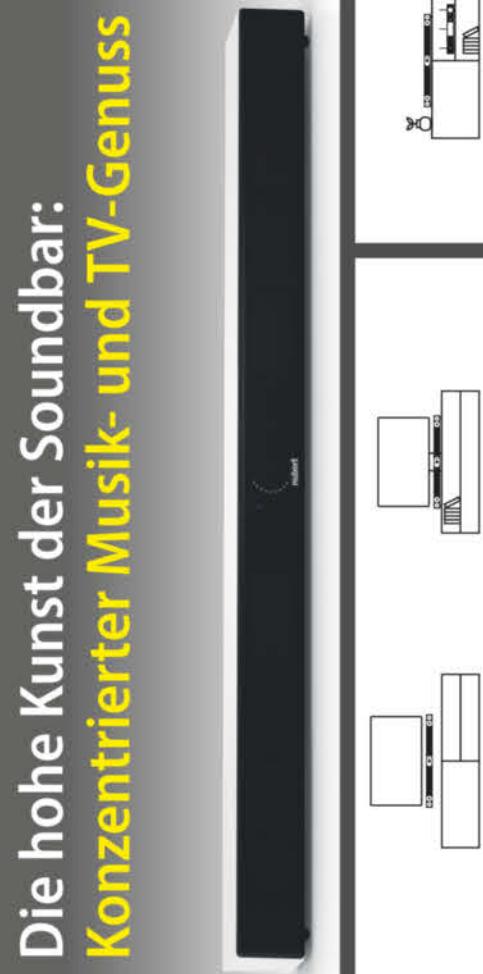
Direkt + günstig vom Hersteller

nuert
®
EHRLICHE LAUTSPRECHER

nuPro AS-2500

- Druckvoll – mitschlankem Gehäuse (nur 8 x 14 x 105 cm)
- Für packende TV-Erlebnisse und vollendeten Musikgenuss
- Brillanter HiFi-Sound – vielseitige, individuelle Klangeinstellungen
- Voice+ für bestes Sprachverständnis
- Breite Klangbühne mit Widesound
- Für Wandmontage oder Lowboard
- nuConnect trx ready: Steckplatz für drahtlose Subwooferanbindung

Die hohe Kunst der Soundbar:
Konzentrierter Musik- und TV-Genuss



anbieten zu können. Im Jahr 2019 plapperete dann ein Microsoft-Mitarbeiter aus, dass Linux in Azure verbreiteter sei als Windows und auch für Infrastrukturdiene in der Cloud dort verwendet wird.

Mit jeder Servergeneration hat Microsoft Hyper-V hochgerüstet. Im Server 2012 lernte es, VMs von Server zu Server zu replizieren und optimierte Datenübertragungen. Im Server 2016 kamen unter anderem verschachtelte Virtualisierung, also VMs in VMs, der Durchgriff auf ausgewählte PCI-Express-Geräte, Container und voneinander mittels Verschlüsselung abgeschirmte VMs hinzu. Mit Server 2019 deckt die Abschirmung auch Linux ab und ersetzt die VMs verschlüsselt auf Wunsch Netzwerkverkehr zwischen VMs.

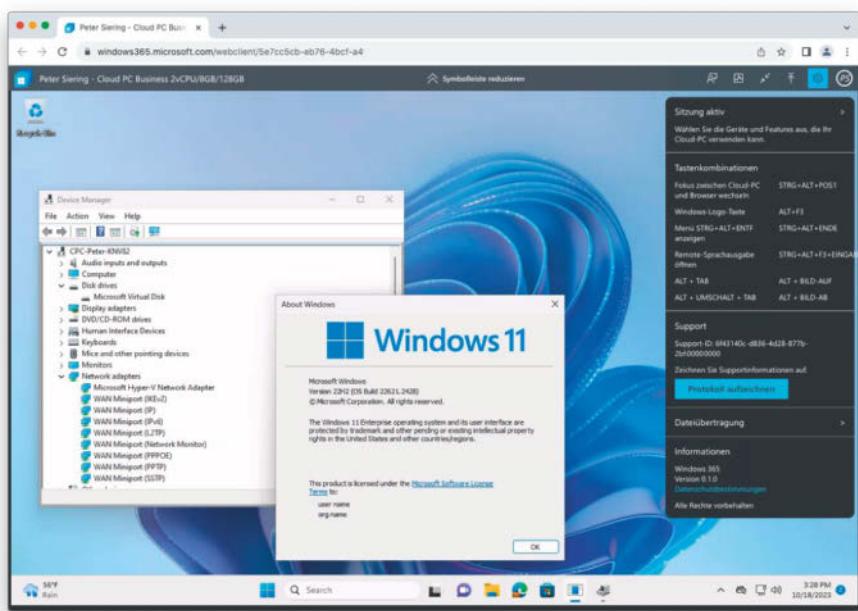
Damit hat Microsoft die Weiterentwicklung von Windows Server auf den Betrieb komplexer Infrastruktur ausgerichtet. Das gilt für die Funktionen allgemein und die Virtualisierung speziell. Es scheint also in keiner Weise fragwürdig, dass man mit Windows große Massen von virtuellen Systemen hosten und verwalten kann – das dürfte Microsoft selbst in Azure auch tun.

Der entstehende Markt

Noch Jahre bevor von einem Cloud-Windows die Rede war, hat Microsoft die Virtualisierung und die Fernbedienung verwendet, um im Markt für „Virtual Desktop Infrastructure“ (VDI) mitzumischen und die Technik so gründlich geschäftlich ausgeweidet: Statt jedem Nutzer einen für viele Anwendungen überdimensionierten eigenen PC auf den Tisch zu stellen, rüstete man die Rechenzentren auf. Den meisten Nutzern fällt kaum auf, dass sie jetzt ein System mit mehreren Leuten teilen.

VDI minimiert auch den Aufwand für Administration und Wartung. An den Arbeitsplätzen können dumme Terminals stehen, die bei einem Ausfall schlicht ersetzt werden. Egal, ob Uralt-PC, privates Notebook des Mitarbeiters oder sogar nur ein Tablet oder Mobiltelefon: alles eignet sich für den Zugriff auf den virtuellen Desktop im Rechenzentrum und erlaubt es sogar, aktive Sitzungen mit geöffneten Anwendungen von einem aufs andere Gerät „mitzunehmen“, etwa die bereits geöffnete Präsentation.

Rund um den Betrieb von VDI sind viele Techniken im Detail entstanden, die darüber entscheiden, was der Nutzer sieht: Erhält er eine eigene VM wie einen eigenen PC? Teilt er mit vielen anderen einen Terminal Server, sind seine Sitzungen wie bei



Im Windows 365 Cloud-PC kumuliert die Technik, die Microsoft nach und nach in Windows eingebaut hat. Entsprechend wenig spektakulär sieht ein im HTML5-Browser angezeigter Desktop aus. Nur das Drumherum verrät, was hier vor sich geht.

einem eigenen PC dauerhaft (persistent) oder erhält er mit jeder neuen Anmeldung einen frischen Desktop (non persistent)? Erhält er nur Zugriff auf einzelne Anwendungen (App-Streaming)? Wo liegen seine Daten? Wie greift er zu, per RDP-Client oder vorgeschalteter Webseite?

Für Nutzer fällt der Unterschied am geringsten aus. Die größten Änderungen gibt es aufseiten der Administration, und zwar sowohl bei der technischen Sicht, als auch der pekuniären und der juristischen: Die Techniker freuen sich über non-persistenten Desktops, weil sich mit einem neuen Image tausende Arbeitsplätze aktualisieren lassen und im Fehlerfall schnell das alte wiederhergestellt ist.

Die Betriebswirte und Controller rechnen mit spitzem Bleistift das Sparpotenzial aus, indem sie nicht mehr einen Benutzer pro PC planen, sondern gleich mehrere drauflossen und so eine bessere Ausnutzung erreichen. Die Juristen, die für die Compliance zuständig sind, freuen sich, dass niemand mehr Geräte mit Datenspeichern aus dem Haus schleppen kann und schätzen die Mittel zur Überwachung der Nutzeraktivität.

RDP genießt nicht den allerbesten Ruf, was die Sicherheit angeht: Das liegt weniger daran, dass das Protokoll selbst unsicher wäre, die Lücken sind längst geschlossen. Aber: Man öffnet einen lokalen Zugriff über das Netzwerk und eventuell sogar das Internet. Der Handel mit RDP-Zugängen blüht in einschlägigen Kreisen, sind sie doch

letztlich ein Zubringer, um fremde Netze hinzunehmen. Das Bewusstsein, dass diese Zugänge etwa mit VPN-Technik zu schützen sind, sickerte erst allmählich ein.

Was heute geht

Das Vermieten virtueller PCs im Rechenzentrum ist mitnichten etwas, was Microsoft exklusiv anbietet: Vor einigen Jahren schon traten spezialisierte Anbieter an, Windows-PCs mit fetter Gamer- oder Workstation-Ausstattung über das Netz zu vermieten. Damit die ihre geballte Grafikleistung bis zum Abnehmer liefern können, greifen sie auf dem Videostreaming entlehnte Techniken zurück: Die an der Hardware platzierte 3D-Hardware kann das Streamen quasi nebenbei erledigen. Gängige Hoster vermieten ebenso Windows-Systeme.

Microsofts Angebote gehen indes einen Schritt weiter, denn das Unternehmen hat schließlich die alleinige Verfügungsgewalt über Windows, kann also Lizenzmodelle nach Belieben gestalten und spezielle Versionen von Windows in seinen Rechenzentren anbieten, auf die niemand anders Zugriff hat, etwa Windows 10 und 11 in Versionen, die mehrere Nutzersitzungen gleichzeitig erlauben. Über wettbewerbsrechtliche Bedenken sei hier mal grob hinweggesehen.

Zwei wesentliche Produkte stehen für cloudbegeisterte Kunden zur Auswahl: Windows 365 ist ein Abo, das die Nutzung von Microsoft konfigurierter Cloud-PCs mit Windows erlaubt; unterm Strich sind

das von Microsoft gehostete VMs. Unternehmen zahlen monatlich einen festen Betrag, der sich nach der Ausstattung der virtuellen PCs richtet. Diese Cloud-PCs lassen sich über die üblichen Managementwerkzeuge wie physische PCs verwalten und einem Nutzer zuordnen.

Azure Virtual Desktop hingegen ist eher ein Baukasten für solche PCs in Microsoft Cloud Azure. Unternehmen bauen hier aus Einzelteilen an ihren Bedarf angepasste virtuelle Systeme. Die Einzelteile sind Rechenressourcen, wie CPUs, Haupt- und Massenspeicher, Lizenzen für die ausgeföhrte Software, Netzwerkbandbreite und -Volumina, aber auch Verwaltungshilfen, etwa für die Anmeldung der Nutzer an die Cloud-PCs.

Aufseiten des Kunden ist für Azure Virtual Desktop viel mehr Know-how gefragt. Und nicht nur das: Die Abrechnung ist deutlich komplizierter. Windows 365 ist quasi das Netflix-Abrechnungsmodell, während Microsoft in Azure sogar den Kleber einzeln abrechnet, der die Teile zusammenhält. Das gleicht Azure aber aus, indem es dort spezialisierte Ressourcen gibt, etwa Grafikkarten zum Rechnen, die für Windows 365 nicht im Angebot sind.

Die einzelteilorientierte Abrechnung hat Vorteile: Wer Azure-Know-how hat, kann sehr leicht kurzfristig und ohne langfristige Verpflichtung Ressourcen für ein Projekt buchen. Ein tiefschlafschlummernder Cloud-PC frisst nur wenig Geld,

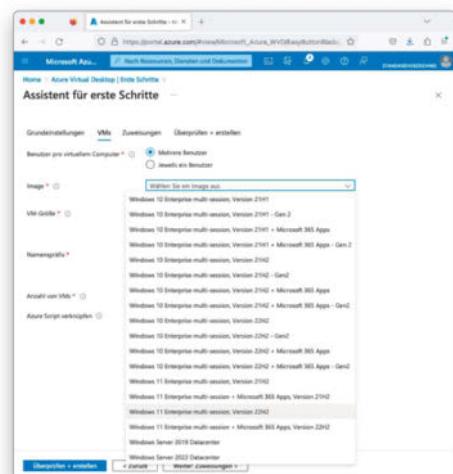
so wie sich ein echter PC im Ruhezustand nur wenig Strom gönnt. So kann, mit Köpfchen benutzt, das komplizierte und grundsätzlich deutlich teurere Angebot eine Sparmaßnahme sein. Aber Achtung: Manche Modellrechnung vernachlässigt, dass VMs nicht nur interaktiv genutzt werden, sondern Rechenzeit auch für Updates und Wartungsarbeiten konsumieren.

Was kommen dürfte

Für den Endkunden, der im Arbeitszimmer einen Desktop-PC stehen hat, um damit gelegentlich zu spielen und einmal im Jahr seine Steuer gegenüber dem Finanzamt zu erklären, ergibt ein solcher Miet-PC nur begrenzt Sinn. Ein dafür ausreichender, gebrauchter PC ist weitaus günstiger und eine einmalige Investition. Die Mietoption hat immerhin den Charme, dass sich andere um Sicherheit und Zuverlässigkeit kümmern.

Bei gewerblichem Einsatz schaut es anders aus: Wer nur sporadisch massive Rechenleistung braucht, würde vielleicht besser mieten. Aber: Die auf Leistung optimierten Angebote sind derzeit nur als „Azure Virtual Desktop“ zu haben und nicht im Flatratepaket „Windows 365“. Und viele selbstständige Grafiker oder Videokünstler, die Azures Möglichkeiten goutieren, dürften an den Einstiegshürden der Konfiguration scheitern.

Auch wenn beide Geschmacksrichtungen der Microsoft'schen Cloud-PCs dem



Beim Einrichten eines „Azure Virtual Desktop“ in Microsofts Cloud ist Know-how gefragt. Zum Ausgleich gibt es spezialisierte virtuelle Maschinen und spezielle Windows-11-Versionen, die mehreren Benutzern gleichzeitig dienen können.

gewerblichen Einsatz vorbehalten sind: Ausprobieren kann man beide, wenn man bereit ist, Microsoft seine Kreditkarten-daten anzuvertrauen. Das scheint dem Unternehmen Nachweis genug für ein gewerbliches Interesse. Unsere Erfahrung dabei tönt schon aus den vorangehenden Absätzen: Azure ist wie ein Besuch im Spielcasino; das Geschenkguthaben ist schnell davon geflogen. (ps@ct.de) ct

Eval-Angebote bei Microsoft: ct.de/yeg1



CodeMeter – Eine Symphonie von Software-Monetarisierungs-Tools

- Komponieren Sie Ihren eigenen Code
- Orchestrieren Sie Ihre Lizenzstrategie
- Stimmen Sie Ihren IP-Schutz genau ab
- Verbreiten Sie Ihr gestaltetes Werk

Klingt einfach, oder?
Und das ist es auch
mit CodeMeter

Treffen Sie uns!



sps

smart production solutions

Halle 6, Stand 428

+49 721 931720

sales@wibu.com

www.wibu.com



**SECURITY
LICENSING**

PERFECTION IN PROTECTION



Bild: KI Midjourney | Bearbeitung c't

Eine Revolution frisst ihre Kinder

Abkehr von klassischen Open-Source-Lizenzen

**Open Source war einmal ange-
treten, um Nutzern die größt-
mögliche Freiheit im Betrieb von
Software zu garantieren und
Software-Titanen den Kampf an-
zusagen. Inzwischen haben die
aber gelernt, die verbrieften
Freiheiten zu ihrem Vorteil zu
nutzen.**

Von Kristian Köhntopp

Es brodelt in der Open-Source-Gemeinde: In jüngerer Vergangenheit lässt sich beobachten, dass etablierte Softwareunternehmen mit großer Nutzerbasis die Freiheiten der Nutzer ihrer Open-Source-Projekte durch einen Lizenzwechsel einschränken. Prominentes Beispiel ist das Unternehmen HashiCorp, das auf Software für Cloud-Infrastruktur spezialisiert ist und seine Kernprodukte unter die BSL v1.1 (Business Source License) gestellt hat, eine formal nicht offene Lizenz. Die Quittung kam prompt: Inzwischen wurde die

Software als OpenTofu geforkt und steht jetzt unter der Schirmherrschaft der Linux Foundation.

HashiCorp ist nicht die erste Firma, die so etwas tut: Red Hat hat vor längerer Zeit die Positionierung von CentOS als freie Alternative zu Red Hat Enterprise Linux (RHEL) geändert und kürzlich die Herausgabe der RHEL-Quelltexte durch eine EULA (End User License Agreement) eingeschränkt, Elastic hat Elasticsearch und Kibana unter die formal nicht offene SSPL (Server Side Public License) gestellt,

und auch MongoDB ist nur noch unter der SSPL zu haben.

Noch viel früher gab es Umzüge von der GPLv2 zur GPLv3, die Einführung der AGPL, und einige Firmen wie Apple haben ein striktes GPLv3-Verbot für ihre eigenen Softwarereprodukte ausgesprochen und migrieren interne Software komplett von der GPL weg. Sind diese Entwicklungen eine Bedrohung für Open Source? Und was sollen diese Lizenzänderungen eigentlich bezeichnen? Um die Konfliktlinien zu verstehen, lohnt sich ein Blick in die Vergangenheit.

Vier Freiheiten und eine Zertifizierung

Open Source und Free Software gab es schon, bevor diese Worte dafür existierten. Doch die eigentliche „Open-Source-Revolution“ begann in den 1990er Jahren mit dem Linux Kernel und den sich daraus ergebenden Distributionen. Der Begriff „Open Source“ wurde von Bruce Perens und Eric Raymond geprägt und in der Open-Source-Definition und den Debian Free Software Guidelines formalisiert.

Beides sind Verfeinerungen und Klarstellungen der ursprünglichen „Vier Freiheiten“ der Mutter aller Open-Source-Lizenzen, der GNU General Public License (GPL), die da lauten:

- Freiheit 0: Die Freiheit, das Programm für jeden Zweck auszuführen.
- Freiheit 1: Die Freiheit, die Funktionsweise des Programms zu untersuchen und eigenen Bedürfnissen der Datenverarbeitung anzupassen. Das impliziert Zugang zum Quelltext.
- Freiheit 2: Die Freiheit, das Programm weiterzugeben. Die GPL fordert aber, dass auch der Quelltext des Programms, der dem weitergegebenen Programm entspricht, angeboten wird.
- Freiheit 3: Die Freiheit, das Programm zu verbessern und diese Verbesserungen unter der originalen Lizenz weiterzugeben. Auch dies impliziert Zugang zum Quelltext und das Recht, diesen zu ändern.

Während die GPLv2 die Lizenz ist, die diese Freiheiten das erste Mal kodifiziert hat, haben sich schnell eine ganze Reihe von Lizenzen entwickelt, die alle „Open Source“ nach den OSI-Richtlinien (Open Source Initiative) oder den DFSG (Debian Free Software Guidelines) sind und von denen einige zwar frei, aber inkompatibel mit der GPL sind. Das hat insbesondere in den späten 90er Jahren viele Usenet-

Newsgruppen dauerhaft mit Diskussionen gefüllt.

Der LAMP-Stack: Eine wilde Mischung von Lizenzen

Der bekannte LAMP-Stack für die Webentwicklung ist auch ein Beispiel für den Lizenzwirrwarr, mit dem man es bei Open Source zu tun haben kann. LAMP beinhaltet den Linux Kernel, den Apache Webserver, die MySQL-Datenbank und die Programmiersprache PHP. All diese Bausteine haben unterschiedliche Lizenzen.

Während der Linux-Kernel GPLv2-lizenziert ist, steht der Apache Webserver unter der Apache License 2. Diese ist freizügig („permissive“), das heißt, es fehlen Bestimmungen, die erzwingen, dass Veränderungen gegenüber dem Originalprogramm unter derselben Lizenz wie das Originalprogramm weitergegeben werden müssen.

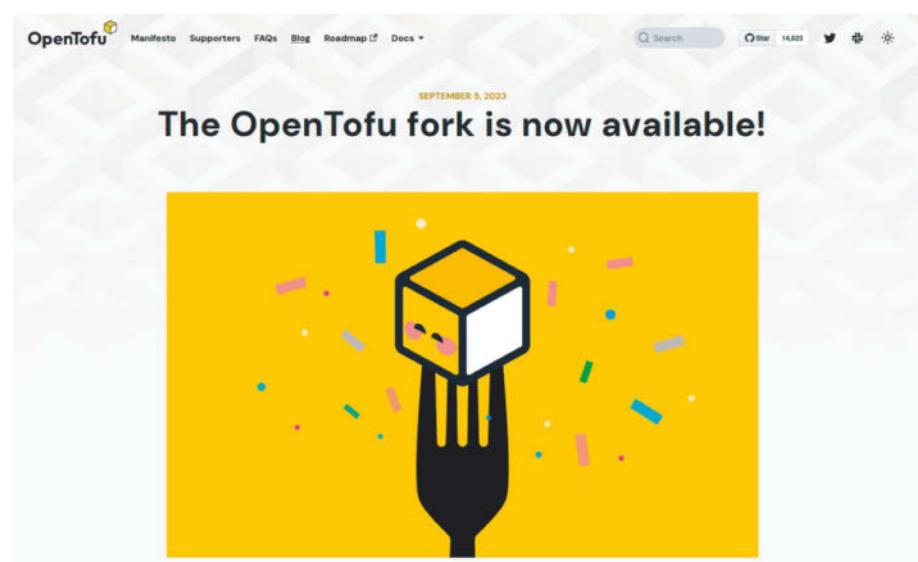
MySQL hat gleich zwei Lizenzen: Die Datenbank kann unter der GPL genutzt werden und unterliegt dann den Bestimmungen und Freiheiten, die oben genannt wurden. Der Urheberrechtsinhaber ist jedoch eine Firma (ursprünglich MySQL, dann SUN und jetzt Oracle), die alle Rechte innehat. Entwickler müssen die Rechte an ihren Patches per Contributor License Agreement (CLA) an Oracle abtreten. Dadurch kann Oracle den Quelltext auch unter einer kommerziellen Lizenz anbieten.

c't kompakt

- In letzter Zeit wenden sich vermehrt Projekte von klassischen Open-Source-Lizenzen ab.
- SaaS-Geschäftsmodelle erschweren die nachhaltige Entwicklung von Open-Source-Software.
- Cloudanbieter richten die Freiheiten, die Open-Source-Lizenzen gewähren, gegen deren Entwickler.

Und PHP steht seit Version 4 unter der PHP-Licence. Auch das ist eine freizügige Lizenz, die Nutzung und Weitergabe von PHP mit oder ohne Modifikationen, als Quelltext oder ausführbares Programm erlaubt, jedoch Einschränkungen hinsichtlich des Gebrauches der Wortmarke „PHP“ enthält.

Das führt zu einer wichtigen Frage der späten 90er Jahre: Was passiert, wenn man Software mit unterschiedlichen Lizenzen mischt und zusammen in einer Distribution verbreitet oder sie in einem Übersetzungsprozess zusammen linkt? Das GNU Projekt hat relativ schnell klar gestellt, dass die Grenzen der Lizenz die Prozessgrenzen sind. Wenn man also ein Programm unter einer anderen Lizenz entwickelt und mit GNU-Software zusammen linkt, sodass das Resultat hinterher in ein-



Nachdem HashiCorp angekündigt hatte, sein Infrastruktur-Werkzeug Terraform unter eine nicht freie Lizenz namens BSL zu stellen, formierte sich schnell Widerstand in der Open-Source-Community. Inzwischen gibt es einen Fork namens OpenTofu (vormals OpenTF).

und demselben Prozess ausgeführt wird, dann geht das legal nur dann in Ordnung, wenn die andere Lizenz mit der GPL kompatibel ist.

Obwohl die GPL eine Open-Source-Lizenz ist, sind nicht viele Lizenzen mit der GPL kompatibel. Das liegt daran, dass die GPL die Freiheit währende Einschränkungen enthält: Man kann die Ausführung der Software nicht einschränken („Nicht in Atomraketen!“) und ein Programm, das GPL-Komponenten enthält, muss auch im Quelltext mit allen Änderungen angeboten werden. Das bedeutet: nicht nur die GPL-Komponenten, sondern alle Komponenten in diesem Prozessraum.

Die Debatte um die „infektiöse“ GPL

Aus der Pflicht zur Veröffentlichung der Quellen ergab sich dann eine weitere Diskussion in den frühen Nullerjahren. Damals wurden Linux und Open Source so groß, dass man sie nicht ignorieren konnte, und eine Menge Start-ups kokettierten mit Open Source. Die GPL war damals zwar die am weitesten verbreitete Lizenz, aber wenn man ein Start-up betreibt, das sich durch Risikokapital finanziert, dann ist es keine attraktive Aussicht für Investoren, alle Quellen der eigenen Software unter der GPL herausgeben zu müssen.

Aus dieser Zeit stammt der Anti-Open-Source-Slogan der „viralen“ oder

„ansteckenden“ GPL, der im Kern jedoch Unsinn ist: Zum einen ist die GPL leicht einzuhegen, denn sie endet ja klar definiert an Prozessgrenzen, andererseits sind die Bestimmungen der GPL sinnvoll, weil sie eine direkte Ausbeutung der Leistung anderer erschwert („Quid pro quo: Wenn Du von Open Source profitierst, dann sollst Du auch auf die gleiche Weise dazu beitragen.“).

GPLv3 gegen Softwarepatente

Venture Capitalists brachten als Reaktion eine weitere Bedrohung für Open Source in Stellung, auf die reagiert werden musste: Softwarepatente. Wenn Start-ups ihre Software unter der GPL herausgeben müssen, können sie ihr geistiges Eigentum eventuell anders einschränken, um später dafür Lizenzgebühren einzutreiben. Damals waren das Softwarepatente, die als reale Bedrohung positioniert wurden.

In der Rechtfertigung für die Version 3 der GPL wird dies auch direkt thematisiert, und die GPLv3 enthält Regelungen, die vereinfacht besagen: „Wenn Du Patente verwendest, um Benutzer von beliebiger GPLv3-lizenzierte Software zu verklagen, dann erlischt Dein Recht, jegliche GPLv3 Software zu verwenden.“ Diese Klausel ist wirksam und sie hat unter anderem dazu geführt, dass Apple jede Form von GPLv3-lizenzierte Software in seinen Produkten vermeidet und sukzessive jegliche GPLv3-

Software in Apple-Produkten durch anders lizenzierte Software ersetzt.

Die Bestimmung, dass die GPL an Prozessgrenzen endet, ist auch für das GNU-Projekt manchmal problematisch, und so gibt es Varianten der GPL und Ausnahmen. Zum Beispiel ist generierter Code von unter der GPL stehenden Programmen nicht GPL-geschützt: Dies ermöglicht Codegeneratoren wie gcc, bison und flex. Und für einige Bibliotheken, darunter die glibc, gibt es die LGPL, auch „Lesser GPL“ genannt, die es erlaubt, die Bibliotheken gegen Nicht-GPL-Software zu linken.

„GPL statt LGPL“ als Waffe

Andersherum haben einige Firmen die GPL als Waffe eingesetzt. Zum Beispiel waren Bibliotheken wie die „libmysql-client.so“ bis einschließlich MySQL 3.23 unter der LGPL verfügbar, aber ab Version 4 stehen sie unter der GPL. Linkt man diese Bibliothek also in sein kommerzielles Programm und will es ausliefern, dann muss man eine kommerzielle MySQL-Lizenz unter dem Dual-Licensing Programm kaufen. MySQL hat das im Slogan „If you are free, we are free. If you are commercial, we are commercial“ zusammengefasst.

Das war eine Zeit lang ein Problem für die Programmiersprache PHP, die ja gegen diese Bibliothek gelinkt hat, um auf die Datenbank im Rahmen des LAMP-Stacks zugreifen zu können, aber eine formal nicht GPL-kompatible Lizenz hat. PHP ist dem Problem mit einer Re-Implementierung des Protokolls als „mysqlnd“ begegnet, und parallel dazu hat MySQL das Problem gelöst, indem sie eine Lizenz-Ausnahme für PHP (und andere) definiert haben.

Weniger GPL, mehr MIT, BSD und Apache

All diese Wirrungen sind ein Grund, warum „moderne“ Open-Source-Software meist nicht mehr die GPL verwendet, sondern andere Lizenzen wie MIT, BSD oder Apache bevorzugt. Sie machen eine Reihe von kapitalfinanzierten Geschäftsmodellen um Open Source einfacher, wenn man eine Firma um ein (oft nur dem Namen nach) Open-Source-Produkt stricken will. Ab 2005 beginnt die Blütezeit von auf Open Source basierenden Geschäftsmodellen. Das hat 15 Jahre lang gut genug funktioniert, um eine Reihe von Projekten groß zu machen.



In den 2000er-Jahren sah sich Softwaregigant Microsoft durch GPL und Open Source bedroht. An Werbeanzeigen wie dieser wird deutlich, dass der Kampf proprietäre versus freie Software früher mit härteren Bandagen ausgetragen wurde.

Einige dieser Projekte sind nicht auf einzelne Firmen als Träger angewiesen: PHP, KDE und Postgres sind zum Beispiel stabile und groß angelegte OSS-Projekte, hinter denen nicht eine einzelne große Firma und ihre Venture-Kapitalgeber stehen – sie alle zeichnen sich durch eine breite und vielfach verankerte Basis von Beitragenden aus.

Seit einigen Jahren zeigen sich Risse: Red Hat, Elastic, MongoDB, MariaDB (ein MySQL-Fork), HashiCorp und viele andere haben ihre Lizenzen geändert oder sich umstrukturiert, um einen bestimmten Use-Case auszuschließen. Da dies eine Verletzung der Freiheit 0 darstellt, sind es damit auch keine Open-Source-Projekte im Sinne der OSI und der DFSG mehr.

Der Elefant im Raum ist die Cloud

Amazon Web Services (AWS) ist ein System und eine Firma, die stellvertretend für ein Geschäftsmodell steht: Es monetisiert den in Unternehmen oft vernachlässigten und unterfinanzierten Aspekt „Operations“ und sieht vor, Software für Dritte nicht zu entwickeln, sondern vorrangig zu betreiben.

Damit ist Amazon sehr erfolgreich: Schon länger laufen in dessen Rechenzentren mehr Neuinstallationen von MySQL (und MongoDB, Postgres oder Elastic und vielen anderen Datenbanken) als lokal beim Kunden („on premise“). Dies ist eine

Klasse von Systemen, die notorisch schwierig zu betreiben sind, weil Fehler beim Betrieb Daten unwiederbringlich ruinieren können. Natürlich kann man eine fehlerhafte Version einer Datenbank auf die vorhergehende Version zurückrollen, aber die Daten, die durch den Fehler zerstört worden sind, sind immer noch dauerhaft weg.

Die systematische Unterschätzung der Kosten und des Aufwandes von Operations macht das Prinzip „Software as a Service“ (SaaS) erfolgreich. SaaS-Anbieter verkaufen nicht nur einen Dienst, sondern im Grunde genommen vornehmlich ein Betriebskonzept für diesen Dienst, der einen reibungslosen und unterbrechungsfreien Betrieb direkt nach der Inbetriebnahme erlaubt. Weil das Betriebskonzept erprobt und vorgeschrieben ist, kann man die Qualität der Dienstleistung standardisieren, messen und dann stabile Verträge mit Dritten darüber eingehen.

Kunden begrüßen das, denn Software zu betreiben ist ein Problem, das mindestens so schwierig ist wie Software zu entwickeln. Viele Kunden wollen sich damit nicht auseinandersetzen. Oft werden dabei Verfügbarkeiten, Antwortzeiten und Kosten in Kauf genommen, die genau so vorher bei den eigenen Inhouse-Teams vollkommen inakzeptabel waren, nur um beispielsweise den Aufgabenkomplex „Betrieb der Datenbanken“ aus dem Haus zu bekommen – das zeigt, wie unange-

nehm Firmen die Auseinandersetzung mit „Operations“ ist.

Dabei ist der Betrieb dieser Software durch AWS nicht durch eine kommerzielle Vereinbarung zwischen dem Hersteller der Software und Amazon gedeckt, sondern erfolgt auf der Grundlage der Freiheit 0 der GPL: Amazon darf die Software zu jedem beliebigen Zweck ausführen, auch als Service für Dritte. In Folge streicht Amazon das Geld für den Betrieb dieser Installationen ein, aber der eigentliche Softwareentwickler, die Herstellerfirma oder -gruppe bekommt davon nichts. Das ist kein nachhaltiges Finanzierungsmodell, aber durch die GPL gedeckt und sehr vorteilhaft für AWS, sodass man dort auch keine Veranlassung sieht, etwas daran zu ändern.

Business Source License als frühe Reaktion

Das ist schon recht früh als Problem erkannt worden: Bei MariaDB findet man die erste Erwähnung der „Business Source License“ (BSL) im Blog von Monty Widenius schon 2013, und MariaDB setzt sie für bestimmte Komponenten seit 2016 ein, mit leichten Veränderungen nach Anregungen von Bruce Perens dann seit 2017 in der Version 1.1. Andere Produkte ziehen nach und gehen ebenfalls unter die BSL: Couchbase, Uptrace, Kurtosis, Sentry, CockroachDB und neuerdings alle HashiCorp-Kernprodukte.



Streiche können teuer werden.

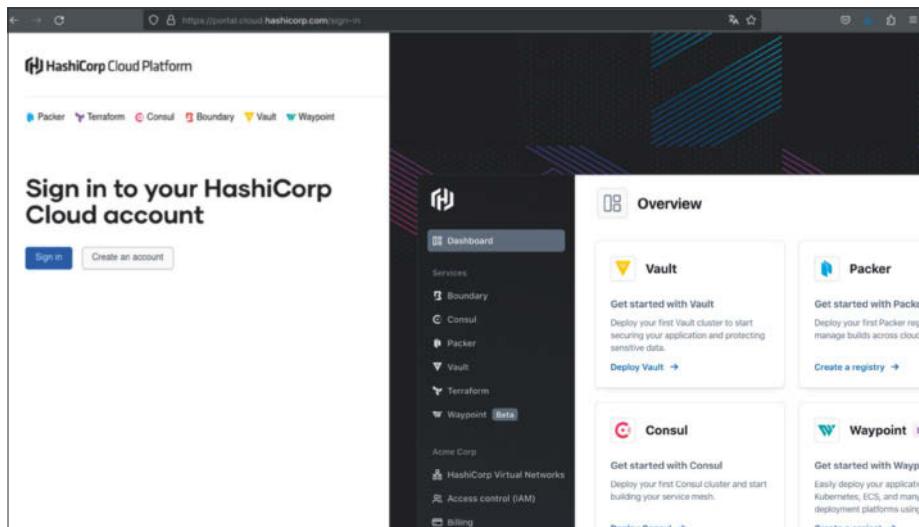
Wie Sie sich gegen Attacken auf Ihre IT-Infrastruktur wappnen können, erfahren Sie in unseren Seminaren und Workshops.

www.cyber-akademie.de



Jetzt Kurse buchen!





HashiCorp betreibt die HashiCorp Cloud Platform (HCP), wo Kunden Terraform und Co. als Managed-Service nutzen können. Der Wechsel zur BSL verbietet es Mitbewerbern, Konkurrenzprodukte auf Basis von Terraform zu entwickeln.

Die Idee hinter der BSL ist, dass die Software weiterhin frei nutzbar und der Quelltext verfügbar bleibt, aber mit Einschränkungen. Die wichtigste Einschränkung meint üblicherweise „AWS darf das nicht nutzen“, formuliert als „Die Software darf nicht als SaaS für Dritte angeboten werden“. Man kann die Software also weiter so verwenden wie zuvor. Man kann sie auch in AWS selbst betreiben oder man kann sie innerhalb der eigenen Organisation für andere Abteilungen bereitstellen.

Aber man wird den Betrieb der Software nicht als Dienst von einem Cloudbetreiber kaufen können, sondern nur vom Ersteller der Software selbst (etwa via AWS Marketplace). Diese Einschränkung dient dazu, die Finanzierung der Entwicklung der Software sicherzustellen, indem AWS die Möglichkeit genommen wird, diesen Geldstrom vollständig für sich selbst abzuzweigen. Die BSL kommt mit einer weiteren Regel, die besagt, dass Codeänderungen spätestens nach vier Jahren unter eine anerkannte Open-Source-Lizenz fallen (meist die Apache-2.0-Lizenz oder die GPL), sodass ältere Versionen der Software auf jeden Fall immer Open Source sind.

Formal ist BSL-Software keine Open Source im Sinne der OSI oder DFSG, weil anders als bei der GPL die Freiheit 0 („Ausführen für welchen Zweck auch immer“) eingeschränkt wird. Man bezeichnet sie deshalb auch als „Source Available“. Auch die von MongoDB, Elastic und Kibana verwendete Server Side Public License verfolgt dieselbe Idee, nur

anders: Die SSPL ist eine AGPLv3 Variante, die ebenfalls die Freiheit 0 einschränkt und verbietet, die Software für Dritte („as a Service“) zu betreiben, es sei denn, man erwirbt eine kommerzielle Lizenz.

Unfreie freie Software

Die Grundidee hinter den Freiheiten der GPL ist Geben und Nehmen: „Du kannst unsere Software nutzen, musst aber selbst gutartig mitspielen.“ Die immer weiter wachsende Stellung von AWS macht es den Erstellern von Open Source zunehmend schwerer, ihren Unterhalt mit der Entwicklungsarbeit zu verdienen – AWS bietet deren Software als Service an und verwebt

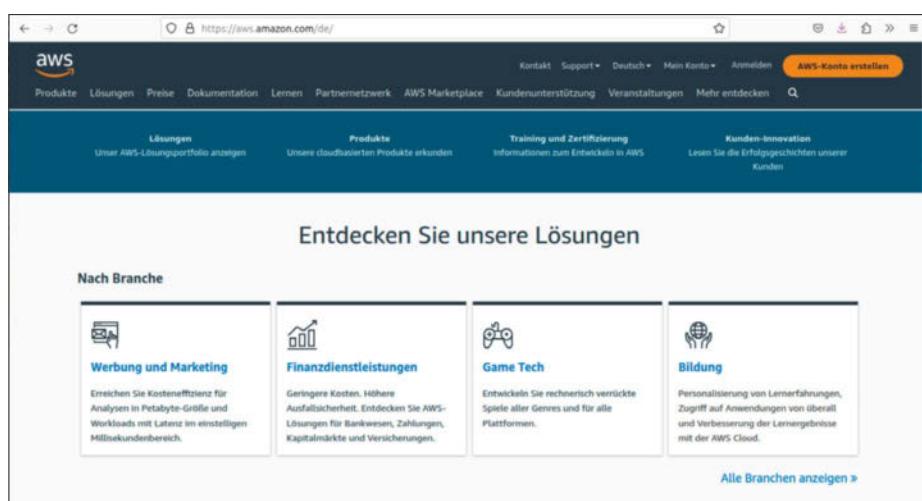
ihn mit weiteren Dienstleistungen. Die Freiheit 0 der freien Software wendet sich hier gegen die Ersteller der Software.

Die BSL und die SSPL sind eine Reaktion darauf. Sie sind formal gesehen keine freie Software nach OSI oder DFSG. Aber vielleicht ist es auch für Open Source wieder einmal an der Zeit, den Veränderungen und den neuen Bedrohungen ins Auge zu sehen und die Lizenzen und die Definitionen der neuen Zeit anzupassen, so wie dies mit der GPLv3 und der AGPL schon einmal geschehen ist.

Vielleicht bewegen sich FSF, OSI und Debian aber auch nicht und beharren auf ihren Definitionen von „freier Software“. Die Entwickler von Software unter der BSL und der SSPL ficht das nicht an: Ihre Software ist jetzt schon „fast frei“ und wegen ihrer speziellen Regel nach einigen Jahren dann auch formal freie Software nach alter Definition. Einen Grund, die BSL oder SSPL abzulehnen, gibt es nicht, ausgenommen Fundamentalismus für freie Software. Aber wenn diejenigen, denen etwas am Open-Source-Gedanken liegt, nicht wollen, dass AWS die einzige Firma mit einem erfolgreichen Open-Source-Geschäftsmodell ist, muss sich etwas ändern. Sonst landen Nutzer schon bald wieder bei langweiliger, traditioneller kommerzieller Software, wie man es früher von Microsoft und Oracle gewohnt war.

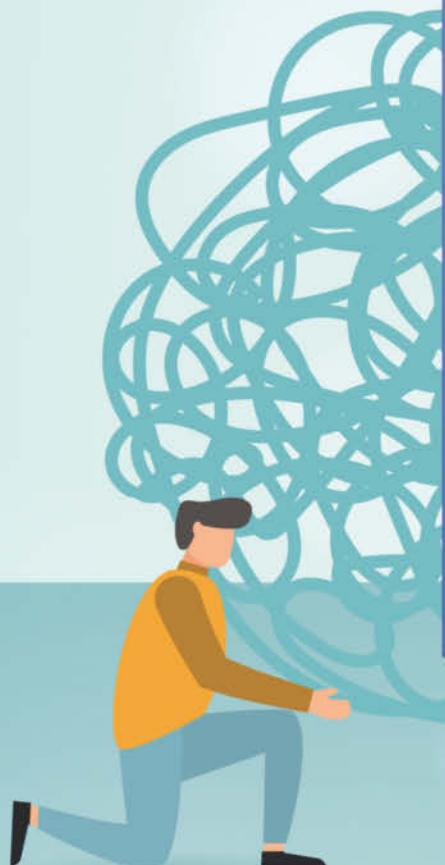
(ndi@ct.de) ct

Liste von Open-Source-Lizenzen nach OSI-Definition, OpenTofu-Manifest: ct.de/yfyh



Immer mehr Unternehmen mieten Ressourcen wie Datenbanken von Cloudanbietern wie AWS an, anstatt eigene Server zu betreiben. Die Entwickler von Open-Source-Software gehen dabei oft leer aus.

Bestens verdrahtet!



inkl.
50
Seiten
Fritzbox

Schnelles WLAN, eine komplette Abdeckung in Ihrem Zuhause und bei Bedarf das eigene VPN. Das c't-Sonderheft deckt alle wichtigen Netzwerkthemen ab mit einem Schwerpunkt rund um den beliebten Fritzbox-Router und seinem stark verbesserten FritzOS:

- Fritzbox einrichten und tunen
- Kaufberatung: die beste Fritzbox für jeden Anschluss
- Mesh & Repeater optimal auswählen
- Mit speziellen VPNs die Privatsphäre schützen
- Netzwerke bauen
- Das bringen Wi-Fi-6E und Wi-Fi-7

Heft für 14,90 € • PDF für 12,99 € • Heft + PDF 19,90 €



shop.heise.de/ct-netzwerke23



Bild: Andreas Martini

FIDO2 wird erwachsen

Login per Passkey: Gekommen, um zu bleiben

Mit dem Rebranding zu „Passkeys“ will die FIDO-Alliance die passwortlose Authentifizierung mit FIDO2 massentauglich machen. Mit Passkeys wird FIDO2 benutzerfreundlicher, denn sie weichen die strenge Hardwarebindung der Spezifikation auf.

Von Kathrin Stoll

Sicherer als ein Login per Passwort und außerdem kinderleicht einzurichten: Die Rede ist von Passkeys. Hinter der passwortlosen Anmeldung steht die FIDO Alliance, ein Industriekonsortium, dem große Techfirmen wie Google, Microsoft, Apple oder Amazon angehören.

Das unter dem Buzzword „Passkeys“ vermarktete Login-Verfahren stützt sich auf den FIDO2-Standard, der bereits im April 2018 eingeführt wurde. Er besteht im Wesentlichen aus zwei Komponenten: dem Web Authentication JavaScript API Standard, kurz WebAuthn, und dem darauf abgestimmten Client to Authenticator Protokoll, kurz CTAP. Im Grunde ist

FIDO2 gemeint, wenn von Passkeys die Rede ist: Laut Andrew Shikiar, dem geschäftsführenden Direktor der FIDO Alliance, ist „Passkeys“ lediglich „der verbrauchergerechte Begriff“ für die passwortlose Authentifizierung mit FIDO2 [1].

Zuerst gab es physische FIDO2-Sicherheitsschlüssel

In der Folge kamen physische Sicherheitsschlüssel mit FIDO2-Unterstützung auf den Markt. Diese FIDO2-Sicherheitsschlüssel sehen meist aus wie klassische USB-Sticks und sichern in der Regel bis zu 25 private Schlüssel. Doch obwohl die

passwortlose Anmeldung per FIDO2-Stick deutlich sicherer ist als die Anmeldung per Passwort, setzte sie sich außerhalb eines kleinen Kreises sicherheitsaffiner Nutzer aus einer Reihe von Gründen nicht durch: Das Verfahren erfordert die Anschaffung zusätzlicher Hardware. Außerdem ist es umständlich: Um den Zugriff auf die Benutzerkonten bei Verlust des Sicherheitsschlüssels nicht zu verlieren, muss man für jeden Dienst einen separaten Login auf mindestens einem zweiten FIDO2-Stick einrichten. Hat man den Stick einmal verloren, ist der Zugriff auf die Benutzerkonten ansonsten dahin. Hinzu kommt, dass nur sehr wenige Dienste das Verfahren überhaupt anboten.

Es folgten die Plattformen

Bereits seit 2019 unterstützt Windows 10 mit Windows Hello offiziell FIDO2. Microsoft war damit der erste Anbieter, der passwortlose Logins auf seiner Plattform nativ unterstützte. Seitdem können Windows-Nutzer ihre Rechner mit Windows Hello als FIDO2-Authenticator einsetzen. Die privaten Schlüssel werden dabei im Trusted Platform Module (TPM) des Rechners abgelegt.

Im September 2022 zog Apple nach und führte in den Betriebssystemen iOS 16 und macOS Ventura die passwortlose Anmeldung per FIDO2 unter dem Buzzword „Passkeys“ ein. Als Authenticator wird dabei ein sicheres Hardwaremodul eingespannt: die Secure Enclave. Dort wird das Schlüsselpaar aus öffentlichem und privatem Schlüssel generiert. Die privaten Schlüssel – die Passkeys – werden allerdings nicht wie bei Windows oder einem FIDO2-Stick im sicheren Hardwaremodul des Geräts gesichert, sondern ausschließlich im iCloud-Schlüsselbund. Die Cloudspeicherung hat zwei Vorteile: Man kann sie über die Cloud zwischen Apple-Geräten synchronisieren und man kann sie wiederherstellen.

Google reihte sich kurze Zeit später ein und gab bekannt, dass Android-Geräte ab Android 9 sowie Rechner ab ChromeOS 109 die passwortlose Anmeldung ebenfalls unterstützen. Wie bei Apple wird das Schlüsselpaar dabei im sicheren Hardwaremodul – bei Android heißt es Secure Element – generiert. Anders als bei Apple werden die privaten Schlüssel jedoch nicht ausschließlich, sondern lediglich optional in der Google-Cloud gesichert. Wer darauf verzichtet, kann seine Passkeys nicht zwischen Geräten des Herstellers synchron

halten und muss sich um eine andere Backupstrategie kümmern. Denn analog zum FIDO2-Sicherheitsschlüssel ist der Zugriff auf die Benutzerkonten im Verlustfall sonst weg.

Zu Beginn waren Verwender der betriebssystemeigenen Authenticatoren weitgehend gezwungen, auch Browser des Betriebssystems zu nutzen. So konnten etwa Mac-Nutzer nur in Safari Passkeys erstellen und sie im iCloud-Schlüsselbund sichern. Mit MacOS Sonoma ändert sich das. In Chromium-basierten Browsern erstellte Passkeys kann man seit dem Update ebenfalls im iCloud-Schlüsselbund sichern.

Nur wer Geräte ohne einen betriebssystemeigenen Authenticator nutzt, etwa einen Rechner mit Linux, schien zunächst außen vor zu sein. Linux-Nutzern blieb zunächst nur die Installation eines Browsers mit Passkey-Unterstützung, aktuell am besten Google Chrome. Chrome unter Linux erlaubt es immerhin, ein Passkey-fähiges Smartphone als externen Authenticator einzurichten, unter Mozillas Firefox kann man bisher lediglich einen FIDO2-Stick benutzen. Mozilla will die Unterstützung für Passkeys in Firefox 120 nachrüsten [2].

Passwortmanager reihen sich ein

Seit diesem Jahr erlauben auch eine Reihe Fremdanbieter-Apps das Einloggen mit Passkeys, darunter 1Password und Dashlane. Die Passwortmanager sind für verschiedene Betriebssysteme verfügbar und mit allen verbreiteten Browsern kompatibel. Sie bieten einen Weg, Passkeys zwischen Geräten verschiedener Hersteller synchron zu halten. Damit heben sie einen zentralen Kritikpunkt an der Pass-

c't kompakt

- Passkeys werden mittlerweile von den Betriebssystemen von Apple, Google und Microsoft und einer Reihe Passwortmanager unterstützt.
- Auch das Zusammenspiel der Betriebssysteme mit den verschiedenen Browsern wird besser.
- Macht jetzt noch das Gros der Webdienste mit, könnte das Internet der Zukunft tatsächlich mit deutlich weniger Passwörtern auskommen.

key-Offensive der FIDO Alliance aus: Wer zum Verwalten seiner Passkeys auf betriebssystemeigene Schlüsselverwaltungen wie den Google Passwortmanager oder den iCloud-Schlüsselbund zurückgreift, kann Passkeys nur zwischen Geräten desselben Ökosystems synchronisieren und wiederherstellen. Kritiker bemängelten, dass das die Nutzer zu stark an die Hersteller binden würde. Zwar kann man die bei Apple, Google, oder Microsoft gesicherten privaten Schlüssel durch Scannen eines QR-Codes auf Geräten verschiedener Hersteller nutzen, es ist jedoch nicht möglich, alle Passkeys auf einmal auf ein anderes Betriebssystem zu übertragen. Anders gesagt: Die Schlüsselverwaltungen der Betriebssysteme bieten keine Möglichkeit zum Exportieren der privaten Schlüssel. In der ursprünglichen WebAuthn-Spezifikation ist das genauso vorgesehen: Sie sieht vor, dass die privaten Schlüssel den sicheren Hardwarespeicher des Geräts nicht verlassen.

Die physischen Sicherheitsschlüssel sehen meist aus wie USB-Sticks und können in der Regel 25 Logins speichern.



Diese Anforderung wird durch die bei Apple forcierte, bei Google freiwillige Cloudspeicherung aufgeweicht. Unter Sicherheitsexperten sorgte das für Kritik. Dieses Verhalten konterkarierte die ganze Idee der Gerätebindung, die FIDO2 so sicher mache.

Auch 1Password oder Dashlane sichern und synchronisieren die privaten Schlüssel verschlüsselt über die Cloud. Die Passwortmanager weichen aber noch in einem weiteren Punkt von der Spezifikation ab: Man kann dort die gesicherten privaten Schlüssel außerdem in eine CSV-Datei exportieren. Das ist praktisch, denn so kann man sie einfach in einen anderen Passwortmanager importieren. Der Nachteil: Ein Angreifer, der es schafft, Ihnen diese Datei zu stehlen, hätte damit alles, was er braucht, um auf Ihre Internetkonten zuzugreifen. Ihren Fingerabdruck oder Ihr Gesicht bräuchte er dazu nicht, denn die biometrischen Daten fließen nicht in das kryptografische Verfahren ein. Sie ersetzen lediglich die Eingabe des Masterpassworts.

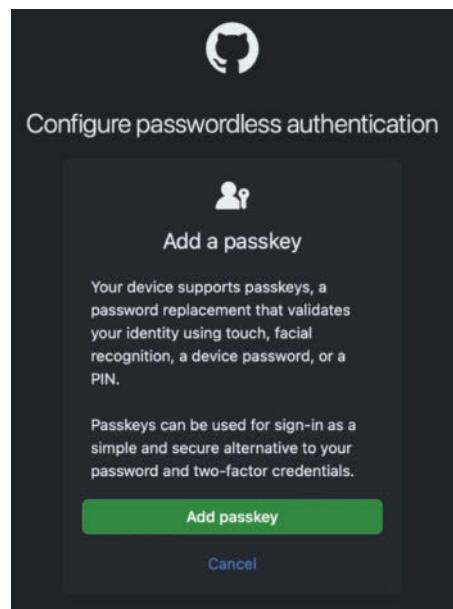
Die Passwortmanager sind damit zwangsläufig noch etwas weniger sicher als die betriebssystemeigenen Schlüsselverwaltungen. Wer das als Argument gegen Passkeys in Feld führen wollte, übersieht allerdings einen entscheidenden Punkt: Auch Passkeys, die in einer Pass-

wortmanager-App mit Exportfunktion gesichert werden, sind sicherer als jedes Login per Passwort. Bequemer sind sie auch. Die FIDO Alliance plant ein Zertifizierungsprogramm für Anbieter von Passwortmanagern, um das Vertrauen der Nutzer zu stärken [1].

Fazit und Ausblick

Das verbrauchergerechte Konzept der Passkeys erweitert die klassische FIDO2-Authentifizierung um die Möglichkeit, dass die privaten Schlüssel nicht mehr an ein Hardwaremodul gebunden sein müssen. Das ist etwas weniger sicher als die ursprüngliche Spezifikation – und doch volle Absicht. Denn Passkeys lösen zwei zentrale Probleme des ursprünglichen Konzepts. Im Unterschied zum klassischen Login per FIDO2 gibt es jetzt Optionen zum Wiederherstellen und Synchronisieren der Login-Daten. Nutzt man einen der Passwortmanager, kann man die privaten Schlüssel sogar exportieren und in eine andere Schlüsselverwaltung importieren. Passkeys sind ein Versuch, Usability, Verfügbarkeit und Sicherheit neu auszubalancieren, um das Gros der Internetnutzer von der passwortlosen Anmeldung mit FIDO2 zu überzeugen.

In Kontexten, in denen mehr Sicherheit gewünscht ist, kann man weiterhin



Auch die Versionsverwaltungsplattform GitHub unterstützt seit einiger Zeit Passkeys.

auf physische Sicherheitsschlüssel zurückgreifen. Ob Passkeys wirklich die Standardmethode zum Einloggen im Web werden, hängt vor allem an den Webdiensten. Derzeit sieht es so aus, als könnte die Rechnung aufgehen. Im Mai hat Google die Anmeldung per Passkey für die Konten der Nutzer aktiviert. Große Webdienste wie PayPal, TikTok, Ebay, GitHub und zuletzt Amazon und WhatsApp haben sich eingereiht. Viele weitere dürften folgen: Google wird offenbar Anfang November eine Android-spezifische API namens Credential Manager veröffentlichen, die es App-Entwicklern laut dem Newsportal TheVerge einfacher machen soll, Passkeys in ihren Apps anzubieten. Es scheint insgesamt, als haben weitere Dienste den Bedarf mittlerweile erkannt. So gab etwa der CTO von Heise Medien gegenüber c't an, dass Passkey-Unterstützung für heise.de auf seiner Wunschliste relativ weit oben stehe. (kst@ct.de) **ct**

So funktionieren Passkeys

Beim Einrichten eines neuen FIDO2-Logins, eines „Passkeys“, erzeugt Ihr FIDO2-Stick oder ein Hardwaremodul Ihres Smartphones, Tablets oder Rechners ein kryptografisches Schlüsselpaar, bestehend aus einem öffentlichen und einem privaten Schlüssel. Der öffentliche Schlüssel ist nicht geheim. Er wird auf dem Server des Webdienstes abgelegt. Der private Schlüssel ist geheim. Er wird sicher in der Hardware oder in einer Cloud verwahrt.

Um die Verwirrung über die Begrifflichkeiten zu komplettieren, bezeichnet das Kunstwort „Passkey“ oft den privaten Schlüssel dieses Schlüsselpaars. Etwa auf der Website der FIDO Alliance selbst. Dort referenziert das Wort „Passkey(s)“ an mancher Stelle „passwortlose Authen-

tifizierung mit FIDO2“, an anderer eben diese privaten Schlüssel.

Loggen Sie sich per Passkey bei einem Webdienst ein, generiert dieser mithilfe des öffentlichen Schlüssels eine kryptografische Challenge. Für die Lösung ist der private Schlüssel nötig. Diesen geben Sie per Fingerabdruck oder Gesichtsscanner, Knopfdruck oder PIN frei. Die Lösung der Challenge, nicht der private Schlüssel selbst, wird an den Webdienst zurückgeschickt und schon sind Sie eingeloggt. Das Verfahren ist Phishing-resistant, weil für das asymmetrische Challenge-Response-Verfahren beide Schlüssel benötigt werden. Den öffentlichen Schlüssel hat nur der echte Webdienst, den privaten Schlüssel haben nur Sie. Die Domain des Webdienstes fließt in das kryptografische Verfahren mit ein.

Literatur

- [1] Kathrin Stoll, „Wir haben den Wendepunkt gerade erreicht“, Andrew Shikiar, geschäftsführender Direktor der FIDO-Alliance, über die Erfolgsaspekte von Passkeys, c't 18/2023, S. 28
- [2] Ronald Eikenberg, Passkeys im Einsatz, Passkeys: So schützen Sie Ihre Accounts mit dem Passwort-Nachfolger, c't 18/2023, S. 14

WebAuthn-Spezifikation,
TheVerge-Berichterstattung: ct.de/yd26

Online-Konferenz – 23. November 2023



TEAM UP!

Teamentwicklung in Zeiten von Remote-Work



... denn Remote-Teamentwicklung schafft neue Herausforderungen

- Wie kann erfolgreiche Teamentwicklung speziell in Remote- und Hybrid-Umfeldern gelingen?
- Wen oder was braucht es dafür?
- Und wie lässt sich das neue Verhalten nachhaltig verankern?

Für viele dieser Fragen haben sich in der Remote-Arbeit Lösungsansätze bewährt, die im Zentrum dieser Online-Konferenz stehen. Ausgewiesene Experten und Expertinnen zeigen erfolgreiche Wege, mit denen Teams ihre Ziele klar definieren und umsetzen können. Team Up! wendet sich an Führungskräfte und Verantwortliche in Projektteams, an (Agile) Coaches & Consultants und an Personalentwicklerinnen.

Jetzt
Tickets
sichern!

teams.inside-agile.de

+++ Außerdem Online-Workshops am 24./25. November, 6./7. Dezember und 7. Dezember 2023 +++

Veranstalter



dpunkt.verlag



MAGAZIN FÜR
PROFESSIONELLE IT

Online-Konferenz – 28. November & 5. Dezember 2023



AGILE LEADERSHIP CONFERENCE

So werden agile Teams besser



Kaum ein Unternehmen kommt heute noch ohne **agile Arbeitsweisen** aus. Der **Leadership Day (28.11.)** und der **Self Leadership Day (5.12.)** behandeln aktuelle Herausforderungen von Führung und Management:

- Wie führt man selbstorganisierte Teams und wieso ist dabei die Selbstführung entscheidend?
- Wie kann man Mitarbeitende beurteilen, wenn die Teamleistung im Fokus steht?
- Braucht es überhaupt noch disziplinarische Führungskräfte?

Die Konferenz richtet sich an **Gruppen-/Team-/Abteilungsleiterinnen und -leiter** sowie erfahrene **Scrum Master/Agile Coaches**.

Wer teilnimmt, sollte ein agiles Grundverständnis mitbringen.

Jetzt
Tickets
sichern!

alc.inside-agile.de

+++ Außerdem Online-Workshops am 29. November und 8. Dezember 2023 +++

Veranstalter



dpunkt.verlag



MAGAZIN FÜR
PROFESSIONELLE IT

Kooperationspartner



Agile Academy



Bild: KI Midjourney | Bearbeitung c't

Arbeitsteilig abkassieren

Die Industrialisierung der Cyberkriminalität

Nach einer Atempause 2022 steigen die Zahlen erfolgreicher Ransomware-Angriffe wieder. Und ein Ende dieser Bedrohung für Unternehmen und Behörden ist nicht in Sicht, denn die Erpressungsbranche professionalisiert sich immer weiter: Spezialisierte Dienstleister bieten Phishing-as-a-Service, geklauten Zugangsdaten oder simpel zu bedienende Verschlüsselungssoftware.

Von Jürgen Schmidt

Mit elektromagnetischen Funkwellen einer Wellenlänge von etwa zwölf Zentimetern kann man Wassermoleküle in Rotation versetzen, was durch Reibung Wärme erzeugt. Das wissen vor allem Physiker und Elektrotechniker, aber praktisch jeder kann davon profitieren. Man geht einfach ins nächstgelegene Kaufhaus, kauft ein Mikrowellengerät, überfliegt vielleicht noch die Bedienungsanleitung und drückt auf den Knopf. Drei Minuten später macht es pling und das Essen ist wie durch Zauberhand heiß.

Dass quasi jeder ein solches Mikrowellengerät kaufen und sich das auch leisten kann, ist ein Resultat der Industrialisierung. Sie hat dafür gesorgt, dass hochkomplexe Dinge arbeitsteilig sowie weitgehend automatisiert hergestellt und als Ware für einen Massenmarkt verkauft werden. Diese Umstellung hat die mensch-

liche Gesellschaft so grundlegend umgekrempelt, dass häufig sogar von der industriellen Revolution gesprochen wird.

Keiner kann alles

Eine ähnlich grundlegende Umwälzung erfährt gerade das organisierte Verbrechen. Denn Cybercrime macht derzeit das Frühstadium der Industrialisierung durch. Früher brauchten Verbrechen im Cyberspace eine Menge Expertenwissen und Fertigkeiten aus vielfältigen Bereichen: Das Finden und Ausnutzen von Sicherheitslücken erfordert häufig Reverse Engineering und Low-Level-Programmierung. Beim Social Engineering helfen Psychologie, Wortgewandtheit und Schauspielkunst. Softwareentwicklung, Netzwerk-Know-how, Kryptografie – die Liste der benötigten Fertigkeiten lässt sich beliebig fortsetzen.

Da kaum jemand Experte in all diesen Bereichen sein kann, schlossen sich Kriminelle schon bald zu Gruppen zusammen, deren Mitglieder sich gegenseitig ergänzten. Und nachdem die Experten Programme und Werkzeuge für die vielfältigen Aufgaben erstellt hatten, lag es nahe, diese mit den weniger versierten Teammitgliedern zu teilen. Oder auch mit befreundeten Gangs – gegen ein kleines Entgelt oder eine passende Gegenleistung natürlich.

Der Trend zur Arbeitsteilung hat dazu beigetragen, dass die systematische Erpressung von Firmen zum Goldesel der Cybercrime-Szene wurde. Die Einnahmen aus diesem dreckigen Geschäft stiegen über Jahre hinweg rasant. 2021 verzeichnete die auf Blockchain-Analyse spezialisierte Firma Chainalysis über 760 Millionen US-Dollar als Eingänge auf Wallets, die sie bekannten Cybercrime-Gruppierungen zuordnet. Das sind wohlgerne nur die erpressten Lösegelder. Die Schäden liegen um ein Vielfaches höher; das Bundeskriminalamt bezifferte die Summe des 2021 durch Ransomware verursachten Schadens allein in Deutschland auf 24,3 Milliarden Euro.

Nur eine Atempause

Im Jahr 2022 gingen sowohl die Zahl der Vorfälle als auch die Einnahmen der Banden nach mehreren Rekordjahren erstmalig deutlich zurück. Doch das lässt sich zumindest zum Teil auf Russlands Angriffskrieg gegen die Ukraine zurückführen. Der sprengte viele der kriminellen Strukturen, die sich vor allem in den Nachfolgestaaten der Sowjetunion gebildet hatten.

So stieß etwa den ukrainischen Mitgliedern der Conti-Bande der flammende Aufruf ihrer Chefs zum gemeinsamen Kampf an Putins Seite recht bitter auf. Kurze Zeit später veröffentlichten Unbekannte riesige Mengen an internen Daten: Von umfangreichen Chatlogs über Bedienungsanleitungen bis hin zum Quellcode der eigentlichen Verschlüsselungssoftware war plötzlich alles offen einsehbar. Die damals größte Ransomware-Bande überlebte das nicht und löste sich kurz darauf auf. Aber auch andere Banden hatten mit solchen Konflikten im eigenen Lager zu kämpfen.

Hinzu kamen Sanktionen der USA, die Zahlungen an bestimmte, in Russland verortete Cyberkriminelle verboten, sodass etwa eine Zahlung an die Lockbit-Bande

nur noch mit einer Ausnahmegenehmigung der Finanzaufsichtsbehörde möglich ist. Und nicht zuletzt haben viele Unternehmen mittlerweile auch tatsächlich Backups, die eine zügige Wiederaufnahme des Produktionsbetriebs auch ohne Entschlüsselung der Daten möglich macht.

Doch bevor sich Freude über ein scheinbar gelöstes Ransomware-Problem breit macht: Es handelt sich wohl um einen Ausreißer. Der Cyberuntergrund hat sich bereits neu sortiert und es sieht stark danach aus, dass die Zahlen sowohl der Angriffe als auch der Lösegeldzahlungen 2023 wieder deutlich zunehmen. Die zunächst nur gefühlte Zunahme der Cybercrime-Aktivitäten bestätigen mittlerweile auch erste Statistiken für das erste Halbjahr 2023, zum Beispiel von Chainalysis (siehe Infografik auf S. 82).

Kriminelle mit Urlaubsanspruch

Entscheidend für die Entwicklung des Cybercrime war auch, dass viele der Banden unternehmerisch tätig wurden und einen Teil ihrer Einnahmen ins Geschäft reinvestierten. Mittlerweile haben die großen Banden Strukturen, die denen eines mittelgroßen Unternehmens erstaunlich ähneln: Die Angestellten haben geregelte Arbeitszeiten, Anspruch auf Krankengeld und Urlaub, sie arbeiten in Teams unter Führung eines mittleren Managements und so weiter.

Ein Großteil dieser Mitarbeiter hat zumindest nach dem eigenen Gefühl nichts

mit kriminellen Aktivitäten zu tun. Da optimiert ein Softwareentwickler etwa das GUI des Ransomware-Builders, das Admin-Interface eines Botnetzes oder die Geschwindigkeit der Verschlüsselung. Andere tragen aus öffentlichen Quellen Informationen zu Firmen zusammen. Selbst Spezialisten für Human Resources leisten sich die Banden. Alles Jobs, die sich so ähnlich auch in der regulären Wirtschaft finden.

Alles „as a Service“

Die treibende Kraft der Weiterentwicklung des Cybercrime-Ökosystems ist nun die bereits angesprochene Industrialisierung. Es gibt inzwischen alles zu kaufen, was man für ein erfolgreiches Cybercrime-Unternehmen benötigt, analog zur normalen Wirtschaft am besten „as a Service“ im Abomodell, oft mit Bezahlung anteilig an den erzielten Einnahmen.

Die Hersteller beziehungsweise Betreiber der Angebote arbeiten vornehmlich im Hintergrund und konzentrieren sich auf die Weiterentwicklung sowie auf die Vermarktung im Untergrund. Die eigentlichen Cybercrime-Aktivitäten überlassen sie weitgehend ihren Kunden, die dann virtuell in die Netze der Opfer einbrechen, dort Daten raustragen beziehungsweise verschlüsseln und Lösegeldforderungen platzieren.

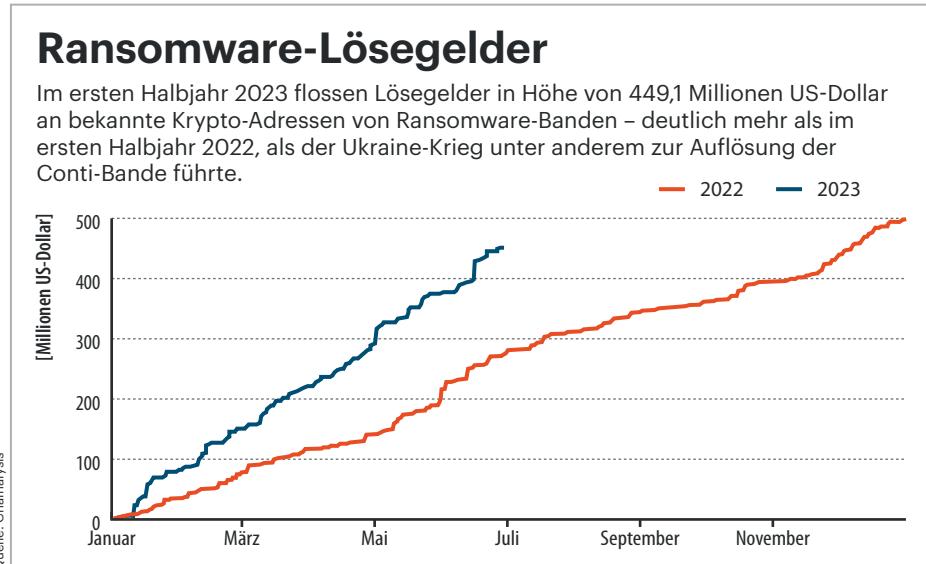
Diese Kunden sind somit die eigentlichen Täter hinter den Cyberangriffen. In der Mehrzahl handelt es sich um wenig gebildete Kriminelle, die oft bereits Erfahrung in anderen Verbrechensparten wie Drogenhandel oder illegalen Wettgeschäften haben und jetzt auf den so überaus lukrativen Ransomware-Zug aufspringen. Aber es gibt auch hochprofessionelle Teams, die viele Jahre Erfahrung mit Cybercrime vorweisen können. Die konzentrieren sich auf die Dinge, in denen sie ihre Stärke sehen, statt Zeit und Energie damit zu verplempern, etwa die 42te Verschlüsselungssoftware zu entwickeln.

Je nach Know-how suchen sich diese Täterschaften verschiedene Opfer aus. Während Newbies eher nach den niedrig hängenden Früchten greifen und bevorzugt bei kleineren Unternehmen mit wenigen Sicherheitsvorkehrungen einsteigen, gehen die geschulten Profis gezielt auf die Jagd nach möglichst lohnenden Zielen. Dafür hat sich der Begriff „Big Game Hunting“ etabliert: Großwildjagd.

Das spiegeln auch die Statistiken der Angriffe wider: Ein Großteil richtet sich gegen kleinere und mittlere Unterneh-

ct kompakt

- Im vergangenen Jahr kassierten Ransomware-Banden wegen des Ukrainekriegs weniger Lösegelder, doch mittlerweile steigen die Zahlen wieder.
- Wichtigster Treiber ist die Industrialisierung des Cybercrime: Zunehmend spezialisierte Dienstleister liefern Malware, Phishing-Kampagnen und vieles mehr von der Stange.
- Wegen der Professionalisierung und unternehmerischer Denkweise können Ransomware-Banden sich mittlerweile sogar Zero-Day-Exploits leisten, was bisher die Domäne staatlich finanziertem Hacker war.



men. So hatten laut einer Analyse der IT-Sicherheitsfirma Arctic Wolf 82 Prozent der Unternehmen, deren Daten im ersten Halbjahr 2023 von Kriminellen veröffentlicht wurden, weniger als 1000 Mitarbeiter. So etwas wird dann bestenfalls in der Lokalpresse erwähnt, aber nie große Schlagzeilen machen. Diese werden beherrscht von Einbrüchen bei bekannten Firmen und Organisationen, bei denen die Lösegeldforderungen dann auch schnell mehrstellige Millionenzahlen erreichen.

Die Phasen eines Angriffs

Um die Entwicklung der Untergrundmärkte weiter zu konkretisieren, hilft ein Einblick in den Ablauf eines typischen Angriffs. Der besteht aus mehreren Phasen, die oft auch von unterschiedlichen Spezialisten und teilweise sogar von verschiedenen Gruppen ausgeführt werden.

1. Initial Access: Die Angreifer verschaffen sich Zugang zum Netz des Opfers.
2. Sie erweitern ihre Rechte und breiten sich im Netz weiter aus.
3. Sie stehlen (exfiltrieren) wichtige Daten, löschen Backups und verschlüsseln schließlich Daten.
4. Lösegeldforderungen werden verhandelt, eventuell eingehende Bitcoins gewaschen, bei Problemen mit der Entschlüsselung gibt es Support.

Für all diese Aufgaben gibt es maßgeschneiderte Produkte von verschiedenen Anbietern. Die wichtigste Rolle spielt Ransomware-as-a-Service (RaaS). Deren Kernstück ist die eigentliche Verschlüsselungssoftware, die der Kunde über einen grafischen Builder an seine Bedürfnisse

anpassen kann, um etwa eine personalisierte Lösegeldforderung einzubauen. In der Regel übernimmt der RaaS-Anbieter auch das Management der zur Entschlüsselung der Daten nötigen Kryptoschlüssel, die er auf einem eigens dafür eingerichteten Server erstellt.

Im Paket enthalten ist typischerweise auch das Waschen der prinzipiell verfolgbaren Bitcoin-Lösegelder, sodass die Affiliates sich nicht mit Krypto-Exchanges und Mixern auskennen müssen. Darüber hinaus erhalten sie auch komplette Playbooks, die etwa gängige Techniken zum Erlangen von Admin-Rechten (Privilege Escalation und Lateral Movement) illustrieren und an konkreten Beispielen erklären. Bei Conti steckte das in einer Datei namens „hanting admins, please review, very helpful!!!.txt“. Und schließlich verhandeln die RaaS-Spezialisten auf Wunsch auch mit den Opfern. Dafür nehmen sie je nach Verhandlungsgeschick und Status der Affiliates 20 bis 70 Prozent der Einnahmen für sich in Anspruch.

Die Opfervermittler

Das bedeutet: Ein angehender Cyberkrimineller muss eigentlich nur noch mit ein paar Mausklicks seine personalisierte Ransomware zusammenklicken und dann passende Opfer finden. Die zu finden, ist die Spezialität der Initial Access Broker (IAB), die Zugänge zu Rechnern beziehungsweise Zugangsdaten verkaufen. Das Angebot reicht von einfachen Benutzernamen/Passwort-Kombinationen im Hunderter- oder Tausenderpack bis hin zu verifizierten VPN-Zugängen bei renommierten Firmen. Mit der Qualität der Ware steigt natürlich der Preis.

IABs investieren auch deshalb einen Aufwand, um Informationen über den jeweiligen Inhaber von Credentials zu sammeln und installieren dazu spezielle Info-stealer und Hintertürprogramme auf dem Rechner. Die sammeln dann alles ein, was sich irgendwie als Identitätsnachweis verwenden lässt: persönliche Daten, im Browser gespeicherte Passwörter, Cookies mit OAuth-Tokens und so weiter. Natürlich kann man solche Infostealer auch einzeln kaufen oder im Abo lizenziieren. Der Infostealer Redline etwa wird seit drei Jahren weiterentwickelt und von mehreren IABs gerne eingesetzt.

Das wichtigste Einfallstor für IABs sind E-Mails, die Empfänger dazu verleiten sollen, über Klicks auf Anhänge oder Links Schadsoftware zu installieren. Eine neue Qualität erhielt das durch Emotet, bei denen die Phishing-Mail scheinbar von Kollegen, Geschäftspartnern oder Freunden kam und sogar konkreten Bezug auf frühere Mails nahm. Die benötigten Daten für diese Fakes sammelte Emotet gezielt durch Outlook-Harvesting ein. Mittlerweile haben das die Konkurrenten von IcedID, Bumblebee und Qakbot auch alle im Programm. Letzterem hat das FBI Ende August die komplette Infrastruktur zerschlagen. Da es jedoch keine Verhaftungen gab, ist es wohl nur eine Frage der Zeit, bis Qakbot wie zuvor auch Emotet wieder voll im Geschäft ist.

Doch IABs haben auch andere Methoden, ihre Infostealer einzuschleusen. Haben Sie sich schon einmal gefragt, wer all diese gecrackten Versionen von Spezialsoftware in Umlauf bringt? Es steckt ja einiges an Arbeit darin, den Kopierschutz auszuhebeln. Oftmals geschieht das im Auftrag eines IABs, der dann dafür sorgt, dass der Anwender mit der gecrackten CAD-Software auch seinen Infostealer installiert.

Malware und Advertising

Im Lauf des vergangenen Jahres gab es immer wieder große Malvertising-Kampagnen einer neuen Qualität. Da kauften IABs wie IcedID bei Google und zum Teil auch Microsoft für richtiges Geld die oberen Anzeigenplätze für Suchergebnisse zu bekannter Software wie Audacity, Kee-pass, Blender3D oder WinRAR. In den letzten Monaten kamen vermehrt Anzeigen für Tutorials zu KI-Tools wie Midjourney, ChatGPT & Co. dazu.

Das Magazin von Fotografen für Fotografen



Fotograf und
c't Fotografie-Redakteur

Thomas Hoffmann, fotografiert am liebsten Landschaft und Natur.

„Für außergewöhnliche Fotos braucht man nicht nur eine verlässliche Ausrüstung, auch Geduld und Zufall spielen eine große Rolle.“

Thomas Hoffmann



Jetzt bestellen:

www.ct-foto.de/minabo



2x c't Fotografie testen

- 2 Ausgaben kompaktes Profiwissen für 14,30 €
- 35 % Rabatt gegenüber Einzelheftkauf
- Inklusive Geschenk nach Wahl
- Wöchentlicher Newsletter exklusiv für Abonnenten

Dank Googles Bemühungen, seinen Anzeigenkunden Traffic zuzuschieben, kann man die ganz oben platzierten Anzeigen kaum von echten Suchergebnissen unterscheiden. Und natürlich bemühen sich die IABs sehr darum, das alles so echt wie möglich aussehen zu lassen: Anzeigen-Text und Downloadseite sind oft vom Original übernommen oder zumindest sehr stark an deren Look-and-Feel angelehnt. Auf Wunsch übernehmen das alles gegen ein kleines Entgelt auch darauf spezialisierte Black-SEO-Profis: Industrialisierung bis in den letzten Winkel.

Damit erscheint dann bei der gewünschten Zielgruppe, die sich natürlich sehr detailliert einstellen lässt, als scheinbar erstes Suchergebnis etwa für WinRAR eine täuschend echte Kopie der Originalseite. Wer nur einen Moment unaufmerksam ist, lädt dort einen trojanisierten Installer herunter und führt ihn danach aus, mit Admin-Rechten natürlich. Und natürlich installiert der auch wie gewünscht WinRAR. Doch im Hintergrund hat er auch klammheimlich etwa den Infostealer Redline aktiviert.

Im Rahmen des Service Level Agreements für Abokunden garantieren dessen Entwickler, dass zu diesem Zeitpunkt keines der relevanten AV-Programme Alarm schlägt. Das soll übrigens keine Werbung für Redline sein; Konkurrenten wie Vidar und Racoon haben ähnliche Angebote. Das Unheil nimmt jedenfalls seinen Lauf, Passwörter und andere Zugangsdaten landen beim IAB, der über Redlines Remote

Administration Toolkit (RAT) auch weiterhin vollen Zugang zum System hat. Er verkauft das dann entweder an einen seiner Stammkunden oder auf entsprechenden IAB-Plattformen weiter und schlussendlich bekommt im schlimmsten Fall der Arbeitgeber des Opfers Besuch von einem skrupellosen Lockbit-Kunden.

Industrialisierte Zero-Days

Ein weiteres wichtiges Einfallstor sind Sicherheitslücken, vor allem in Diensten, die direkt aus dem Internet zu erreichen sind. Klassischerweise ging die größte Gefahr da von bekannten Sicherheitslücken aus, gegen die es bereits Updates gibt. Denn mit diesem Update legt der Hersteller die Information offen, wo genau sich die Lücke befindet und oft sogar auch, wie sie sich ausnutzen ließe. Und natürlich gibt es im Untergrund Spezialisten, die diese Informationen aus den Patches extrahieren und dazu passende Exploits schreiben, die dann bösartigen Code einschleusen und ausführen. Und viel zu oft sind sie dabei schneller als die Admins, die die schützenden Patches installieren müssten.

Während das auch weiterhin passiert, hat das sich drehende Industrialisierungsrad neue Angriffsformen für den Massenmarkt Cybercrime erschlossen. Bei sogenannten Zero-Day-Lücken, kurz Odays, spüren die Angreifer selber neue, bis dato unbekannte Sicherheitslücken auf und schreiben dafür Exploits. Gegen die gibt es dann keinen einfachen Schutz. Das war aufgrund der erforderlichen Fertigkeiten

und des damit verbundenen Aufwands – das Entwickeln eines voll funktionsfähigen, stabilen Exploits dauert oft Wochen oder sogar Monate – lange Zeit staatlich finanzierten Angreifern vorbehalten.

Doch die florierende Cybercrime-Wirtschaft macht es möglich, dass Banden entweder solche Spezialisten anheuern oder zumindest Geschäftsbeziehungen zu Gruppen unterhalten, die keine Hemmungen haben, ihnen solche Zero-Day-Exploits zu verkaufen. So hat die Ransomware-Bande C1Op bereits mehrere Zero-Day-Lücken in File-Transfer-Software systematisch ausgenutzt: in Accellions File Transfer Appliance, Fortras GoAnywhere Managed File Transfer und im Frühjahr schließlich in der Datei-Transfer-Software MOVEit von Progress.

Dazu hat die Bande sogar ihre Vorgehensweise umgestellt: Das herkömmliche Prozedere, systematisch ein Opfer nach dem anderen abzuarbeiten, hätte nicht so richtig funktioniert. Denn das Einbrechen, Auskundschaften, Erobern und Verschlüsseln der Daten benötigt seine Zeit. Und ab dem ersten erfolgreichen Angriff müssen die Erpresser damit rechnen, dass ihr wertvoller Zero-Day-Exploit bekannt und das Einfallstor verschlossen wird. Im besten Fall hätten sie ein paar Dutzend Opfer auf diesem Weg kompromittieren können, von denen dann vielleicht ein oder zwei gezahlt hätten.

MOVEit: Einer der größten Raubzüge

Also änderten sie den Modus Operandi: Für den MOVEit-Coup scannten sie zunächst das gesamte Internet nach anfälligen Servern und sammelten Informationen über ihre zukünftigen Opfer, wohl, um eine priorisierte Liste der vermutlich zahlungsbereiten Opfer zu erstellen. Danach arbeiteten sie diese Liste möglichst schnell ab, brachen innerhalb kürzester Zeit bei allen Opfern ein und stahlen alles an Daten, was sie dort erreichen konnten. Mit zeitraubendem Verschlüsseln hielten sie sich nicht auf; nicht einmal Hintertüren wurden auf den geknackten Servern gefunden. Es ging lediglich um den Datenaub.

Bei der nachfolgenden Erpressung drohte C1Op denn auch ausschließlich mit der Veröffentlichung dieser gestohlenen Daten. Das ist eine Vorgehensweise, die mittlerweile Schule macht. Sie löst das bekannte Verschlüsseln von wichtigen Daten keineswegs ab, sondern ergänzt es um eine weitere Erpressungsoption, die je nach Si-



Die Entwickler von Ransomware-as-a-Service (RaaS) bieten ihren Kunden grafische Builder, mit denen diese die Software an ihre Bedürfnisse anpassen.

tuation alternativ oder auch ergänzend ins Spiel kommt [1].

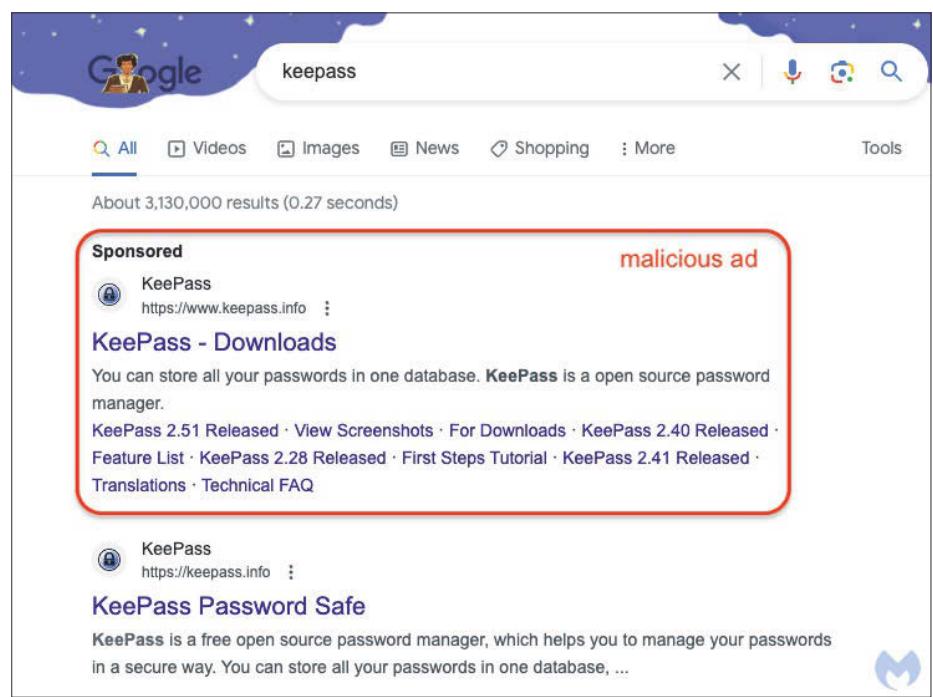
Für die eigentliche Erpressung ließ sich Cl0p dann wieder Zeit. Schließlich mussten für eine effiziente Verhandlungsführung und maximale Drohkulisse zunächst in jedem Einzelfall hunderte oder gar tausende GByte an Daten gesichtet werden. Im Oktober, über vier Monate nach den Angriffen, wurden immer noch neue Opfer bekannt; mittlerweile beläuft sich deren Zahl laut Emsisoft auf stolze 2342 (ct.de/yn9z). Der Angriff auf MOVEit-Nutzer ist damit einer der größten Erpressungsraubzüge der Cybercrime-Geschichte. Selbst wenn die Prozentzahl der zahlenden Opfer im einstelligen Bereich liegen sollte, dürfte sich der Raubzug für Cl0p mehr als gelohnt haben. Und die Konkurrenz hat bestimmt keine Hemmungen, ein so erfolgreiches Konzept nachzuahmen.

KI-Phishing als nächstes großes Ding?

Die Industrialisierung treibt die Entwicklung von Cybercrime auf eine Art und Weise voran, dass wir die Auswirkungen heute noch gar nicht überschauen können. Solange weiter solche Summen in diesen Markt fließen, wird sich die Weiterentwicklung sogar noch beschleunigen. Da tauchen ständig neue Angebote auf, etablierte Dienste erfinden sich neu oder verbessern sich signifikant. Letztes Beispiel: Phishing-as-a-Service. Die Zeiten der rädebrechenden nigerianischen Prinzen sind vorbei.

Bei aktuellen Angeboten nennen die Auftraggeber nur noch das Ziel. Für das erstellen die Dienstleister dann die passende Phishing-Seite aus einem Template-Baukasten, natürlich bei einem Bullet-Proof-Hoster, der alle Beschwerden ignoriert und anonyme Bezahlung akzeptiert. Ein Botnetz des Dienstleisters verteilt die Mails, deren Links zum Zeitpunkt des Verands nicht blockiert werden. Das garantiert der Anbieter, der auch detaillierte Statistiken über Klickraten liefert. Die Empfängerliste stellt der Dienstleister nach Wünschen des Kunden zusammen.

Und der erste Anbieter, der die Phishing-Texte mit einer guten KI erstellen lässt, die sie womöglich auch noch nach einer Recherche auf den Empfänger persönlich zuschneidet, wird den Markt aufrollen. Denn die Kunden wechseln durchaus die Anbieter, je nachdem, wer gerade die besten Konditionen oder die neuesten Features bietet.



Wer bei Google einfach nur nach seriöser, populärer Software sucht, kann sich Malware einfangen. Denn Kriminelle buchen immer wieder Topwerbeplätze bei der Suchmaschine. Die Links landen dann ganz oben auf der Ergebnisseite.

Teilweise nutzen sie auch verschiedene Angebote parallel und setzen etwa je nach Szenario die gerade besser passende Verschlüsselungssoftware ein: im Allgemeinen Black Basta, weil man da die besten Konditionen ausgehandelt hat, Lockbit, wenn es besonders schnell gehen muss und Alphv/BlackCat beim nächsten Opfer, weil deren Ransomware sich besser für Angriffe auf VMWare-ESXi-Server eignet. Lockbit wirbt zwar tatsächlich mit seiner hohen Geschwindigkeit beim Verschlüsseln, das Beispiel ist jedoch zur Illustration des Prinzips frei erfunden.

Die Vorteile der Standard-Malware

Damit kann man aus forensischer Sicht allein aus der am Tatort vorgefundenen Ransomware so gut wie nichts mehr über die dahinterstehenden Täter sagen. Man weiß damit weder, ob es sich um Profis oder Newbies handelt, noch kann man daraus eine typische Vorgehensweise ableiten. Es handelt sich um industriell gefertigte Massenware, die es an jeder Ecke zu kaufen gibt. Das ist übrigens sogar ein Grund, warum immer öfter auch Profis den Einsatz von Malware von der Stange einer selbstgeschriebenen Lösung vorziehen: Sie wollen Spuren vermeiden, die eventuell zu ihnen zurückverfolgbar wären.

Und damit schließt sich ein Kreis: Früher haben sich Cybercrime-Banden die Tricks und Techniken der staatlichen Elite-Hacker abgeschaut. Zero-Day-Exploits und Hands-on-Keyboard-Angriffe, bei denen sich Angreifer tatsächlich selbst auf den kompromittierten Rechnern und im Netz umsehen, waren zunächst deren Domäne und wurden erst in den letzten Jahren auch bei normalen Cybercrime-Vorfällen üblich. Heute gehen selbst staatliche Hacker im Cybercrime-Untergrund auf Shopping-Tour – ähnlich wie sie ganz selbstverständlich ihre Mikrowelle im Kaufhaus kaufen. (cwo@ct.de) ct

Der Autor Jürgen Schmidt leitete viele Jahre lang das Ressort Security bei c't und ist nun Senior Fellow Security bei Heise. Mit heise Security Pro baut er eine Community für IT-Professionals auf, in der aktuelle Angriffe, Probleme der täglichen Praxis sowie aktuelle und künftige Technik diskutiert werden. Weitere Informationen dazu finden Sie unter heise.de/heisec-pro.

Literatur

[1] Olivia von Westernhagen, GPCodes Erben, Geschichte der Ransomware, Teil 3: Erpressung als Dienstleistung, c't 11/2023, S. 150

Erwähnte Statistiken: ct.de/yn9z



Bild: EU-Kommission

Jetzt wird umgesetzt

Was die neuen Digitalgesetze für Unternehmen und Bürger ändern werden

Eine ganze Batterie von Gesetzen und EU-Verordnungen entfaltet ab 2024 ihre Wirkung. Insbesondere die EU-Plattformregulierung ändert eine Menge, fast immer zum Vorteil für die Verbraucher. Ein Blick voraus.

Von Falk Steiner

Hört man sich dieser Tage in den Entscheidungsgremien der Europäischen Union um, spürt man allerorten

hektische Betriebsamkeit. Das Motto könnte lauten: so viele Acts, so wenig Zeit! Die EU-Kommission unter ihrer Präsidentin Ursula von der Leyen hatte sich zum Antritt digitalpolitisch sehr viel vorgenommen. Eine Datenstrategie sollte erarbeitet und umgesetzt werden, vor allem aber wollte man den turbokapitalistischen Plattformmarkt einhegen – und so die EU als leuchtendes Vorbild etablieren.

Nun wird es knapp, denn die Wahl zum neuen EU-Parlament naht. Im Mai und Juni 2024 wählen die Bürgerinnen und Bürger in den 27 Mitgliedsstaaten ihre Vertreter und damit indirekt mit der EU-Kommission auch eine neue Regierung. Das

bedeutet: Spätestens Ende März startet die Wahlkampfphase, dann steht der legislative Betrieb der Union de facto still. Viele Gesetzesvorhaben, die bis dahin nicht unter Dach und Fach sind, sterben einen stillen Tod. Deshalb schraubt man derzeit eifrig an endgültigen Versionen von wichtigen neuen Verordnungen wie dem Medienfreiheitsgesetz und dem KI-Gesetz.

Digitalpolitisch wird 2024 auch ein Jahr werden, in dem viele der bereits beschlossenen EU-Gesetze zeigen müssen, was sie wert sind. Besonders im Fokus von Konzernen, aber auch von Bürgerinnen und Bürgern steht wohl der Digital Services Act (DSA) – er wird im Laufe des

kommenen Jahres seine volle Wirkung entfalten. Und das auf gleich zwei Ebenen: Zum einen werden sich besonders große Plattformen wie Facebook, X, YouTube und TikTok genauer an die Buchstaben des Gesetzes halten müssen [1].

Aber auch fast alle kleineren Anbieter von Plattformen und Hostingdiensten müssen die Grundvorgaben des DSA erfüllen. Alle sozialen Netzwerke, Video-plattformen, Empfehlungsdienste, Webforen und Kommentarbereiche fallen ab dem 17. Februar 2024 unter den DSA. Bis dahin muss jeder EU-Mitgliedsstaat eine unabhängige Behörde benennen, die dort dafür zuständig ist, den DSA durchzusetzen und die beteiligten Stellen zu koordinieren. Nach dem Willen der Bundesregierung soll dieser sogenannte „nationale Koordinator“ in Deutschland bei der Bundesnetzagentur angesiedelt werden.

Illegal nicht mehr egal

Gemäß DSA gilt ab Februar 2024: Wann immer ein Anbieter über nach nationalem Recht möglicherweise illegale Inhalte – also etwa Urheberrechtsverstöße, Beleidigungen oder Straftaten nach dem Betäubungsmittelgesetz – informiert wird, muss er die Meldung prüfen und gegebenenfalls zeitnah den Zugang zu den Inhalten unterbinden. Außerdem muss er den betroffenen Nutzer darüber informieren und Widerspruch ermöglichen. Bei besonders gravierenden Straftaten, die etwa eine Gefahr für Leib und Leben darstellen, müssen Anbieter zudem die zuständigen Behörden informieren.

Was vielen Nutzern und Anbietern noch nicht klar ist: Diese Pflichten werden sämtliche Onlinedienste treffen, die nutzergenerierte Inhalte veröffentlichen. Man denke etwa an Steam-Diskussionen, Buchrezensionsplattformen oder auch die Kommentarforen von heise online. Alle diese Dienste müssen ab dem 17. Februar leicht verständliche Meldeprozesse vorhalten. Tun sie das nicht, drohen Bußgelder. Nutzer können potenzielle Verstöße beim nationalen DSA-Koordinator anzeigen.

Der DSA bringt aber noch mehr Änderungen: So verbietet er jegliche personalisierte Werbung auf Plattformen, die sich – etwa mittels Targeting anhand von minderjährigentypischen Merkmalen – gezielt an Unter-18-Jährige wenden. Auch das Profiling mittels besonders sensibler Daten nach Artikel 9 Abs. 1 DSGVO ist künftig verboten. Ethnische Herkunft, politische Meinungen, religiöse oder welt-

anschauliche Überzeugungen, Gewerkschaftsangehörigkeit, genetische Daten, biometrische Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben sind damit für Werbetreibende in der EU tabu.

Mit dem DSA sind ab Februar zudem irreführende Oberflächengestaltungen ausdrücklich verboten. Wohl jeder kennt die grauen „Widersprechen“-Schaltflächen und die grünen „Akzeptieren“-Buttons für Einwilligungen zur Datenverarbeitung oder zum Abschluss von Verträgen. Das Verbot umfasst aber auch andere Methoden, etwa Werbeanzeigen, die wie nutzergenerierte Inhalte aussehen. Allerdings enthält der DSA einige Unschärfen. Was am Ende unter die „Dark Pattern“-Regelung fällt, werden in den kommenden Jahren Gerichte klären müssen.

Relevant für Verbraucher sind auch die bislang eher wenig beachteten Vorschriften für Onlinemarktplätze wie Amazon, die ab Februar greifen werden: Sie sollen den Kampf gegen Betrug und Produktfälschungen künftig erleichtern. Anbieter, die ihre Produkte auf Marktplätzen einstellen, müssen sich gegenüber der Handelsplattform identifizieren sowie eine Telefonnummer, Zahlungskonten, Registereinträge und eine Konformitätsklärung für ihre Produkte abgeben.

Einige der Daten, wie Name, Anschrift und Kontaktmöglichkeiten, stehen dann über die Marktplatzbetreiber auch deren Kunden zur Verfügung, also den Verbrauchern. Die Betreiber müssen grundsätzlich prüfen, ob Anbieter ihren Informationspflichten nachkommen. Außerdem sollen sie stichprobenhaft kontrollieren, ob Anbieter auch illegale Produkte im Portfolio haben, also etwa Produktfälschungen. Wenn ein Anbieter gegen diese Regeln verstößt, muss ihn der Marktplatzbetreiber sperren.

Ge- und Verbote für Gatekeeper

Ergänzend zur Plattformregulierung im DSA wird ab März 2024 auch der Digital Markets Act (DMA) für jeden Bürger spürbar in den Markt eingreifen. Mit dem DMA beschreitet die EU-Kommission einen neuen Weg des Wettbewerbsrechts: Anstatt Konzerne für Fehlverhalten im Nachhinein zu sanktionieren (ex post), definierter sie in Art. 5 und 6 des DMA einen Katalog von Pflichten und Verboten, dem sich sogenannte Torwächter (Gatekeeper) zu unterwerfen haben.

ct kompakt

- Ab 2024 müssen Konzerne wie Google und Amazon ihre Angebote verbraucherfreundlicher gestalten. Die Verordnungen DSA und DMA werden dann voll wirksam.
- Auch weitere neue Vorschriften bringen Änderungen, etwa die USB-C-Pflicht.
- Wegen der Europawahl wird 2024 in der EU legislativer Stillstand vorherrschen. Neue Gesetze sind dann nicht zu erwarten.

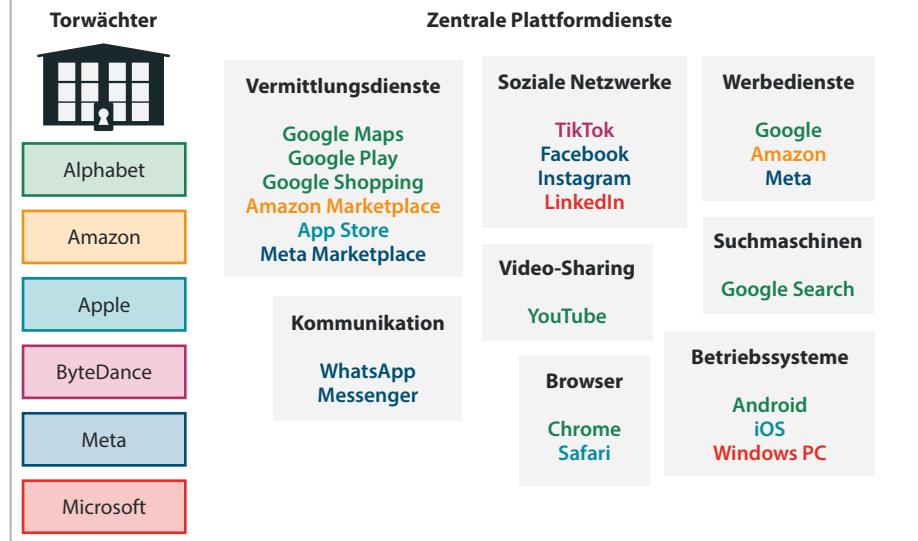
Diese Torwächter designiert die EU-Kommission selbst. Am 6. September hat sie die ersten sechs verbindlich benannt [2]. Demnach müssen sich Alphabet, Amazon, Apple, ByteDance, Meta und Microsoft den DMA-Vorschriften stellen – allerdings gilt diese Pflicht nicht für sämtliche Dienste der Konzerne (siehe Abbildung). Ab Benennung hat ein Torwächter dem DMA zufolge sechs Monate Zeit, die neuen Anforderungen umzusetzen, die aus seiner marktbeherrschenden Stellung resultieren. Für die genannten Konzerne endet diese Frist folglich am 6. März 2024.

Wer Torwächter ist, darf seine eigenen Produkte und Dienstleistungen auf der eigenen Plattform nicht besserstellen als die von anderen Anbietern. So hat die EU-Kommission Googles Suchmaschine als Torwächter benannt, was für Google heißt: Ab März kontrolliert die Kommission, ob Google-eigene Dienste in den Suchergebnissen prominenter platziert werden als andere. Die hervorgehobene Darstellung von Google-Shopping-Ergebnissen zuungunsten anderer Preisvergleicher etwa könnte bald der Vergangenheit angehören. Fraglich ist auch, ob Kartendienste von anderen Anbietern nicht gleich prominent wie die von Google Maps platziert werden müssen.

Das erklärte Ziel, Verbrauchern mit dem DMA mehr Auswahl zu gewährleisten, betrifft aber nicht nur Plattformen, sondern auch die Installation von Software und deren Integration ins Betriebssystem. Alles, was nicht unabdingbarer Bestandteil von Betriebssystemen ist, muss der Nutzer deinstallieren können. Der Betriebssystemhersteller muss außerdem Alternativen zulassen. Noch ist nicht ab-

Torwächter-Benennungen

Am 6. September 2023 hat die EU-Kommission bekanntgegeben, welche Dienste von der DMA-Regulierung umfasst sind.



sehbar, wie diese Pflicht umzusetzen ist. Auswahlfenster könnten eine größere Rolle spielen, etwa zum Austausch des Webbrowsers, wie man sie aus dem Browser-Krieg einstmals von Windows kannte.

Ebenso könnte die EU-Kommission monieren, wie Googles Suche im Android-Betriebssystem integriert oder die Browser-Engines in Android und iOS verzahnt sind. Auch bei den App-Stores stehen Änderungen durch den DMA an. Apple zum Beispiel verhindert bislang, dass Nutzer Apps auch aus anderen Quellen installieren können als über den eigenen Store.

Genau dieses, von Apple bekämpfte, sogenannte Sideloadung ist ein erklärtes Ziel des DMA: Bislang hat Apple nicht nur technisch den Daumen auf allem, was in iOS geladen wird. Insbesondere kassiert das Unternehmen bei jedem Kauf mit. Hier will die EU-Kommission ansetzen und neue Geschäftsmodelle innerhalb des Apple-Ökosystems ermöglichen.

Es könnte also sein, dass 2024 zumindest innerhalb der EU erstmals andere Stores auf Apple-Geräten verfügbar werden. Vorläufig noch nicht als Gatekeeper benannt ist übrigens iPadOS, die Schwester von iOS für größere Endgeräte. Das allerdings könnte sich in den kommenden Monaten ändern, sodass die Sideloadung-Pflicht auch bei iPads greifen würde.

Geförderter Datenaustausch

Auch den Messenger-Diensten widmete sich die EU-Kommission im DMA, insbe-

sondere der Vormachtstellung von Metas Services WhatsApp und Facebook Messenger. Die beiden fallen unter die speziellen Regeln für „Over-The-Top-Messenger“. In mehreren Schritten sollen solche „rufnummernunabhängigen interpersonalen Kommunikationsdienste“ zur Interoperabilität mit Dritten gezwungen werden: Sie müssen künftig eine Kommunikation ihrer Nutzer etwa zu Threema, Signal, Telegram, Wire, Apples iMessage oder Matrix-basierten Kommunikations-apps ermöglichen.

Im ersten Schritt sollen sich die Gatekeeper-Dienste WhatsApp und Facebook Messenger 2024 für Einzelchats nach außen öffnen: Nutzer anderer Messenger müssen an einzelne Nutzer dieser Dienste Textnachrichten, Bilder, Videos und andere angehängte Dateien verschicken und von diesen empfangen können. In einem zweiten Schritt (2025) sollen auch Gruppenchats interoperabel sein. Nach vier Jahren tritt dann 2027 eine dritte Stufe in Kraft: Dann müssen auch Sprach- und Videoanrufe Messenger-übergreifend möglich sein. Die Öffnungspflicht greift allerdings nur, wenn sich überhaupt ein anderer Dienst meldet, der gerne mit diesen Gatekeepern kompatibel kommunizieren möchte – Signal und Threema schließen das für die Meta-Angebote bislang aus.

Weitgehend unbemerkt von der Öffentlichkeit ist mit dem Data Governance Act (DGA) seit dem 24. September ein

weiter Baustein in der EU-Datenstrategie wirksam. Das Gesetz soll zusammen mit dem noch nicht verabschiedeten Data Act (DA) den Austausch von Daten im europäischen Binnenmarkt ankurbeln und ähnlich wie der DMA hermetische Ökosysteme aufbrechen, allerdings im öffentlichen Sektor. Die EU-Kommission beschreibt das Verhältnis der beiden Gesetze zueinander so: „Während der Data Governance Act die Prozesse und Strukturen zur Erleichterung des Umgangs mit Daten schafft, klärt der Data Act, wer unter welchen Bedingungen einen Mehrwert aus Daten schaffen kann.“

Ab 2024 sollen dazu „Datenvermittlungsdienste“ nach Art. 2 DGA als Drehscheiben treuhänderisch und gegen Bezahlung Daten zur Weiterverwendung anbieten können. Außerdem definiert der DGA einen „Datenaltruismus“ und Organisationen, die sich diesem verschreiben. Hierbei geht es um Datendrehscheiben, die Daten fürs Allgemeinwohl öffentlich verfügbar machen. Als Beispiele nennt die EU-Kommission die Corona-Datenspende-App und die Smart-Citizen-Plattform, die es Bürgern ermöglicht, Daten über Lärmpegel und Luftverschmutzung in ihrem Haus auszutauschen, die über Sensoren gesammelt werden.

USB-C wird verpflichtend

Eine weitere praxisrelevante Neuregelung der EU greift zwar de facto erst Ende 2024, dürfte aber schon zuvor Auswirkungen haben: die USB-C-Pflicht. Akkubetriebene Geräte wie Spielekonsolen, Tablets, Smartphones, Kopfhörer, Digitalkameras, Tastaturen und E-Book-Reader müssen dann mit dem normierten Stecker ausgestattet sein. Bei In-Ear-Stöpseln reicht es, wenn die Ladeschale USB-C nutzt.



Zertifizierte Datenvermittlungsdienste und Altruismus-Organisationen dürfen sich ab 2024 mit Logos auf der Website zieren.

Cybersicherheit in der Industrie 4.0 und kritischer Infrastruktur

23. November 2023

Sicherheit ist oberstes Gebot, wenn es um die IT für Produktionsanlagen geht. Cyberangriffe nehmen zu. Vernetzte Systeme sind die neue Realität. Doch der Bedarf an Security-Experten bleibt ungedeckt. **Sind Sie bereit, diese Lücke zu schließen?**

Unser Webinar bereitet Sie auf die Herausforderungen vor. Erfahren Sie von unseren Experten die Best Practices und die Chancen sowie Risiken, die damit einhergehen. In interaktiven Sessions und praktischen Übungen können Sie das Gelernte direkt anwenden.

Lernen Sie, die fünf größten Cybersicherheitsprobleme in modernen Produktionsanlagen zu bewältigen: **OT-Security, Netzwerksicherheit, Lieferkettensicherheit, Cloud-Computing und die Rolle der Mitarbeitenden.**



Jetzt Tickets sichern:

heise-academy.de/webinare/cybersec623





Glasfaserstrecken in der Straße gibt es inzwischen vielerorts, jetzt soll sie auch ins Haus. Auch deshalb entfällt 2024 das Nebenkostenprivileg für Koaxialkabelanschlüsse.

Zwar greift diese Regelung erst spät im Jahr – doch die meisten Hersteller haben die Entwicklung bereits eingepreist und berücksichtigen die Standards schon jetzt. Das eine oder andere Mini-USB-Kabel darf also im kommenden Jahr in Richtung Verwertungshof entsorgt werden.

Weitere EU-Gesetze, die gerade final verhandelt werden, dürften 2024 noch keine größere Rolle spielen, etwa die Verordnung über Künstliche Intelligenz (AI Act) oder der Cyber Resilience Act (CRA). Beide sind als eher mittelfristig wirksame Gesetze angelegt, ersteres, um rechtliche Rahmenbedingungen für den KI-Einsatz festzulegen, und zweiteres, um Cybersicherheitsanforderungen für vernetzte Endgeräte zu definieren, von der Waschmaschine über den smarten Fernseher bis zur IP-Überwachungskamera.

VDS-Debatte und IT-Sicherheitspflichten

In Deutschland könnte 2024 eine Entscheidung darüber fallen, ob sich die Bundesregierung endgültig von der anlasslosen Vorratsdatenspeicherung verabschiedet und stattdessen auf das sogenannte Quick-Freeze-Verfahren setzt. Damit würde, so der Vorschlag von Bundesjustizminister Marco Buschmann (FDP), im Verdachtsfall eine Anordnung an die Provider gehen, vorhandene Daten ab diesem Zeitpunkt für die Strafverfolger „einzufrieren“ und ihnen auf einen späteren richterlichen Beschluss hin zur Verfügung zu stellen. Bundesinnenministerin

Nancy Faeser (SPD) ist das eigentlich zu wenig: Sie will mindestens, dass Provider IP-Adressen auf Vorrat speichern.

Eine Veränderung steht auch im Bereich der Netzwerk- und Informationssicherheit an: Mit der Umsetzung der europäischen NIS2-Richtlinie sollen alle Mitgliedsstaaten 2024 neue „rechtliche Maßnahmen zur Steigerung des Gesamtniveaus der Cybersicherheit in der EU“ einführen, wie die EU-Kommission schreibt. Insbesondere wird sich dadurch in Deutschland der Kreis der verpflichteten Institutionen massiv ausweiten – von weniger als 10.000 auf etwa 30.000 Stellen.

Diese müssen dann, so viel ist auch schon vor Verabschiedung des deutschen Gesetzes klar, Cybersicherheitsvorfälle an das Bundesamt für Sicherheit in der Informationstechnik (BSI) melden. Außerdem werden sie sich im Bereich Detektion und Abwehr von Angriffen besser aufstellen müssen. Die NIS2-Richtlinie enthält überdies Vorschriften für Geschäftsführungen von Unternehmen: Ihnen dürfte es künftig schwerer fallen, sich aus Kostengründen vor IT-Sicherheitsmaßnahmen zu drücken.

Mobilfunk und Breitband

Last, but not least stehen auf nationaler Ebene 2024 Weichenstellungen für die digitale Infrastruktur an. So muss sich die Bundesnetzagentur festlegen, ob sie für 2028 auslaufende Frequenzbereiche im Mobilfunk eine Verlängerung unter Auflagen festlegt – die von der Bonner Aufsichts-

behörde derzeit bevorzugte Variante. Alternativ stünde eine neue Frequenzauktion an.

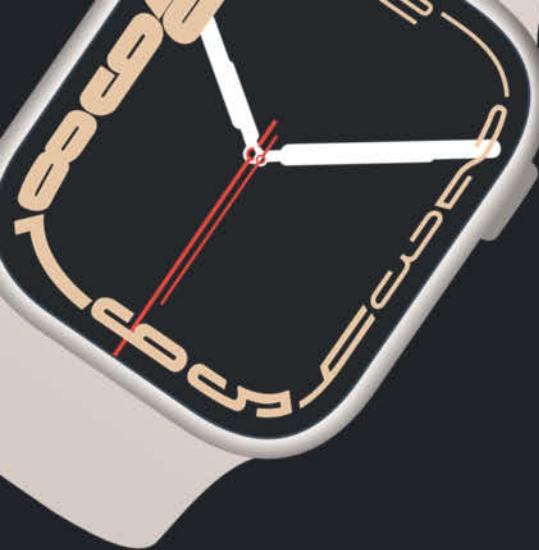
Sollte sich die Behörde für die Auflagenvariante entscheiden, könnte sich das positiv auf die Netzverfügbarkeit auswirken, bis hin zu einem Roaming von Anbietern untereinander. 2024 dürfte sich die Netzqualität ohnehin verbessern, weil die Anbieter bereits existierende Pflichten erfüllen müssen.

Beim Glasfaserausbau wird sich eine gesetzliche Änderung langsam bemerkbar machen: Mitte 2024 fällt das sogenannte Nebenkostenprivileg für Kabelanschlüsse endgültig weg. Das bedeutet, dass Mieter ihren Weg des Fernsehempfangs frei wählen können, also auch dann, wenn die Immobilie über Koaxialkabel erschlossen ist. Bislang durften Hauseigentümer den Kabelanschluss grundsätzlich über eine Umfrage mit den Betriebskosten abrechnen, ob er nun genutzt wird oder nicht. Das ist künftig nur noch dann erlaubt, wenn entsprechende Vereinbarungen mit den Mietern vorliegen. Ab Mitte 2024 gilt: Jeder Mieter darf selbst Verträge abschließen und über das Internet in die Röhre schauen oder sich etwa eine DVB-T2-Antenne hinstellen.

(hob@ct.de) **ct**

Literatur

- [1] Falk Steiner, Online-Riesen an der Leine, Digital Services Act ist für Meta, X & Co. wirksam, c't 21/2023, S. 30
- [2] Holger Bleich, Freier Eintritt, Neues EU-Wettbewerbsrecht für Digitalkonzern scharfgeschaltet, c't 22/2023, S. 36



MIT Mac & i IMMER DER ZEIT VORAUS

+ Geschenk
nach Wahl



z. B. 10 € Amazon.de-Gutschein oder Apple-Watch-Ständer

2x Mac & i mit 35 % Rabatt testen!

Mac & i – Das Magazin rund um Apple

- Tipps & Praxiswissen
- Hard- und Softwaretests
- Reports und Hintergründe

Für nur **16,80 € statt 25,80 €**

(Preis in Deutschland)

Jetzt bestellen:

www.mac-and-i.de/minabo



Mac & i. Das Apple-Magazin von c't.

Vertrackt

Aktuelle Trends beim Verbraucherschutz



In unserer Rubrik „Vorsicht, Kunde“ berichten wir in jeder Ausgabe über einen besonders eklatanten Fall der Missachtung von Verbraucherrechten. Dabei tauchen immer mal wieder neue Muster der Verfehlungen auf, insbesondere wenn Gesetze geändert wurden oder aktuelle Entwicklungen wie Pandemie oder Inflation ihren Tribut fordern. Manch unguter Trend hält sich jedoch auch über viele Jahre ungebrochen.

Von Tim Gerber

Die Corona-Jahre haben dem Onlinehandel einen kräftigen und weiter anhaltenden Boom beschert. Eine Folge davon ist ein Mangel an Fachkräften im Bereich Lager und Logistik, wie eine Auswertung von rund 45 Millionen online geschalteten Stellenanzeigen der Jahre 2019 bis 2023 durch die Bertelsmann-Stiftung zeigt. Die Kunden bekommen dies durch längere Liefer- und Reaktionszeiten im Kundenservice zu spüren.

Der in Ausgabe 23 dieses Jahres geschilderte Fall eines mangelhaften E-Bikes aus dem Online-Store von Aldi-Süd markiert dabei einen Trend, der sich auch im kommenden Jahr fortsetzen dürfte: Erst war die begehrte Ware wochenlang nicht lieferbar und das für die Sommerzeit gedachte Fahrrad kam erst im Herbst. Dann stellte sich im nächsten Sommer heraus, dass der Akku nicht ansatzweise die spezifizierte Reichweite erreichte. Schließlich

wurde die Reklamation wegen Urlaubs und Personalmangels wochenlang nicht bearbeitet.

Zu allem Unglück kam dann auch noch eine weitere typische Mängelerscheinung hinzu: nicht lieferbare Ersatzteile. Auf ein endloses Warten muss man sich in solchen Fällen aber nicht einlassen. Der Verkäufer muss in jedem Fall mangelfreie Ware liefern. Kann er die Mängel an der gelieferten Ware nicht innerhalb vernünftiger Fristen abstellen, muss er eben neue, mangelfreie Ware liefern.

Wahlrecht ausüben

Eigentlich hat der Kunde von Anfang an die Wahl, ob er eine Reparatur (im Gesetzesjargon Nachbesserung genannt) oder eine Neulieferung wünscht. An die Wahl des Kunden ist der Verkäufer gebunden. Tritt ein Sachmangel auf, sollte man dem Verkäufer am besten schriftlich und un-

missverständlich mitteilen, welche Form der Nachbesserung man wünscht.

Die mangelhafte Ware muss man dem Verkäufer bei beiden Varianten zunächst zur Prüfung des Mangels überlassen und ihm eine angemessene Frist dafür einräumen. Was angemessen ist, richtet sich danach, wie die Ware und der Mangel konkret beschaffen sind und wie aufwendig sich die Prüfung für den Verkäufer gestaltet. Unter zwei Wochen kommt man da kaum weg, bei einem komplexen Produkt wie einem Pkw steht die Rechtsprechung dem Verkäufer auch deutlich mehr Zeit zu.

In für sie kniffligen Fällen versuchen Verkäufer gern, sich ganz aus der Affäre zu ziehen, indem sie dem Kunden einfach den Kaufpreis erstatten. Rechtlich zulässig ist das ohne Einverständnis des Kunden nicht. Denn der zwischen Kunden und Lieferanten geschlossene Kaufvertrag ist weiterhin gültig.

Die Position des Kunden ist da deutlich besser: Er kann vom Kaufvertrag zurücktreten, wenn der Verkäufer einen Sachman gel nicht in angemessener Zeit beseitigt.

Preistreiber

Von steigenden Preisen sind manche Telekommunikationsunternehmen offenbar auch nach der x-ten Zinserhöhung noch immer überrascht. Jedenfalls lassen sie nichts unversucht, bei ihren Kunden nachträglich höhere Tarife abzukassieren. Dabei buhlt man lange nach dem Preis schock in Folge des Krieges in der Ukraine unverändert noch immer mit Lockangeboten um Kunden. Anschließend möchte man ihnen mit Verweis auf die hohen Energiekosten die satten Rabatte wieder streitig machen. Dem Treiben haben allerdings Verbraucherschützer mithilfe der Gerichte einen Riegel vorgeschoben.

Provider 1&1 wollte dennoch bei einem Kunden seine erhöhten Entgelte kassieren, obwohl man dessen Widerspruch gegen die Preiserhöhung schriftlich akzeptiert und eine Vertragsfortsetzung zu den bisherigen Konditionen zugesichert hatte. Nachdem wir den Fall für Ausgabe 22/2023 recherchiert hatten, entpuppte sich das dann wie so oft als pures Missverständnis.

IT-Pannen

Immer wieder machen Kunden auch IT-Pannen großer Unternehmen zu schaffen.



Auch hier dürfte der Fachkräftemangel sein Übriges tun. Rasche Problemlösungen sind meist nicht in Sicht. Prominentestes Beispiel für diesen Trend war die Deutsche Bahn mit ihren Vertriebsproblemen beim Deutschlandticket. Die regelmäßige Bereitstellung des Abo-Tickets in der App zum Beginn eines Monats klappt mal und mal auch nicht oder nur manuell durch Eingabe der Abo-Nummer.

Was klappt, ist der monatliche Einzug, gern auch mehrmals für denselben Fahrgast.

Ebenso reibungslos erfolgt dann das – freilich vollkommen grundlose – automatisierte Inkasso. In zwei Fällen mussten wir in den vergangenen Monaten über das unangemessene Vorgehen des Staatskonzerns berichten, stellvertretend für sehr viel mehr Betroffene, die sich an uns gewandt hatten. Typisch nicht nur bei der Bahn ist, dass E-Mails nicht beantwortet wurden.

Auch dafür dürften Kosten und zusehends auch ein Mangel an Personal ursächlich sein. Immer mehr Unternehmen versuchen inzwischen, bereits aus Kunden-Chats bekannte Algorithmen auch zur Beantwortung von Kundenmails einzusetzen. Gern verkauft man das trendig als KI, auch wenn die Systeme weit davon entfernt sind, auch nur Spuren von Intelligenz aufzuweisen.

Bezos' verschlossene Auster

Aber auch wo noch Menschen Hand anlegen, ist nicht sicher, dass der Umgang mit dem Kunden von Intelligenz bestimmt wird. Einen Sonderpreis in Ignoranz verdient sich regelmäßig der Online-Händelsriese Amazon. Bis zum Jahresbeginn sahen sich viele Kunden noch mit augenscheinlich von Kriminellen innerhalb der Logistikkette des Konzerns verursachten Fehllieferungen konfrontiert: Statt hochwertiger Smartphones kamen minderwertige Zahnpasta oder Duschgels an. Und oftmals versuchte Amazon, die betroffenen Kunden die Zeche für den Betrug zahlen zu lassen: Man forderte die Kunden auf, für eine Kaufpreiserstattung den „richtigen“ Artikel einzusenden, den sie aber nie erhalten hatten.

Über die Hintergründe solcher Pannen schweigt sich Amazon notorisch aus. Immerhin: Zuschriften zu solchen Fällen sind inzwischen deutlich spärlicher geworden. Wir schlussfolgern daraus, dass Amazon das Problem inzwischen halb-

wegs im Griff zu haben scheint. In einem anderen Fall (Ausgabe 21/2023) hatte der Kunde über den Amazon-Marktplatz ein gebrauchtes Buch für zirka 10 Euro verkauft, es abgeschickt und einen Zustellungs nachweis vorgelegt. Trotzdem wollte Amazon ihm den Erlös vorenthalten. Erst auf Nachfrage unsererseits bekam der Kunde dann doch noch sein Geld. Nachvollziehen lassen sich solche Vorgänge nicht, Amazon bleibt auch da wie stets jeden Erklärungsversuch schuldig. Sonderlich vertrauenserweckend ist das nicht.

Im Rausch der Glasfaser

Auch der immerhin in vielen Regionen erfreulich voranschreitende Glasfaser ausbau fordert mitunter das ein oder andere Opfer. Meist ist es die mangelnde Kommunikation zwischen den vielen an der Realisierung beteiligten Köchen – alter und neuer Provider, Tiefbau firma und Subsubunternehmen –, die den Betroffenen den Appetit auf den Highspeed-Anschluss gründlich verderben. In dem Fall, über den wir in Ausgabe 5/2023 berichteten, hatte O2 für seinen Glasfaserkunden dessen bisherigen DSL-Vertrag gekündigt, obwohl der Glasfaseranschluss noch gar nicht fertiggestellt war. Im System des Providers war er aber fälschlich vom Bauunternehmen als fertig hinterlegt. Ergebnis: DSL abgeschaltet und kein Internet.

Freilich hört in solchen Fällen niemand darauf, was der Kunde berichtet. Darauf ist „das System“ nicht vorbereitet. Alles läuft automatisch und im Einzelfall halt unaufhaltsam schief. Angesichts der hunderttausende Haushalte und Unternehmen, deren Umstellung auf Glasfaser noch ansteht, wird weiterhin mit solchen Pannen zu rechnen sein. Glücklicherweise scheinen sie jedoch gemessen an der Zahl der Anschlüsse relativ selten zu sein. Das lässt ausnahmsweise hoffen. (tig@ct.de) ct



Foto: Markus B.

Obwohl der Glasfaseranschluss eines Kunden alles andere als fertiggestellt war, kündigte sein Provider bereits den bisherigen DSL-Anschluss.



Europäisch oder amerikanisch bezahlen?

Mit Girocard und Wero gegen Visa und Mastercard

„Click to Pay“ von Visa und Mastercard lockt zukünftig mit simpler Bedienung. Doch in puncto Datenschutz und Kosten haben die digitale Girocard und das geplante Wero Vorteile. Wir stellen drei neue Bezahlwege für den Onlinehandel vor, die Ihnen 2024 über den Weg laufen dürften.

Von Markus Montz

Verbraucher kaufen immer mehr online ein, auch wenn die Zahl der Einkäufe und das Umsatzvolumen nach dem Coronaboom 2020/21 erst einmal rückläufig war. Zugleich wissen viele deutsche Verbraucher nicht nur einen hohen Bedienkomfort zu schätzen, wie ihn beispielsweise PayPal bietet. Auch Themen wie Datenschutz, Sicherheit und Kosten spielen bei der Wahl des Zahlungsmittels eine Rolle. Das ist auch Banken und Zahlungsdienstleistern bewusst, die in diesem heiß umkämpften Markt bestehen und mit neuen Features Kunden anziehen möchten.

Auf der einen Seite steht die digitale Girocard, mit der man mittlerweile in ersten Onlineshops bezahlen kann. Auf der anderen Seite haben sich die US-Riesen Mastercard und Visa in Position gebracht. Ihr neuer Standard „Click to Pay“ ist nach eigenem Bekunden einsatzreif und bei ersten Onlinehändlern bereits zu finden. Wir vergleichen beide und beleuchten die Pläne eines dritten Akteurs, der europäischen Bezahlinitiative (European Payments Initiative, EPI). Sie will 2024 ihr Wallet „Wero“ an den Start bringen.

Digitale Girocard

Die Girocard, ehemals EC-Karte, hat sich in Deutschland seit etlichen Jahren als beliebtestes elektronisches Zahlungsmittel im stationären Handel etabliert. Bei den meisten Banken und Sparkassen bekommen Kunden sie kostenlos zum Girokonto, zudem können sie damit anders als mit Visa und Mastercard an fast allen Ladenkassen in Deutschland bezahlen. Überdies bleiben die (ohnehin sparsam fließenden) Daten in Deutschland. Händler kostet die Girocard pro Zahlung deutlich weniger als die Karten der US-Netzwerke. Allerdings war sie lange Zeit nicht für den Onlinehandel geeignet.

Das hat die Deutsche Kreditwirtschaft (DK), der Spitzenverband der deutschen Banken und Sparkassen, ab Mitte der 2010er Jahre geändert. Seither können die Kreditinstitute technisch eine digitale Version der Girocard erzeugen, also einen digitalen und zweckgebundenen Zwilling der Karte. Kunden brauchen dann nur noch ein Smartphone oder anderes geeignetes Gerät zum Bezahlen.

Für NFC-fähige Android-Smartphones bieten die Sparkassen, die Volks- und Raiffeisenbanken sowie weitere Genossenschaftsbanken bereits seit 2018 Bezahl-Apps für den stationären Handel an: „Mobiles Bezahlen“ bei den Sparkassen, „Pay“ bei den VR-Banken und „Digitales Bezahlen“ bei einigen Sparda-, PSD- und anderen Genossenschaftsbanken. Der Kunde aktiviert darin zunächst mit wenigen Handgriffen eine digitale Girocard und ist anschließend sofort startklar.

Die Sparkassen haben die Girocard außerdem in das Wallet des iPhones integriert; seit 2021 kann man sie über

Apple Pay auch im Onlinehandel nutzen. Allerdings muss der Händler nicht nur die Girocard akzeptieren, sondern auch Apple Pay.

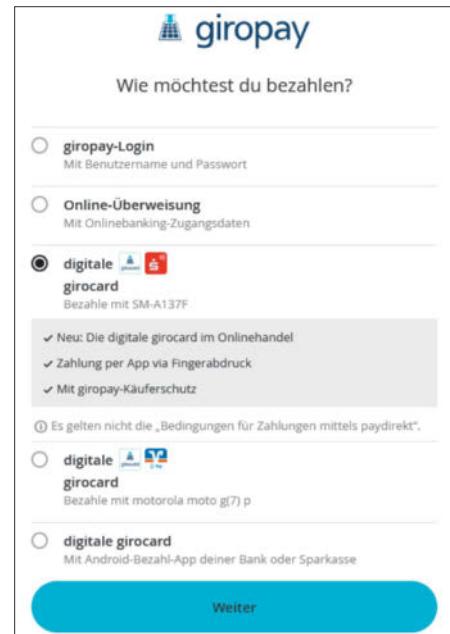
Seit Ende 2022 sind auch die Apps von Sparkassen und Genossenschaftsbanken technisch in der Lage, Onlinezahlungen abzuwickeln. Dazu hat die DK die digitale Girocard mit dem runderneuerten Onlinebezahlsystem Giropay verknüpft [1, 2]. Seither hat die digitale Girocard zwei wichtige Fortschritte gemacht. Erstens sind Zahlungen mit ihr nicht mehr auf Einkäufe auf dem Smartphone beschränkt, sondern funktionieren auch im Browser am PC. Zweitens können Kunden der Sparkassen und Genossenschaftsbanken anstelle der Giropay-App die Bezahl-App ihres Instituts nutzen, wenn die mit der digitalen Girocard zahlen.

Bedienung yeah, Akzeptanz meh

Im Versuch zeigt sich schnell, dass der hauseigene DK-Dienstleister Paydirekt, zusammen mit Giropay ein einfach und schnell zu bedienendes System gebaut hat. Um die digitale Girocard bei Onlinehändlern zu nutzen, checkt man als Kunde mit Giropay (!) als Bezahlmethode aus. Anschließend wählt man die digitale Girocard seines Instituts, bekommt eine Push-Nachricht in die Bezahl-App und gibt diese per Fingerabdruck oder Gesichtsscan frei. Dazu muss man beim Onlinehändler eingeloggt sein, damit dieser im Hintergrund die Mailadresse an Giropay übergeben kann. Wenn man das nicht möchte, kommt man trotzdem ans Ziel; am PC scannt man mit der Bezahl-App einen QR-Code, am Handy oder Tablet leitet Giropay die Anfrage an die Bezahl-App.

Da Giropay und Girocard dem Kosmos der deutschen Kreditwirtschaft angehören, bleiben die Daten in beiden Fällen in Deutschland. Es fließen grundsätzlich nur Daten, die für die Zahlung oder Rückabwicklung erforderlich sind. Bei einigen Händlern können Warenkorb-inhalte dabei sein; laut Datenschutzerklärung von Giropay jedoch nur dann, wenn es zur Betugsprävention oder für Rückabwicklungen erforderlich ist. Bei uns passte das nicht. Giropay will die Daten aber löschen, sobald sie nicht mehr dafür benötigt werden. An Dritte, etwa zu Marketingzwecken, gehen die Daten nicht.

Die Apps sind laut DK gut gegen unbefugten Zugriff geschützt und die digitale Karte kann man nur mit dem Gerät ver-



Im Auswahlmenü von Giropay kann man direkt eine Push-Zahlung in der Bezahl-App seiner Sparkasse oder Volksbank anstoßen.

wenden, auf dem sie liegt. Gefahren drohen in erster Linie, wenn Cyberkriminelle Nutzer durch psychologische Tricks zu Fehlern verleiten und diese den Tätern beispielsweise in deren Bezahl-App eine Karte freischalten [3]. Für Probleme mit Händlern gibt es einen Käuferschutz. In summa kann man die digitale Girocard damit nach unserem Eindruck weitgehend bedenkenlos und komfortabel im Onlinehandel nutzen.

Doch bisher bieten nur eine Handvoll Onlineshops Zahlungen mit der digitalen Girocard an. Wir haben mit dem Dienst aufladen.de getestet, denn große und weithin bekannte Händler fehlen noch. Das Zögern vieler Onlinehändler hängt Experten zufolge auch damit zusammen, dass hinter der digitalen Girocard eben nicht nur die Girocard selbst steckt, sondern zusätzlich das System von Giropay. Das führt bei Händlern, aber auch bei Kunden noch zu Verwirrung. Beiden ist der Unterschied häufig nicht klar, Händler müssen zudem zwei sonst getrennte Komponenten gedanklich verbinden.

Baustelle Nummer zwei sind die Privatbanken. Selbst Girocard-treue Institute wie Commerzbank, Deutsche Bank oder Postbank bieten die digitale Girocard noch nicht an, ganz zu schweigen von den Direktbanken Comdirect, Consors, DKB und ING. Sie stellen mittlerweile Visa-Debitkarten als primäre Karte zum Girokon-

ct kompakt

- Volksbanken und Sparkassen bringen ihre datenschutzfreundliche und komfortable digitale Girocard endlich in den Onlinehandel.
- Click to Pay ist simpel, kostet Kunden aber indirekt mehr Geld und verlangt noch wichtige Vorsichtsmaßnahmen.
- Als europäische Bezahlmethode soll Wero starten und hohen Datenschutz und niedrige Kosten gewährleisten.



Sobald man die Zahlung mit der digitalen Girocard in der Bezahl-App seiner Bank oder Sparkasse bestätigt hat, ist der Prozess abgeschlossen.

to aus [4]. Diese Karten lassen sich leicht in Click to Pay von Visa und Mastercard integrieren. Die ING ist außerdem stark in der European Payments Initiative (EPI, beide siehe unten) engagiert.

Ebenso schauen Apple-User noch in die Röhre, sofern sie keine Sparkassenkunden sind. Zwar wäre es technisch möglich, die Apps von Sparkassen und Genossenschaftsbanken unter Verzicht auf die NFC-Bezahlfunktion für die Ladenkasse auf iPhone & Co. zu bringen. Aufgrund ihres Streits mit Apple um die Öffnung der NFC-Schnittstelle auf iPhones für Fremdienste ist das jedoch in naher Zukunft nicht zu erwarten.

Löst die DK diese Probleme nicht, werden auch die geplanten neuen Funktionen wenig Wirkung entfalten. Eine Altersverifikation ist technisch bereits einsatzbereit, ebenso eine Verknüpfung mit Bonuspunkteprogrammen. Der Handel wünscht sich außerdem Funktionen, die bei Kreditkarten längst zum Standard gehören, wie Reservierungen, Nachbelastungen für Hotels und Autovermietungen sowie regelmäßig wiederkehrende Zahlungen, also Abos. Es ist aber kein Geheimnis mehr, dass daran bereits gearbeitet wird.

Click to Pay

Die Konkurrenz schlägt nicht. Visa und Mastercard und übrigens auch American Express und Discover haben unter dem Titel „Click to Pay“ die neue Industriespezifikation „EMV Secure Remote Commerce“ marktreif gemacht. Die Idee: Kunden müssen beim Bezahlen mit der Kredit- oder Debitkarte keine Kartennummern, Ablaufdaten und Namen mehr eintippen. Stattdessen genügt beim Checkout im Onlineshop eine Mailadresse. Mit dieser Mailadresse hat man vorher beim Kartennetzwerk ein eigens dafür vorgesehenes Nutzerkonto eröffnet und bereits Namen, Adresse und Mobilfunknummer hinterlegt.

Für die Karte selbst erzeugt das Kartennetzwerk ähnlich wie in den Wallets von Apple Pay oder Google Pay Token, also Pseudo-Kartennummern. Händler oder Zahlungsabwickler bekommen nur diese Token in ihre Systeme. Die Verknüpfung mit der echten Kartennummer stellen die Kartennetzwerke her und leiten Zahlungen anschließend an die Bank des Händlers weiter.

Bei Mediamarkt und Saturn können Inhaber von deutschen Visa-Karten bereits mit Click to Pay zahlen, Mastercard nannte uns als Partner unter anderem Norwegian Airlines und Qatar Airways. Beim Checkout wählt man dazu die Option „Kartenzahlung / Click to Pay“. Wichtig: Die im Kundenkonto hinterlegte Mailadresse muss mit der Mailadresse im Click-to-Pay-Konto übereinstimmen.

Bei unserem Testlauf enthielt die Bedienoberfläche noch erkennbare Provisionen. Nach dem Klick auf „Bezahlen“ poppte ein Fenster mit den Logos von Visa und dem Zahlungsabwickler Computop auf, der das Verfahren technisch umsetzt. Darin wählten wir die gewünschte Karte aus und bestätigten die Zahlung. Beim ersten Mal schickte Visa zunächst einen Bestätigungscode per Mail und SMS, fragte anschließend nach dem Sicherheitscode auf der Karte und authentifizierte per 3-D Secure. Bei Versuch zwei entfielen diese Zwischenschritte. Für den Bestätigungscode hatte Visa nun auf unseren Wunsch ein Cookie gesetzt und bei 3-D Secure griff eine Ausnahmeregel (Zahlung unter 30 Euro). So machte der Prozess einen komfortablen Eindruck.

Ganz so sehr wie die digitale Girocard können wir Click to Pay aber noch nicht empfehlen. Zwar verhindert das Kartentoken, dass die Kartendaten in falsche Hände geraten. Allerdings genügt für das Login ins Click-to-Pay-Nutzerkonto derzeit die Mailadresse, an die Visa oder Mastercard anschließend ein Einmalpasswort schickt. Man muss das Mailkonto, dessen Mailadresse man mit Click to Pay verknüpft hat, also unbedingt durch eine Zwei-Faktor-Authentifizierung schützen. Das Login wie auch der Checkout dürften

Zahlungsart

Zahlungsinformationen geben Sie im letzten Schritt ein.

Amazon Pay - Nutze Dein Amazon Konto
amazon pay

Finanzierung - 0 % bis zu 10 Monate für myMediaMarkt-Kunden

17,64 €/Monat
für 10 Monate mit 9.9% Finanzierung
Finanzierungsgesamtbetrag 176,40 €

Bezahlen Sie schnell und unkompliziert in kleinen Monatsraten. Die Vermittlung Ihrer Finanzierung erfolgt ausschließlich für den Finanzierungspartner BNP Paribas S.A. Niederlassung Deutschland, Standort München: Rüdesheimer Straße 1, 80686 München

Bitte beachten Sie, dass eine Finanzierung nur möglich ist, wenn Sie bereits 18 Jahre alt sind.

Paypal
PayPal

Kartenzahlung / Click to Pay

VISA

Kauf auf Rechnung (jetzt bestellen, später bezahlen)

Vorkasse

Giropay

Händler, die Click to Pay anbieten, zeigen dies an der Kasse bei den Kreditkartenzahlungen an.

aber noch überarbeitet werden, zumal Visa und Mastercard ohnehin biometrische Authentifizierungsverfahren vorantreiben.

Beim Thema Datenschutz haben wir ebenfalls Bedenken, solange das Nutzerkonto mit den persönlichen Daten bei den Kartennetzwerken selbst liegt. Zwar nutzen Visa und Mastercard die Daten ausweislich ihrer Datenschutzerklärungen nicht für Marketingzwecke. Dafür haben beide ihren Hauptsitz in den USA und der neue Datenschutzrahmen mit der EU hat den Härtetest vor dem EuGH noch nicht bestanden.

Die Erfahrung zeigt, dass Visa und Mastercard datenschutzrechtlich heikle Dinge lieber an die kartenausgebenden Banken delegieren. Schließlich kennen beide Netzwerke auch für normale Kreditkartenzahlungen nur Karten- und Tokennummern und reichen Zahlungsanfragen lediglich an die Bank des Kunden weiter. Ob und wann die Banken die Nutzerverwaltung von Click to Pay übernehmen, ist aber unklar.

Wenn Zahlungsdienstleister und Banken erst einmal die Schwächen behoben haben, ist das Potenzial von Click to Pay enorm. Kartenzahlungen werden nicht nur deutlich einfacher. Denkbar sind auch Komfortfunktionen, zum Beispiel Express-Checkouts wie bei PayPal oder Amazon Pay bei gleichzeitig hohem Sicherheits- und Datenschutzniveau. Zusammen mit der steigenden Zahl an Debitkarten von Visa und Mastercard haben die beiden Netzwerke gute Chancen, bei ihren Kunden zu punkten.

Wero: Europas Antwort?

Neben der digitalen Girocard und Click to Pay macht sich mit der European Payments Initiative (EPI) ein drittes Bezahlsystem startklar. Das Konsortium europäischer Banken und Zahlungsdienstleister will eine europaweite Alternative zu den US-Konzernen schaffen. Nach internen Richtungsstreitigkeiten sind neben zwei Zahlungsabwicklern noch 14 Banken oder Bankenverbände übrig, darunter aus Deutschland die Sparkassen-Finanzgruppe, die genossenschaftliche DZ Bank und die Deutsche Bank.

Das Konsortium entwickelt derzeit ein europaweit nutzbares Wallet namens Wero. Nach Auskunft von EPI soll Wero Mitte 2024 zunächst in Belgien, Deutschland und Frankreich verfügbar sein. Wero wird direkt an das Bankkonto der Nutzer

angebunden, Zahlungen fließen von Konto zu Konto. Kunden können Wero auf Wunsch in ihr Onlinebanking im Browser oder in die Smartphone-Banking-App integrieren. EPI hat außerdem eine eigene App angekündigt. In der ersten Ausbaustufe soll man sich damit „Person to Person“ (P2P) sowie „Person to Professional“ (P2Pro) manuell Geld schicken können, vergleichbar mit PayPal.

Darauf will EPI nach eigener Aussage „in kurzer Zeit“ Online- und Mobilzahlungsfunktionen für Onlineshops folgen lassen. Im Lauf des Jahres 2025 sollen auch Zahlungen an der Ladenkasse möglich sein, was von den Händlern und ihren Zahlungsabwicklern abhängt. Die EPI plant, Wero den Händlern mit niedrigen Entgelten schmackhaft zu machen. Für Kunden sollen Abonnements, Ratenzahlungen und Reservierungen möglich sein. Das Konsortium nennt außerdem „Später zahlen“-Optionen, Bonuspunkteprogramme sowie digitale Identitätsdienste als denkbare Zusatzfunktionen.

Da EPI seinen Prototypen erst für Anfang 2024 angekündigt hat, kann man zum Komfort, zu Datenschutz und der Sicherheit noch nichts sagen. Zu den Werbeversprechen gehört, dass Nutzerdaten in der EU bleiben. Wenn Wero ähnlich komfortabel wird wie das Schweizer Twint und Dänemarks Mobile Pay und wenn man damit auch auf Flohmärkten oder in kleinen Geschäften bezahlen kann, stehen seine Chancen nicht schlecht. Genau diese Nische erreichen die Kreditkartenfirmen nämlich nicht, während PayPal sich dort wegen seiner hohen Entgelte nicht etablieren kann, mit denen der Dienst die Händlermargen drückt. Ein weiterer unmittelbar einleuchtender Vorteil für Kunden ist, dass das Verfahren auch in anderen EU-Ländern funktioniert, ohne dass sie sich Gedanken über die richtige Karte oder einen Datenfluss über den Atlantik machen müssen.

Fazit

Die digitale Girocard und Wero werden es schwer gegen die geballte Markt- und Technologiemacht von Visa, Mastercard & Co. und gegen deren Marketingaufwand haben. Deren Abläufe im Onlineshopping sind bei vielen Kunden eingespielt – und werden mit Click to Pay radikal einfacher. Unsere Sicherheits- und Datenschutzbedenken beim Prototypen dürften sich im Endzustand zudem zerstreuen. Dass schnell weitere Händler hinzukommen,



Um mit Click to Pay zu zahlen, wählt man nach dem Checkout einfach die gewünschte Karte aus, muss keine Nummern eintippen und oft auch keine weiteren Schritte mehr gehen.

ist abzusehen. Die Frage ist lediglich, wann Visa und Mastercard die noch deutlichen Rückstände beim Datenschutz und der Sicherheit des Nutzerkontos angehen.

Genau dort punktet die digitale Girocard, deren Bedienkomfort PayPal und anderen mittlerweile ebenbürtig ist. Doch guter Datenschutz und Komfort helfen ihr nicht, wenn nicht mehr Banken mitziehen und man nur in wenigen Shops damit bezahlen kann, zumal kaum welche im Ausland darunter sind. Umso wichtiger wäre es, dass der dritte Spross im Bezahlgarten, nämlich Wero von EPI, nun endlich Früchte abwirft, damit Verbraucher auch europaweit eine vergleichbar datenschutzfreundliche, komfortable und preisgünstige Alternative zu den US-Konzernen erhalten, zu denen neben Visa und Mastercard auch PayPal zählt. Deren hohe Entgelte treiben nämlich auch die Verbraucherpreise hoch. (mon@ct.de) **ct**

Literatur

- [1] Markus Montz, Europäisch oder amerikanisch? Elektronisches Bezahlen am Scheideweg, c't Jahresrückblick 2022, S. 148
- [2] Tobias Weidemann, Renovierung erfolgreich, Warum man mit Giropay nun besser zahlt als mit PayPal oder Kreditkarten, c't 4/2023, S. 122
- [3] Markus Montz, Neue Masche beim Kartentrug, c't 26/2022, S. 37
- [4] Markus Montz, Kleines Logo, große Wirkung, Mit welcher Debitkarte Sie besser bezahlen: Girocard oder Visa und Mastercard, c't 22/2022, S. 114



Bild: Rudolf A. Blaha

Nutzervölker auf Wanderung

Social Media in Zeiten der X-Emigration und der Kurzvideos

Wer macht das Rennen als X/Twitter-Nachfolger: Das verteilte Fediverse, Metas Threads, das ebenfalls Teil des Fediverse werden soll, oder ein anderer Konkurrent? Bluesky etwa erhält derzeit viel Zulauf. Und wie entwickelt sich die Lage bei TikTok und seinen Konkurrenten?

Von Jo Bager, Greta Friedrich und Eva-Maria Weiß

X, wie Twitter seit Juli heißt, geht es glänzend. Das behauptet zumindest die im Juni von Elon Musk als CEO eingesetzte neue Chefin Linda Yaccarino. 90 Prozent der größten Werbekunden würden inzwischen wieder Anzeigen schalten, schon Anfang 2024 werde man schwarze Zahlen schreiben.

Elon Musk hatte im November 2022 sogar das Ziel von einer Milliarde Nutzer für 2024 ausgegeben, was eine Verdopplung im Vergleich zu den damaligen Zahlen bedeuten würde. Musk hat Großes mit X vor, aus dem Kurznachrichtendienst soll eine Universal-App mit Audio-, Video-, Zahlungs- und Bankfunktionen werden.

Laut neutralen Statistiken läuft es allerdings nicht so gut. So lag der Web-

Traffic von X im August zehn Prozent unter dem Wert des Vorjahrs und die Downloadzahlen der App liegen auf einem Zehnjahrestief, so die Zahlen der Marktforschungsunternehmen Similarweb und Apptopia.

Die Nutzer verlassen die Plattform und das hat viele Gründe, zum Beispiel neu eingeführte Einschränkungen wie das Leselimit. Zahlende Nutzer bekamen Anfang Juli pro Tag nur 8000 Beiträge zu sehen, nichtzahlende sogar nur 600. Angeblich sollte diese Maßnahme der Bekämpfung von Bots dienen. Links zu Konkurrenten oder zu Medien, mit denen Elon Musk im Clinch liegt, hat X zeitweise blockiert oder den Aufruf über den URL-Verkürzer t.co verzögert. Zuletzt hat Elon

Musk im September angekündigt, dass X für alle Nutzer kostenpflichtig werden soll.

X hat diese Einschränkungen schnell wieder über Bord geworfen und es ist wohl auch sehr unwahrscheinlich, dass die Plattform jemals eine Bezahlschranke einziehen wird. Dennoch nagen solche Maßnahmen an der Geduld und der Restloyalität der Nutzer. Die sind es leid, dass elementare Funktionen mal verfügbar sind, mal nicht, und sie nicht wissen, woran sie mit X eigentlich sind.

Ein weiterer Grund für den Exodus vieler Nutzer ist Musk selbst. Der Milliardär hat die Führung zwar offiziell abgegeben, bestimmt aber aufgrund seiner riesigen medialen Präsenz, unter anderem mit rund 160 Millionen Followern, das Erscheinungsbild von X. Musk hatte angekündigt, wöche Einflüsse auf X, etwa der „Mainstream-Medien“, zurückzudrängen und die Plattform für mehr Meinungsfreiheit zu öffnen, also: seine Meinung und mehr Stimmen aus dem rechten Spektrum.

Sender wie BBC, NPR, PBS sowie CBC wurden so zeitweise als „staatlich“ oder „öffentlich finanziert“ gebrandmarkt, eine Kennzeichnung, die sonst Medien aus Ländern wie China oder Russland vorbehalten ist. Ins Bild passt auch, dass Musk Donald Trump wieder auf seine Plattform gelassen hat. Trump, der nach dem versuchten Staatsstreich 2022 gesperrt worden war, postet mittlerweile wieder auf X.

Musk selbst provoziert immer wieder durch Grenzüberschreitungen in seinen Posts, etwa indem er antisemitische Klischees bedient („Georg Soros hasst die Menschheit.“). All das führt zu einer Atmosphäre auf X, in der rassistische und antisemitische Hetze sowie Fake News gedeihen. So ist X unter den großen sozialen Netzwerken das mit Abstand schlechteste, wenn es um Maßnahmen und Vorgaben geht, die gegen Desinformation rund um den Klimawandel wirken. Das ergab eine im September veröffentlichte Untersuchung der Climate Action Against Disinformation Coalition.

Viele Nutzer fühlen sich daher auf X nicht mehr wohl und sehen sich nach Alternativen um. Und davon gibt es reichlich.

Threads

Im Juli startete die Facebook- und Instagram-Mutter Meta ihren Twitter-Konkurrenten Threads, allerdings bislang nicht in

ct kompakt

- Viele Nutzer fühlen sich auf X nicht mehr wohl.
- Die Wechselwilligen wählen aus diversen vergleichbaren Netzwerken.
- Social-Media-Dienste gucken sich Funktionen der Konkurrenz ab.

Europa. Das liegt vermutlich daran, dass Threads in der aktuellen Form im Konflikt mit dem Digital Markets Act steht (siehe S. 86). Der schreibt einflussreichen IT-Konzernen vor, dass sie einen Dienst nicht innerhalb eines anderen Dienstes aus eigenem Hause bevorzugen dürfen.

Trotz der Einschränkung in Europa wurde Threads in der übrigen Welt schnell zu einem großen Erfolg. Meta meldete nach rund vier Tagen bereits mehr als 100 Millionen eröffnete Konten. In der ersten Woche verzeichnete Threads bis zu 50 Millionen täglich aktive Nutzer. Das hat auch damit zu tun, dass die Registrierung per Instagram-Account besonders einfach ist. Entsprechend lässt sich ein Threads-Account bislang auch nicht unabhängig vom Instagram-Konto löschen. Die Zahl der täglich aktiven Nutzer ist nach der ersten Euphorie wieder gesunken und hat sich derzeit bei etwa 10 Millionen eingependelt.

Meta hat die Bedienoberfläche stark an die von X angelehnt. Allerdings erlaubt Threads bis zu 500 statt 280 Zeichen pro Beitrag. Wegen der Instagram-Verzahnung findet man schnell Nutzer, denen man folgen kann. Zudem lassen sich Posts zu einer Insta-Story hinzufügen. Auch die Funktion der blauen Haken von verifizierten Accounts hat Threads von X übernommen.

Allerdings fehlen im Vergleich viele Funktionen, etwa die Suche, eine Timeline oder öffentlich einsehbare Follower-Listen. Derzeit gibt es auch keine Hashtags, sie sollen in naher Zukunft folgen. Außerdem lässt sich der Dienst derzeit ausschließlich über Mobilgeräte und nicht per Browser verwenden.

Meta hat angekündigt, Threads an das ActivityPub-Protokoll anzubinden, also an das Fediverse. Damit soll man „eines Tages“ sogar seine Follower mitnehmen können, wenn man von Threads auf eine andere Plattform umzieht. Das wäre eine

ungegewohnte Offenheit für einen Konzern, dessen Apps auch aufgrund von Lock-in-Effekten erfolgreich sind.

In der Community des Microblogging-Dienstes Mastodon ist man uneinig darüber, ob man diesen Schritt begrüßen sollte. Die Befürchtung: Threads könnte seinen Client um proprietäre Funktionen erweitern, um Nutzer dorthin zu locken, und später die Verbindung zum Fediverse wieder kappen.

Fediverse und Mastodon

Eines der beliebtesten Ziele von X-Emigranten war und ist Mastodon. Rund 1,7 Millionen täglich aktive Nutzer zählt das verteilte Netzwerk. Es kommen weiterhin neue Nutzer hinzu, aber langsam. Mastodon steht im Ruf, Neulingen den Einstieg nicht gerade einfach zu machen.

Das haben auch die Mastodon-Entwickler erkannt und das Netzwerk beim letzten großen Update um viele Neuerungen erweitert, die die ersten Schritte erleichtern sollen. Das hat die erstmalige Anmeldung deutlich vereinfacht; Vorschläge und Tipps sollen Neulingen unter die Arme greifen.

Alle Nutzer können nun von der erweiterten Suche profitieren – die aber anders als in anderen Netzen nicht einfach für alle Inhalte aktiv ist. Stattdessen müssen Teilnehmer der Indizierung zustimmen, wenn eigene öffentliche Inhalte in den Suchergebnissen anderer auftauchen sollen.

Bluesky

Der 2019 von Twitter-Gründer Jack Dorsey angestoßene Dienst mit dem blauen Himmel scheint derzeit ein Momentum zu haben, insbesondere hierzulande. Global meldete der Betreiber im September eine Million Nutzer. Das ist immer noch eine Größenordnung weniger als Threads und das Fediverse. Aber die Zahlen steigen schnell und kontinuierlich an. Daran ändert auch die Zugangsbeschränkung nichts: Man benötigt einen Invite-Code, um hineinzukommen. Wer einmal drin ist, erhält etwa alle zehn Tage einen solchen Code, mit dem er jemanden einladen kann.

Viele ehemals auf Twitter und Mastodon gut vernetzte Personen tummeln sich bereits bei Bluesky, manche Anwender haben die Plattform schon als neue Social-Media-Heimat auserkoren. Auch erste Medien wie der Spiegel und Politiker wie die SPD-Chefin Saskia Esken und der

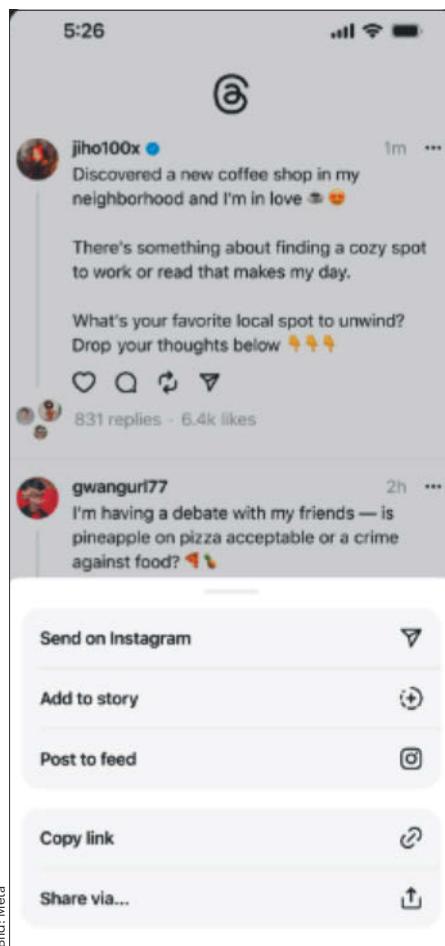


Bild: Meta

Threads ist eng mit Instagram verschränkt.

Grüne Konstantin von Notz sind bei Bluesky.

Die Beliebtheit des Dienstes hat offenbar damit zu tun, dass sich Twitter-Nutzer in Bluesky sofort zurechtfinden, aber auch damit, dass sie sich hier willkommener fühlen als im Fediverse. c't-Autor Volker Weber hatte im September seine Bluesky-Follower gefragt, warum sie den Dienst anderen Netzwerken vorziehen. Unter den Antworten hat die von Johann van de Bron viel Zustimmung erhalten: „Auf Mastodon waren wir Eindringlinge. Hier sind wir selbst die Ureinwohner und gehen mit unseren Twitter-Gewohnheiten niemandem auf den Sack.“

Viele Schnittstellen des Dienstes liegen ebenso offen wie die Nutzerdaten, also Postings, Follows, geblockte und stummgeschaltete Accounts und mehr. Ein kleines Ökosystem externer Anwendungen greift darauf zu. Dazu zählt zum Beispiel der RSS-Reader Flipboard, der bereits als Fediverse-Client dienen kann. Mit der Mehrspaltenoberfläche von deck. blue werden sich viele Nutzer anfreunden,

die früher Tweetdeck für Twitter genutzt haben.

Bluesky lässt dem Nutzer die Wahl des Algorithmus für seine Timeline. Unter „Discover new feeds“ stehen schon reichlich Feeds aus der Community zur Wahl, auch zu spezielleren Themen wie „Gardening“ oder „Cat pics“. Mit dem Web-Client SkyFeed klickt jedermann sehr einfach in einem grafischen Editor einen individuellen Feed zusammen, der sich anschließend auch für die Allgemeinheit freigeben lässt.

Eine zentrale Funktion fehlt allerdings bisher: Bluesky soll eines Tages wie das Fediverse auch als verteiltes Netzwerk fungieren. Bislang kann man sich nur bei dem Server bsky.social des Betreibers anmelden.

Artifact

Immer wieder taucht der Name Artifact unter den Kandidaten für einen Twitter-Nachfolger auf. Das mag damit zu tun haben, dass Kevin Systrom und Mike Krieger den Dienst gestartet haben, die Gründer von Instagram. Allerdings fungierte das nur als App für Android und iOS verfügbare Artifact zunächst eher als News- Aggregator, wenn auch als ein ausgefeilter.

Nutzer können sich aus einer vorgegebenen Auswahl an Medien zu bestimmten Themen ein Nachrichtenpaket schnüren lassen. Medien, bei denen man ein

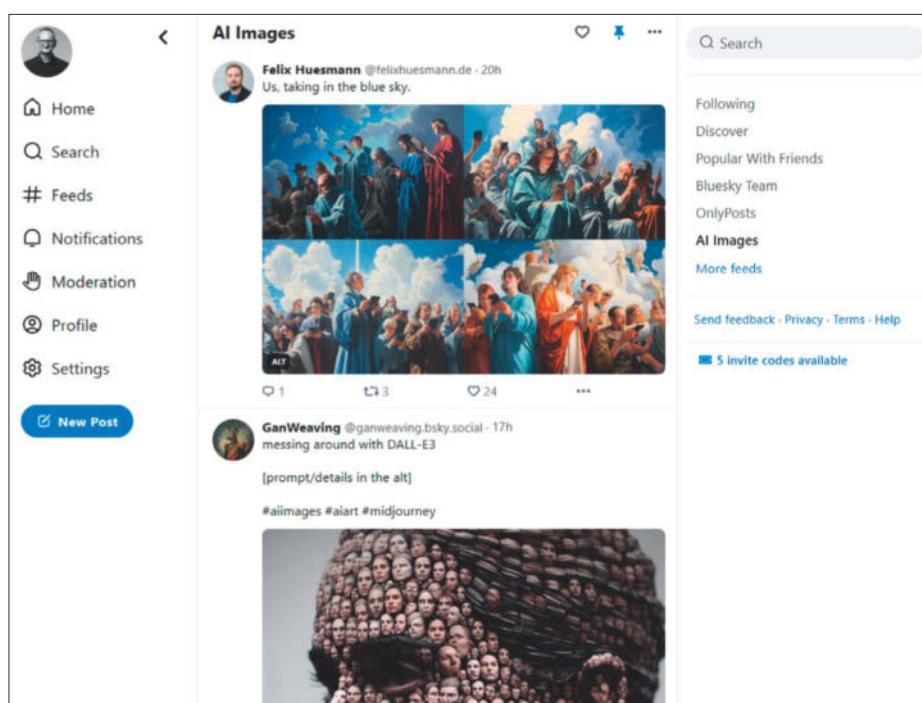
Abonnement unterhält, bevorzugt Artifact bei der Auswahl. Zudem lernt der Dienst aus der Nutzung und schneidet das Nachrichtenangebot auf die Interessen des Anwenders zu. Die App kann Meldungen auch per KI zusammenfassen.

Erst nach und nach ziehen soziale Funktionen in Artifact ein. Zunächst gab es die Kommentarfunktion. Seit Ende September können Nutzer eigene Beiträge mit Titel, Text und Fotos veröffentlichen. Die Beiträge werden wie Nachrichten anderer Medien behandelt. Sie erhalten eindeutige URLs, mit denen man sie auf verschiedenen Apps und Diensten teilen kann.

Der ganze Rest

Neben Threads, Mastodon/Fediverse, Bluesky und Artifact werden etliche weitere Dienste als X-Nachfolger oder -Konkurrenten gehandelt, die aber nicht eine so große Aufmerksamkeit erfahren wie die genannten. Eine Auflistung im Schnell-durchlauf: Post.news zielt speziell auf Journalisten. Ein Micropayment-System soll bewirken, dass guter Content honoriert wird. In Zukunft soll man damit auch Artikel hinter Paywalls bezahlen können.

Pebble.is, früher T2 (Twitter 2), ist ein sehr einfacher Twitter-Klon mit 280-Zeichen-Posts. Hive orientiert sich sowohl an Twitter als auch an Instagram. Bei Hive greifen keine Algorithmen in die Reihenfolge des Contents ein. Immerhin andert-



Bei Bluesky können Nutzer eigene Timelines definieren, die zum Beispiel nur Posts mit KI-Bildern anzeigen.

halb Millionen Nutzer zählt das Netzwerk. Nostr setzt wie das Fediverse mit Mastodon und Bluesky auf Föderation, spielt aber mit ein paar zehntausend Nutzern keine große Rolle.

Die Nachmacher

Von den Umwälzungen bei Twitter wollen aber nicht nur Twitter-ähnliche Dienste profitieren, auch andere soziale Netzwerke fühlen sich offenbar inspiriert: So führte die Kurzvideoplattform TikTok im Juli die Funktion „Textbeiträge“ ein, mit der Nutzer sich schriftlich mitteilen statt sich vor der Kamera auszutoben. Sie können dabei wie gewohnt Musik hinterlegen und die Texte sogar im Duett erstellen. Dies dürfte aber eher introvertierte TikTok-Nutzer ansprechen als flüchtende Xer.

Sich voneinander inspirieren zu lassen, ist anscheinend ein Trend in der Welt der Social-Media-Dienste. Nach Instagram hat im Sommer auch WhatsApp die Funktion „Kanäle“ eingeführt, die stark an Telegramkanäle erinnert. Auch hier kann eine Einzelperson oder Organisation Nachrichten an eine Follower-Gruppe verteilen, die nur mitliest.

Viele WhatsApp-Nutzer sind aber nach wie vor genervt von den Kanalvorschlägen, die WhatsApp ungefragt bei den Statusmeldungen einblendet und die man nicht entfernen kann. Entsprechend vervollständigt Google die Suchanfrage „whatsapp kanäle“ immer noch zuerst mit Begriffen wie ausblenden, entfernen, löschen, deaktivieren oder ausschalten.

Immerhin nimmt Meta den Datenschutz offenbar ernster als Telegram: Kanalbetreiber sehen weder die Profilfotos noch Telefonnummern von Mitgliedern und auch sonst kann niemand sehen, wer wem folgt. Telegram glänzt dafür bei den Stories, die es im Sommer neu eingeführt hat: Bei jeder neuen Story können Nutzer hier entscheiden, ob Screenshots erlaubt sind und wer genau die Story (nicht) sehen darf. Letzteres kann man sogar später wieder ändern.

YouTube-Werbung erfolgreicher als gedacht?

Nicht immer führt das Klonen fremder Funktionen zum Erfolg. YouTube sah sich durch den großen Erfolg von TikTok bedroht – und startete mit YouTube Shorts kurzerhand eine eigene Rubrik für Kurzvideos. Im September berichtete die Financial Times, dass hochrangige YouTube-Mitarbeiter Shorts als Gefahr für

YouTubes Geschäftsmodell sehen: Wenn mehr Menschen die kurzen Filmhappen konsumieren, sehen weniger die Werbung vor, während und nach den längeren Videos, die ja eigentlich YouTubes Spezialität und Garant für Werbeeinnahmen sind.

Zumindest für US-Teenager ist YouTube offenbar nach wie vor das beliebteste soziale Netzwerk, gefolgt von TikTok. Das zeigt eine Umfrage von Precise TV und Giraffe Insights im Frühjahr 2023 in den USA, befragt wurden 1000 Teenager zwischen 13 und 17 Jahren und ihre Eltern. Gefragt nach ihrer Lieblings-Social-Media-App antwortete die Mehrheit der Jugendlichen (27 Prozent), das sei YouTube. Platz zwei belegte TikTok (21 Prozent), Platz drei ging überraschend an Facebook (20 Prozent). Instagram folgte mit 18 Prozent auf dem Fuß, Snapchat lag mit 6 Prozent weit dahinter.

Auch um die Werbung bei YouTube steht es laut der Analyse nicht so schlecht wie hochrangige Mitarbeiter fürchten: 45 Prozent der Jugendlichen erklärten, dass sie sich an Anzeigen auf YouTube erinnern. Damit führt YouTube die Liste haus hoch an, Facebook erreicht 28, TikTok und Instagram je 23 Prozent. Ein Viertel der Befragten fand die YouTube-Anzeigen außerdem besonders ansprechend, für TikTok, Facebook und das Fernsehen empfand das weniger als jeder Fünfte so. Sechs von zehn Jugendlichen gaben sogar

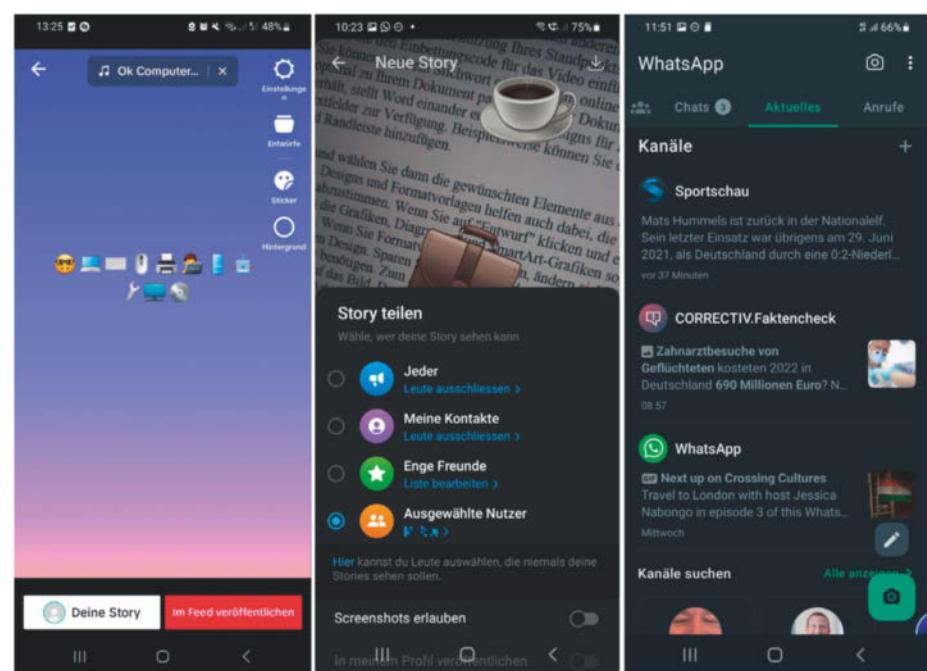
an, YouTube-Anzeigen auch mal aktiv anzusehen, statt sie zu überspringen.

BeReal gefangen im Mittelmaß

Vor allem bei der Generation Z, also den zwischen 1995 und 2010 Geborenen, ist auch die Foto-App BeReal nach wie vor beliebt. Statista fragte zwischen April 2022 und März 2023 über 10.000 Menschen in Deutschland im Alter von 18 bis 58 Jahren, aufgeteilt in drei Generationen: „Welche Social-Media-Plattformen nutzen Sie regelmäßig?“ Daraufhin nannten 18 Prozent der Gen Z BeReal, bei der Gen Y waren es neun und bei der Gen X nur ein Prozent. Doch auch bei der Gen Z sind neun soziale Netzwerke beliebter als BeReal, Instagram führt die Liste mit 79 Prozent an.

Zwar kopieren große Player wie TikTok (mit Now) und Instagram (mit Candid) das Konzept von BeReal, das auf spontanen, ungestellten Fotos beruht. Doch das Original steckt offenbar in der Mittelmäßigkeit fest. Immerhin konnte laut einer weiteren Erhebung von Statista im Juni 2023 knapp ein Viertel der Menschen in Deutschland zwischen 18 und 64 Jahren BeReal. Doch mit diesen 24 Prozent landet die App weit hinten auf Platz 15 von 18 – immerhin überholte sie aus dem Stand Jodel und Clubhouse. Befragt wurden gut 1200 Menschen.

Vielleicht bleibt BeReal eher klein, weil es zurückhaltender ist als die Konkur-



Fast ein Ringtausch: Die großen Social-Media-Dienste lassen sich gern voneinander inspirieren (von links: TikTok-Textbeiträge, Telegram-Stories, WhatsApp-Kanäle).



Kommentar: Niemand braucht die eine einzige Twitter-Alternative

Von Eva-Maria Weiß

Bei Facebook waren sie alle, erst die Studierenden, dann die Eltern, die Kids, die Unternehmen, Medien und Politiker. Ein solches soziales Netzwerk wird es nicht mehr geben. Die Suche nach der einen einzigen Twitter-Alternative ist deshalb überflüssig. Der Nachwuchs ist schon längst breit gestreut unterwegs. Ganz nach dem Motto: Jedem Tierchen sein Pläsierchen.

Twitter ist nie Social Media für alle gewesen. Aber es war wichtig für Unternehmen, die ihre News und Blogbeiträge dort posteten, und Journalisten, die alles schön im Blick be-

halten konnten. Ersatz könnten Newsletter sein. Twitter war auch lustig für kollektives Erleben, wenn im Fernsehen die Bachelorette den Querschläger spannend statt idiotisch fand. Sobald Bluesky Hashtags kann, könnte es die Plattform der Wahl dafür werden. Denn genau diese Bubble scheint dorthin umzuziehen.

Instagram und TikTok haben schon vor langer Zeit für eine Differenzierung gesorgt. Es ist kein Geheimnis, dass Jugendliche und junge Erwachsene nicht auf der Plattform ihrer Eltern unterwegs sein wollen. „Für TikTok bin ich zu alt, das versteh

ich nicht“, sagt man ab Mitte 30 und lacht.

Neue Netzwerke kommen hinzu, Clubhouse und WT.Social sind gescheitert, Donald Trump hatte ein Truth Social gestartet, er postet übrigens wieder Beiträge bei X. Schließlich haben sich die Messenger in die Social-Media-Welt gemischt. Während nämlich bei Instagram und anderen Content vor allem von Creatorn und Influencern kommt, posten die anderen ihre ungeschönten Fotos aus dem Alltag in den Statusmeldungen von WhatsApp und Signal. Jedem Bildchen sein Plätzchen, in verschiedenen Apps.

renz. Die App bietet wenige Beschäftigungsmöglichkeiten: Einmal am Tag fordert sie ihre Nutzer auf, ein Foto zu machen, und dann können diese die Bilder ihrer Freunde und anderer Nutzer sehen und darauf reagieren, mehr nicht. Im Gegensatz zu Konkurrenzmedien kann man bei BeReal weder viral gehen noch versucht die App, ihre Nutzer möglichst lange an sich zu fesseln [3].

Ausblick

Ein soziales Netzwerk ist aus Sicht der Nutzer ein Stück Heimat im Netz, ein lange gewachseses Gebilde, in dem man regelmäßig seine Beziehungen pflegt. Nach einer Studie von Faktenkontor verbringen Deutsche pro Woche im Schnitt knapp 21 Stunden in sozialen Netzwerken.

Seine Onlineheimat gibt man nicht so schnell auf. Das haben auch die Streikführer bei Reddit gemerkt. Bei dem Netzwerk haben die ehrenamtlichen Administratoren viele Subreddits gesperrt, also Untermen. Damit haben sie gegen ein neues Geschäftsmodell protestiert, das ihnen in letzter Konsequenz die Arbeit erschwert hätte. Der Protest ist aber letztlich verpufft, weil die Nutzer auf andere Subreddits ausgewichen sind.

Ist der Leidensdruck groß genug, sind Anwender allerdings sogar mehrfach bereit, sich nach Alternativen umzuschauen: Zuerst weg von Elon Musks X hin zu Mastodon, dann weiter zu Bluesky, weil sie mit dem verteilten Netzwerk fremdelten. Wobei sich viele der potenziell Wechselwilligen schwertun, ihre Zelte auf den alten Plattformen komplett abzureißen, und ihre Accounts dort erst einmal weiterlaufen lassen.

Rund um Bluesky gibt es viele Fragen, etwa die nach dem Geschäftsmodell. Der Blogger André Vatter weist zurecht darauf hin, dass der Niedergang von Twitter viel mit ungenügender Moderation zu tun hat. Und was ist bei Bluesky zu erwarten? Er fragt: „Wer betreibt das Netzwerk und wie wird es reguliert? Wem ist Bluesky PBLLC gegenüber rechenschaftspflichtig? Warum haben weder die Hauptseite des Netzwerks (für unangemeldete Nutzer), noch die Seite der gemeinnützigen Gesellschaft oder die Seite („© 2023 Bluesky, PBC. All rights reserved.“) des AT Protocols eine Datenschutzerklärung? Es gibt Hinweise zur Privacy Policy für das Netzwerk, in der aber das Akronym „GDPR“ nicht einmal vorkommt – es hapert ein wenig an Respekt für europäisches Datenschutzrecht,

oder?“ Der Hype rund um Bluesky kann auch ganz schnell wieder verpuffen und die Anwenderschaft zur nächsten Onlineheimat weiterziehen.

Viele soziale Netzwerke gleichen sich ohnehin einander an. Wechselseitig kopieren sie Funktionen der Konkurrenz: Mal wünschen die Nutzer sich das explizit, mal meinen die Anbieter, dass die Funktion ihrem Netzwerk gut steht. Indem sie Neulinge wie BeReal nachahmen, umklammern die Etablierten ihre Nutzer und halten die Neuen klein. Wie im Fall von YouTube müssen sie dabei aber aufpassen, ihren Markenkern nicht zu verraten oder unter Imitationen zu verbuddeln.

(jo@ct.de) **ct**

Literatur

- [1] Jo Bager, Greta Friedrich: Die Macht der Netzwerke, Mehr Vielfalt, schnell wachsende Neulinge: das Social-Media-Jahr 2022, c't 2022 Jahresrückblick, S. 164
- [2] Volker Weber: Die Möchtegern-Erben, Ein Blick auf diverse potenzielle Twitter-Nachfolger, c't 13/2023, S. 124
- [3] Kathrin Stoll: Social Media ungeschönt, c't 5/2023, S. 86

Links zu den zitierten Statistiken und c't-Artikeln finden Sie hier: ct.de/yp8m



IT-Kompetenz, die bleibt: Tech-Talente finden und fördern durch Weiterbildung

Die Herausforderungen für Entscheider sind groß und vielfältig: Sie bestehen darin, geeignetes Personal zu gewinnen, zu fördern und zu halten. Mitarbeiter erwarten, dass Unternehmen kontinuierlich in sie investieren. Mit passender Qualifizierung werden Wissenslücken geschlossen – eine Win-Win-Situation für alle Seiten.

IT-Weiterbildung ist DER Schlüssel zum Erfolg.

Die heise academy bietet Lerninhalte zu allen aktuellen und relevanten IT-Themen, die jetzt und in Zukunft wichtig sind. Wir sind IHR Partner für professionelle IT-Weiterbildung. Bauen Sie als Unternehmen mit passender Qualifizierung Zukunftskompetenzen auf.

Die richtige Zeit für IT-Weiterbildung ist genau jetzt.



**Jetzt downloaden:
Key Facts zur IT-Weiterbildung**
heise-academy.de/whitepaper-it-weiterbildung





Bild: Mohssen Assanimooghaddany/dpa

Apps statt Ärzte

Wie Lauterbach das Gesundheitswesen digital heilen will

Das deutsche Gesundheitssystem krankt seit Jahren: Es fehlen Medikamente für Kinder, Pflegekräfte für Senioren, Fachärzte und Allgemeinmediziner in der Stadt und vor allem auf dem Land. Gesundheitsminister Karl Lauterbach sieht die Ursachen in mangelnder Effizienz und rückständiger Digitalisierung. Letzterer will er einen kräftigen Schub geben: mit Sammlung und Freigabe von Forschungsdaten, Ausbau von Medizin-Apps und Gesundheitskiosken für Arme.

Von Detlef Borchers

Bundesgesundheitsminister Karl Lauterbach hat zur großen Aufholjagd bei der Digitalisierung geblasen. Ab 2024 soll das elektronische Rezept bundesweit gelten. Ab 2025 sollen die Krankenkassen die elektronische Patientenakte (ePA) „für alle“ herausgeben. Sie sammelt die Gesundheits- und Behandlungsdaten der Patienten und teilt sie mit Krankenkassen, Forschungszentren und Pharmaunternehmen. Mithilfe künstlicher Intelligenz, die die Daten auswertet, sollen die Kassen zum Beispiel Krankheiten frühzeitig erkennen und Versicherte gegebenenfalls zum Arzt schicken.

Um sein Spar- und Umverteilungsprogramm durchzusetzen, ringt Lauterbach mit den Ländern um eine Krankenhausreform. Einzelne Krankenhausstandorte sollen verkleinert oder ganz geschlossen werden. Wo in ärmeren Regionen und Stadtteilen Hausärzte und Krankenhäuser

fehlen, sollen Gesundheitskioske mit medizinisch geschulten Fachkräften die Notversorgung übernehmen. Ausgebildete Ärzte bekommen die Patienten dort allenfalls später per Videokonferenz zu sehen.

Damit der Umbau und die weitere Digitalisierung funktionieren, muss aber zunächst das Grundgerüst der Digitalisierung, die Telematikinfrastruktur (TI), auf Version 2.0 erneuert werden. Dazu wandelt Lauterbach die für die Telematik zuständige Projektgesellschaft Gematik in eine Digitalagentur um. Der bisherige Gematik-Chef Markus Leyck Dieken wurde bereits abgelöst; er verlässt das Unternehmen zum Jahresende.

Null Vertrauen

Der wichtigste Baustein der laufenden Digitalisierung des Gesundheitswesens trägt den Namen Zero Trust. Dahinter verbirgt

sich ein Sicherheitsprinzip, das jede Handlung unter den Vorbehalt stellt, dass alle Akteure, Daten und Institutionen sicher identifiziert sind. Das Konzeptpapier für eine Zero-Trust-Architektur im Gesundheitswesen der Zukunft zielt darauf ab, die bisherige TI-Struktur komplett neu zu gestalten. Die Rolle der bisherigen Hardware übernimmt künftig Software: Die problembehafteten VPN-Konnektoren entfallen.

Später sollen die Heilberufsausweise der Ärztinnen und Ärzte sowie die elektronischen Praxisausweise (SMC-B-Karten) der Konnektoren durch ID-Systeme ersetzt werden, „um unabhängig von der dafür notwendigen Spezialhardware wie Kartenterminal und Konnektor auf Anwendungen der TI und auf weitere Anwendungen mit Gesundheitsbezug zugreifen zu können“, beschreibt die Gematik die neue mobile Freiheit.

Auch beim Zugriff über das offene Internet und über private mobile Endgeräte soll die TI 2.0 den wichtigen Schutz medizinischer Daten gewährleisten. Das Konzept dafür hat die Bundesdruckerei zusammen mit ihren Töchtern Genua GmbH und D-Trust sowie der Compugroup Medical und dem Fraunhofer-Institut für Angewandte und Integrierte Sicherheit (AISEC) entwickelt. Bezuglich der Leistungsfähigkeit der TI 2.0 haben sich die Beteiligten hohe Ziele gesetzt: „Konkret muss die Lösung für mehr als 80 Millionen Versicherte und über 200.000 Leistungserbringer praktikabel, das heißt ohne störende Antwort-, Lauf- oder Reaktionszeiten nutzbar sein.“ Gegenüber der heutigen TI mit ihren ständigen Störungen wäre das ein echter Fortschritt.

Tschüss, Gesundheitskarte

Auf Seiten der Versicherten wird die Gesundheitskarte durch einen elektronischen Ausweis (eID) ersetzt, der bei einem sektoralen Identity Provider (IDP) wie zum Beispiel Verimi/T-Systems registriert ist. Patienten können ihn zum Beispiel in einer App auf dem Smartphone präsentieren. Spätestens ab dem 1. Januar 2024 müssen die Krankenkassen nach dem Gesetz zur Modernisierung der Pflege und zur Digitalisierung (DVPMG) auf Wunsch der Versicherten einen solchen Ausweis ausstellen. Ab dem 1. Januar 2026 muss die eID alternativ zur Gesundheitskarte in allen Zusammenhängen als Versichertennachweis akzeptiert werden.

Als erste Krankenkasse hat die Barmer mit ihrer ID-Wallet-App die Zulas-

sung der Gematik erhalten. Die App nutzt den IDP-Anbieter Verimi, der die Daten in einer deutschen Cloud von T-Systems speichert. Um in der Arztpraxis nachzuweisen, dass er bei der Barmer versichert ist, muss ein Patient in der App seine PIN eingeben und die Eingabe mit einem biometrischen Faktor (Fingerabdruck oder Face-ID) bestätigen. Das Einsticken der eGK in ein Lesegerät entfällt.

Was die Authentifizierung erleichtert, verkompliziert allerdings andere neue Lösungen wie die Nutzung der eGK für das elektronische Rezept. Einfach mit der eGK eines Familienmitglieds ein eRezept in der Apotheke einlösen? Das geht nicht. Für dieses Szenario soll eine neue elektronische Stellvertreterfunktion geschaffen werden.

Datenschutzrisiko Verimi

Die eID soll nicht auf das Gesundheitswesen beschränkt bleiben. 26 deutsche Unternehmen (darunter T-Systems) sind an Verimi beteiligt. Sie alle wollen von dessen ID-Wallet profitieren. Denn eine solche elektronische Brieftasche verwahrt nicht nur die Versichertenkarte der Patienten und den Heilberufsausweis des Arztes. Verimi will in dem ID-Wallet staatliche, medizinische und kommerzielle Ausweise kombinieren – vom Führerschein über den elektronischen Personalausweis bis hin zur Bankkundenkarte.

In der Vergangenheit fiel Verimi bereits durch mehrere gravierende Pannen auf: 2019 drohte Verimi ein Lizenzentzug der Bafin, nachdem Mitarbeiter berichteten, Transaktionszahlen des Bezahlsystems „Verimi Pay“ künstlich in die Höhe getrieben zu haben. Im September 2021 teilte die Verimi GmbH der Berliner Datenschutzbeauftragten mit, dass sie aufgrund eines Programmierfehlers seit April 2019 eine unbekannte Anzahl von bis zu



Bild: Kay Nietfeld/dpa

Bundesgesundheitsminister Karl Lauterbach treibt die von seinen Vorgängern verbockte Digitalisierung des Gesundheitssystems voran.

c't kompakt

- Gesundheitskarten, Kartenleser, Konnektoren: Die ganze teure Hardware soll bald Elektroschrott werden. An ihre Stelle treten Software und elektronische Kennungen.
- Alle Gesundheitsdaten möglichst aller Patienten sollen in elektronischen Akten gesammelt und hundert Jahre lang für Forscher, Versicherungen und Pharmaunternehmen zugänglich sein.
- Auf dem Land und in ärmeren Stadtteilen sollen Gesundheitskioske entstehen, in denen man Ärzte allenfalls am Bildschirm sieht.

450.000 Nutzernamen und Passwörtern im Klartext in Logfiles gespeichert habe. Zuletzt hebelte der Sicherheitsexperte Martin Tschirsich im Jahr 2022 das Verimi-Verfahren der Foto-ID mit einfachsten Methoden aus.

Diese Pannenhäufung macht das Verimi ID-Wallet zu keinem vertrauenswürdigen Ort, um dort wichtige Dokumente abzulegen. Bei einer weiteren Datenschutzpanne könnten diese Dokumente in die falschen Hände geraten.

Strafen für Ärzte und Hersteller

Aber zurück zum Umbau der TI: Neben dem Konzept hat das federführende Konsortium einen Stufenplan für den Aufbau der Zero-Trust-Architektur entwickelt. Ein Demonstrator soll die Leistungsfähigkeit dieser TI 2.0 darstellen und beispielsweise zeigen, wie Ärzte und Versicherte sich mit ihrer ID anmelden. Unklar bleibt allerdings, wie der Übergang vom heutigen System gelingen soll. Lapidar heißt es: „Die Migration von der bestehenden TI1.0 zur vorgeschlagenen Architektur muss in einem folgenden Konzept betrachtet werden.“

Viele der geplanten Neuerungen sind noch nicht im Detail beschrieben. Klar ist, dass es auch in der TI 2.0 das eRezept und die elektronische Patientenakte geben wird. Alles Weitere, noch nicht konkretisierte firmiert im Konzeptpapier unter der Abkürzung WANDA, die für weitere Anwendungen für den Datenaustausch in der Telematikinfrastruktur steht.

Den Anfang macht das E-Rezept, das ab dem 1. Januar 2024 verbindlich das



Bild: Sebastian Gollnow/dpa

Länder wie Australien sind bei der Digitalisierung des Gesundheitssystems schon viel weiter: Dort erbeuteten Erpresser 2022 angeblich gut geschützte Daten von 9,7 Millionen Versicherten der Medibank und versilberten sie im Darknet.

Papierrezept ablösen soll. Der Termin zum Jahresbeginn dürfte allerdings kaum zu halten sein. Denn zuvor müssen sich Bundestag und Bundesrat mit dem Gesetz befassen. Ab April 2024 könnten die Krankenkassen Ärzte mit einer Kürzung ihrer Vergütung um ein Prozent bestrafen, wenn sie weiterhin das alte Muster-16-Rezept ausstellen und die Umstellung auf das E-Rezept verzögern. Die Sanktionen greifen allerdings nur, wenn die Technik bis dahin reibungslos funktioniert. „Sollte es mit einzelnen IT-Anbietern Probleme geben, dann werden wir diese Hersteller einbestellen. Denn es kann nicht sein, dass das System nicht funktioniert, weil einzelne Anbieter hier einfach die Modernisierung nicht schaffen“, erklärte Lauterbach in einer Berliner Arztpraxis.

Man darf gespannt sein, wie das Gesundheitsministerium die einzelnen Softwarehäuser in die Pflicht nimmt, wenn es bei der Technik für das E-Rezept am Zusammenspiel der verschiedenen Komponenten scheitert. Schließlich lobt der Minister gerne die Anbietervielfalt im Gesundheitswesen. Eine mögliche Sanktionierung der Hersteller wäre ein Novum, könnte aber beim Übergang zur TI 2.0 eine wichtige Rolle spielen.

Vorsorge-Checks der Kassen

Ein genauerer Blick lohnt sich auch bei der elektronischen Patientenakte, die laut Gesetzentwurf „für alle“ angelegt werden

soll. Tatsächlich richten die Krankenkassen ihren Versicherten ab dem 15. Januar 2025 automatisch eine ePA ein und befüllen diese mit den ihnen vorliegenden Daten über Diagnosen und Krankenhausaufenthalte, sofern sie von den Versicherten keinen Widerspruch erhalten. Dies gilt sowohl für die gesetzlichen Krankenkassen als auch für die privaten Krankenversicherungen, bei denen rund zehn Prozent der Bevölkerung versichert sind.

Darüber hinaus sollen die Krankenkassen die Möglichkeit erhalten, ihre Versicherten auf Gesundheitsrisiken hinzuweisen, die sie aus den vorhandenen Daten ableiten. Nach dem Willen des Gesundheitsministeriums sollen sie Versicherten anhand solcher Analysen auch empfehlen, einen Arzt einer bestimmten Fachrichtung aufzusuchen. Dies soll ohne ärztliche Diagnose geschehen.

Der Entwurf zum Gesundheitsdatennutzungsgesetz (GDNG) beschreibt diese Datenauswertung durch die Kassen als „datengestützte Auswertungen zum individuellen Gesundheitsschutz ihrer Versicherten“. Die Datenanalyse soll auch mit Hilfe künstlicher Intelligenz erfolgen. Als Beispiele nennt der aktuelle Gesetzentwurf die Früherkennung seltener Erkrankungen und Krebsrisiken sowie die Prüfung der Arzneimitteltherapiesicherheit (AMTS).

Die niedergelassenen Ärzte sehen in den Plänen eine Bevormundung. So for-

dert der Mediziner Stefan Streit, ein Kritiker des digitalisierten Gesundheitswesens, genau das Gegenteil: ein im Grundgesetz verankertes Diskriminierungsverbot für Krankheiten und Krankheitsdaten. „Die Konfliktlinie der Digitalisierung liegt zwischen Krankenkassen und Bundesgesundheitsministerium auf der einen und Patienten und Ärzten auf der anderen Seite, auch wenn manche die lieber zwischen Ärzten und Patienten sähen“, schreibt er in einer seiner Streitschriften.

Datensammlung

Hat ein Versicherter in der Arztpraxis oder in der ePA-App dem Zugriff auf seine Patientenakte zugestimmt, sind Ärzte verpflichtet, die in ihren Praxisverwaltungssystemen vorhandenen Daten wie Medizinische Informationsobjekte (MIO) – dazu gehören Mutterpass oder Impfpass-, Medikationspläne, Laborbefunde sowie die Patientenkurzakte in die ePA zu kopieren. Damit ein vollständiger Medikationsplan vorliegt, müssen Ärzte nach dem Gesetz zur Beschleunigung der Digitalisierung des Gesundheitswesens (DigIG) auch alle Daten von E-Rezepten in die ePA kopieren.

Außerdem müssen Ärzte ihre Patienten explizit auf „Risiken und Nebenwirkungen“ der Daten hinweisen. Dies gilt für Informationen über HIV-Infektionen, psychische Erkrankungen, Schwangerschaftsabbrüche und Ergebnisse genetischer Untersuchungen. Denn die Versicherten sollen künftig möglichst viele Daten für Forschungszwecke freigeben. Immerhin können sie einzelne Angaben auch temporär sperren, wenn sie sich in anderen Zusammenhängen womöglich nachteilig für die Versicherten und ihre Nachkommen auswirken. Denn die Identität der Patienten wird lediglich durch eine Pseudonymisierung verschleiert, bei der etwa Krankenversicherungsnummer, Name und Adresse und das Geburtsdatum durch ein Kohortendatum ersetzt werden. Die genaue Methode wird derzeit noch diskutiert. Eine vollständige Anonymisierung, bei der sich die Person nicht mehr oder nur unter erheblichem Aufwand zuordnen lässt, findet jedoch nicht statt.

Über den Tod hinaus

Die Krankheitsdatenforschung, die nach Karl Lauterbachs Vorstellung durch künstliche Intelligenz unterstützt werden könnte, dürfte zum größten Umbau des deutschen Gesundheitswesens führen. Aus

dem solidarischen Gesundheitssystem wird ein System, das in der Lage ist, bestimmte Gruppen und Krankheiten zu diskriminieren. Der Jurist Dirk Heckmann, der an der TU München die Professur für Recht und Sicherheit der Digitalisierung innehat, bezeichnet dies weniger dramatisch als „dynamisch lernendes Gesundheitswesen“.

Aus Schweizer Perspektive (wo ebenfalls an einer ePA für alle gearbeitet wird) argumentiert Felix Gutzweiler von der Stiftung Sanitas Krankenversicherung in einem Kommentar für den Tagesanzeiger so: „Solidarität im digitalen Gesundheitswesen umfasst mehrere Dimensionen, darunter finanzielle, Verhaltens-, Datensolidarität und Monitoring-Solidarität. Während die finanzielle Solidarität allgemein bekannt ist, sind die Konzepte der Datensolidarität und der Monitoring-Solidarität weniger vertraut. Dass Datenmessung und -teilung ebenfalls Solidaritätsaspekte beinhalten, ist noch wenig kommuniziert.“ Und weiter: „Es ist von entscheidender Bedeutung, das Bewusstsein dafür zu schärfen, dass das Teilen von Gesundheitsdaten ein Akt der Solidarität ist und sowohl individuell als auch gesamtgesellschaftlich von Nutzen sein kann.“

Zu diesem Solidaritätsverständnis passt, dass das GDNC hierzulande die Aufbewahrungsfrist für Forschungsdaten von 30 auf 100 Jahre verlängert, um eine generationenübergreifende Forschung zu ermöglichen. Das bedeutet, dass auch die Enkel und Urenkel noch Auswirkungen zu spüren bekommen, wenn Oma und Opa der Freigabe ihrer Gesundheitsdaten nicht widersprechen. Das könnte sich positiv auf die Erforschung von Erbkrankheiten auswirken – aber auch negativ auf Vertragsbedingungen und Krankenkassenbeiträge.

Kioske im Ghetto

Auch die von Lauterbachs Ministerium geplanten Gesundheitskioske zeigen den Trend zu einer Medizin, die bestimmte Gruppen diskriminiert. Sie sollen in „sozial benachteiligten Regionen und Stadtteilen“ ein niedrigschwelliges Angebot für Versicherte sein, sich untersuchen und beraten zu lassen. Ausdrücklich ist davon die Rede, dass dort angestellte Pflegekräfte mit „Heilkundekompetenz“ arbeiten sollen und Verbandswechsel, Wundversorgung und subkutane Injektionen übernehmen. Lauterbach selbst bezeichnete dieses Angebot in einem Interview als „Community Health Nursing“.



Bild: Sebastian Gollnow/dpa

Ärzte greifen künftig mehr zu Maus und Tastatur als zum Stethoskop, um Patienten an Gesundheitskiosken per Videosprechstunde zu beraten.

Damit würden die geplanten Kioske Aufgaben übernehmen, für die früher Hausärzte oder Pflegedienste zuständig waren. Ärzte sind in den Kiosken nicht vor Ort. Bis 2026 sollen die Kioske aber zu Anlaufstellen für „assistierte Telemedizin“ ausgebaut werden: Per Videocall können Patienten dann einen Arzt hinzuziehen. Auch Apotheken sollen zu solchen Anlaufstellen werden.

Die Einrichtung der Kioske ist zwar Sache der Kommunen, die die als benachteiligt definierten Gebiete festlegen. Die gesetzlichen Krankenkassen sollen jedoch 74,5 Prozent der Kosten übernehmen, die privaten 5,5 Prozent. Bei tausend geplanten Kiosken macht das 375 Millionen Euro pro Jahr aus.

Apps für Risikobereiche

Seit zwei Jahren gibt es digitale Gesundheits-Apps (DiGA), seit einem Jahr digitale Pflege-Apps (DiPA). Sie gelten als Medizinprodukte niedriger Risikoklasse, deren Kosten die Krankenkassen auf Antrag übernehmen. Beispiele dafür sind ein Diabetes-Tagebuch zur Dokumentation der Blutzuckerwerte oder eine App gegen Panikattacken.

Die Einstufung und Bewertung der Apps erfolgt durch das Bundesamt für Arzneimittel und Medizinprodukte (BfArM). In den nächsten Jahren sollen DiGAs und DiPAs bis zur Integration in Disease-Management-Programme (als eDMP) ausgebaut werden und auch hö-

here Risikoklassen umfassen. In Zukunft soll auch das „Internet der medizinischen Dinge“ mit einbezogen werden. Das Spektrum reicht von Wellness- und Fitness-Trackern der Versicherten bis hin zu Geräten mit medizinischer Indikation, die dem Medizinproduktegesetz unterliegen, weil sie beispielsweise in der Teilekardiologie Vitalfunktionen überwachen.

Präzision oder Gängelung?

Aus Sicht des Gesundheitsministeriums ebnet die Digitalisierung den Weg in die personalisierte Medizin. Befürworter sprechen von Präzisionsmedizin. Kritiker hingegen sehen sie als Einstieg in ein umfassendes System, alle medizinischen Handlungen zu kontrollieren und zu überwachen. Einige von ihnen, wie der Arzt Bernd Hontschik in seinem Buch „Heile und Herrsche!“, gehen so weit, einen Weg vom solidarischen Gesundheitswesen über die ökonomisierte Gesundheitswirtschaft hin zu einer Gesundheitsherrschaft zu prophezeien, in der alle Versicherten digital überwacht werden.

Deshalb sei abschließend noch einmal Karl Lauterbach zitiert. „Statt die Patienten und ihre Bedürfnisse in den Mittelpunkt zu stellen, wurden von der Politik in der Vergangenheit technokratische Lösungen umgesetzt, die die Seele der Versorgung auffressen“, erklärte er auf dem Deutschen Ärztetag 2023. Dafür bekam er Beifall. (hag@ct.de) **ct**

Die Entdeckung der Langsamkeit

Die wichtigsten Baustellen bei der Digitalisierung der Verwaltung

Egal ob es um Onlineanträge geht, um die Rentenübersicht oder um digitale Identitäten: Die meisten staatlichen Digitalisierungsprojekte hinken den ursprünglichen Zeitplänen weit hinterher. Auch die Kosten sind beachtlich, der Nutzen bleibt bislang aber arg überschaubar.

Von Christian Wölbert

Nach „Wumms“ und „Doppel-Wumms“ nun also der „Deutschlandpakt“: Mit dem im September vorgestellten Maßnahmenpaket will Bundeskanzler Olaf Scholz (SPD) das Land „schneller, moderner und sicherer“ machen. Geht es nach ihm, sollen Länder und Kommunen sich unter anderem dazu verpflichten, 15 häufig genutzte Verwaltungsverfahren bis Ende 2024 flächendeckend zu digitalisieren. Und zwar „Ende-zu-Ende“, also nicht nur die Anträge, sondern auch die Prozesse in den Behörden. Auf der Liste stehen unter anderem die Wohnsitzummeldung sowie die Anträge auf Personalausweis, Führerschein, Elterngeld und Wohngeld.

Der Grund für den Vorstoß des Kanzlers: Bislang kommtt Deutschland bei der Digitalisierung seiner Verwaltung trotz Milliardenausgaben kaum vom Fleck. Obwohl allein der Bund im Jahr 2020 drei Milliarden Euro für die Entwicklung von Online-Anträgen lockergemacht hatte, sind aktuell nur 136 Verfahren flächendeckend online. Laut Onlinezugangsgesetz (OZG) hätten es bis Ende 2022 schon 575 sein müssen.

Außerdem steht der schwierige Teil der Aufgabe erst noch bevor: Rund 100 der

136 Onlinedienste stammen vom Bund, sind also per se flächendeckend verfügbar. Für den Großteil des Rests sind die Kommunen zuständig. Diese Dienste müssen jeweils in zig Landkreisen und Städten eingeführt werden, was den Gesamtaufwand vervielfacht. Hinzu kommt, dass viele Onlinedienste, wenn sie denn verfügbar sind, bislang kaum genutzt werden.

Stadt für Stadt, Dorf für Dorf

Beobachten lässt sich das zum Beispiel bei der **digitalen Wohnsitzanmeldung**. Da sich bundesweit pro Jahr etwa 5,5 Millionen Menschen ummelden, würde ein flächendeckender, praktischer Onlinedienst sowohl Bürger als auch Behörden massiv entlasten. Im Oktober 2021 stellte der Bund 18,8 Millionen Euro aus dem erwähnten 3-Milliarden-Topf zur Verfügung, damit das Bundesland Hamburg einen solchen Onlinedienst für möglichst alle Bundesländer und Kommunen entwickelt. Am 1. September 2022 ging der Dienst in Hamburg online. Mitte Oktober dieses Jahres, also über ein Jahr später, war allerdings lediglich eine weitere Stadt angebunden, und zwar Lübeck.

Aktuell binde man 52 „Pilotkommunen“ aus 13 Bundesländern an, erklärte die Hamburger Senatskanzlei auf Anfrage von c't. Da es bundesweit etwa 5200 Meldebehörden gibt, bleibt also noch einiges zu tun, bis man sich flächendeckend online ummelden kann. Landauf, landab muss jedes Städtchen den Hamburger Onlinedienst in seine Abläufe einbauen, Mitarbeiter schulen, seine Webseiten umbauen und so weiter.

Onlinequote von 1,5 Prozent

Außerdem ist der föderale Flickenteppich nicht das einzige Problem bei der Digitalisierung. Mal sind die Softwareentwickler

überlastet, mal fehlen technische Standards. So können zum Beispiel den Hamburger Dienst zur Wohnsitzummeldung aktuell nur Alleinstehende nutzen. Das liegt laut Hamburg am Standard XMeld, über den der Onlinedienst mit der Wegzugs- und der Zuzugsmeldebehörde kommuniziert. Dieser Standard unterstützt erst im Release 3.4, das für Mai geplant ist, die Ummeldung von Familien.

Häufig bremsen auch rechtliche Vorgaben. So erinnert die Wohnsitzummeldung an einen mehrtägigen Hürdenlauf: Wenn man seine neue Adresse eingetippt hat, schicken die Behörden erst einmal einen Brief dorthin, um zu überprüfen, ob man wirklich dort wohnt. Hat man einen Code aus dem Brief online eingegeben, darf man zu Hause mit dem Smartphone via NFC die Adressdaten auf seinem elektronischen Personalausweis aktualisieren. Zu guter Letzt kommt ein Adressaufkleber per Post, den man auf den Aufweis klebt. Aber das ist nicht alles: Mieter müssen auch noch eine Wohnungsgeberbescheinigung auf Papier besorgen, einscannen und hochladen.

Da ist es kein Wunder, dass sich laut Senatskanzlei in Hamburg seit September 2022 erst circa 4000 Personen online umgemeldet haben. Das entspricht 1,5 Prozent der insgesamt 260.000 jährlichen An- und Ummeldungen in dem Bundesland. Für deutsche Verhältnisse wirkt diese Quote sogar vergleichsweise gut: Bei der digitalen Kfz-Anmeldung lag der Onlineanteil 2021 bei desaströsen 0,6 Prozent.

Das zweite Mammutprojekt

Mit den föderalistischen Strukturen haben auch die Macher der **Registermodernisierung** zu kämpfen. Dieses Projekt ist aus Sicht vieler Experten noch wichtiger als die Onlineanträge, weil es die Basis für eine bürgerfreundliche Verwaltung schaffen soll. 51 zentrale und dezentrale behördliche Datenbanken („Register“) sollen künftig auch die Steuer-ID speichern, damit die Behörden automatisiert Daten austauschen können, wenn Bürger dem zustimmen. Das Ziel: Künftig soll man nicht mehr bei jedem Antrag erneut Dokumente wie die Geburtsurkunde hochladen müssen.

Aufgrund der zersplitterten Registerlandschaft, zum Beispiel den erwähnten 5200 Meldebehörden, ist das Projekt aber kaum überschaubar. Für den Zeitraum von 2021 bis 2025 schätzen Bund und Länder

die Kosten auf circa zwei Milliarden Euro. Das Potenzial für Verspätungen und Kostensteigerungen ist gewaltig.

Lindner sammelt IBANs

Aber auch wenn der Bund allein zuständig ist, geht es nicht unbedingt schnell. Ende August 2022 verkündete Finanzminister Christian Lindner (FDP), dass die Bundesregierung nicht allen **Bürgern Geld überweisen** kann, weil der Bund die IBAN der meisten Menschen nicht kennt und diese auch noch nicht zusammen mit der Steuer-ID speichert. Das Einsammeln und Verknüpfen dauere „mal eben 18 Monate“, trug Lindner vor.

Inzwischen hat Lindner die rechtlichen Voraussetzungen für das Projekt geschaffen: Das Bundeszentralamt für Steuern (BZSt) darf die IBANs der Bürger speichern, Familienkassen und Banken sollen sie zuliefern. Die Banken müssen die Kontonummer übermitteln, wenn Bürger sie dazu auffordern, zum Beispiel per Onlinebanking. Die Familienkassen müssen die Daten von Kindergeldempfängern automatisch übermitteln. Dem BZSt folge können die ersten Daten „voraussichtlich“ im Dezember 2023 fließen.

Einen Zeitpunkt, zu dem die IBANs aller oder zumindest der meisten Bürger vorliegen sollen, wollte das BZSt auf An-

frage von c't nicht nennen. „Dies hängt von verschiedenen Faktoren ab, insbesondere davon, wann und welche konkreten Geldleistungen vom Gesetzgeber beschlossen werden.“ Finanzminister Lindner zufolge wird der Bund „schon 2025“ technisch in der Lage sein, das im Koalitionsvertrag versprochene Klimageld auszuzahlen. Ob das Geld tatsächlich fließt, ist aber unklar, weil die Ampelparteien sich über Startzeitpunkt und Höhe des Klimagelds nicht einig sind.

Kein Durchblick bei der Rente

Gemächlich voran geht es auch bei der **digitalen Rentenübersicht**, die Bürgern auf einen Blick zeigen soll, mit wie viel Geld sie im Alter rechnen können – vom Staat, aber auch von privaten Vorsorgeanbietern. Versprochen wurde ein solches Portal schon 2019 von der damaligen Großen Koalition. Die erste öffentliche „Pilotphase“ startete allerdings erst im Juni 2023. Und das Einloggen lohnt sich für Bürger bislang eher nicht, denn aktuell liefern nur wenige Altersvorsorgeanbieter freiwillig Daten zu. Die Bundesregierung will alle Anbieter dazu verpflichten, doch ab wann diese Pflicht greift, steht noch nicht fest: An einer Verordnung dafür werde gearbeitet, heißt es bei der Deutschen Rentenversicherung.

Eine weitere Dauerbaustelle des Bundes sind die **digitalen Identitäten**. Das Innenministerium bastelt seit Jahren an der Smart-eID, mit der man Ausweisdaten sicher im Smartphone speichern können soll. Kostenpunkt von 2021 bis Ende 2023: 89,8 Millionen Euro. Für Bürger ist dabei noch nichts herausgekommen. Der Handyausweis startet nach vielen Terminverschiebungen wohl frühestens im Winter und dann voraussichtlich auch nur für ein paar Samsung-Modelle. Weitere 50 Millionen Euro steckt das Wirtschaftsministerium in diverse Forschungsprojekte zu digitalen Identitäten.

500 Millionen für Bildungsplattform

Im Vergleich zur **nationalen Bildungsplattform** wirken aber selbst Smart-eID & Co. günstig. Die Plattform soll keine neuen Lerninhalte schaffen, sondern im Wesentlichen bestehende Onlinebildungsangebote vernetzen und Lernenden ermöglichen, ihre Zeugnisse und Ergebnisse verschlüsselt abzulegen. Trotz dieser überschaubaren Ziele sieht das Bildungsministerium (BMBF) 500 Millionen Euro für die Entwicklung vor. Mit dieser Summe könnte man zum Beispiel die umstrittenen Kürzungen beim Elterngeld ein Jahr lang hinausschieben.

Aus Sicht des Softwareentwicklers und Netzpolitaktivisten Henning Tillmann ist die Bildungsplattform ein Beispiel für ein „sehr überfinanziertes“ Digitalprojekt. „Das Problem ist, dass in der Politik häufig noch zu wenig Digitalexpertise vorhanden ist und Geldmittel nicht nach Bedarf, sondern nach politischen Gesichtspunkten vergeben werden“, sagt er. „Digitale Bildung war während Corona eins der großen Aufregerthemen, also wurde für die nationale Bildungsplattform schnell sehr viel Geld bereitgestellt.“ Es gebe andererseits aber auch unterfinanzierte Digitalprojekte – etwa den Sovereign Tech Fund, der Open-Source-Projekte unterstützt.

Heftige Kritik an den teuren Bildungsplattform-Plänen kommt auch vom Bundesrechnungshof. Die Bundesländer würden längst an ähnlichen Plattformen arbeiten, argumentieren die Haushaltsprüfer. Das Bildungsministerium habe sich über bestehende Planungen „hinweggesetzt“, was eine „eklatante Verletzung des Grundsatzes der Wirtschaftlichkeit“ darstelle. Aufgrund fehlender Konzepte und Abstimmungen drohten nun „Förderruinen“.

(cwo@ct.de) **ct**

Neues Zuhause? Online anmelden.



© Senatskanzlei/Christian O. Bruch

Elektronische Wohnsitzanmeldung – aktuell verfügbar in Hamburg und Lübeck

Glückliche Nordlichter: In Hamburg und Lübeck kann man seinen Wohnsitz online ummelden, anderswo müssen Bürger weiter zum Amt.

Schöne neue Hardware-Welt

Hardware-, Mobil- und Netzwerk-Trends

Hardware im Wandel: Intel krempelt die x86-Architektur um, hybride Prozessoren und Rechenhelfer ohne CPU erkämpfen sich ihren Markt. Gleichzeitig versprechen neue Technologien brillantere Displays, noch kompaktere Mobilgeräte und flottere Netze.

Von Georg Schnurer

Nach dem für die CPU-Hersteller desaströsen Jahr 2023 kommt Bewegung in den Markt: AMD bringt in Kürze Mobilprozessoren mit unterschiedlichen CPU-Kernen heraus. KI-Beschleuniger ziehen auch in Desktop-Prozessoren ein und Intel setzt künftige Prozessoren aus Chiplets zusammen. Gleichzeitig will Intel mit x86-S alte Zöpfe bei der x86-Architektur abschneiden. Parallel dazu verlieren klassische CPU-Kerne vor allem bei Supercomputern an Bedeutung – spezialisierte Rechenbeschleuniger übernehmen immer mehr Aufgaben und die CPU wird zum Beiwerk. Und klar, die bislang wenig umweltverträgliche Chip-Produktion soll nun auch schrittweise „grüner“ und der Universal Serial Bus (USB) noch universeller werden. Ab Seite 112 weisen wir Ihnen einen Weg durch die teils gegenläufigen Entwicklungen.

Matter, ein universeller und offener Standard fürs smarte Zuhause, verspricht endlich Ordnung ins ausufernde Gewirr verschiedener herstellerspezifischer und zueinander inkompatibler

Smart-Home-Systeme zu bringen. Ob und wie Matter die zer splitterte Smart-Home-Welt einen kann, erfahren Sie ab Seite 140.

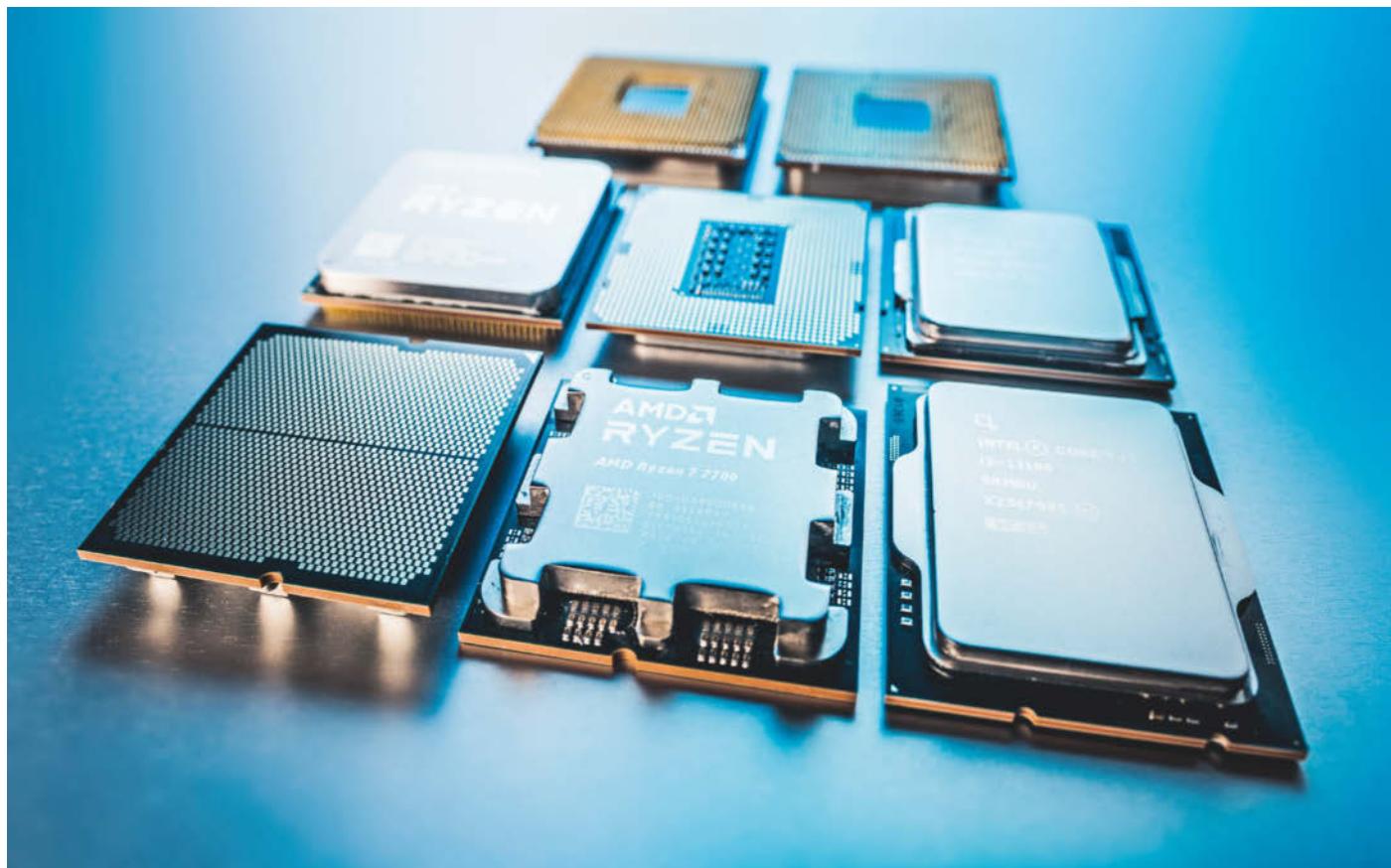
Darf es ein wenig flacher sein? Oder lieber langlebig und reparaturfreundlich? Bei Smartphones stellen gegenläufige Trends die Käufer vor die Wahl. Doch damit nicht genug: Auch die Mobilfunktechnik entwickelt sich rasant weiter. Im Fokus steht allerdings nicht mehr nur eine immer höhere Datenübertragungsrate, es geht vor allem bei neuen 5G-Systemen auch um eine idealerweise globale Netzabdeckung: Klassische Funkmasten werden durch Satelliten ergänzt, aber auch durch Drohnen und hoch fliegende Sendestationen – das soll zum einen für hohe Bandbreiten in Ballungsgebieten sorgen, zum anderen Empfang auch im Nirgendwo ermöglichen. Wie das alles funktioniert und was ein Smartphone können muss, um davon zu profitieren, lesen Sie ab Seite 144.

Mixed-Reality-Headsets (MR) versprechen die Vereinigung von Augmented Reality (AR), also das Überlagern realer optischer und akustischer Sinneseindrücke mit Computerinhalten, und Virtual Reality (VR), das komplett Eintauchen in eine computergenerierte Welt. Mit VR taucht man in digital erzeugte Welten ein (Immersion), zum Beispiel bei 3D-Lehrgängen oder Spielen, die AR blendet digitale Objekte über die reale Welt, etwa Bedienoberflächen von Programmen oder Montagehinweise für reale Gerätschaften. Apples groß angekündigte Computerbrille Vision Pro soll neuen Schwung in die MR-Welt bringen. Ob die sündhaft teure Computerbrille ein Rohrkrepierer wird oder mit einem lange erwarteten Knall eine neue Epoche einläutet – bei der Meinungsbildung hilft unser Beitrag ab Seite 156. (gs@ct.de) **ct**





| | |
|---|------------------|
| Hardware-, Mobil- und Netzwerk-Trends | Seite 110 |
| Hybrid-Zeitalter: Die CPU-Pläne von AMD und Intel | Seite 112 |
| Kluge Chips: x86-Prozessoren mit KI-Einheiten | Seite 116 |
| Ohne Altlasten: CPU-Mikroarchitektur x86-S | Seite 120 |
| Die APUs kommen: Das Ende der klassischen Prozessoren? | Seite 124 |
| Ökochips: Wie die Umweltlasten der Halbleiterfertigung sinken sollen | Seite 126 |
| Eierlegende Wollmilchsau: Wie es mit USB4 weitergeht | Seite 132 |
| Kristallwerkstatt: Displaytechnik für Smart-TVs | Seite 136 |
| Großes Versprechen: Der herstellerübergreifende Smart-Home-Standard Matter ... | Seite 140 |
| Smartphone-Trends: Flacher und teurer, aber auch langlebiger und offener | Seite 144 |
| Globale Mobilfunkabdeckung: Wie das geht, welche Smartphones sich eignen | Seite 148 |
| Kein Ende in Sicht: Hype um Handheld-Konsolen | Seite 154 |
| Spatial Computing: Wie Apples MX-Headset Vision Pro arbeitet | Seite 156 |
| Netzbeschleunigung: Wie WLAN und Ethernet schneller werden | Seite 160 |



Aufbruch ins Hybrid-Zeitalter

Die Pläne der CPU-Hersteller AMD und Intel für 2024

Bald bringt auch AMD Mobilprozessoren mit unterschiedlichen CPU-Kernen auf den Markt. Zudem ziehen KI-Beschleuniger ein und Intel setzt künftige Prozessoren aus Chiplets zusammen.

Von Christof Windeck

Das Jahr 2023 wollen die Prozessorhersteller AMD und Intel vermutlich so schnell wie möglich vergessen. Nach dem überraschenden PC-Boom der Jahre 2020 und 2021 – ausgelöst von der Home-

office-Welle wegen der Corona-Lockdowns – brachen die Verkaufszahlen von Desktop-PCs und Notebooks drastisch ein. Nun keimt Hoffnung, dass 2024 die Umsätze wieder steigen. Neue CPU-Generationen sollen das Wachstum antreiben, vor allem bei Notebooks. Künftige Prozessoren bekommen außer schnelleren und effizienteren CPU-Kernen auch spezialisierte KI-Rechenwerke.

Notebook-Wandel

Für viele Notebook-Käufer ist der jeweils eingebaute Prozessor nicht (mehr) die wichtigste Komponente. Denn viele Alltagsaufgaben bewältigt auch eine billigere CPU einer älteren Generation flott. Daher können andere Eigenschaften für die Kaufentscheidung wichtiger sein: lange

Akkulaufzeit, niedriges Gewicht, ein farbstarkes, helles Display im 16:10-Format, komfortables Docking per USB-C oder Thunderbolt oder ein starker 3D-Grafikchip für PC-Spiele.

Auf diese Entwicklung haben sich die Notebookhersteller eingestellt. Sie lassen Geräte mit älterer Technik tendenziell länger auf dem Markt: als billigere Alternativen zu neuen Modellen mit stärkeren Prozessoren. So verkauft etwa Apple weiterhin das drei Jahre alte MacBook Air mit M1-CPU, obwohl es seit Mitte 2022 auch eines mit M2 gibt.

Das führt wiederum dazu, dass manch neue Familie von Mobilprozessoren nur noch relativ starke Versionen umfasst, etwa welche mit mindestens sechs CPU-Kernen, also ab den Klassifizierungen

Ryzen 5/Core i5 über 7/i7 bis zu 9/i9. Schwächere Varianten lassen die CPU-Hersteller weg oder frischen sie in längeren Abständen auf, etwa Ryzen 3/Core i3, Athlon, Pentium und Celeron. Das stiftet Verwirrung, weil Prozessoren mehrerer Generationen gleichzeitig auf dem Markt sind. Und wer die jeweils neueste Prozessortechnik haben will, muss sich auf hohe Preise einstellen.

Notebooks: AMD

AMD hatte im Herbst 2022 ein neues Namensschema für die Ryzen-Mobilprozessoren ab 2023 verkündet: Alle neuen Versionen eines Jahres tragen Produktnummern, die mit derselben Zahl anfangen. 2023 wurde das Jahr der Ryzen-7000-Familien, wobei in den meisten davon aber gut abgehängte Technik steckte. Hinter dem vermeintlich neuen Ryzen 7035 verbergen sich eigentlich Varianten des Ryzen 6000. Und beim schwächeren Ryzen 7020 setzt AMD sogar noch auf Zen-2-Kerne, die schon 2019 im Ryzen 3000 erschienen.

Als echte Neuheiten mit Zen-4-Kernen kamen 2023 nur die Baureihen Ryzen 7040 (Phoenix) und Ryzen 7045 (Dragon Range). Letztere zielt auf fette Gaming-Notebooks und mobile Workstations. Der Ryzen 7040 ist für häufiger verkaufte flache (Business-) Notebooks gedacht; in einigen Versionen steckt der KI-Beschleuniger Ryzen AI, siehe dazu auch Seite 116.

AMD-Chefin Lisa Su hatte den Ryzen 7040 zwar bereits im Januar auf der CES angekündigt, doch Notebooks mit diesen Chips kamen erst Monate später auf den Markt, zunächst auch nur mit den stromdurstigeren „H“-Typen. Erst seit dem Sommer findet man auch welche mit Ryzen 7040U – und zwar meistens teurere Geräte oberhalb von 1000 Euro.

Auf der CES 2024 dürfte AMD einige Ryzen-8000-Vertreter mit den Codenamen Strix Point und Strix Halo vorstellen, die dann wohl frhestens im Mai 2024 in Notebooks auftauchen werden. Und dann werden es vermutlich genau wie 2023 zunächst wieder nur die stärksten Versionen für teurere Notebooks sein, also die H- oder HS-Varianten der Leistungsklasse 35 bis 45 Watt.

Vom Ryzen 8000 erwarten wir den Umstieg von der Zen-4- auf die Zen-5-Generation, weiterhin mit Speicher-Controllern für DDR5 oder LPDDR5. Ob TSMC die Zen-5-Chiplets mit N4- oder N3-Technik produziert, lässt AMD bisher offen.

2024 dürfte auch AMD Hybridprozessoren mit starken Performance-(P-)Kernen sowie mehreren Effizienz-(E-)Kernen herausbringen, so wie es Intel seit 2022 tut und wie es bei ARM-Chips für Smartphones seit Jahren üblich ist. Schon jetzt ist eine Variante des Ryzen 7040 mit Zen-4- und Zen-4C-Kernen aufgetaucht, aber noch nicht offiziell. Der Leistungsunterschied dieser Kernversionen scheint geringer zu sein als bei Intels P- und E-Kernen.

Als weitere Produktvariante bestückt AMD manche Prozessoren mit einem zusätzlichen, aufgestapelten Chiplet, das den L3-Cache deutlich vergrößert. Das beschleunigt vor allem PC-Spiele, aber etwa auch manche (DAW-)Musiksoftware. Mit dem Ryzen 9 7945HX3D verkauft AMD auch einen Mobilprozessor mit vergrößertem L3-Cache. Ein solcher ist auch in der mobilen CPU-Generation Ryzen 8000 zu erwarten.

Fraglich ist, ob AMD auch bei Notebook-Chips PCI Express 5.0 (PCIe 5.0) bringt; das verspricht dabei nur vernachlässigbare Vorteile, anders als bei Servern und Workstations. Bisherige PCIe-5.0-SSDs sind viel zu stromdurstig für Notebooks.

AMD dürfte auch die integrierten Grafikprozessoren (IGP) überarbeiten, man munkelt von RDNA 3.5. Die IGP könnte theoretisch auch durch mehr Speicherkanäle für LPDDR5X-RAM profitieren. Allerdings sitzen in teureren (Ga-

c't kompakt

- Künftige x86-CPUs von AMD und Intel enthalten KI-Beschleuniger, wie man sie von Smartphone-SoCs schon lange kennt.
- Nach AMD setzt nun auch Intel Prozessoren aus mehreren Chiplets zusammen, was flexiblere Konfigurationen ermöglicht.
- Das Notebook-Angebot wandelt sich, Geräte mit Prozessoren mehrerer Generationen sind gleichzeitig auf dem Markt.

ming-)Notebooks üblicherweise diskrete GPUs und in billigeren Notebooks eher wenige LPDDR5X-Chips. Die höhere RAM-Datentransferrate könnte aber auch KI-Software beschleunigen.

Unterhalb der starken Zen-5-Mobilprozessorfamilien Strix Point und Strix Halo könnten überarbeitete Zen-4-Serien mit Ryzen-8000-Familiennamen kommen.

Notebooks: Intel

Am 14. Dezember startet Intels Mobilprozessorgeneration Meteor Lake, die mehrere wesentliche Neuerungen bringt. Intel nennt sie nicht mehr Core i, sondern Core Ultra. Diese Prozessoren sind aus

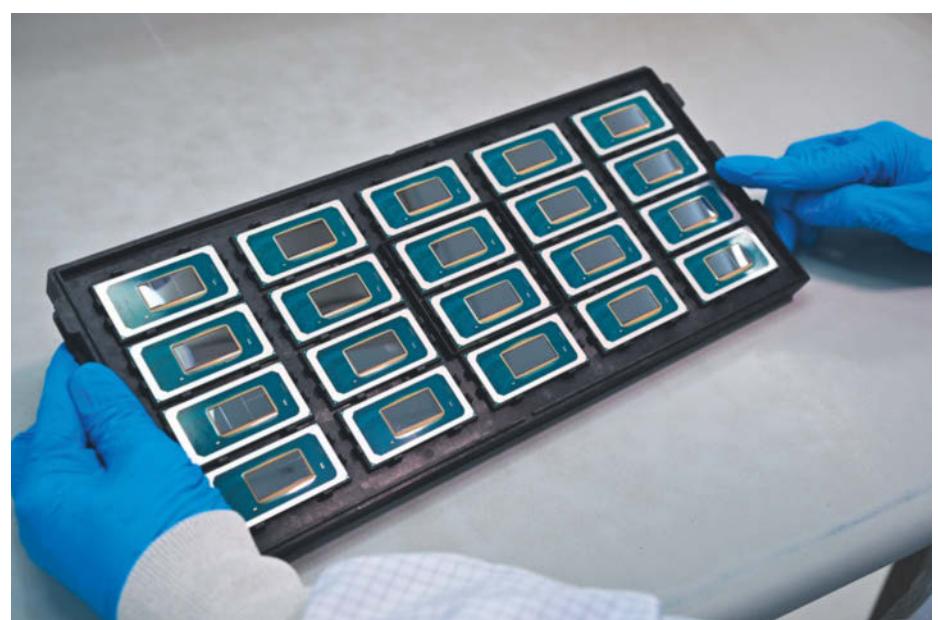
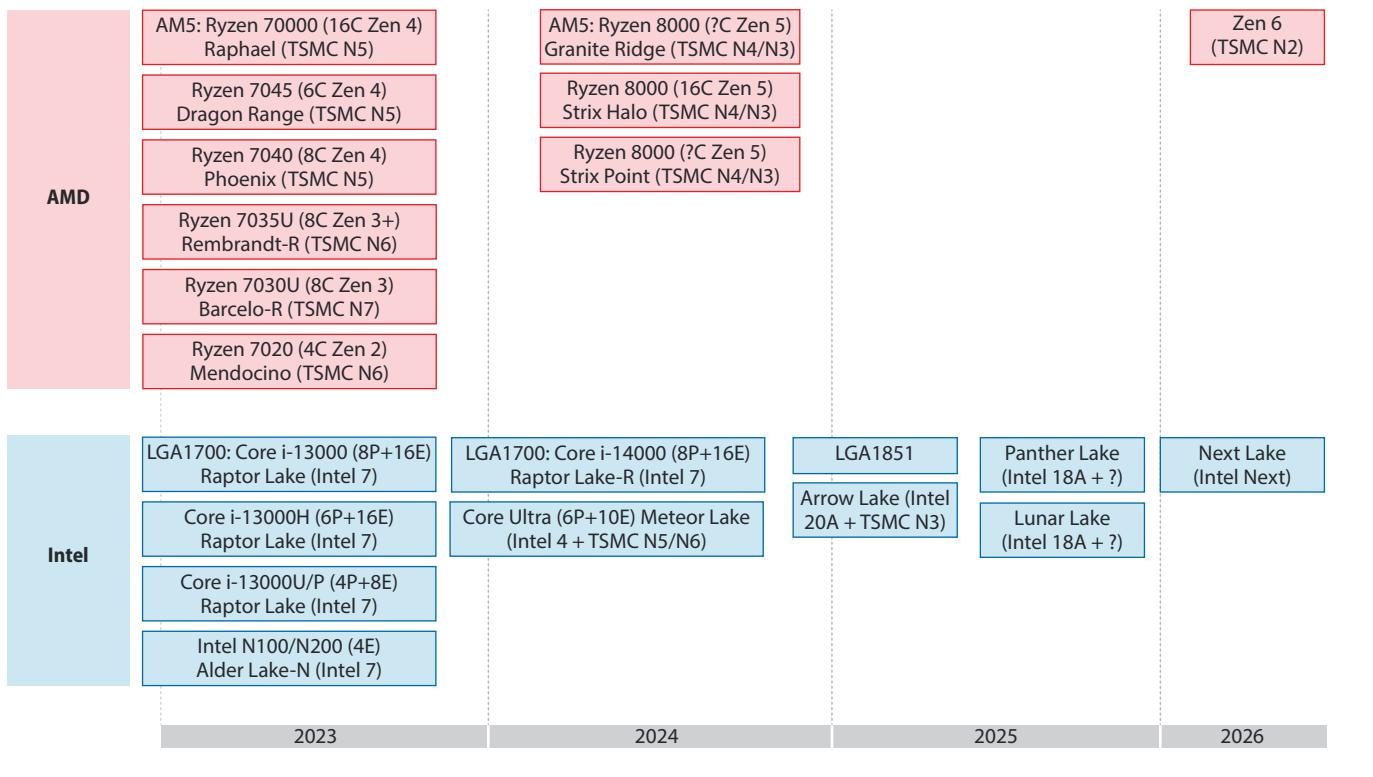


Bild: Intel

Nur wer ganz genau auf die Core-Ultra-Prozessoren starrt, kann erkennen, dass sie jeweils aus mehreren Chiplets zusammengesetzt sind.

PC-Prozessoren von AMD und Intel bis 2026

AMD plant 2024 mit der Ryzen-Generation 8000 und Zen 5, gefertigt von TSMC mit N4- oder N3-Technik. Intel setzt ebenfalls auf Chiplets und kauft bei TSMC zu, nutzt aber auch die neuen Fertigungsverfahren Intel 4, Intel 20A und Intel 18A.



mehreren Chiplets zusammengefügt, die Intel aber „Tiles“ (Kacheln) nennt, und das CPU-Chiplet entstammt der jüngsten Intel-Fertigungstechnik „Intel 4“. Das Verfahren erzeugt zwar eher 7-Nanometer-Strukturen, könnte aber dank einer Reihe von Optimierungen mit TSMC N5/N4 mithalten. Das GPU-Chiplet lässt Intel wiederum bei TSMC (N5) fertigen. Intel baut ebenfalls einen KI-Beschleuniger namens „NPU“ ein.

Eine größere Auswahl von Meteor-Lake-Notebooks ist erst 2024 zu erwarten; zur CES 2024 könnte Intel noch weitere CPU-Varianten ankündigen. Die Chiplets ermöglichen viele Kombinationsmöglichkeiten, also eine breite Produktpalette.

Intel befindet sich in einer waghalsigen Aufholjagd, in die Intel-Chef Pat Gelsinger Dutzende Milliarden US-Dollar pumpt. Das Unternehmen wandelt sich zum Auftragsfertiger mit angeschlossener CPU-Abteilung und will bei der Fertigungstechnik bis spätestens 2026 an TSMC vorbeiziehen.

Auch bei den CPU-Generationen soll es Schlag auf Schlag gehen und noch 2024

soll mit Arrow Lake ein erster Prozessor mit der Fertigungstechnik „Intel 20A“ kommen. Bisher ist unklar, ob Arrow Lake auf Meteor Lake folgt oder diese Serie eher ergänzt. Auf Arrow Lake sollen später Lunar Lake und Panther Lake folgen, beide in der Fertigungstechnik Intel 18A, die größere Verbesserungen verspricht. Denn die Transistoren haben eine neuartige Bauform namens RibbonFET, die andere Firmen Gate All Around (GAA) nennen, und die „Verdrahtung“ innerhalb des Chips wird optimiert durch Backside Power Delivery alias PowerVia. Wie groß der Leistungssprung wird, bleibt jedoch abzuwarten.

Für billige Notebooks hatte Intel bereits Ende 2022 die „N“-Serie (N100, N200, Core i3-N305) als Ablösung der zuvor verbreiteten Celeron N und Pentium N vorgestellt. Diese N-Typen bestehen nur aus den E-Cores der Alder-Lake-Prozessoren, sind deutlich leistungsfähiger als ihre Vorgänger und die 6-Watt-Vierkerner ermöglichen im Prinzip auch lüfterlose Notebooks und Mini-PCs. Während es Mini-PCs mit N100 schon in größerer Auswahl gibt, erscheinen damit bestückte

Notebooks aber nur allmählich. Ein Nachfolger für 2024 ist nicht in Sicht.

Desktop-PCs: AMD

Ende 2022 hatte AMD mit den Ryzen 7000 (Zen 4) die neue Prozessorfassung AM5 für DDR5-Speicher eingeführt. Sie soll bis 2025 oder 2026 aktuell bleiben, aber 2024 erwartet man neue Zen-5-Ryzens der Serie 8000. Vermutlich bleibt es bei maximal 16 starken P-Kernen. Ob AMD auch den Desktop-Prozessoren E-Kerne gönnt, ist fraglich. Die Zen-4-Kerne von AMD sind kompakter als die bisherigen P-Kerne von Intel und AMD setzt schon länger auf Chiplets. Daher kann AMD wohl auch nur mit verbesserten P-Kernen sehr hohe Multithreading-Performance liefern. Außerdem hat AMD ja noch das Ass mit dem aufgestapelten L3-Cache im Ärmel und es sind auch X3D-Varianten des Ryzen 8000 zu erwarten.

Für noch stärkere Desktoprechner der Workstation-Klasse sind die Ryzen 7000 Threadripper vorgesehen, die wir eigentlich schon Mitte 2023 erwartet hatten. Sie sind mit den Zen-4-Epycs 9004 (Genoa) verwandt.

Desktop-PCs: Intel

Kürzlich hat Intel die Core-i-Serie 14000 als kaum verbesserte Version des Ende 2022 vorgestellten Core i-13000 „Raptor Lake“ für LGA1700-Mainboards herausgebracht. Der Umstieg auf die 14. Core-i-Generation lohnt sich bei Desktop-PCs nur bei wenigen Exemplaren wie dem Core i7-14700K, der mehr E-Kerne hat als sein Vorgänger. Nach einem BIOS-Update laufen die 14000er auch in LGA1700-Boards.

2024 ist mit Arrow Lake-S (der dann wohl auch als Core Ultra kommt) ein größerer Leistungssprung zu erwarten. Denn Intel wird dann wohl auch bei Desktop-Prozessoren auf Chiplets setzen. Diese neuen Prozessoren brauchen neue Mainboards, man munkelt von der Fassung LGA1851. Dabei dürfte Intel auf die Option für DDR4-RAM verzichten, man braucht also stets DDR5-Module.

Wegen der Chiplet-Fertigung könnte auch Intel Varianten mit stark vergrößertem L3-Cache bringen. Vermutlich wird die Anzahl der P-Kerne bei maximal acht Stück bleiben, die aber bei der Rechenleistung pro Taktzyklus zulegen dürften. Dazu kommen dann möglicherweise mehr E-Kerne, etwa bis zu 24 statt bisher höchstens 16 und ebenfalls mit höherer Rechenleistung pro Takt. Dadurch ist erheblich mehr Multithreading-Performance zu erwarten.

ARM-Joker

AMD und Intel fertigen beide x86-Prozessoren und haben daher gemeinsame Konkurrenten, nämlich Produzenten von ARM-Prozessoren. Apple hat mit seinen MacBooks mit hauseigenen „M“-Prozessoren mit ARM-Kernen seit Ende 2020 erhebliche Marktanteile erobert. Doch während der M1 ein Paukenschlag war, brachte der M2 2022 nur kleinere Verbesserungen. Daher konnten AMD und Intel deutlich aufholen. Bis zum Redaktionsschluss dieser c't-Ausgabe hatte Apple noch kein M3-MacBook angekündigt.

Microsoft und Qualcomm bemühen sich schon seit mehr als einem Jahrzehnt, Windows-Notebooks mit ARM-Chips in den Markt zu drücken. Das scheiterte jedoch, weil die ARM-SoCs von Qualcomm bisher schlichtweg zu lahm waren und es obendrein Nachteile bei der Kompatibilität mit x86-Software und -Treibern gibt. Wieder einmal verspricht Qualcomm, dass sich das bald mit einer

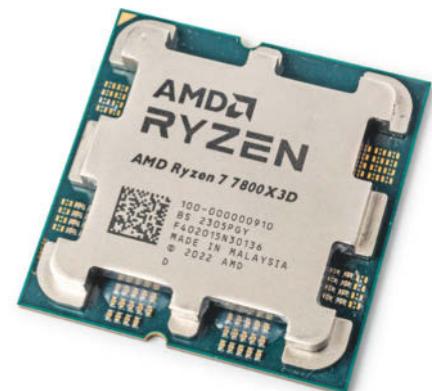
neuen Generation von ARM-Chips ändern könnte – aber das lässt sich aus heutiger Sicht nicht einschätzen. Umgekehrt behaupten AMD und Intel, die Effizienz ihrer Prozessoren so stark verbessern zu wollen, dass sie Apples „M“-Typen in einigen Jahren überholen werden. Warten wir's ab.

Server-Rennen

AMD hat mit den Serverprozessoren Epyc 9004 (96 Zen-4-Kerne, Genoa), Epyc 9004X (96 Zen-4-Kerne, großer L3-Cache) und Epyc 97x4 (128 leicht abgespeckte Zen-4C-Kerne, Bergamo) erhebliche Marktanteile erobert. Für günstigere Server erschien nun auch noch der Epyc 8004 „Siena“ mit 64 Zen-4C-Kernen, der auf Mainboards mit der Fassung SP6 passt und nur acht statt zwölf DDR5-Speicherkanäle anbindet.

Im Vergleich zu dieser Epyc-Phalanx hat es Intels Xeon-SP Gen 4 (Sapphire Rapids) schwer, weil er höchstens 60 Kerne hat. Nur wenn angepasste Software die zahlreichen eingebauten Beschleuniger unter anderem für Verschlüsselung oder KI-Funktionen nutzt, kann Intels Xeon punkten. Daran ändert auch der für Dezember angekündigte Xeon-SP Gen 5 „Emerald Rapids“ nichts Grundsätzliches, er hat ein paar Kerne mehr und größere Caches.

2024 will Intel aber aufdrehen, als Erstes mit dem „Sierra Forest“ mit bis zu 288 Effizienzkernen (E-Cores). Das ist



Der Achtkerner Ryzen 7 7800X3D unterscheidet sich vom Ryzen 7 7700X durch ein zusätzliches Chiplet, das die Kapazität des L3-Cache verdreifacht.

nicht nur eine Ansage an AMD, sondern auch an den AmpereOne mit 192 ARM-Kernen.

Kurz darauf soll „Granite Rapids“ mit vermutlich bis zu 120 P-Kernen erscheinen, aber auch bis zu 500 Watt Leistungsaufnahme. Auch Intel will zwei CPU-Fassungen bedienen, LGA7529 für zwölf Kanäle und weiterhin LGA4677 für acht.

2024 könnte AMD mit mehreren Varianten von „Turin“-Epcys kontern. Man erwartet bis zu 128 Zen-5-Kerne und 192 Zen-5C-Kerne pro CPU.

Das kommende Prozessorjahr wird spannend, weil mehrere neue Fertigungsverfahren in Großserie ins Rennen gehen.

(ciw@ct.de) ct



MacBook Air

15: Apple ist bislang der einzige Hersteller, der Notebooks und Desktop-PCs mit ARM-Prozessoren in großen Stückzahlen verkauft.

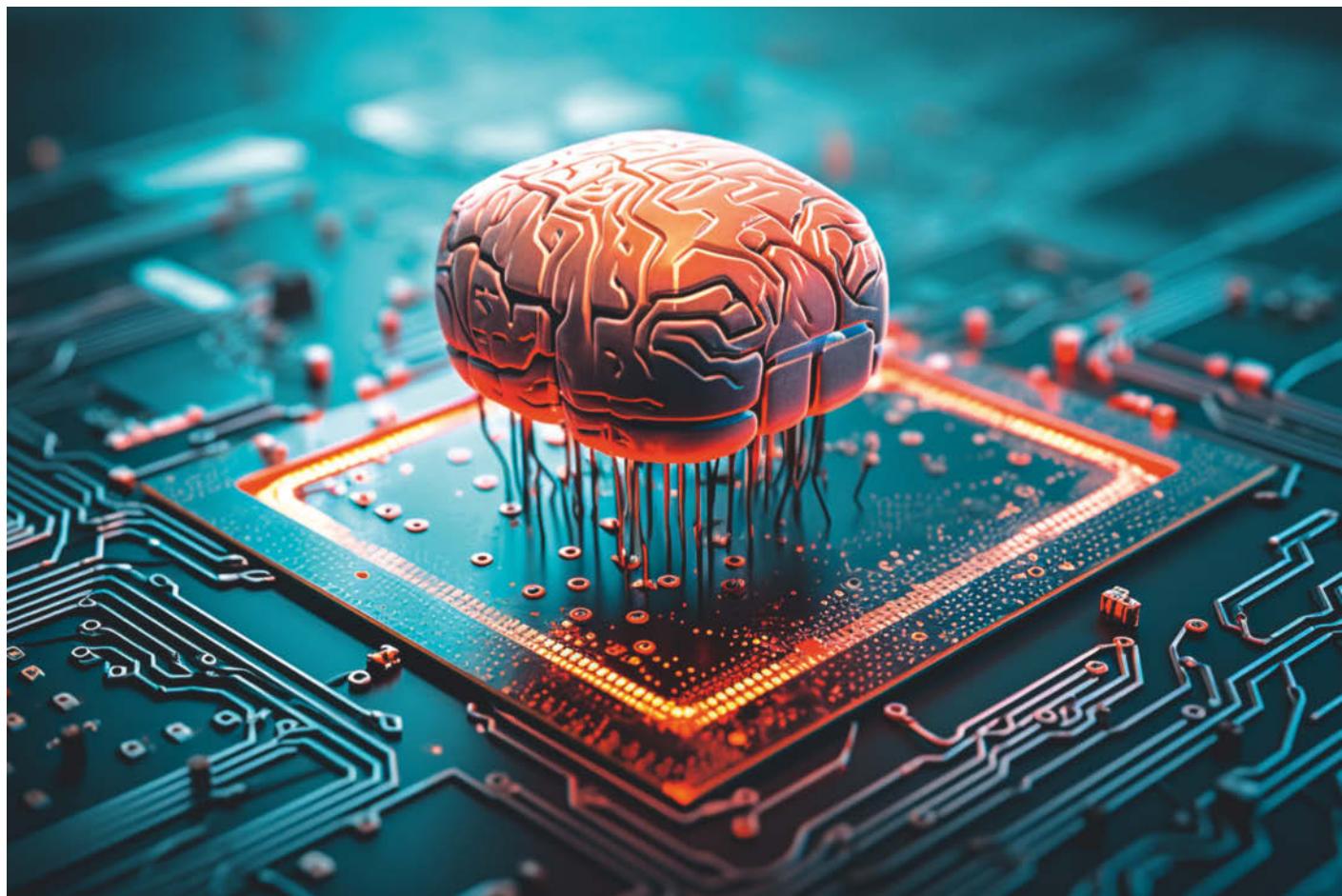


Bild: KI Midjourney | Bearbeitung: c't

Der Grips in Chips

x86-Prozessoren mit KI-Einheiten

Künstliche Intelligenz ist die Hoffnung nahezu der gesamten IT-Industrie. Das Kürzel KI wird inflationär überall drangezappt, wo irgendetwas rechnet oder speichert. Wir schauen uns die KI-Erweiterungen in x86-Prozessoren von Intel und AMD etwas genauer an.

Von Carsten Spille

Für manche dystopisch besetzte, für einige der Maschine gewordene Heiland und Löser aller Probleme, ist „Künstliche Intelligenz“ ein ebenso altes wie schwammiges Schlagwort: Muss eine KI selbst denken können und kreativ sein? Oder reicht es, wenn sie Informationen innerhalb eines vorgegebenen Zusammenhangs stumpf wiederkäut wie etwa eine Datenbankabfrage? Eine solche selbst denkende oder „echte“ KI wird als Artificial General Intelligence (AGI) bezeichnet, ist bisher aber nur Science Fiction.

Zwischen diesen beiden Extremen liegt der aktuelle Stand zumindest der öffentlich zugänglichen Entwicklung. KI-Systeme verknüpfen bekannte Daten zum

Teil sinnvoll neu, sofern die Anfragen passend gestellt sind, und geben damit Antworten auf viele Anfragen. Die Resultate können durchaus beeindrucken, wenn die KI zum Beispiel ein bekanntes Van-Gogh-Gemälde im Stile Picassos erzeugt oder brauchbaren Programmcode generiert. Sie übertrifft damit aber nicht die Summe ihrer Informationsbestandteile, sodass eine KI, in deren Datenbasis Picasso fehlt, dessen Stil auch nicht nachbilden könnte – was ein Mensch zugegebenermaßen auch nicht könnte, aber ein Picasso konnte diesen Stil eben doch aus sich heraus erfinden.

Den Hype um KI lassen sich auch Hersteller von PC-Hardware nicht entgehen

und hängen das „KI“-Schildchen an viele ihrer Produkte. Mit im Boot sind dabei auch AMD und Intel. AMD hat in ausgewählten, sprich einigen wenigen, Ryzen-7040-Modellen „RyzenAI“ bereits eingebaut, Intel will ab Dezember 2023 mit der CPU-Generation „Meteor Lake“ nichts weniger als das Zeitalter der KI einläuten.

Doch wie soll das mit winzigen Zusatzschaltkreisen überhaupt gehen? Wir haben gelernt, KI braucht große Mengen Rechenleistung und viel Speicherplatz – und doch werben selbst Smartphones damit. Dazu vorab ein wenig Erklärung.

Große Modelle, große Leistung

Der aktuell „heiße Scheiß“ sind generative KI-Systeme, die wie zum Beispiel ChatGPT auf sogenannten Large Language Models (LLMs) fußen. Diese tragen ihren Namen nicht zu Unrecht, denn um ein Modell zu erstellen, anhand dessen die KI-Algorithmen ihre Schlüsse ziehen, sind Dutzende Milliarden an Beispielen nötig, die nicht nur für die Trainingsberechnung berücksichtigt werden, sondern dabei auch viel Speicherplatz zusätzlich zu den Daten belegen.

Ein extrem vereinfachtes Beispiel für so einen Parameter kann dabei die Verknüpfung eines Katzenbildes mit dem Wort Katze sein. Da es nicht nur bei Bildern, sondern speziell bei semantischen Wortgruppen in menschlicher Sprache viele Milliarden solcher Bedeutungspaares gibt, die alle miteinander verknüpft sind, kann schnell der Eindruck entstehen, ein KI-System wäre nahezu allwissend. Schließlich beantwortet es die meisten Anfragen korrekt. Wenn die KI keine konkrete Antwort weiß, schnippelt sie sich zur Not auch irgendwas aus Versatzstücken zusammen, was zuweilen für Belustigung sorgt. Manchmal klingen die falschen Antworten aber so überzeugend, dass man sich im Vertrauen auf „die KI“ auch fies in die Nesseln setzen kann, wie es einem US-Anwalt widerfuhr, dem eine KI-Anfrage nach Präzedenzfällen frei erfundene Gerichtsverfahren lieferte.

Damit man ein KI-Modell ohne eigenen Supercomputer nutzen kann, werden die semantischen Paare gewichtet, das sogenannte Training. Grob vereinfacht wird zum Beispiel einem Bild von einem Löwen außer Tier und Natur auch das Wort Katze zugeordnet, weil es ein Tier aus der Familie der Katzen ist. Die Beziehung von Löwenbild und Katze ist allerdings schwächer als die eines Katzenbildes

und Katze. Diese Gewichtungen werden ebenfalls gespeichert, sodass bei der Anwendung des fertig trainierten Modells nur noch der Gewichtungsbaum durchlaufen werden muss, um zum gewünschten Ergebnis zu gelangen. Das geht schneller, und auch ein einfacheres, sprich billigeres und energieeffizienteres Rechenwerk als ein Prozessorkern kann das leisten: Hier kommen die KI-Einheiten ins Spiel, auf die wir gleich eingehen.

KI: VNNI, NPU, IPU und VPU – alles klar?

Künstliche Intelligenz hat schon einen langen Weg hinter sich. Dabei haben sich einige Abkürzungen angesammelt, die alle irgend etwas mit KI zu tun haben. NPU etwa steht für Neural Processing Unit, das „I“ in IPU für Intelligence. Das bei Intel bis vor kurzem noch gehandelte Kürzel VPU kommt von Visual Processing Unit, einer Spezialschaltung der israelischen Tochterfirma Movidius, die in größeren Versionen auch schon als Beschleunigerkarten oder -USB-Sticks verkauft wurden.

Eine frühe Variante waren die Vector Neural Net Instructions, kurz VNNI. Diese vier Befehle samt Unterstützung für das BFloat16-Datenformat sind seit 2019, also seit der Cascade-Lake-Generation, in Intels Xeon-Server-Prozessoren integriert. VNNI wurde zunächst als Teil von AVX512 konzipiert und in die Beschleunigerarchitektur Knights Mill eingebaut.

Da man für KI-Berechnungen, in diesem Fall neuronale Netze, häufig Matrixmultiplikationen zweier 16-Bit- oder gar nur 8-Bit-Werte benötigt und eine darauf folgende Addition mit 32-Bit-Genauigkeit, kann man in den breiten AVX512-Einheiten prima mehrere gleichzeitige Berech-

nungen mit geringerer Präzision kombinieren. Anstelle der üblichen 32-Bit-Werte passen zum Beispiel doppelt so viele FP16-Inhalte oder viermal so viele INT8-Zahlen. Außerdem wird das Speichersystem entlastet, wenn man für teils mehrfach genutzte Operanden weniger wertvollen Speicherplatz in den kleinen Caches braucht.

Wenn KI-Aufgaben auf solchen spezialisierten Rechenwerken laufen, benötigen sie in der Regel entweder für dieselbe Ausführungsgeschwindigkeit wesentlich weniger Energie oder sind bei gleicher Leistungsaufnahme viel schneller. Meistens liegt der Unterschied nicht bei wenigen Prozenten, sondern sogar bei ganzzähligen Faktoren. Für so einen Gewinn in Sachen Effizienz lohnt sich auch die zusätzlich belegte Siliziumfläche.

AMDs RyzenAI

Einen Schritt weiter gehen spezialisierte Schaltungen für KI-Berechnungen. Wie oben beschrieben, sind speziell für die Anwendung von fertig trainierten Modellen, das sogenannte Inferencing, Recheneinheiten mit geringerer Präzision ausreichend. Sie profitieren außerdem von jeweils kleinen, lokalen Speicherblöcken, die die Daten schnell abliefern und wieder entgegennehmen können. Dafür können im typischen KI-Einsatz mehr solcher Berechnungen zugleich laufen, als normale Desktop-Prozessoren Kerne haben.

Eine KI-Anwendung profitiert also von spezialisierter Hardware – und genau die findet jetzt auch ihren Weg in weitere x86-Prozessoren. Den Anfang macht dabei AMDS mobile Ryzen-7040-Reihe, von deren sieben Modellvarianten die schnellsten fünf ab dem Ryzen 5 7640U eine sogenannte RyzenAI-Einheit bekommen. Ryzen 5 7540U und Ryzen 3 7440U gehen diesbezüglich leer aus. Überdies müssen die OEMs, also die Notebookhersteller, die KI-Einheit im BIOS freischalten und – so unbestätigte Gerüchte – dafür eine Lizenz bei AMD erwerben.

Als Basis für RyzenAI nutzt AMD eine Entwicklung der zugekauften FPGA-Tochter Xilinx. Die Firma hält sich zwar bedeckt, um welche Schaltungen genau es sich dabei handelt, aber es dürfte sich um eine Variante der Xilinx AI-Engine handeln. Die Technik dahinter nennt AMD XDNA.

Ein Satz von zwanzig hoch spezialisierten sogenannten Tiles (zu Deutsch Kacheln), also identisch wiederholten

c't kompakt

- KI-Anwendungen sind ein Riesen-thema und werden auch den Heim-anwender früher oder später errei-chen.
- Spezialisierte Schaltungen lassen KI-Algorithmen wesentlich schneller und effizienter laufen.
- AMD hat mit RyzenAI bereits KI-Einheiten in einigen Mobil-CPUs, Intel will mit der NPU in Meteor-Lake im Dezember nachziehen.

Schaltungsblöcken, übernimmt die Berechnungen für KI-typische Datenformate von INT8 bis INT32 und BFloat16. Fünf weitere Tiles verbinden die Rechenknöpfchen über ein Network-on-Chip und Einheiten für den direkten Speicherzugriff. Vier parallele KI-Streams sollen die RyzenAI-Engines anfangs verarbeiten. Dank feinstufiger Taktanpassung und -abschaltung schlucken die XDNA-Engines bei Nichtbenutzung so gut wie keine Leistung. Laufen sie hingegen unter Voll-dampf, sollen sie 10 Billionen KI-Operationen pro Sekunde (TeraOps) schaffen.

AMDs für KI-Anwendungen gedachtes Framework Unified AI Stack läuft zwar auf Ryzen-7040-Prozessoren, konnte die dedizierte XDNA-Einheit anfangs aber nur eingeschränkt nutzen, weshalb die meisten Algorithmen zunächst doch wieder unter viel Energieeinsatz von den CPU-Kernen beackert werden.

Bei Intel gibts NPU

Der KI-Block ist bei Intels kommenden Meteor-Lake-Prozessoren im SoC-Chiplet untergebracht und ist damit nicht auf eine dicke Grafikeinheit oder eine große Phalanx CPU-Kerne angewiesen. Anders als AMD greift Intel bei seiner Neural Processing Unit, kurz NPU, nicht auf eine Entwicklung seiner FPGA-Sparte Altera zurück, sondern auf Technik der ebenfalls

Nichts weniger als das Zeitalter der KI will Intel mit den Meteor-Lake-Prozessoren und ihrer integrierten KI-Einheit einläuten.

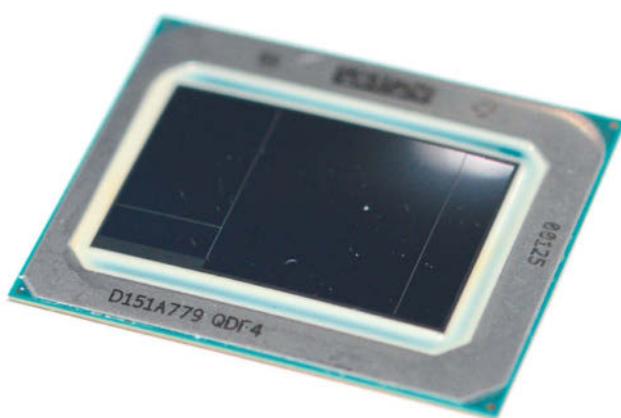


Bild: Carter Spille

zugekauften Firma Movidius. Diese hieß VPU, wie ursprünglich auch die Engine in den Meteor-Lake-Prozessoren. Sie wurde vor der offiziellen Ankündigung aber noch schnell in NPU umbenannt. Die NPU soll anders als bei AMDs Mobil-CPUs in allen Meteor-Lake-Varianten vorhanden sein.

Anhand eines Szenarios für den KI-Bildgenerator Stable Diffusion 1.5 illustriert Intel die Vorteile, KI-Algorithmen auf verschiedenen Hardwaretypen laufen zu lassen, von den Prozessorkernen über die integrierte Grafikeinheit bis zur spezialisierten NPU. Wenn die letztgenannte für Teilaufgaben eingespannt wird, soll das System im Vergleich zu den CPU-Kernen des Meteor-Lake-Prozessors fast um den

Faktor 7,8 effizienter sein: Im Schnitt schluckt das System dann 10 statt 40 Watt und ist knapp doppelt so schnell fertig. Läuft alles auf der integrierten Grafikeinheit, geht es noch einmal 30 Prozent schneller, aber die Leistungsaufnahme steht bei 37 Watt, also nur minimal sparsamer als mit den CPU-Kernen. Am schnellsten ist die Kombination aus GPU und NPU, nämlich noch einmal 22 Prozent flinker als die GPU-Version, sie braucht aber satte 30 Watt und schafft in der Effizienzbetrachtung nur Faktor 5. Für Geräte, die mit ihrem Akku den Tag auskommen sollen, ist der Sieger in diesem Szenario also schnell gefunden und auch der Grund, warum Intel die Fläche für die NPU-Schaltkreise im SoC-Tile investiert.

Über die genauen KI-Fähigkeiten schweigt Intel derzeit, brüstet sich aber mit einer breiten Unterstützung kommender Software, darunter sind der Microsoft 365 Copilot, Windows Studio Effects und andere (s. Kasten).

Fazit

KI kommt nicht erst, sie ist bereits da. Ihre Nutzung wird sich in den Alltag von PC-Anwendern integrieren, ob man das nun merkt wie bei einem Chatbot oder es im Hintergrund verborgen bleibt, wie bei Freistellfunktionen bei der Fotobearbeitung oder bei weichgezeichneten Hintergründen in Videocalls.

Damit diese Aufgaben schnell und effizient erledigt werden, bauen AMD und Intel Spezialschaltkreise in aktuelle Prozessoren, auf denen die KI-Algorithmenstromsparend laufen, ohne die Vordergrundaufgabe zu stören. Ob RyzenAI oder Intel-NPU dabei besser oder auch nur flexibler sind, lässt sich derzeit kaum sagen – im Idealfall merkt man überhaupt nicht, dass gerade die KI-Einheit schuftet. (csp@ct.de) **ct**

Software, die KI-Einheiten nutzt

Zu den wenigen Möglichkeiten, die KI-Einheiten schon heute zu nutzen, gehören die Windows Studio Effects – eine Sammlung von Filtern für Videogespräche, die etwa den Hintergrund unscharf rechnet, den Blickkontakt auf den Gesprächspartner korrigiert oder zappelige Nutzer per Auto-Framing immer in der Bildmitte hält. Das ist alles nicht neu, gehört aber jetzt zum regulären Funktionsumfang von Windows 11. Die Filter sind systemweit aktiv, sodass die energieeffiziente Nutzung der KI-Einheit nicht einzeln in Teams, Zoom, Webex & Co. eingepflegt werden muss.

Die Windows Studio Effects tauchen nur dann in den Windows-Einstellungen auf, wenn eine dedizierte KI-Einheit an Bord ist. Bestimmte Ryzen-7040-Prozessoren mit XDNA sind die ersten passenden x86-CPUs, werden aber nicht

mehr lange die einzigen bleiben: Intel will noch vor Jahresende die nächste Prozessorgeneration Meteor Lake starten, die dann eine NPU getaufte KI-Einheit enthält.

Intel hat die etwas längere Vorbereitungszeit gut genutzt und wartet neben den Studio Effects mit einigen zusätzlichen Kooperationen und Framework-Unterstützung aus eigenem Hause auf. Außer der Windows-Schnittstelle DirectML und ONNX RT, die auch AMDs RyzenAI ansprechen, gibt es bei Intel die Inference-Engine von OpenVino, einem schon bewährten API.

Darüber hinaus kooperiert Intel in Sachen KI mit Branchengrößen wie Adobe und mit Spezialanbietern wie der Streamingsoftware OBS, Xsplit, den Blender-Entwicklern oder der verbreiteten Spiele-Engine Unreal.

Es gibt **10** Arten von Menschen.
iX-Leser und die anderen.



Jetzt Mini-Abo testen:
3 Hefte + Bluetooth-Tastatur
nur 19,35 €

www.iX.de/testen



www.iX.de/testen



49 (0)541 800 09 120



leserservice@heise.de





Weg mit alten x86-Zöpfen

Das bringt Intels kommende CPU-Mikroarchitektur x86-S

Intel plant, Altlasten der x86-Architektur zu entsorgen: Künftige x86-Prozessoren werden reine 64-Bit-CPUs sein. Damit ändert sich so einiges, um Sicherheit und Zuverlässigkeit zu verbessern.

Von Andreas Stiller

Um es gleich vorweg zu sagen: Intels radikaler Verjüngungsschnitt der x86-Architektur ist überfällig. Die vielen Erb- lasten schleppten nicht nur Unmengen an Fehlern ein, sondern auch Einfallstore für Schadsoftware. Allein das 2013 endlich be- erdigte A20-Gate, mit dem man „einfach so“ physische Adres- sierungen verändern konnte, taucht in Hunderten der sogenannten „Specification Updates“ auf, die Intel seit dem berüchtigten FDIV-Bug des Pentium-Prozessors für alle CPUs herausgibt. Nun sollen viele weitere Altlasten im Orkus verschwinden. Ironie der Ge- schichte: Die ausgerechnet von Erzkonkur-

rent AMD gegen Intels IA64-Pläne einge- führte x86-64-Technik wird der Goldstan- dard bei x86, während ein großer Teil von Intels Entwicklungsarbeit in der Ablage rund landet.

Bislang ist die im April 2023 von Intel veröffentlichte „X86-S External Architec- tural Specification“ zwar nur ein Vor- schlag („Proposal“) und Intel warnt Fir- men davor, auf dieser Basis schon eigene

Entwicklungen zu beginnen. Doch es ist ziemlich klar, wohin der Weg führt. Und ARM hat es vorgemacht: Die ersten ARMv9-Designs sind längst im praktischen Einsatz und arbeiten als reine 64-Bitter.

ct Hardcore

16 Bit über Bord

Das geplante Großreinemachen bei x86 dürfte zwar relativ wenige Transistoren einsparen, aber viel Aufwand für die Validierung und den Test von Prozessorentwürfen. Das wiederum erleichtert Optimierungen. Zu den Highlights von x86-S gehört der komplette Abschied von der 16-Bit-Datenverarbeitung, sei es Real Mode, 16 Bit Protected Mode oder VM86. Auch der 32-Bit Real Mode wird beerdigt, der – ursprünglich undokumentiert – etwa unter DOS die Möglichkeit eröffnete, volle 4 GByte (2³² Byte) zu adressieren. Er wurde seinerzeit scherhaft auch „Unreal Mode“ genannt, bei c't hieß er „Big Mode“. Er war im Grunde überflüssig, weil man 32-Bit-Befehle unter DOS ohnehin über die Präfixe 0x66 und 0x67 nutzen konnte, allerdings beschränkt auf 1 MByte Adressraum.

Ganz tot ist die 32-Bit-Technik bei x86-S nicht, denn der 32-Bit Protected Mode bleibt – aber mit erheblicher Einschränkung: nur noch im User-Modus (Ring 3) als sogenannter 32-Bit Submode (Compatibility Mode). Dadurch laufen übliche 32-Bit-Programme in den meisten Fällen problemlos, wie jetzt im 64-Bit-Modus etwa unter Windows 10/11 auch. Das Booten von 32-Bit-Betriebssystemen wird bei x86-S jedoch nur noch in einer virtuellen Maschine (VM) möglich sein.

Adieu, Segmentierung!

Einige weitere Streichungen wird man kaum bemerken: Zum einen adressieren x86-S-CPUs ausschließlich ohne Segmentierung, also „flach“. Das letzte 32-Bit-Betriebssystem von Rang, das segmentierte Adressierung verwendete, war Windows 95. Die dafür nötigen Segmentregister beziehungsweise Selektoren bleiben zwar erhalten, haben bis auf zwei aber keine

x86-Prozessoren: Ausführungsmodi

| Operating Mode | ab CPU-Typ (Jahr) | bei x86-S |
|------------------------------|-------------------|-----------|
| 16-Bit Real | 8088 (1979) | – |
| 16-Bit Address | 8088 (1979) | – |
| 16-Bit Protected | 80286 (1982) | – |
| 16-Bit VM86 | 80386 (1985) | – |
| 32-Bit Real | 80386 (1985) | – |
| 32-Bit Protected | 80386 (1985) | – |
| System Management Mode (SMM) | 80386SL (1990) | ✓ |
| 32-Bit Submode (Ring 0) | Opteron (2003) | – |
| 32-Bit Submode (Ring 3) | Opteron (2003) | ✓ |
| 64-Bit Submode | Opteron (2003) | ✓ |

ct kompakt

- Künftige x86-S-Prozessoren verarbeiten keinen 16-Bit-Code mehr und booten nur noch im UEFI-Modus.
- 32-Bit-Software läuft weiterhin so, wie es jetzt schon unter 64-Bit-Betriebssystemen klappt.
- Die Vereinfachungen räumen potenzielle Sicherheitsrisiken aus dem Weg und erleichtern Veränderungen der Mikroarchitektur.

Bedeutung mehr: Ihre Basis ist jeweils null und ihr Limit „unendlich“. Ausnahmen sind die Segmentregister FS und GS, weil Betriebssysteme sie als bequeme Hilfsmittel zumeist fürs Thread-Management nutzen, nämlich als eine Art fester Pointer. Daher hat sie AMD seinerzeit auch in den x86-64 Long Mode übernommen. Dabei ist jedoch nur der Basiswert des Deskriptors von Bedeutung, das Limit ist immer 2⁶⁴. Anfangs konnte man FS und GS 64-bittig allein über Maschinenregister (MSR) setzen, später vereinfacht mit den neuen Prozessorbefehlen wrfsbase, rdfsbase, wrgsbase, rdgsbase, damit auch Prozesse mit User-Rechten die 32- oder 64-bittige Basis setzen oder auslesen können. Diese Erweiterung muss jedoch über ein Bit im Steuerregister (Control Register) CR4.16 explizit erlaubt sein. AMD zog später mit den Zen-Prozessoren nach. Als kleines Bonbon am Rande: Über FS und GS war es bei AMD auch im 32-Bit-Modus im Ring 0 möglich, auf den 64-Bit-Adressbereich zuzugreifen.

Noch weit weniger als die Segmentierung wird man die Sicherheitsringe 1 und 2 vermissen. Von den bedeutenden Betriebssystemen hatte lediglich OS/2 mehr als die zwei Ringe 0 (Supervisor) und 3 (User) genutzt.

Eine I/O Permission Map zum Freischalten von I/O-Befehlen für den User-Modus gibt es bei x86-S nicht mehr. String-I/O-Befehle (ins, outs) sind sogar komplett perdu, sie generieren nur noch einen Undefined Opcode Error. Man braucht sie heutzutage auch nicht mehr, weil Streaming I/O üblicherweise über die I/O Memory Management Unit (IOMMU) und Direct Memory Access (DMA) erfolgt. Ansonsten muss man I/O-Befehle eben (auf Ring 0) in einer Schleife abarbeiten.

Als Interrupt-Controller unterstützt x86-S nur noch X2APIC, den Intel 2011 mit Sandy Bridge (Core i-2000) einführt und den Programmierer mit MSR-Befehlen konfigurieren. Der Vorgänger XAPIC (seit Pentium 4) verwendet weit aus schnelleres Memory Mapped IO (MMIO). Der X2APIC ist jedoch weit zukunftsrichtiger für bis zu 2³² Prozessoren beziehungsweise Kerne ausgelegt, während sich XAPIC auf 256 beschränkt. Um die Sache für ältere Software einfacher zu machen, kann X2APIC bei Bedarf und bei „Kleinsystemen“ von bis zu 256 Kernen den alten XAPIC via VMX emulieren und das soll auch bei x86-S so bleiben. Ein Interrupt „Delivery Mode“ namens Extint, der noch zusätzliche externe Controller benötigte, wurde indes gestrichen.

Flexible Seitentabellen

Eine x86-S-CPU läuft grundsätzlich mit Paging und wandelt dabei alle logischen Speicheradressen mit Übersetzungstabellen und dem Translation Lookaside Buffer (TLB) in physische (RAM-)Adressen um. Apropos „logische Adressen“: Der Begriff hat sich zwar eingebürgert, aber es wäre eigentlich genauer, virtuellen Adressen zu sprechen. Letztlich teilt dabei das Betriebssystem einem Programm beziehungsweise einem Thread einen bestimmten zusammenhängenden Adressbereich zu. Der verteilt sich aber möglicherweise auf unterschiedliche, nicht zusammenhängende Speicheradressen der eingebauten RAM-Chips; ja, er muss nicht einmal physisch vorhanden sein. Im Hintergrund kümmert sich das Betriebssystem im Zusammenspiel mit dem BIOS und einer Fülle an Hardwarefunktionen darum, diesen virtuellen Adressbereich auf physisches RAM abzubilden, setzt dabei Zugriffsrechte durch und steuert, wie die CPU-Caches Zugriffe auf diesen RAM-Bereich puffern.

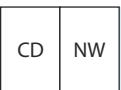
Control Register CR0 bei x86-Prozessoren

| Bit | Label | Beschreibung | Inhalt bei x86-S |
|-----|-------|-----------------------|------------------|
| 0 | PE | Protected Mode Enable | 1 |
| 1 | MP | Monitor Co-Processor | 1 |
| 2 | EM | x87 FPU Emulation | 0 |
| 3 | TS | Task Switched | variabel |
| 4 | ET | Extension Type | 1 |
| 5 | NE | Numeric Error | 1 |
| 16 | WP | Write Protect | variabel |
| 18 | AM | Alignment Mask | [Maske] |
| 29 | NW | Not Write-Through | 0 |
| 30 | CD | Cache Disable | variabel |
| 31 | PG | Paging | 1 |

Cache-Attribute für Speicherbereiche

Steuerregister in x86-Prozessoren legen für jede logische Speicheradresse beziehungsweise Page (Speicherseite) fest, ob und wie die CPU-Caches die dort abgelegten Daten puffern sollen. Dazu dienen die Register CR0, CR3, CR4 sowie die Page Attribute Table (PAT). Alternativ gibt es 256 Memory Type Range Registers (MTRR).

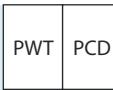
Register CR0



CR0-Bits 29 & 30 steuern
Caching des gesamten RAM
CD: Cache Disable / NW: Not Write-Through

| CD | NW | Bedeutung |
|----|----|--|
| 0 | 0 | Normalbetrieb: Write Through (Shared Cache Lines), Write Back (andere) |
| 0 | 1 | ungültig |
| 1 | 0 | Non Fill Cache Mode (speicherkohärent) |
| 1 | 1 | nach Init: Non-Evict Mode (inkohärent) (nicht bei x86-S) |

Register CR3



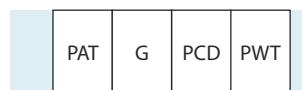
CR3-Bits 3 & 4 Page Cache Disable und
Page Write Through (PCD/PWT)
steuern Caching des page Directory

Register CR4



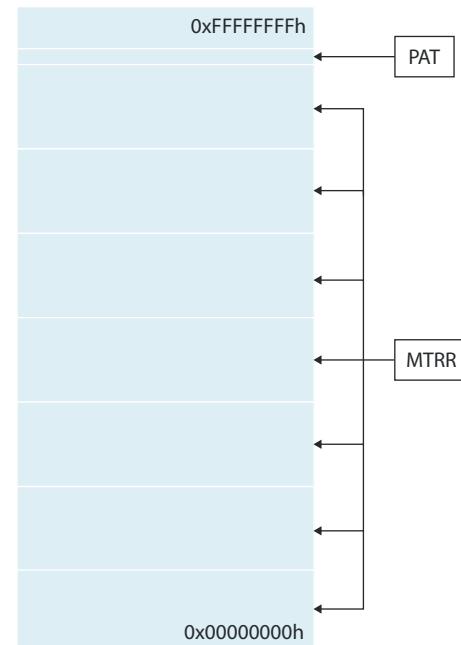
CR4-Bit 7 Paging Global Extensions (PGE)
schaltet Global Pages mit G-Flag frei

Page Directory oder Page Table Entry (PTE)



PCD/PWT steuern
Caching jeder Page
G-Flag (Global) steuert
TLB-Flushing auf Page-Ebene

Physische RAM-Adressen



Die komplexen Hardwarefunktionen moderner Prozessoren zur Umrechnung der Adressbereiche virtueller Maschinen machen effiziente Virtualisierung überhaupt erst möglich. Damit schützen die Prozessoren aber auch einzelne VM-Adressbereiche voneinander ab, um Malware-Angriffe zu erschweren, bis hin zu separater Verschlüsselung der Speicherbereiche parallel laufender VMs.

Wegen der immer weiter wachsenden virtuellen Adressbereiche der Prozessoren musste im Laufe der Zeit auch die Zahl der hierarchisch aufgebauten Page Level erhöht werden. Ab 57 Bit logisch braucht man derer fünf und daraus folgt, dass die Speicherverwaltung bei jedem Zugriff alle fünf Level abklappern müsste (Page Walk), bis sie schlussendlich den Seiteneintrag mit der physischen RAM-Adresse und den Zugriffseigenschaften auslesen kann. Das würde für Maschinenverhältnisse quasi unendlich lange dauern, aber zum Glück gibt es in den Prozessoren einen Translation Lookaside Buffer (TLB), der für die jeweiligen Seiten die Übersetzung zwis-

senschspeichert, sodass ein Page Walk nur nötig wird, wenn nichts im Puffer steht.

Für Prozessoren, die virtuelle Adressen mit 57 Bit Länge und mehr verarbeiten (was das Konfigurationsregister CPUID(7) ECX.16 = 1 ausweist), soll x86-S einen neuen „4L/5L Page Switch“ anbieten, den man via MSR verwaltet. Damit ließe sich auf Thread-Basis entscheiden, ob sie Seitentabellen mit vier oder fünf Levels nutzt. Schreibt der Thread eine Basisadresse für die Seitentabellen ins MSR CR3_LOAD_4LPGTBL, wird CR3 mit der Adresse geladen und das Bit CR4.5L gelöscht. Ein Schreibzugriff auf MSR CR3_LOAD_5LPGTBL setzt es. Intel bleibt diesbezüglich in der Dokumentation aber sehr vage und erwähnt etwa CR4.5L. Ob das ein neues Bit im CR4-Register ist oder nur eine andere Bezeichnung für CR4.LA57 (Bit 12) wie im aktuellen Intel-Software Developer's Manual, weiß man nicht.

Eine wichtige Neuerung von x86-S betrifft den Bootvorgang. Bislang bootete ein System grundsätzlich zunächst im 16-Bit

Real Mode, schaltet dann in den 32-Bit Real Mode und je nach (UEFI-)BIOS dann in den 64-Bit-Modus. Ein x86-S-Prozessor startet gleich im 64-bittigen Long Mode, den Intel 2004 von AMD übernommen hat. Beim Init ist es die Firmware Interface Table, die die vom BIOS vorgesetzten Werte der zum Start im 64-Bit Page Mode nötigen Steuerregister CR0, CR3 und CR4 enthält. Speicher steht dem Bootloader aber schon vor der RAM-Initialisierung zur Verfügung, weil er den CPU-Cache zum RAM umdefinieren kann. Dazu dient das (undokumentierte) MSR 0x2E0 (Cache-as-RAM), Intel nennt es Non Evict Mode (NEM). Das wird wohl auch x86-S beibehalten. UEFI ist Pflicht, vermutlich in 64 Bit, aber dazu hat sich Intel noch nicht ausgelassen.

Virtualisierung

Ein Memory Type Range Register (MTRR) legt für einen darin beschriebenen Bereich physischer Speicheradressen fest, ob der Cache Zugriffe darauf puffern darf, ob der Cache dabei mit Write Back oder Write Thru arbeiten soll, ob Schreibzugriffe ge-

bündelt werden dürfen (Write Combining) oder ob sie ganz verboten sind. Allerdings sind MTRR für diese Attribute eigentlich schon seit dem Pentium III überflüssig, weil man sie seither mit einer Erweiterung für die Seitentabellen namens Page Attribute Table (PAT) für jede Speicherseite (Page) festlegen kann. Nun konkurrieren beide Methoden und es gibt bestimmte Regeln, wer wann die Priorität hat.

Fixed MTRRs sind noch obsoletter, weil sie sich auf den DOS-Adressbereich von 640 bis 960 KByte beziehen, wo in Urzeiten ISA-Karten ihren Speicher einblendeten.

Für die Virtualisierung müssen die Programmierer von VMware, Microsoft und Co. je nach unterstütztem Gastsystem für x86-S mehr oder weniger hohen Aufwand treiben, zum Teil können sie wohl auch auf alte Vollemulation zurückgreifen. Mit 64-Bit-Gästen, die an Long Mode gewöhnt sind, sollten sich nur wenig Probleme ergeben. Lediglich ein paar Kleinigkeiten sind zu berücksichtigen, etwa solche, die sich durch das Booten in 64-Bit ergeben. Hier ist das andere Nachrichtenformat bei der Initialisierung „SIP16“ zu berücksichtigen.

Halbwegs bequem gehts auch mit 32-Bit-Gästen, die mit Flatmode und mit Ring 0 und 3 arbeiten, so wie Windows ab XP oder 32-Bit-Linux. Der 32-Bit-Submodus hilft bei der Virtualisierung erheblich, allerdings erfordert die fehlende Unterstützung von Ring 0 dann doch etwas Emulationsaufwand mit Schatten-Seitentabellen (Shadow Page Tables). Die meisten Betriebssysteme können zumindest mit dem Interrupt-Controller XAPIC umgehen. Sollte das Gastsystem jedoch noch auf den Uralt-Interruptcontroller 8259 zurückgreifen, der über I/O-Ports angesprochen wird, dann gibts General Protection Faults GP(0), die der Hypervisor abfangen und behandeln kann. Das gilt auch für alle I/O-Zugriffe im User-Modus, die x86-S ebenfalls nicht mehr erlaubt. Der Hypervisor kann sie abfangen und den Zugriff dann in seinem 64-Bit-Ring 0 durchführen. Gleiches gilt für die nicht mehr unterstützten String-I/O-Befehle.

32-Bit-Befehle mit Präfix, die eine 16-Bit-Adressierung ergeben würden, werden ebenfalls mit GP(0) oder bei Sprungbefehlen mit „unbekannter Befehl“ bestraft. Auch das kann der Hypervisor einfach behandeln.

Schwieriger wirds bei Gästen, die à la OS/2 mit zusätzlichen Ringen arbeiten. Intel unterscheidet zwischen Gästen, die

mehr oder weniger Emulation erfordern, und solchen „uncommon cases“, die Voll-emulation erfordern; zu letzteren dürfte DOS gehören. Und schließlich gibt es auch solche, die gar nicht virtualisiert werden können. Die für alte DOS-Spiele so beliebte DOS-Box dürfte indes kaum Probleme bereiten, weil sie schon seit einiger Zeit in einer 64-Bit-Version im Long Mode läuft.

Konfigurationsregister

Die neuen Funktionen und insbesondere die Streichungen spiegeln sich in den diversen Konfigurationsregistern der CPU wider. Das fängt mit CPUID an. Ein Feature-Bit namens LEGACY_REDUCED_OS_ISA soll das Vorhandensein von x86-S anzeigen, ein weiteres die Verwendung von SIP16 beim Init. Auch für das neue 4L/5L Page Switching ist hier zukünftig ein Bit vorgesehen.

Im CRO-Register sind derzeit insgesamt elf Konfigurationsbits definiert, von denen unter x86-S aber sieben festgelegt sind: Protected Mode (PE) und Paging (PG) sind grundsätzlich eingeschaltet, Not Write Through (NW) und FPU-Emulation (EM) ausgeschaltet. Zusammen mit CRO.30 Cache Disable (CD) ergibt sich als einzige sinnvolle Konfiguration im Betrieb daraus das normale Caching mit Write Back und der Cache lässt sich während der Initialisierung abschalten. Die Caching-Strategie hängt ansonsten von der jeweils angesprochenen Page ab und wird wie erwähnt von PAT oder MTRR gesteuert.

Das CR4-Register legt die Abkehr von VM86 Mode (VME), Protected Mode Interrupts (PVI) und die Größe der Seiten-einträge auf 8 Bytes (PAE) fest. Im erweiterten Feature-Register EFER sind Long Mode (LMA, LME), Syscall (SCE) und das No Execution Bit (NE) dauerhaft gesetzt und im 64-bittigen Flags-Register (RFlags) die Bits für IOPL, VM, VIF und VIP fix auf null gesetzt.

Fazit

Von den Neuerungen durch x86-S wird der überwiegende Teil der PC-Nutzer nichts bemerken. Darauf achten müssen außer Programmierern systemnaher Software nur Leute, die noch sehr alte 32-Bit-Betriebssysteme installieren möchten oder ihre Rechner partout im BIOS-Modus starten wollen.

Es ist schon lustig, dass mit Pat Gelsinger einer der wichtigsten Architekten, der vor fast 40 Jahren für das Design der erfolgreichen 80386-CPU verantwortlich zeichnete, nun als Intel-Chef bereit ist, die

Abschaffung vieler obsoletter Funktionen einzuleiten.

Es ist absehbar, dass x86-S zunächst nur in wenigen Prozessoren auf den Markt kommen wird – etwa für Server –, während andere CPU-Typen weiterhin als „normale“ Prozessoren auch Uralt-Unterstützung bieten. Wenn das erfolgreich ist, werden sich die Marktanteile schnell verschieben. Wann AMD folgt, muss man sehen. (ciw@ct.de) **ct**

Intel-Whitepaper zu x86-S: ct.de/yf9q

Control Register CR4 bei x86-Prozessoren

| Bit | Label | Beschreibung | Inhalt bei x86-S |
|-------|----------|---|------------------|
| 0 | VME | Virtual 8086 Mode Extensions | 0 |
| 1 | PVI | Protected-Mode Virtual Interrupts | 1 |
| 2 | TSD | Time Stamp Disable | variabel |
| 3 | DE | Debugging Extensions | variabel |
| 4 | PSE | Page Size Extensions | variabel |
| 5 | PAE | Physical Address Extensions | 1 |
| 6 | MCE | Machine Check Exception | variabel |
| 7 | PGE | Page Global Enabled | variabel |
| 8 | PCE | Performance Monitoring Counter Enabled | variabel |
| 9 | OSFXSR | Betriebssystem (OS) unterstützt FXSAVE und FXRSTOR | variabel |
| 10 | OSXME | OS unterstützt Unmasked SIMD Floating Point Exceptions | variabel |
| 11 | UMIP | User Mode Instruction Prevention | variabel |
| 12 | LA57 | 57 Linear Addresses, 5 Page level | variabel |
| 13 | VMXE | Virtual Machine Extensions Enable | variabel |
| 14 | SMXE | Safer Mode Extensions Enable | variabel |
| 15 | SEE | SGX Enclave Instructions Enable | variabel |
| 16 | FSGSBASE | Freischalten der Instruktionen RDGSBASE, RDGSBASE, WRGSBASE, WRGSBASE | variabel |
| 17 | PCIDE | PCID Enable | variabel |
| 18 | OSXSAVE | XSAVE und Processor Extended States Enable | variabel |
| 19 | KL | Key Locker Available | variabel |
| 20 | SMEP | Supervisor Mode Execution Protection Enable | variabel |
| 21 | SMAP | Supervisor Mode Access Prevention Enable | variabel |
| 22 | PKE | Protection Key Enable | variabel |
| 23 | CET | Control Flow Enforcement Technology Enable | variabel |
| 24 | PKS | Protection Keys for Supervisor-Mode Pages Enable | variabel |
| 25 | UINTR | User Interrupts Enable | variabel |
| 26-63 | RES | Reserviert | |

Weg mit der CPU!

Rechenbeschleuniger mit eingebauten CPU-Kernen für Supercomputer

In vielen (KI-)Supercomputern leisten Rechenbeschleuniger den Löwenanteil der Arbeit. Für die wenigen untergeordneten Aufgaben, für die noch herkömmliche Prozessoren nötig sind, reicht eine Handvoll Kerne. Separate CPUs fallen daher in solchen Systemen immer häufiger weg.

Von Christof Windeck

Zurzeit kommen die ersten Server-APUs auf den Markt. Diese Kombiprozessoren vereinen im gleichen Gehäuse Rechenbeschleuniger sowie Allzweckprozessoren mit ARM- oder x86-Befehlsatz. Den Anfang machen AMDs Instinct MI300A und Nvidias GH200 „Grace Hopper“. Intel wird wohl folgen und plante das ursprünglich 2025 für „Falcon Shores“, nun vielleicht erst später. Solche Server-APUs versprechen für leistungsstarke Supercomputer erhebliche Vorteile bei Kosten, Energieeffizienz und Platzbedarf.

Die Abkürzung „APU“ für Accelerated Processing Units prägte AMD vor rund zwölf Jahren und sprach seinerzeit gerne und viel auch von der Heterogeneous System Architecture (HSA). Gemeint war, dass die bei Mobilprozessoren schon lange üblichen integrierten Grafikprozessoren nicht mehr als bloße 3D-Beschleuniger und Videodecoder dienen, sondern auch Allzweckrechenleistung beisteuern. Dann rechnen sowohl die x86- als auch die GPU-Einheiten gemeinsam, bilden also ein heterogenes System.

Das klappt tatsächlich; über Programmierschnittstellen wie OpenCL und CUDA lässt sich praktisch jede aktuelle GPU – ob in der CPU integriert oder nicht – zum Rechnen einspannen. Doch grösse-

re Vorteile ergeben sich erst, wenn es um eine wirklich starke GPU geht, etwa um einen KI-Rechenbeschleuniger mit mehreren Teraflops (TFlops) Leistung. Die gibt es bisher für PCs und Notebooks nicht – mit einer prominenten Ausnahme: Apples M-Prozessoren. Denn während x86-APUs schlichtweg zu schwache GPU-Teile haben, gönnt Apple der GPU sehr viel Siliziumfläche und bindet extrem schnellen (LPDDR5X)-Speicher an.

Fast wie bei Apple

Apple hebt noch einen weiteren grundsätzlichen Vorteil des Konzepts hervor und nennt das Unified Memory Architecture (UMA). Damit ist gemeint, dass sowohl die CPU- als auch die GPU-Kerne sowie eventuell andere eingebaute Beschleuniger auf den gemeinsamen Arbeitsspeicher in einem gemeinsamen Adressraum zugrei-

fen. Aufwendiges Umkopieren von Daten etwa aus dem CPU-Speicher in den lokalen Speicher einer Grafikkarte entfällt. Das spart Zeit und Energie für Datentransfers etwa per PCI Express (PCIe).

Die UMA-Idee ist alles andere als neu und bei den bereits erwähnten Mobilprozessoren von AMD und Intel sowie auch bei Smartphone-SoCs seit Langem üblich. Doch erst mit superschnellem RAM bringt UMA größere Vorteile. Apple bindet dazu zahlreiche Low-Power-DDR5-SDRAM-Chips (LPDDR5X) parallel an einen M-Chip an. Damit geht ein Nachteil einher: Genau wie beim verlötzten, schnellen GDDR-Speicher auf Grafikkarten lässt sich bei M-Chips der Speicher nicht nachträglich erweitern. Flexibel erweiterbarer Hauptspeicher etwa aus gängigen DIMMs bietet zwar viel höhere Kapazitäten, bei Servern derzeit etwa bis zu 3 TByte pro CPU. Doch selbst die 12 RAM-Kanäle eines AMD Epyc 9004 bringen es zusammen auf höchstens 460 GByte/s, während Apples M2 Ultra bis zu 800 GByte/s schafft. Highend-Grafikkarten wie die GeForce RTX 4090 mit GDDR6X-RAM erreichen rund 1 TByte/s und KI-Rechenbeschleuniger wie der Nvidia H100 Hopper dank High Bandwidth Memory (HBM3) mehr als 3 TByte/s.

AMD setzt den Instinct MI300A aus zahlreichen Chiplets und HBM3-Spei-

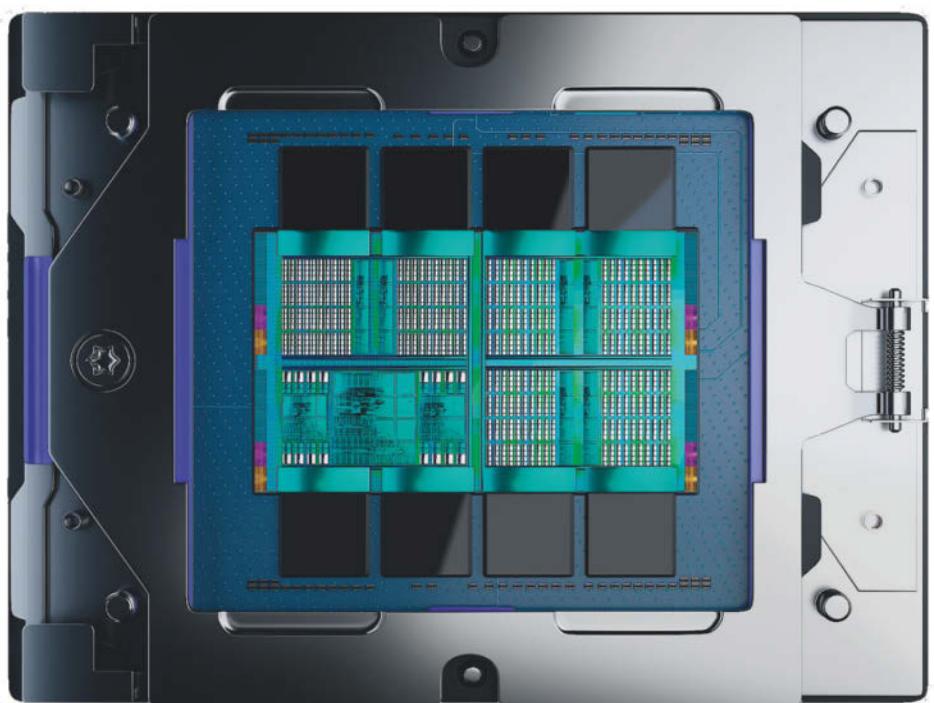


Bild: Lawrence Livermore National Laboratory/ISC

Der AMD Instinct MI300A hat nicht nur mehrere Hundert schnelle Recheneinheiten, sondern auch 24 x86-Kerne vom Typ Zen 4. Man sieht sie links unten im bläulichen Teil des Chips.

cherstapeln zusammen. Die 24 CPU-Kerne und mehrere Hundert GPU Compute Units greifen gemeinsam auf 128 GByte Speicher zu. Die 24 Zen-4-Kerne entsprechen grob einem Epyc 9224.

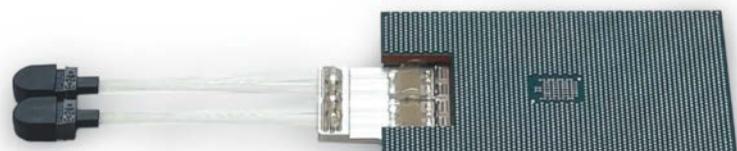
Etwas anders baut Nvidia den GH200 auf. Der Hopper-Teil – also der Rechenbeschleuniger – verfügt über 96 GByte HBM3 und kommuniziert über seine integrierte NVLink-C2C-Schnittstelle mit dem Grace-Chip mit 72 ARM-Kernen, der wiederum 480 GByte LPDDR5X-RAM anbindet. Dieser Speicher ist zwar auch schnell, aber langsamer als HBM3. Beim GH200 geht es also nicht um UMA. Doch über ein zweites schnelles NVLink-4-Interface mit 900 GByte/s (450 GByte/s pro Richtung) lässt sich der Hopper-Teil direkt mit bis zu 255 anderen GH200-Knoten koppeln. Dadurch entsteht ein Cluster mit mehr als 24 TByte superschnellem RAM, in das die Daten auch sehr großer KI-Modelle passen. Für 2024 hat Nvidia eine neue Version des GH200 mit 141 GByte noch schnellerem HBM3E angekündigt. Mit den selbst entwickelten Grace-Prozessoren mit ARM-Kernen emanzipiert sich Nvidia zudem von den x86-CPU-Firmen AMD und Intel.

Welche CPU-GPU-Mischung für Falcon Shores geplant ist, verrät Intel bisher nicht. Beim aktuellen Xeon GPU Datacenter Max alias Ponte Vecchio übt Intel bereits, wie man Kombiprozessoren aus mehr als 45 Chiplets zusammensetzt – ebenfalls inklusive HBM. Bei Falcon Shores werden dann eben auch einige CPU-Tiles eingebaut. Als schnelle Schnittstelle verwendet Intel bisher den Cache-kohärenten Compute Express Link (CXL), der auf PCI Express aufsetzt. PCIe 5.0 x16 schafft 64 GByte/s pro Richtung, also nur ein Siebtel von NVLink 4.

Intel verkaufte vor einigen Jahren allerdings bereits Xeons mit integrierten Controllern für die Schnittstelle Infini-band. Für die Zukunft plant Intel, Controller für optische (Glasfaser-)Interconnects direkt in Prozessoren zu integrieren. Solche Anläufe gab es allerdings auch schon in der Vergangenheit, ohne dass sie sich bisher durchgesetzt haben.

APU-Vorteile

Die integrierten CPU-Kerne der Server-APUs bieten viel weniger Rechenleistung als der jeweilige GPU-Teil. Doch die integrierte CPU ermöglicht den Aufbau von Clusterknoten, die keine weiteren x86- oder ARM-Prozessoren benötigen. Denn



Gemeinsam mit Intel entwickelt Ayar Labs einen FPGA-Chip mit direkt integriertem optischem Transceiver, der 4 Tbit/s überträgt.

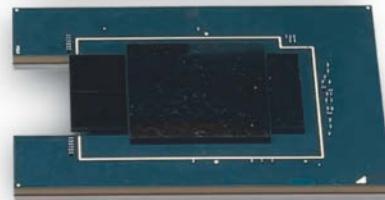


Bild: Ayar Labs

eine Server-APU lädt selbst ein (UEFI)-BIOS, startet das Betriebssystem und stellt PCIe-Lanes oder andere schnelle Schnittstellen bereit. Dazu kommt der Vorteil des gemeinsamen Adressraums (UMA).

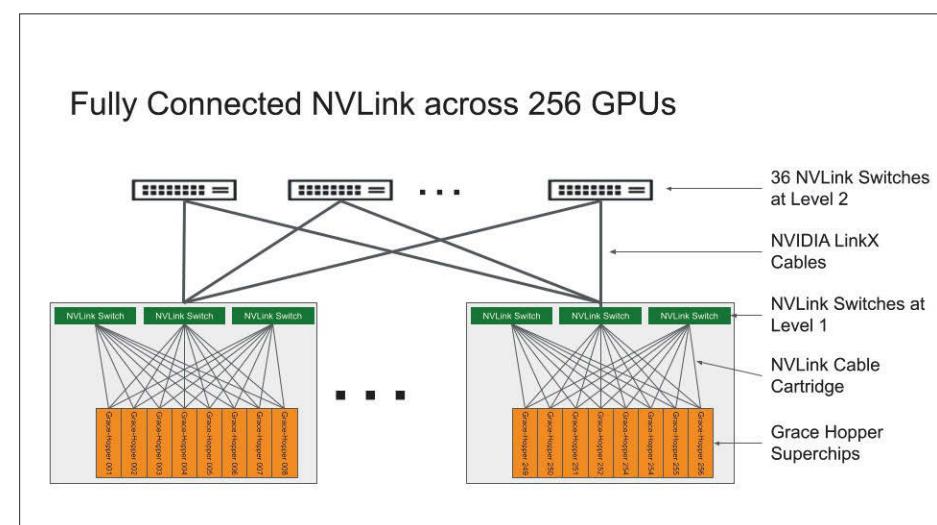
Das Weglassen separater Allzweckprozessoren spart zudem Kosten, Platz und Energie. Denn schnelle Datentransfers fressen Strom – und zwar tendenziell desto mehr, je länger die Leitungen sind. Auch jede zusätzliche Umsetzung von Signalen braucht Energie. Daher baut Nvidia NVLink direkt in den GH200 ein. Alternativ kann man andere Hostadapter per PCIe an die APU anschließen. So macht es etwa die HPE-Sparte Cray, die ihren eigenen Slingshot-Interconnect für die Dragonfly-Topologie verwendet. Sie dient etwa zur leistungsfähigen Verbindung unterschiedlicher Partitionen von Supercomputern, etwa von Teilen mit und ohne Rechenbeschleuniger.

Derzeit werden einige Supercomputer aus Server-APUs aufgebaut, etwa mit dem AMD MI300A das 2-Exaflops-System „El Capitan“ am Lawrence-Livermore-Nationallabor der USA. Nvidias GH200 zielt eher auf kommerziell genutzte KI-Superrechner, laut Nvidia investieren unter anderem Google, Meta und Microsoft in „Superpods“ vom Typ DGX GH200.

Fazit

Rechenbeschleuniger mit integrierten CPU-Kernen sind keine Allzwecklösung für Server, sondern zielen auf Supercomputer und Rechencluster für KI und High-Performance Computing (HPC). Die HPC-APUs werden gewöhnliche Serverprozessoren auf absehbare Zeit nicht verdrängen, sondern dort ergänzen, wo sie ihre speziellen Vorteile ausspielen.

(ciw@ct.de) ct





Ökochips

Wie die hohen Umweltlasten der Halbleiterfertigung sinken sollen

Chiphersteller zeigen gerne ihre blitzblanken Reinräume, doch die Produktion von Halbleitern frisst Unmengen an Energie und Ressourcen und erzeugt klimaschädliche Abgase. Das soll sich ändern.

Von Christof Windeck

Halbleiterbauelemente sind meistens winzig und bestehen im Wesentlichen aus einem Stückchen Silizium. Damit ähneln sie im Prinzip einer kleinen Glasscherbe. Doch Chips haben ein schmutziges Geheimnis: Ihre Fertigung verursacht eine vergleichsweise riesige Umweltbelastung. Und die summiert sich, weil pro Jahr weltweit mehr als eine Billion Bauelemente produziert werden, von der winzigen LED bis hin zum Frisbee-großen KI-Superchip. Der Chip-Ausstoß soll zudem deutlich wachsen, mehrere Staaten fördern den Aufbau neuer Chipwerke (Fabs) mit Milliardensummen. Grundsätzlich sind sich Chiphersteller, Behörden und Politiker daher einig, dass die Halbleiterfertigung ökologischer werden muss. Dabei lassen sich manche Probleme relativ leicht lösen und für andere sind immerhin Lösungsmöglichkeiten bekannt. Doch es gibt auch einige harte Brocken.

Chip-Aufbau

Um zu verstehen, welche Teilschritte der Chipfertigung besonders hohe Umweltlasten verursachen, hilft ein kurzer Blick auf die Technik, die wir unter ct.de/yzzz ausführlich erklärt haben. Als Beispiel dient die sogenannte CMOS-Technik, mit der Firmen wie Intel, Samsung und der weltgrößte Chip-Auftragsfertiger TSMC Halbleiter produzieren. Die Abkürzung CMOS steht für Complimentary Metal-Oxide Semiconductor und bezeichnet eine Kombination aus Feldeffekttransistoren (FETs), die die Basis der Logikschaltungen in den meisten Prozessoren, Grafikprozessoren und Mikrocontrollern bilden. Die CMOS-Fertigung erzeugt feinste Strukturen auf Scheiben (Wafern) aus hochreinen Silizium-Einkristallen, wobei viele Schichten übereinander gestapelt werden. Die Strukturierung jeder einzelnen Schicht erfolgt durch fotolithografische Verfahren: Man über-

zieht den Wafer mit lichtempfindlichem (Foto-)Lack, auf den anschließend ein Laser durch eine Maske die jeweils gewünschten Strukturen projiziert. Dann wird in einem chemischen Verfahren der Fotolack entwickelt, woraufhin sich die belichteten Stellen mit Lösemitteln beseitigen lassen. Der nun folgende Ätzprozess entfernt das nicht mehr von Lack geschützte Material. Es folgt ein Reinigungsschritt mit Reinstwasser (Ultrapure Water, UPW).

Komplexe moderne Chips mit mehreren Milliarden Transistoren und vielen Schichten durchlaufen das beschriebene Strukturierungsverfahren Dutzende Male. Um eine weitere Schicht über den bereits vorhandenen zu erzeugen, bringt man hauchdünne Lagen aus neuem Material auf. Das geschieht beispielsweise durch Abscheidung aus der Gasphase (Chemical Vapor Deposition, CVD). Metallschichten für leitende Verbindungen lassen sich durch sogenanntes Sputtern erzeugen. Dazu kommt der Wafer zusammen mit einem Block des gewünschten Metalls in eine Vakuumkammer. Der Beschuss mit Ionen zerstäubt (to sputter = zerstäuben) die Oberfläche des Metallkörpers und durch passende elektrische Ladung wandern die Metallteilchen auf die Oberfläche des Wafers.

Viele Schichten brauchen weitere Bearbeitungsschritte, manche müssen beispielsweise erhitzt werden (Annealing), benötigen eine (chemische) Politur oder eine Dotierung, die die elektrischen Eigenschaften des halbleitenden kristallinen Materials beeinflusst. Dazu bringt man beispielsweise Bor- oder Phosphorionen in die Schicht ein, etwa mittels Diffusion oder Ionenimplantation, wiederum im Vakuum.

Energiefresser

Die Beschreibung der wesentlichen Fertigungsschritte lässt schon erahnen, dass sie enormen Aufwand benötigen. Vor allem der Energiebedarf ist hoch. Bleiben wir zunächst bei der eigentlichen Chip-Fab. Die Bearbeitung der Chips erfolgt in Reinräumen, deren Raumluft keine Staubpartikel enthalten darf. Dazu sind gewaltige, sehr stromhungrige Belüftungs- und Filtersysteme nötig. Viele Fertigungsschritte finden im Vakuum statt: Jedes Mal, nachdem neue Wafer etwa in eine Sputter-Maschine geladen wurden, muss diese von Vakuumpumpen erneut entlüftet werden. Auch diese Vakuumpumpen nehmen viel

Leistung auf. Ebenso die leistungsstarken Laser, die das für die Fotolithografie nötige Licht erzeugen.

Große Chip-Fabs benötigen Leistung in der Größenordnung von 100 Megawatt (100 MW beziehungsweise 100.000 Kilowatt/kW). Branchenexperten schätzen, dass moderne Fertigungsanlagen rund 1,1 bis 1,5 Kilowattstunden (kWh) Energie pro Quadratzentimeter Chipfläche umsetzen. Für einen einzigen 30-Zentimeter-Wafer mit rund 700 Quadratzentimetern Fläche sind das also 0,7 bis 1,05 MWh, und große Chip-Fabs verarbeiten über 300.000 Wafer pro Jahr – dann stehen 315 Gigawattstunden auf dem Zähler.

Manche Fabs betreiben eigene Gaskraftwerke, um Kosten zu sparen. Aus Gas selbst erzeugter Strom war in Deutschland bis zum Ukrainekrieg deutlich billiger als Netzstrom. Die Eigenproduktion hat auch den Vorteil, von externen Störungen unabhängig zu sein. Doch bei der Stromerzeugung aus Erdgas werden pro kWh rund 350 Gramm CO₂ frei.

Den Strom für Chip-Fabs aus Wind-, Sonnen- oder Wasserkraftwerken zu beziehen senkt die CO₂-Bilanz der Chipfertigung deutlich. Intel nannte als einen der Gründe für die Ansiedlung der neuen

c't kompakt

- Die Produktion von Halbleitern benötigt viele Ressourcen und stößt große Mengen an Treibhausgasen aus.
- Alle großen Chiphersteller wollen ihre Ökobilanz verbessern, zum Beispiel durch Windenergie.
- In den nächsten Jahren wächst die Fertigungskapazität für Chips jedoch rasant.

Chip-Fab in Magdeburg die große Kapazität von Windenergie in Sachsen-Anhalt. Schon jetzt könnte das Bundesland im Prinzip genügend grünen Strom für den Betrieb der Fab bereitstellen.

Vorprodukte

Der Stromverbrauch der Fabs ist jedoch nur ein Teil der Energiemenge, die in Chips steckt. Denn für die Produktion sind Hunderte verschiedener Materialien und Chemikalien nötig, viele davon in hochreiner Qualität. Die Herstellung solcher

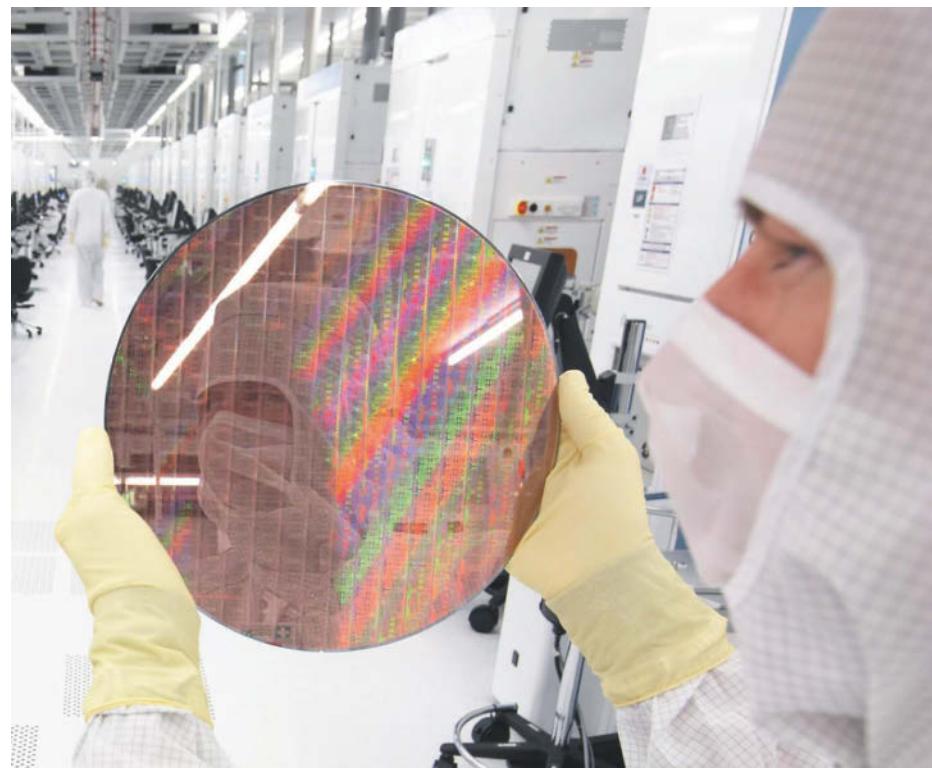


Bild: GlobalFoundries

Halbleiterchips entstehen auf Wafern, die aus Silizium-Einkristallen bestehen. Auf einen 300-Millimeter-Wafer passen je nach Fläche der einzelnen Chips wenige Dutzend bis mehrere Zehntausend Stück.

Substanzen verschlingt wiederum gigantische Mengen Energie.

Schauen wir als Beispiel auf die Produktion der Wafer, die Zulieferer aus großen Einkristallen aus hochreinem Silizium schneiden. Diese Kristalle entstehen zwar aus dem billigen Rohstoff Quarzsand (im Wesentlichen Siliziumdioxid, SiO_2), aber in einem extrem aufwendigen Fertigungsprozess. Schon das Baggern des Sands kostet Energie und zerstört die Landschaft. Dazu kommt der Transport ins Schmelzwerk. Dort wird der Sand in elektrisch beheizten Öfen mit Zusatzstoffen geschmolzen, um Rohsilizium (Si) zu erzeugen. Das erfordert pro Kilogramm Silizium rund 15.000 kWh Strom.

Und das ist erst der vergleichsweise sparsame Anfang der Verarbeitungskette. Wafer für Halbleiterbauelemente beste-

hen aus zu 99,9999999 Prozent reinem Material (9N-Reinheit). Um die Fremdstoffe zu entfernen, lässt man das Silizium mit Chlorwasserstoff (HCl) zu flüssigem Trichlorsilan reagieren und destilliert es anschließend. Aus der hochreinen Flüssigkeit erzeugt man dann wieder festes Silizium. Letzteres wird danach auf über 1400 Grad Celsius erhitzt, um aus der Schmelze einen riesigen Einkristall zu züchten, was Wochen dauern kann. Anschließend zerstellt ihn eine Säge mit feinen diamantbeschichteten Drähten in Scheiben – die Wafer –, die danach noch einzeln poliert werden müssen. Ältere Schätzungen gehen davon aus, dass in jedem fertigen Wafer schließlich rund 2000 kWh Energie stecken – ähnlich viel, wie ein deutscher Einpersonenhaushalt im Mittel pro Jahr an Strom verbraucht.

Chemiebomben

Bei der Chipfertigung kommen noch Hunderte andere hochreine Materialien zum Einsatz, die jeweils ähnlich aufwendig erzeugt werden müssen wie die Wafer. Die Chipfertigung nutzt mehr als die Hälfte aller chemischen Elemente, die das Periodensystem verzeichnet. Zulieferer wie Applied Materials erzeugen chemische Verbindungen mit besonderen Eigenschaften, um Halbleiter mit neuen Funktionen herstellen zu können. Ein Beispiel sind Materialien, die sich je nach Orientierung ihrer Kristallstruktur unterschiedlich stark ätzen lassen (anisotropes Ätzen).

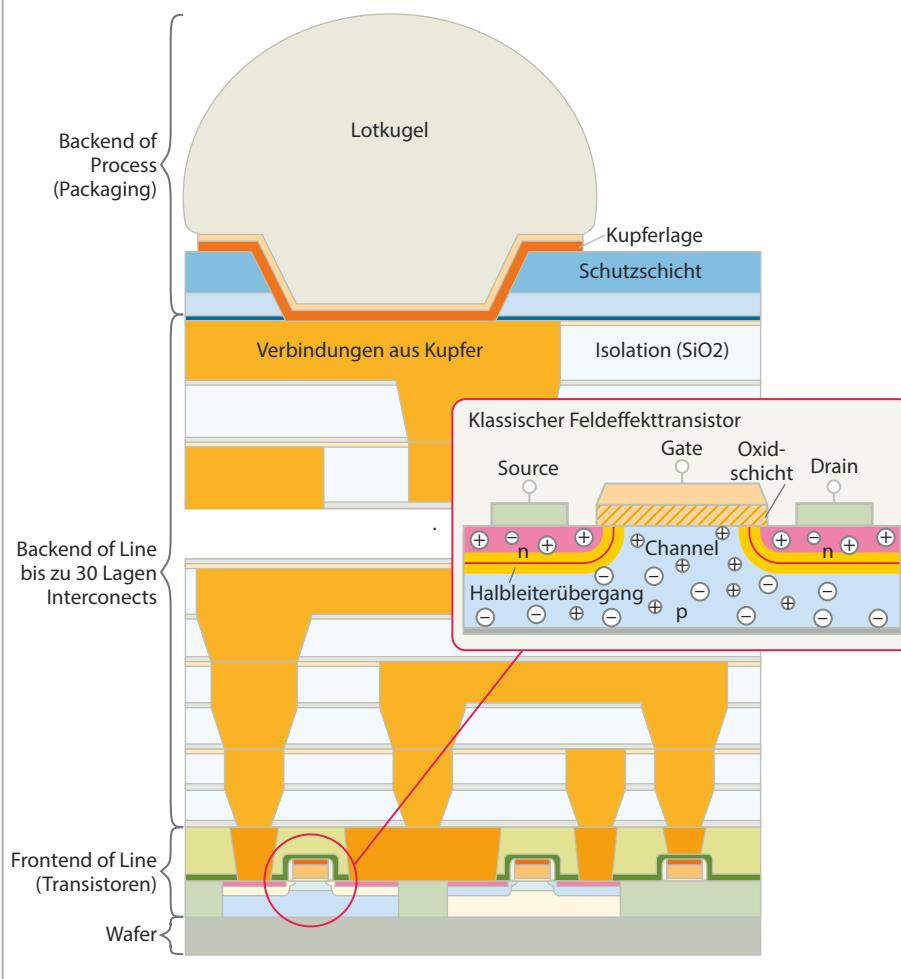
Einige dieser Stoffe – vor allem Edelmetalle – verursachen noch deutlich höhere Umweltlasten als die Waferproduktion, weil die nötigen Erze aus Lagerstätten mit geringer Konzentration stammen. Man muss also sehr viel Gestein abbauen und verarbeiten. Das führt zu großen Landschaftsschäden in den Abaugebieten, bei einigen Materialien entstehen problematische Abfälle wie radioaktive Substanzen. Und auch der Transport braucht viel Energie, weil die Rohstoffe aus Lagerstätten rund um den Erdball stammen.

Ein weiterer erheblicher Umweltbeziehungsweise Klimafaktor der Chipfertigung sind Prozessgase. Denn einige davon haben ein extrem hohes Treibhauspotential, man spricht vom Global Warming Potential (GWP). So arbeiten Chiphersteller etwa beim Plasmaätzen mit Hexafluorethan (C_2F_6), das 12.340-mal so viel zur Klimaerwärmung beiträgt wie dieselbe Menge Kohlendioxid (CO_2); ein Kilogramm Hexafluorethan in der Atmosphäre wirkt also so schädlich wie 12,3 Tonnen CO_2 . Außerdem benötigen Chip-Fabs Tetrafluormethan (CF_4), Fluoroform (CHF_3), Octafluorcyclobutan (C_4F_8) und Stickstofftrifluorid (NF_3).

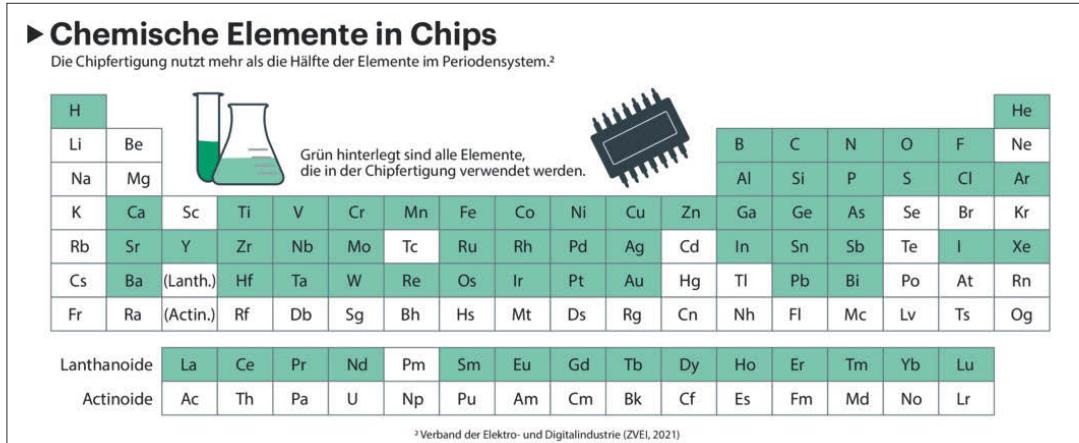
Die Verbindungen gehören zur Gruppe der „Ewigkeitschemikalien“, die man auch Perfluorverbindungen (PFC) beziehungsweise per- und polyfluorierte Alkylsubstanzen (PFAS) nennt. Sie bauen sich in der Umwelt nur sehr langsam ab, weshalb die EU sie eigentlich verbieten will. Doch in der Halbleiterproduktion lassen sich viele PFAS bisher nicht durch andere Stoffe ersetzen, unter anderem weil viele Fertigungsschritte enorme Präzision verlangen. So werden beim Ätzen mancher Schichten nur wenige Atomlagen abgetragen und das auch noch von sehr widerstandsfähigen Materialien wie Silizium

Schichtaufbau eines Prozessors

Ganz unten, direkt auf dem Wafer sitzen die einzelnen Transistoren. Darüber folgen viele Schichten mit Leiterbahnen aus Kupfer, die einzelne Transistoren verbinden und versorgen. Je weiter es in Richtung der Anschlüsse (Lotkugeln) geht, desto dicker werden die Leiterbahnen.



Die Halbleiterfertigung nutzt mehr als die Hälfte der Elemente, die das Periodensystem aufführt.



oder – bei modernen Leistungshalbleitern – von Siliziumcarbid (SiC), das wegen seiner großen Härte auch als Schleifmittel (Carborund) dient.

Die schädlichen (PFC-)Prozessgase müssen aus der Abluft der Fabs entfernt werden, man spricht von „PFC Abatement“, also PFC-Minderung. Das trifft es ganz gut, denn die Reinigung lässt noch rund 1 Prozent vieler Gase durch – und selbst dieser kleine Rest führt zu vergleichsweise hohen Klimaschäden. Dabei ist der Aufwand fürs PFC Abatement groß: Manche Anlagen arbeiten mit Plasmafackeln, deren Hitze die PFCs in andere Verbindungen zerlegt, sodass sie anschließend von Absorptionsmaterialen aufgenommen werden können. Die PFAS-Emissionen der Halbleiterfertigung sind allerdings deutlich geringer als die mancher Chemiewerke und bisher kommen PFAS sogar noch tonnenweise in Alltagsprodukten zum Einsatz, etwa als fettabweisende Beschichtung auf Einwegverpackungen.

Die Chipfertigung benötigt relativ viel Wasser zur Prozesskühlung und vor allem für zahlreiche Reinigungsschritte in der Produktion. Dabei kommt wie erwähnt Reinstwasser zum Einsatz, das aufwendig aus Rohwasser erzeugt werden muss. Auch die Produktion von Edelgasen wie hochreinem Neon, das manche der Laser für die Lithographiesysteme benötigen, schluckt viel Energie.

Schwere Lieferkette

Viele Chiphersteller veröffentlichen bereits relativ detaillierte Umweltbilanzen, die Emissionen nach dem üblichen Schema in verschiedene „Scopes“ klassifizieren. Scope 1 umfasst die direkten Umweltwirkungen der eigentlichen Fabs, etwa durch die ausgestoßenen Gase. Bei Scope 2 geht es beispielsweise um Emissionen, die bei

der Erzeugung des zugekauften Stroms entstehen. Scope 3 beschreibt Umweltbelastungen, die durch die Lieferkette, also etwa Vorprodukte und Rohstoffe (vorgelagert) und Nutzung (nachgelagert) der Produkte entstehen. Bei Intel sind gehören zu den nachgelagerten Belastungen der Stromverbrauch der verkauften Chips und der Aufwand für ihre spätere Entsorgung.

Laut einer aktuellen Studie der Boston Consulting Group (BCG) entfallen 42 Prozent der Umweltlasten eines Chips auf die Lieferkette. Chiphersteller müssen also mit ihren Zulieferern kooperieren, damit diese umweltschonendere Verfahren einsetzen, alternative Materialien nutzen und dafür andere Fertigungstechniken entwickeln. Das ist ein sehr langwieriger Prozess.

Es gibt aber auch bei vielen Chip-Rohmaterialien relativ leicht umsetzbare Möglichkeiten, um die Umweltbelastungen zu reduzieren. So könnte man zur energieintensiven Herstellung der Wafer grünen Strom einsetzen. Einige Firmen verlagern solche Prozesse deshalb in Länder wie

Norwegen und Island, in denen elektrischer Strom aus Wasserkraft sehr billig ist. Allerdings sind Fachkräfte dort noch schwieriger zu finden als in dicht besiedelten Industrieländern, in denen wiederum Ökostrom knapp und teuer ist.

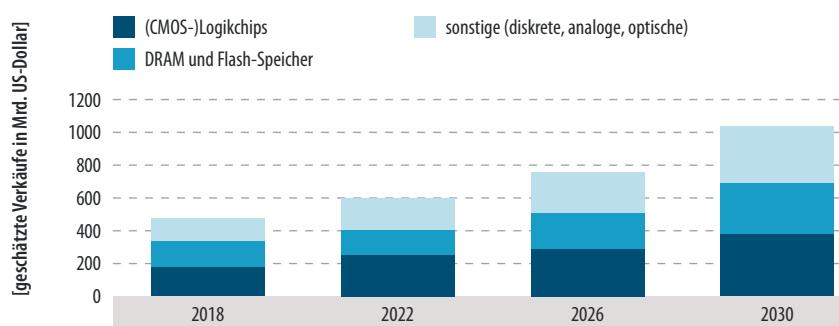
Schonvorschläge

Um die Umweltbelastung der Halbleiterfertigung zu mindern und im Idealfall klimaneutral werden zu können, sind gewaltige Veränderungen nötig. Dabei hilft der Halbleiterindustrie ihr rasanter Fortschritt, den das berühmte Moore'sche Gesetz (Moore's Law) beschreibt: Ungefähr alle zwei Jahre lässt sich die Anzahl der elektronischen Funktionen pro Quadratzentimeter Chipfläche verdoppeln. Das liegt im Wesentlichen an den Fortschritten bei der Strukturverkleinerung. Dazu wiederum müssen immer wieder neue, verbesserte Verfahren und Maschinen eingesetzt werden.

Die Halbleiterindustrie ist es daher gewohnt, relativ schnell auf neue Technik umzusteigen. Deshalb sollte auch die Ein-

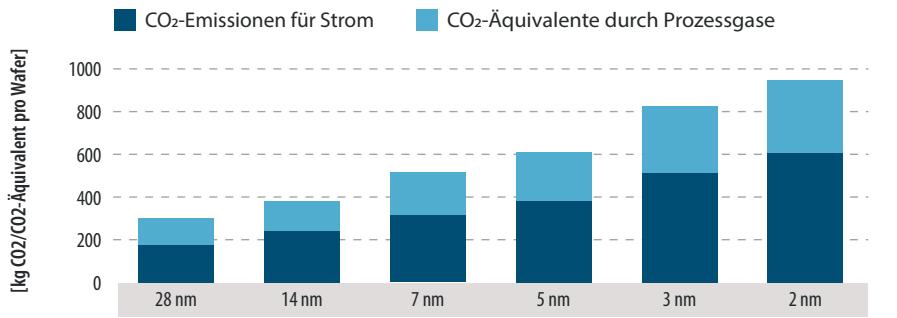
Wachstumsprognose Halbleiter

Der Umsatz mit Halbleitern und deren produzierte Anzahl steigen in den kommenden Jahren weiter stark an.



CO₂-Emissionen pro Wafer

Je kleiner die Chip-Strukturen, desto größer die Anzahl an Bearbeitungsschritten. Damit wächst der Bedarf an Energie und Prozessgasen pro Wafer.



führung weniger umweltschädlicher Technik zügig gelingen. Das im belgischen Löwen ansässige, europäische Halbleiterforschungsinstitut IMEC kooperiert dazu im Projekt Sustainable Semiconductor Technologies and Systems (SSTS) mit Chipherstellern und -entwicklern. Die Grundidee ist die Erfassung von Energieaufwand und Umweltwirkung sämtlicher Fertigungsschritte von Halbleitern. Anhand dieser Daten können Chipentwickler ihre Schaltungen und schonendere Alternativen dazu bewerten. Ziel ist es, jeden Funktionsblock eines Chips so auszulegen, dass er sich möglichst umweltschonend produzieren lässt.

Zurzeit laufen international viele Forschungsprojekte, die die Umweltbelastun-

gen bei der Produktion von Solarzellen für die Photovoltaik mindern sollen. Denn dabei werden im Vergleich zur Halbleiterproduktion viel größere Mengen an Silizium verarbeitet. Die Herstellung und Verarbeitung von Solarwafern ist zwar viel weniger aufwändig als die von Chipwafern, weil Solarzellen sehr simple Strukturen haben und viel geringere Anforderungen an Kristallstruktur und Reinheit stellen. Doch manche Verbesserungen für Solarwafer könnte auch die Chipindustrie umsetzen.

Im Prinzip trägt auch das Moore'sche Gesetz dazu bei, die Umweltbelastung der Chipfertigung zu mindern. Denn alle zwei bis drei Jahre passt ein Chip mit denselben technischen Funktionen auf die Hälfte seiner bisherigen Siliziumfläche.

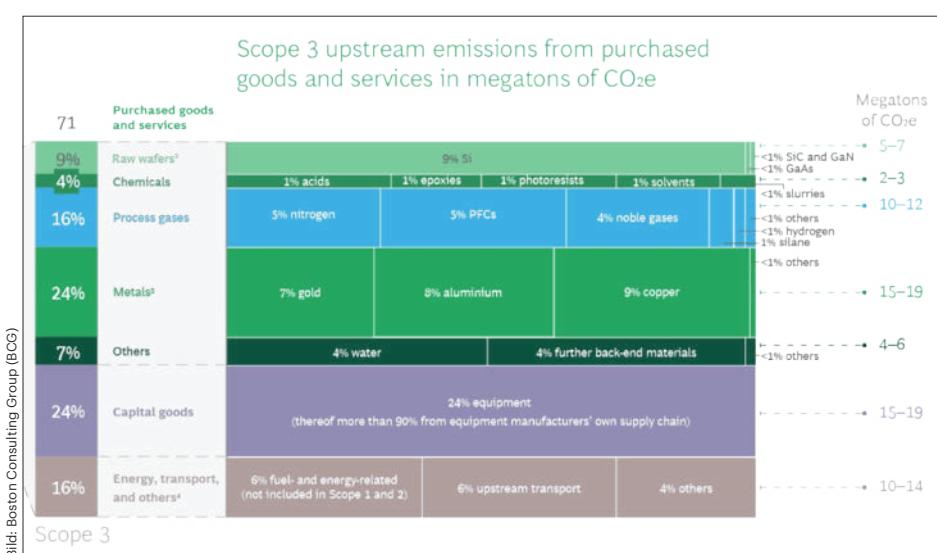
Doch neue Produkte haben fast immer auch mehr Funktionen und die Anzahl der jährlich produzierten Chips wächst kontinuierlich. Neue Chip-Fabs ersetzen meistens nicht etwa alte Werke, sondern kommen neu hinzu und weiten die Produktionskapazitäten aus. Damit steigen auch die Umweltlasten. Obendrein sind viele neue Chip-Fabs für kleinere Strukturen als bisher ausgelegt. Und je winziger die Strukturen, desto höher der Energieaufwand pro Wafer in der Fertigung. Denn die feineren Strukturen benötigen typischerweise mehr Lithografie-Durchgänge und bekommen auch weitere Funktionslagen.

Konkrete Maßnahmen

Drei Maßnahmen zur Umweltschonung lassen sich dennoch relativ einfach umsetzen: Energie sparen, grünen Strom verwenden und Recycling. Mit steigenden Strompreisen lohnen sich typischerweise zahlreiche kleinere Stromsparmaßnahmen, die zusammengenommen einige Prozentpunkte bringen können. Für mehr regenerative Energien helfen Investitionen in Solar- und Windkraftanlagen – das dauert zwar und verschlingt anfangs Geld, aber der Lösungsweg ist bekannt, und grüner Strom aus eigenen Erzeugungsanlagen vor Ort hat den zusätzlichen Vorteil, dass die Abhängigkeit von schwankenden Strompreisen und langfristig auch die Energiekosten sinken.

Recycling klappt nicht bei allen Materialien, aber ein gutes Beispiel ist die zunehmende Wiederaufbereitung von Wasser. Intel plant, dadurch den Wasserverbrauch drastisch zu reduzieren. Das ist an wasserarmen Standorten wie Phoenix (Arizona), Rio Rancho (New Mexico) und Kiryat Gat (Israel) auch wirtschaftlich eine gute Investition in die Zukunft und stärkt die Resilienz der Fabs gegen Versorgungsstörungen. Intel will bis 2030 sogar „netto wasserpositiv“ werden, bezieht dabei aber Investitionen in die Wasseraufbereitung außerhalb der Fabs mit ein. Das klingt ein bisschen nach den umstrittenen „klimaneutralen“ Fertigungsverfahren, bei denen man weitermacht wie bisher, aber für Kompensationsmaßnahmen an anderen Orten bezahlt.

Um die Umweltwirkung der Chipfertigung zu verbessern, wollen Hersteller wie Infineon besonders wirksame Anlagen zum erwähnten PFC Abatement einsetzen.



Rund 42 Prozent der bei der Chipfertigung anfallenden Treibhausgase entfallen auf die Vorprodukte, davon wiederum etwa 24 Prozent auf die benötigten Maschinen und je 9 Prozent auf Wafer und Kupfer.

Harte Brocken

Im Prinzip könnte man Rohstoffe aus alten Halbleiterbauelementen rückgewinnen. Zumindest in Europa wird bereits ein nennenswerter Teil des Elektroschrotts eingesammelt und als Sekundärrohstoff relativ teuer gehandelt. Recycler schreddern vorsortierte Platinen und Bauteile und schmelzen sie ein, um aus vergoldeten Steckverbindern, Beschichtungen und internen Kontakten in Chips teure Edelmetalle wie Gold und auch Kupfer zu gewinnen. Für Silizium lohnt sich das bisher nicht, denn das Rohsilizium aus alten Chips wäre stärker verunreinigt als der viel billigere Quarzsand. Für die ins Silizium eingebrachten Zusatzstoffe lohnt sich die Wiederaufbereitung erst recht nicht, da sie nur in winzigen Mengen und in geringerer Konzentration drinstecken als in den jeweiligen Bergwerken und anderen Lagerstätten.

Um nennenswerte Teile der eigentlich Chips zu recyceln (und nicht nur die Platinen), sind deutlich verbesserte Separierungsverfahren nötig. Die werden derzeit erforscht, aber wiederum vor allem für Solarzellen. Wenn es dafür eines Tages gute Verfahren gibt, könnte man sie vielleicht auch für Mikrochips anpassen.

Umweltschäden, die der Bergbau zur Rohstoffgewinnung verursacht, lassen sich nicht komplett vermeiden, sondern nur reduzieren oder kompensieren. Ob das tatsächlich geschieht, lässt sich nach bisherigen Erfahrungen nur in demokratisch regierten Ländern kontrollieren, die sich internationalen Umweltschutzregeln unterwerfen. Einige Rohstoffe für Vorprodukte der Chipfertigung kommen jedoch aus autoritär regierten Staaten. China übt damit sogar Druck auf westliche Chipfirmen aus und reduziert beispielsweise den Export von Gallium, das in Galliumnitrid-(GaN)-Wafern für besonders effiziente Schalttransistoren steckt.

Fazit

Die Halbleiterfertigung hat einen erheblichen ökologischen Fußabdruck und die Menge der jährlich produzierten Chips steigt weiter stark an. Im Vergleich zu anderen Branchen wie der Stahlindustrie und fossilen Energieverbrauchern steht die Chipfertigung allerdings nur für einen kleinen Anteil am weltweiten Ausstoß von Treibhausgasen. 2021 verursachte alleine die Verbrennung fossiler Energie rund 34 Milliarden Tonnen CO₂. Laut einer Studie stießen die Chipfertiger 2021 „nur“ rund 75 Millionen Tonnen CO₂ und CO₂-Äquiva-



Bild: Globalfoundries

Die Globalfoundries-Fab in Dresden betreibt eigene Gaskraftwerke, die man im Hintergrund der großen Gebäude für Reinräume erkennt.

lente aus, hatten also weniger als 3 Promille Anteil am gesamten Aufkommen. Dennoch steht die Branche unter Druck, ihre Umweltbelastung zu senken; darauf drängen auch viele ihrer direkten Kunden. Gesetzliche Regelungen nehmen die Industrie zudem immer strenger in die Pflicht.

Es gibt bereits einige positive Signale, wie Chips grüner werden könnten. So entstehen Fabs für Chips und Vorprodukte an

Standorten, an denen regenerativ erzeugter Strom günstig ist. Außerdem werden Fertigungsverfahren stetig verbessert und es gibt Projekte, um die Umweltbelastung von Fertigungsverfahren schon beim Design von Chips zu berücksichtigen. Trotzdem wird es sehr lange dauern, bis die Chipfertigung wirklich nachhaltig ist. (ciw@ct.de) **ct**

Hintergrundinformationen: ct.de/yzz



Bild: Globalfoundries

In den großen Reinräumen der Chiphersteller stehen zahlreiche Maschinen, die riesige Mengen an Strom benötigen.



Quo vadis, USB?

Wie es mit USB4 weitergeht

USB4 verspricht, die eierlegende Wollmilchsau zu sein, weil unzählige Datenprotokolle über dasselbe USB-C-Kabel fließen, das gleichzeitig noch den Notebookakku rasant füllt. Doch während USB4 noch nicht bei gängiger Peripherie angekommen ist, werfen USB4 2.0 und Thunderbolt 5 schon ihre Schatten voraus.

Von Florian Müssig

Die Spezifikation von USB4 ist nicht mehr allzu neu; das Standardisierungsgremium USB Implementers Forum (USB-IF) hat sie bereits anno 2019 verabschiedet. Anders als bei vorherigen USB-Versionen haben sich die Ingenieure keine gänzlich neue technische Basis ausgedacht, sondern Thunderbolt 3 als diese genommen und ihr nur Feinschliff wie die Unterstützung von Hubs angedeihen lassen [1]. Beide Standards liefern deshalb bis zu 40 Gbit/s. Der größte Unterschied: Bei USB4 sind die Ansteuerung von DisplayPort-Monitoren oder das Weiterreichen von PCIe-Signalen nur optional, während Thunderbolt sie verpflichtend voraussetzt.

Kompatible Host-Controller findet man inzwischen in allen gängigen Mobilprozessoren von AMD (ab Ryzen 6000),

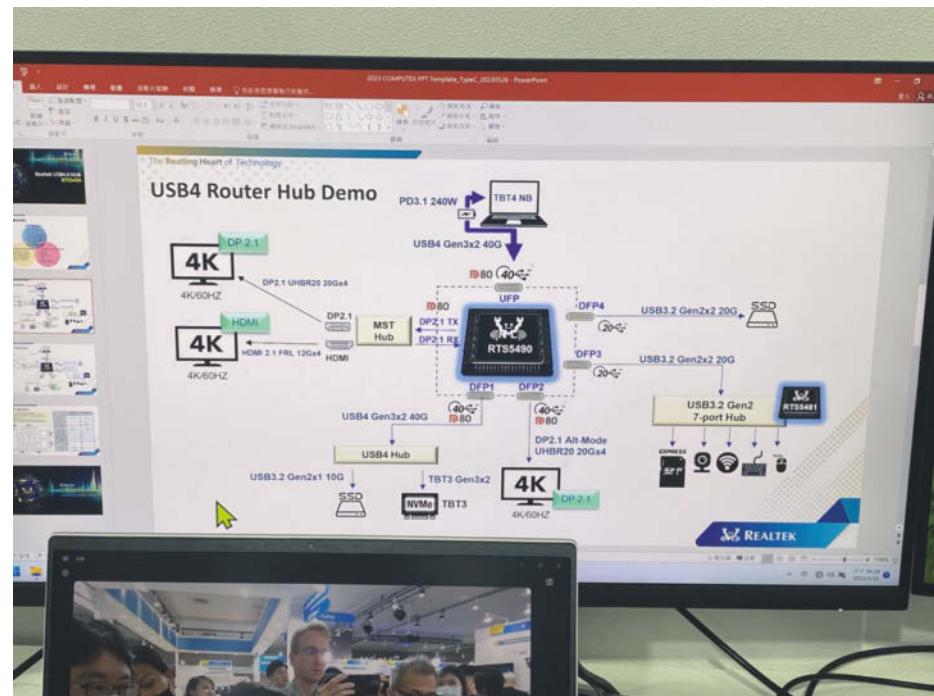
Apple (ab M1) und Intel (ab 12. Core-i-Generation). Sie sind allerdings nur bei höherpreisigen Notebooks tatsächlich nutzbar, weil die Universalität der USB4-Schnittstelle das Systemdesign komplexer und damit teurer macht. Dort macht der theoretische Unterschied Option (USB4) oder Pflicht (Thunderbolt) praktisch keinen Unterschied, denn es gehört zum guten Ton schlicht dazu, dass USB-C-Buchsen DisplayPort-Signale ausgeben können. Und das Weiterreichen von PCIe ist eine Microsoft-Anforderung, um das Chaos einzudämmen: Wenn ein Notebook mit Windows verkauft wird und eine USB4-Buchse haben soll, muss diese auch PCIe für externe GPUs oder SSDs bieten. Sonst gibt es kein Windows-Logo und damit auch keine damit verbundenen Marketingzuschüsse.

Die meisten Peripheriehersteller nutzen bislang allerdings die Tatsache aus, dass USB4 rückwärtskompatibel zu Thunderbolt 3 ist – weshalb es praktisch weiterhin nur Thunderbolt-Docks gibt und noch keine „echten“ USB4-Docks und USB4-SSDs. Das ist aus Nutzersicht erst mal nicht weiter schlimm, denn hinsichtlich der maximalen Datenrate von 40 Gbit/s liegen USB4, Thunderbolt 3 und auch Thunderbolt 4 (die Erweiterung von Thunderbolt 3 um die USB4-spezifischen Neuerungen) praktisch gleichauf.

Docks

Dennoch gibt es Haken: Thunderbolt-Docks findet man in unzähligen High-End-Ausprägungen für Schreibtische, aber nicht als kompakte und günstige Reise-Docks. Das will VIA mit seinem End-Point-Controller VL830 ändern: Der Chip spricht USB4 (40 Gbit/s) mit dem Host, gibt an seinen USB-Ports aber bestenfalls USB 3.2 mit 20 Gbit/s weiter. Die restliche Bandbreite steht für DisplayPort-Signale zur Verfügung, worüber dann mehrere 4K-Monitore angesteuert werden können, oder einer mit noch höherer Auflösung oder flotteren Bildwiederholraten als 60 Hz. Das klappt bei derzeit gängigen Adapters von USB-C auf DisplayPort oder HDMI nicht – wegen der geringeren nutzbaren USB-3.x-Bandbreite, die zudem noch fix auf Lanes aufgeteilt ist. Der VL830 ist laut VIA fertig entwickelt; jetzt sind die Hersteller von Docks gefragt.

Für größere Dockingstationen, die natives USB4 und nicht wie alle bisherigen



USB4 kann vieles; entsprechend komplex ist das, was für Dockingstationen gedachte Hub-Controller wie Realteks RTS5490 leisten müssen.

Thunderbolt 3 sprechen, hat Realtek den RTS5490 entworfen. Das ist der erste USB4-Hub-Controller, der sämtliche USB4-Funktionen unterstützt und weiterreicht – was in einem beeindruckend komplexen Blockschaltbild mündet. Realtek hat den RTS5490 bereits als funktionsstüchtigen Prototypen demonstriert, aber zur Massenproduktion noch keine Worte verloren. Anders als Docks mit VL830 dürften solche mit RTS5490 wohl im hohen Preisbereich bisheriger Thunder-

bolt-Docks liegen, also jenseits von 250 Euro.

Für externe SSDs sind noch keine nativen USB4-Controller in Sicht. Klar gibt es längst welche, die einen Thunderbolt-Port enthalten und an USB4-Hosts mit voller Geschwindigkeit laufen. Doch damit die externen SSDs auch an USB-3.x-Hosts funktionieren, muss bislang ein weiterer Controller auf die Platine und ein Umschalter, der nach dem Anstecken je nach Host-Fähigkeit den richtigen Controller auswählt – das macht sie komplex und teuer [2]. JMicron gab auf Nachfrage zu Protokoll, dass die Entwicklung von USB4-Storage-Controllern bereits läuft, doch Produkte seien erst ab 2024 zu erwarten.

USB4 zum Nachrüsten

ASMedia hat Mitte 2023 auf der Computex mal wieder seinen schon länger angekündigten USB4-Host-Controller ASM4242 ausgestellt. Auf eine Steckkarte verfrachtet lassen sich mit ihm etwa Desktop-PCs um zwei USB4-Ports ergänzen. Doch während Renesas vor gut zehn Jahren den Chipsatzherstellern mit seinem Stand-alone-USB-3.0-Controller zuvorgekommen ist und dieser in hohen Stückzahlen eingebaut wurde, findet man jetzt schon viele High-End-PCs und -Notebooks, die Thunderbolt 4 (Intel, Apple) beziehungsweise USB4 (AMD) mittels in



MSI wird eine Steckkarte bringen, die ASMedias USB4-Host-Controller ASM4242 enthält. Sie reicht per DisplayPort eingespeiste Videosignale per USB-C weiter, und dank sechspoligem Zusatzstromstecker versorgt sie USB-C-Peripherie mit bis zu 100 Watt.

E-Books im heise Shop

Generell portofreie Lieferung für Heise Medien- oder Maker Media Zeitschriften-Abo-Neu-Abonnenten oder ab einem Einkaufswert von 20 € (innerhalb Deutschlands). Nur solange der Vorrat reicht. Preisänderungen vorbehalten. E-Books können einem DRM-Schutz unterliegen.



FSP hat bereits ein USB-C-Netzteil fertig entwickelt, das bis zu 180 Watt (36 Volt, 5 Ampere) liefert.

den CPUs integrierten Controllern an Bord haben. Die Zielgruppe für den ASM4242 ist also viel kleiner, was dessen Entwicklungsverzögerungen erklären könnte. Nichtsdestotrotz soll er bis Jahresende in Massenproduktion sein.

Steckkarten mit ASM4242 bringen zwei USB-C-Buchsen und zwei DisplayPort-Eingänge; die dort eingespeisten Videosignale kommen dann ebenfalls aus den USB-C-Buchsen – der Signalweg verläuft also von der Grafikkarte in die USB4-Erweiterung und von dort dann zum Monitor. Eine der ersten Erweiterungskarten wird von MSI kommen, die intern auch noch einen sechspoligen Stromanschluss vorsehen, damit einer der beiden USB-C-Ports ein angeschlossenes Gerät mit 100 Watt aufladen kann (der zweite Port schafft 27 Watt).

Stärkere Netzteile

Apropos Stromversorgung: Auf der Computex waren an vielen Ständen USB-C-Netzteile zu sehen, die mehr als 100 Watt auf ihren Typenschildern führten. Dort sollte man allerdings genauer hinsehen: Bei Netzteilen, die mehr als einen Ausgang haben, addieren die Hersteller deren Maximalwerte gerne – aus einer einzelnen USB-C-Buchse bekommt man in solchen Fällen nie die volle Leistung. Immerhin: An einem Anschluss stehen nicht mehr nur 100 Watt (20 Volt), sondern immer häufiger auch 140 Watt (28 Volt) bereit. Das ist die erste Stufe der

Erweiterung Extended Power Range (EPR), mit der die Schwester-Spezifikation USB-PD (Power Delivery) bei weiterhin 5 Ampere Maximalstrom die bisherige 100-Watt-Begrenzung knackt.

Den nächsten Schritt mit 180 Watt (36 Volt) haben wir auf der Messe einzig bei einem Prototyp von FSP entdeckt, doch Framework will seinen Laptop 16 noch vor Jahresende ausliefern – mit 180-Watt-USB-C-Netzteil im Karton. Das lässt hoffen, dass 2024 auch leistungsstarke Notebooks anderer Hersteller bei der Stromversorgung auf USB-C statt Rundstecker setzen, zumal der Standard ja auch eine noch höhere Ausbaustufe mit 240 Watt (48 Volt) vorsieht. Das reicht selbst für das Gros an Gaming-Notebooks aus.

Kleine Randnotiz: Im c't-Labor waren zuletzt schon mehrere performante Notebooks mit USB-C-Netzteilen, die mehr als 100 Watt lieferten. Sie verwendeten allerdings proprietäre Erweiterungen, welche bei 20 Volt mehr als 5 Ampere über das Kabel schicken. Das funktioniert nur für jeweils eine spezifische Kombination aus Netzteil und Notebook, aber nicht in beliebigen oder über Herstellergrenzen hinweg, was wiederum den Reiz von USB-C ausmacht. Das einzige uns bislang bekannte Notebook, dessen Netzteil schon heute EPR-konform über 100 Watt geht, ist Apples MacBook Pro 16.

Dass man ein 140-Watt-Netzteil daran wiederum nur per MagSafe- und

nicht per USB-C-Kabel voll ausfahren kann, hat einen sehr banalen Grund, wie uns das USB-IF verraten hat. Wir hatten bislang spekuliert, dass es damals noch keine ICs gab, die den erweiterten Modus beherrschen, doch das war laut USB-IF nicht das Problem: Apples Einkäufer hätten zu dem Zeitpunkt, als sie die Verträge mit Zulieferern klargemacht haben, schlicht keine USB-C-Buchsen gefunden, die bereits auf mehr als 20 Volt ausgelegt gewesen wären, wie es der Standard vorschreibt.

Ein weiterer Knackpunkt ist, dass Notebooks seit Ewigkeiten mit 19, 19,5 oder 20 Volt geladen werden. Stärkere USB-C-Netzteile bringen das Gefüge ins Wanken, denn für 28 Volt oder mehr müssen die Schaltkreise von der USB-C-Buchse bis zur (auf dem Mainboard integrierten) Ladeelektronik anders ausgelegt werden – zwar kein Hexenwerk (siehe Framework Laptop 16), aber Neuland für die Hersteller.

Doppelte und dreifache Geschwindigkeit

Die Ende 2022 bekannt gegebene Spezifikationserweiterung mit der kruden Bezeichnung USB4 2.0, die die Datenraten über die bislang üblichen 40 Gbit/s erhöht, braucht ebenfalls noch ein Weilchen, bis sie im Handel ankommt. Das sei etwas für 2024 und danach, meinten die Vertreter des USB-IF. USB4 2.0 sei zudem weni-

ger für noch flottere externe SSDs oder ähnliche Peripherie gedacht, sondern für High-End-Systeme mit mehreren Monitoren, die mit hohen Auflösungen und rasanten Bildwiederholraten einen enormen Bandbreitenbedarf haben.

Aus diesem Grund sieht USB4 2.0 nicht mehr nur eine Geschwindigkeitsverdopplung auf synchrone 80 Gbit/s vor, sondern auch einen asynchronen Betriebsmodus, der in eine Richtung – genannt Monitor – 120 Gbit/s schiebt, aber in Gegenrichtung nur die altbekannten 40 Gbit/s liefert, wie sie schon Thunderbolt 3 vorsah.

Während bei USB4 2.0 die hinzugekommenen Neuerungen in etablierter Manier nur optional sind, macht Thunderbolt erneut Nägel mit Köpfen: Der Nachfolger von Thunderbolt 4 setzt nicht nur den neuen 80-Gbit/s-Modus sowie den asynchronen 120-Gbit/s-Modus als verpflichtend voraus, sondern auch eine EPR-konforme Stromversorgung mit mindestens 140 Watt. Obendrein hat er den viel eingängigeren Namen – nämlich schlicht und ergreifend Thunderbolt 5.

(mue@ct.de) 

Literatur

- [1] Florian Müsseg, USB, die Vierte, Spezifikation für USB 4 finalisiert, c't 20/2019, S. 26
 - [2] Florian Müsseg, USB-SSD, noch schneller, Orico MTQ-40G: die erste SSD mit USB4 im Test, c't 22/2022, S. 62

Thunderbolt™ 5 Leaps Ahead

| | | Thunderbolt™5 | Thunderbolt™4 | USB4* | USB 3/DP |
|-------------------------|--|--|--|-----------------------------|-----------------------------|
| Unrivaled Simplicity | One universal computer port | • | • | - | - |
| | Required speed for Universal cables up to 2 meters in length | 120 Gbps | 40 Gbps | 20 Gbps | - |
| | Accessories with four Thunderbolt ports | • | • | - | - |
| Exceptional Performance | PC speed requirements | 800Gbps & 120 Gbps | 40 Gbps | 20 Gbps | 10 Gbps |
| | PC video requirements | Dual 6K | Dual 4K | One monitor (No Minimum BW) | One monitor (No Minimum BW) |
| | PC data requirements ³ | PCIe 6.4 Gbps USB3 - 10 Gbps | PCIe 32 Gbps USB3 - 10 Gbps | - | - |
| | Required PC charging on at least one computer port | Required up to 140W available up to 240W | Required up to 100W available up to 140W | - | - |
| | Required PC wake from sleep when computer is connected to a Thunderbolt dock | • | • | - | - |
| | Minimum PC port power for accessories | 15W | 15W | 7.5W | 4.5W |
| | Thunderbolt Networking ⁴ | 64 Gbps | 32 Gbps | - | - |
| Reliable Connectivity | Mandatory certification for all shipping computers, accessories and cables | • | • | - | - |
| | Cable testing and quality audits for Thunderbolt cable manufacturers | • | • | - | - |
| | Required Intel® VT-d-based DMA protection | • | • | - | - |
| | USB4 specification | Compliant | Compliant | Compliant | Compatible |

1031

Thunderbolt 5 hat gegenüber USB4 2.0 nicht nur den eingängigeren Namen, sondern setzt Neuerungen wie die höheren Geschwindigkeitstufen mit 80 oder gar 120 Gbit/s verpflichtend voraus. Sie sind nötig, um zwei Monitore mit mehr als 4K-Auflösung oder höheren Bildwiederholraten als 60 Hertz (oder beides über ein Kabel anzusteuern.



Forscher schätzen, dass in 90% der von ihnen untersuchten Fälle von Cyberangriffen, das Active Directory involviert ist! Mit dieser aktualisierten und erweiterten Neuauflage des **iX Kompakt zur AD-Sicherheit** können Sie sich dringend benötigtes Fachwissen zum Schutz vor Ransomware aneignen:

- ▶ Denken wie ein Hacker – Angriffe verstehen und verhindern
 - ▶ Forensische Analyse von Vorfällen und Angriffen
 - ▶ Microsofts Schichtenmodell: Tiers festlegen und abschotten, privilegierte Zugriffe absichern
 - ▶ Marktübersicht: Werkzeuge für die AD-Absicherung

Heft für 29,50 € • Digital für 27,99 €
Heft + Digital 41,50 €



Generell portofreie Lieferung für Heise Medien- oder Maker Media Zeitschriften-Abonnenten oder ab einem Einkaufswert von 20 € (innerhalb Deutschlands). Nur solange der Vorrat reicht. Preisänderungen vorbehalten.

 heise Shop



Kristallwerkstatt

Aktuelle und künftige Displaytechnik für Smart-TVs

Dieser Beitrag wirft einen Blick auf die Technik in Fernsehern und beleuchtet, wodurch sich günstige von teureren TVs unterscheiden und was der Nobelpreis damit zu tun hat.

Von Ulrike Kuhlmann

Bei dem Begriffswirwar, das die TV-Hersteller in der Vergangenheit angezettelt haben, kann man ganz schön ins Schleudern kommen: LCD, LED, QLED, QNED, Nanocell, OLED, QD-OLED, WOLED. Dabei ist es viel einfacher: Die ersten fünf Verfahren nutzen Flüssigkristalltechnik, also LCD (Liquid Crystal Display), die letzten drei organische Leuchtschichten, sind also OLED-TVs.

Samsung erfand mit LED-TV (Light Emitting Device) vor etlichen Jahren die erste verwirrende Wortschöpfung für seine LCD-TVs mit LED-Backlight. Das koreanische Unternehmen erntete dafür in der Fachwelt viel Kritik, aus Marketing-sicht war der Namenscoup aber erfolgreich. Aktuelle LCD-Fernseher nutzen farbkonvertierende Nanopartikel, um sattere Farben zu erzeugen. Die winzigen Quantenpunkte (Quantum Dot, QD) mit milliardstel Metern Durchmesser sitzen im Backlight des LC-Displays und sind genau das Material, für dessen Entdeckung und Synthese der diesjährige Chemie-Nobelpreis vergeben wurde. Die Nanopartikel modulieren das Licht des LED-Backlight und heben es auf eine andere Wellenlänge, ändern also dessen Farbe.

Für die Bilderzeugung in den mit Quantenpunkten arbeitenden TVs ist wie gehabt der Flüssigkristall zuständig. Das scherte die Marketingabteilung von Sam-

sung aber nicht, sie schuf mit QLED-TV eine weitere technisch unsaubere Produktbezeichnung. Das hatte wiederum mit den erstarkten organischen Displays zu tun, die Samsung seinerzeit nicht im Programm hatte – QLED erinnert nicht zufällig an OLED. Mehr zu den OLEDs weiter unten.

Beleuchtungsfragen

Die LED-Hintergrundbeleuchtung ist aktuell eines der wichtigsten Unterscheidungskriterien für LCD-TVs. Es gibt vier Backlight-Varianten. Direct-LED ohne lokale Ansteuerung ist die billigste: Bei ihr sitzen relativ große Leuchtdioden gleichmäßig im Displayrücken verteilt. Die LEDs können nicht einzeln gedimmt werden, sondern leuchten im Verbund.

Beim Edge-LED-Backlight sitzen stattdessen etwas kleinere LEDs aufgereiht an einer Kante des Displays, oft an der unteren. Meistens werden auch hier

die LEDs nur im Verbund angesteuert. Einige TVs konnten die LED-Streifen in groben Segmenten separat ansprechen und damit eine lokale Abdunklung erzielen. Das führte meist zu unschönen Artefakten im Bild, weil sich bei bewegten Objekten am Schirm die Übergänge zwischen den streifenartigen, sehr großen Zonen störend bemerkbar machten. Inzwischen findet man kaum noch TVs, die diese Variante lokaler Abdunklung tzen.

Stattdessen setzen viele Hersteller heute auf LC-Displays mit Full Array Local Dimming. Beim FALD sind kleinere Dioden gleichmäßig im Rücken verteilt und lassen sich in voneinander unabhängigen Gruppen ansteuern. Wie gut dieses lokale Dimmen funktioniert, hängt von der Größe der schachbrettartig angeordneten LED-Felder ab. Weil die meisten Hersteller die Anzahl der LED-Cluster nicht preisgeben, bestimmen wir die Menge der dimmbaren Zonen anhand von Testbildern – je mehr Cluster es sind, desto feiner werden helle und dunkle Flächen im Bild gegeneinander abgegrenzt. Das Local Dimming benötigt man vor allem für HDR-Inhalte (High Dynamic Range), darunter beispielsweise kontraststarke Filme, in denen sich die funkelnden Sterne sauber am dunklen Nachthimmel abzeichnen und gleichzeitig helle Fackeln vor dunklen Felswänden flackern.

Noch feiner gelingt Local Dimming mit Mini-LEDs. Diese winzigen Leuchtdioden sitzen wie beim FALD gleichmäßig verteilt im Displayrücken, es gibt aber deutlich mehr Zonen und die LEDs sind flacher. Dadurch fällt das gesamte Display wesentlich schlanker aus. LCD-TVs mit Mini-LEDs gehören derzeit zu den besten und teuersten Flüssigkristallbildschirmen. Die Herausforderung dieser Backlight-Variante ist die homogene Ausleuchtung der Gesamtfläche; sie erfordert ein möglichst ähnliches Verhalten aller LEDs und einen gründlichen Abgleich des Bildschirms während der Produktion.

Organisches

Die Königsklasse sind derzeit TVs mit einem organischen Display: OLED-TVs. Es gibt sie inzwischen von allen großen Herstellern, die Produktion der großen organischen Panels liegt dagegen nur in den Händen zweier Unternehmen: LG Display ist seit vielen Jahren der Marktführer und seit gut einem Jahr mischt auch Samsung Display mit. Sony nutzt OLED-Panels von beiden, alle anderen TV-Her-

steller (bis auf Samsung) bislang ausschließlich WOLED-Technik von LG.

Während LCDs immer eine Hintergrundbeleuchtung benötigen und der Flüssigkristall lediglich als Lichtventil vor dem Backlight dient, leuchten OLEDs selbst. Die organische Schicht wird dazu mit einem Strom beaufschlagt, der sie zur Lichtemission anregt. Für jeden Bildpunkt ist ein eigener Transistorverbund zuständig. Dieses pixelgenaue Dimmen schafft man nicht mit einem LED-Backlight.

Zusätzlich sind OLEDs weitgehend blickwinkelunabhängig, während die Darstellung am LCD sich mehr oder weniger stark verändert, wenn man schräg von der Seite aufs Display schaut. Ein weiterer Vorteil ist die sehr kurze Zeit, in der sich die Helligkeit der Bildpunkte verändern lässt. Die rasanten Schaltzeiten der OLEDs sorgen für scharfe Konturen in Bewegtbildern, was sowohl für schnelle Videosequenzen als auch fürs Gaming wichtig ist.

Die Nachteile: OLEDs sind teurer in der Herstellung und man muss sie pfleglich behandeln, damit sich Bildinhalte nicht partiell festsetzen. Das sogenannte Einbrennen entsteht, weil die Leuchtschicht wie jedes organische Material mit der Zeit altert; sie leuchtet dann bei gleichem Stromdurchfluss nicht mehr so hell wie am Anfang. Das macht sich an Stellen im Bild bemerkbar, die über lange Zeit heller leuchten mussten. Deshalb sollte man Standbilder nicht dauerhaft am OLED-TV anzeigen, es also beispielsweise nicht als digitalen Bilderrahmen nutzen. Das verhindern die meisten TV-Hersteller aber ohnehin durch Abschaltautomatiken, die bei Standbildern nach wenigen Minuten einen dunklen Bildschirmschoner ak-

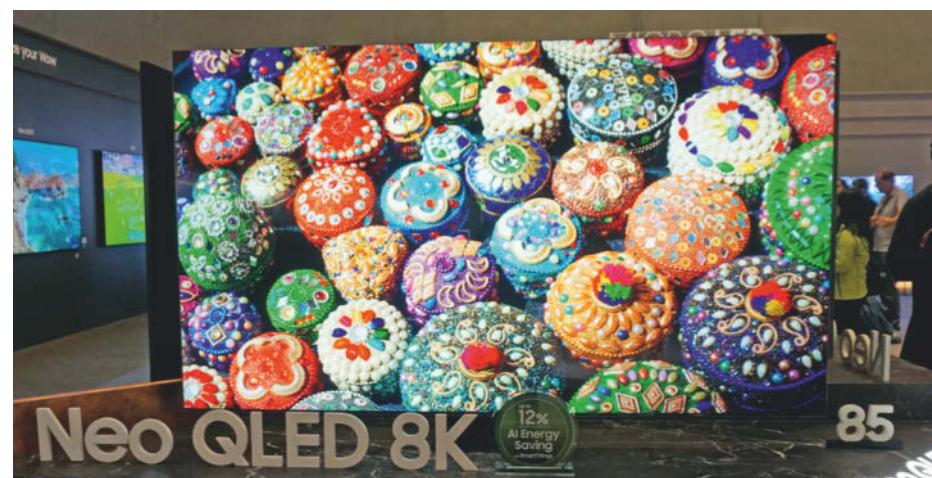
c't kompakt

- LCD- und OLED-Technik bleiben vorerst die führenden Displaytechniken in Smart-TVs.
- Die Art des LED-Backlight entscheidet bei LCDs über die HDR-Qualität.
- Mikro-LED-Displays werden erst mit neuen Produktionstechniken günstiger.

tivieren. Zudem betreiben die Hersteller ihre organischen Displays mit viel Lichtreserven und sie haben diverse Techniken entwickelt, die der frühzeitigen Alterung entgegenwirken (siehe dazu auch [1]).

WOLED vs. QD-OLED

LG und Samsung verwenden unterschiedliche Techniken in ihren OLED-TVs. Bei LG erzeugen organische Schichten großflächig „weißes“ Licht an den Subpixeln, darüber sitzende rote, grüne und blaue Filter bringen Farbe in die Darstellung. Zusätzlich gehört ein weißes Segment ohne Farbfilter zu jedem Bildpunkt. Man nennt die OLEDs von LG auch WOLED. Samsung arbeitet dagegen mit einer blau leuchtenden organischen Schicht, in einem Substrat darüber wandeln Quantenpunkte das blaue Licht in rotes und grünes um. Entsprechend nennt man die Samsung-Technik auch QD-OLED. Da die Lebensdauer respektive Effizienz der aktuell genutzten blau fluoreszierenden Emitterschicht begrenzt ist, möchte Samsung auf phosphoreszierende Emitter wechseln, wie sie bei der US-amerikani-



Samsung nutzt in seinen Neo QLED genannten LCD-TVs ein fein dimmbares Mini-LED-Backlight und Quantenpunkte für einen großen Farbumfang.



Hersteller wie Panasonic, Metz und Roku integrieren in ihre Smart-TVs unterschiedliche Betriebssysteme; Panasonic beispielsweise das eigene „My Home Screen“, aber auch Fire TV und Google TV.

schen OLED-Materialschmiede Universal Display Corp. (UDC) in der Entwicklung sind. Mit den deutlich effizienteren blauen Leuchtstoffen könnte Samsung ein oder zwei organische Layer einsparen, was die Produktion vereinfachen würde.

Während sich die Leuchtdichte von LCDs durch ein kräftigeres LED-Backlight steigern lässt, muss man bei OLEDs zu anderen Maßnahmen greifen, LG etwa zu einer Tandem-Struktur mit doppelt ausgelegten OLED-Stacks im Display. Das ist allerdings teurer und bleibt deshalb vorerst den TV-Topmodellen vorbehalten. Zusätzlich verbessert LG die Lichtausbeute mit einem MLA genannten Linsenraster. Samsung möchte die Leuchtdichte durch die oben erwähnten phosphoreszierenden Leuchtstoffe erhöhen und die Effizienz steigern, indem die Nanopartikel direkt auf dem Dünnfilm des OLED-Stacks platziert werden; beides senkt zu-

gleich die Kosten. Derzeit sind die QD-OLED-Panels deutlich teurer als die WOLED-Panels von LG Display, was viele TV-Hersteller davon abhält, in ihren Geräten QD-OLED-Panels zu nutzen.

Während LG seine OLED-TVs in ultrahoher Auflösung bis 8K entsprechend 7680 × 4320 Pixeln und mit großen Diagonalen bis 97 Zoll produziert, gibt es Samsungs QD-OLEDs bislang mit maximal 4K-Auflösung entsprechend 3840 × 2160 Pixeln und bis maximal 83 Zoll Diagonale. Auf den neuen Fertigungsstraßen sind die Produktionsmengen bei Samsung noch überschaubar, weshalb der Hersteller im kommenden Jahr auch OLED-Panels des koreanischen Mitbewerbers LG Displays einbauen will. Im eigenen Lager herrscht aber eine gewisse Skepsis gegenüber diesem Deal. Angesichts des rückläufigen TV-Absatzes könnte sich Samsung stattdessen auf die eigenen Spitzen-

produkte aus der Neo-QLED-Familie konzentrieren, also vermehrt LCD-TVs mit Mini-LEDs und Quantenpunkten anbieten.

Mini, Mikro, Nano

Zugleich feilt Samsung wie alle großen Hersteller an Displays, in denen winzige Leuchtdioden als Bildpunkte dienen. Diese echten LED-TVs vereinen alle Vorteile der LCD-Technik mit denen der OLEDs: Sie sind superhell, flink, kontraststark, blickwinkelunabhängig und zeigen dank anorganischer Leuchtmittel so gut wie keine Alterungseffekte. Allerdings ist die Fertigung großer Mikro-LED-Displays bisher extrem teuer. So soll der von Samsung zur IFA gezeigte 89-Zoll-Fernseher mit Mikro-LEDs etwa 100.000 Euro kosten. Zum Vergleich: Das 98-zöllige Samsung QLED-TV Q80C mit LCD-Technik und Mini-LEDs im Backlight ist für 7000 Euro zu haben. Ähnlich saftig sind die Preise für LGs Magnit genannte Mikro-LED-Displays oder Sonys Crystal-LED-Serien, die nicht für Unterhaltungselektronik gedacht sind, sondern als große Anzeigentafeln genutzt werden. Das Marktforschungsinstitut DSCC erwartet, dass Samsung in diesem Jahr weniger als 1000 Mikro-LED-TVs ausliefern wird.

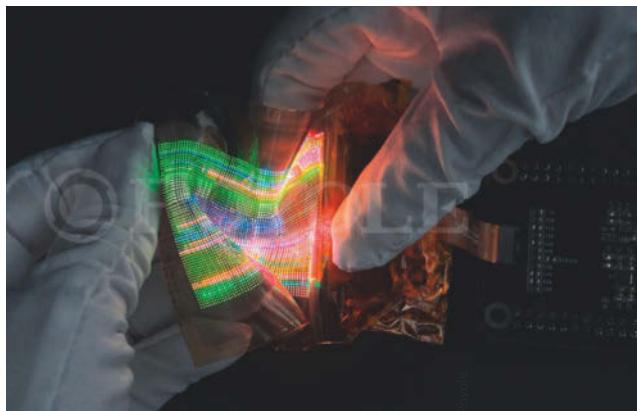
Schuld ist die extrem aufwendige Produktion der Mikro-LED-Displays. Die winzigen Dioden – sie sind zwischen wenigen Mikrometern bis 100 Mikrometer groß – müssen von 9-zölligen Saphir-Wafern auf das Displaysubstrat transferiert werden. Dieser Pick & Place genannte Vorgang ist langwierig, denn für ein 4K-Display müssen mindestens 24 Millionen



In besseren Fernsehern stecken auch in den kommenden Jahren OLEDs oder LCDs mit Mini-LEDs im Backlight; Displays aus Mikro-LEDs bleiben vorerst zu teuer.



Philips setzt wie viele andere TV-Hersteller auf WOLED-Panels von LG; Alleinstellungsmerkmal ist die in Philips-TVs integrierte Ambilight-Wandbeleuchtung.



Mit Mikro-LEDs lassen sich flexible und transparente Displays herstellen, da die Winz-LEDs kaum Fläche auf dem Trägersubstrat belegen.

LEDs transferiert werden, üblicherweise deutlich mehr, weil man pro Bildpunkt Reserven vorsieht. Weil der Prozess zugleich fehleranfällig ist, kann man keine großen Flächen bestücken. Deshalb werden die LED-TVs aus kleinen LED-Kacheln zusammengesetzt. Diese müssen nahtlos aneinanderstoßen und dürfen sich in ihren optischen Eigenschaften nicht wesentlich unterscheiden, weil man andernfalls die Kachelung störend wahrnehmen würde. Auf den einzelnen Kacheln erfolgt für eine homogene Ausleuchtung jeweils ein Helligkeitsabgleich und anschließend ein Gesamtabgleich für das zusammengesetzte Display.

Da jede Kachel eine feste Größe und LED-Anzahl hat, ist die Auflösung der Mikro-LED-TVs an die Schirmdiagonale gebunden: Je größer, desto mehr Pixel hat das Display. Erst ab 75 Zoll lässt sich derzeit 4K-Auflösung realisieren. Die LEDs erfordern hochkomplexe Backplanes und zur Ansteuerung noch mehr Transistoren als im OLED, was die Kosten ebenfalls in die Höhe treibt. Das klingt vorerst nicht nach einer erschwinglichen Displaytechnik, weshalb man mit dem nächsten Fernsehkauf besser nicht auf die Mikro-LED-TVs warten sollte. Die Technik wird derzeit für große Werbeflächen und in der Videoproduktion für riesige virtuelle Kulissen genutzt.

Selbstorganisierend und selbstleuchtend

Die Produktion winziger Mikro-LEDs funktioniert bereits, ihre Vereinzelung und Anordnung als Bildpunkte auf dem Displaysubstrat wirft aber noch Probleme auf. Perspektivisch könnten die Hersteller dafür selbst organisierende Techniken wie Fluid Self Assembly (FSA) nutzen. Dabei werden Millionen LEDs gleichzeitig mit einer Flüssigkeit auf das Substrat aufge-

bracht, wo sie selbstständig ihre Plätze finden. Auf dem Substrat befinden sich dazu an den Pixelpositionen winzige Vertiefungen aus Elektroden, die mit den Steuertransistoren des Displays verbunden sind. Über diese werden die LEDs nach der Positionierung elektrisch angeschlossen. Präsentiert wurden solche selbst organisierenden Verfahren auf der Display Week.

Außerdem gibt es verschiedene Verfahren, RGB-Mikro-LEDs zu erzeugen. Meist werden sie mit blau leuchtenden Mikro-LEDs und darauf abgebrachten farbkonvertierenden Nanopartikeln realisiert. Alternativ sind Mikro-LEDs denkbar, die je nach Ansteuerung rot, grün oder blau leuchten und über eine sequenzielle RGB-Emission wie einem DLP-Beamer Mischfarben erzeugen. Mit einem solchen Verfahren erregte das britische Start-up Porotech auf der Display Week Aufsehen.

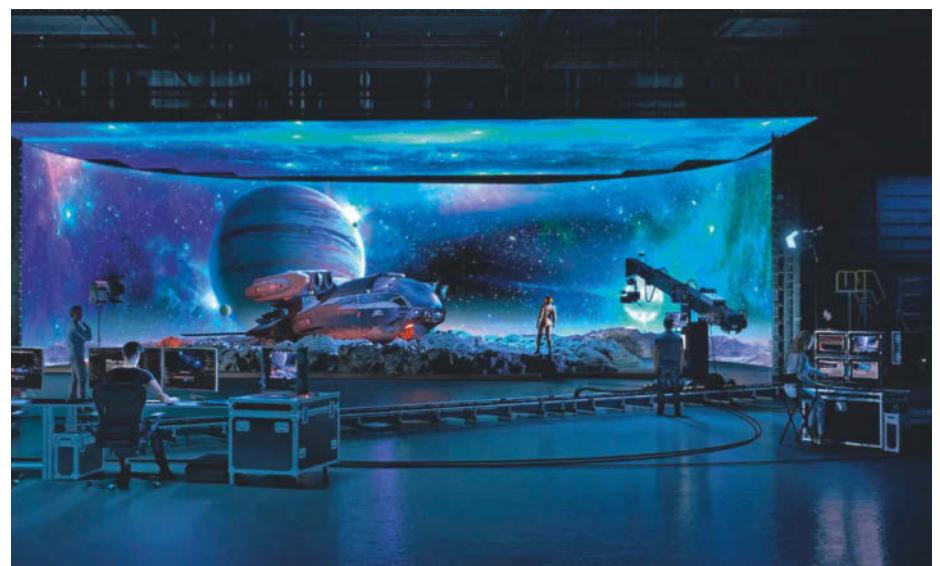
Ebenfalls noch im Forschungsstadium befinden sich selbstleuchtende Nanopartikel, also Quantenpunkte, die nicht phototonisch, sondern elektrisch zur Lichtemission gebracht werden. Die Elektrolumineszenz von Quantenpunkten würde kostspielige Vereinzelungsprozesse erübrigen, da man die Leuchtschicht wie beim OLED flächig auftragen kann. Bislang ist es aber nicht gelungen, farbstarke und stabile elektrolumineszierende QDs in den drei (RGB) Primärfarben zu entwickeln, die sich für eine Massenproduktion eignen.

Durchsichtig

Weil die Mikro-LEDs derart winzig sind, belegen sie auf großen Substraten sehr wenig Fläche. In der Folge können große Bildschirme aus Mikro-LEDs durchsichtig und auch flexibel sein, sofern sie auf dünnen Trägersubstraten sitzen und Polymer-elektronik nutzen. Damit schließen sie zu OLED-TVs auf, die es ebenfalls in transparenten Ausführungen gibt. Der Haken: Beide Varianten können kein Schwarz darstellen beziehungsweise es scheint der Hintergrund an den Stellen im Bild durch, die kein Licht emittieren, mithin schwarz wiedergeben sollen. Für künftige Videoschirme an Fensterscheiben ist das aber kein Problem: Entweder man schaut im Dunklen oder fährt ein Rollo hinter den transparenten Bildschirm. (uk@ct.de) **ct**

Literatur

- [1] **Ungeliebte Schattenbilder: Maßnahmen gegen das Einbrennen am OLED-TV**, c't 23/2018, S. 16



Sony nutzt seine Crystal-LED-Videowände als virtuellen Hintergrund in der Filmproduktion.



Bild: Berti Kolbow-Lehradt

Auf dem Weg zum besseren Smart Home

Der herstellerübergreifende Smart-Home-Standard Matter materialisiert sich langsam

Mit Unterstützung bekannter Markenhersteller und erster Gebäudeausrüster zieht Matter ein Jahr nach der Grundsteinlegung endlich in Smart Homes ein. Dennoch muss noch einiges geschehen, bis der neue Standard das Versprechen einlöst, die zersplitterte Automationswelt zu vereinen.

Von Berti Kolbow-Lehradt

Das im Oktober 2022 in Version 1.0 veröffentlichte Kommunikationsprotokoll Matter soll die abgeschotteten Herstellersysteme der Smart-Home-Welt miteinander verbinden. Dass die Tech-Giganten Amazon, Apple, Google und Samsung es vorantrieben, nährte von Beginn an die Erwartung, dass sich für die Nutzer die Frage schlagartig erübriggt, ob neue Geräte zu den vorhandenen Komponenten und Funkstandards passen.

Doch in den ersten Monaten nach der Premiere passierte wenig: Nur von einigen kleinen Herstellern kamen tatsächlich Matter-kompatible Geräte, viele Firmen beließen es bei Absichtserklärungen oder

verschoben versprochene Firmware-Updates – darunter Branchengrößen wie Philips Hue und IKEA. Mit Belkin zog sich ein Produzent sogar schon wieder zurück, bevor er überhaupt eine Komponente auf den Markt gebracht hatte. Zudem unterstützte Matter nur wenige Gerätekategorien und lief recht unrund, wodurch sich endgültig Enttäuschung breit machte. Es schien, als sei Matter am Ende, bevor es richtig angefangen hatte.

Mittlerweile hat die beteiligte Herstellergemeinschaft jedoch die Kurve gekriegt: Neue Matter-Spezifikationen haben technische Stolperfallen beseitigt und die Auswahl an Gerätekategorien erhöht. Viele

Neuvorstellungen von Zubehör stützen die Hoffnung, dass die Masse der Smart-Home-Produkte künftig Matter unterstützt. Aktualisierte Software für Bestandsgeräte bringt den Standard auch ohne Neukauf in die Haushalte. Und nicht zuletzt entfaltet der Einstieg von Branchengrößen die Sogwirkung, die ein neuer Standard benötigt, um sich durchzusetzen.

Hue-volution

Dass Matter die Smart-Home-Welt in kurzer Zeit komplett revolutioniert, glaubt inzwischen jedoch fast niemand mehr. Aber: Dass Philips Hue mit seinen smarten Leuchtmitteln inzwischen Matter unterstützt, darf man als wichtiges Signal dafür werten, dass sich das Zusammenspiel diverser Herstellermärken langsam, aber stetig verbessert.

Im September rüstete ein Firmware-Update die Hue-Bridge automatisch mit Matter auf. Dadurch reicht das LAN-Gateway gekoppelte ZigBee-Bestandsprodukte an ein Matter-Netzwerk durch. Dies gilt für hunderte Komponenten bis zur ersten Hue-Birne von 2012.

Um bei der Hue-Bridge die bisherige Verknüpfung mit einer Plattform wie Apple HomeKit oder Amazon Alexa durch die mit Matter zu ersetzen, muss man sie dort zuerst entkoppeln und dann durch das Einlesen eines QR-Codes neu verbinden. Andernfalls drohen Doppeleinträge mit Störfaktor.

Wer aktuell bereits mit dem Zusammenspiel der Systeme zufrieden ist, kann sich den Wechsel zu Matter zwar sparen. Anders sieht es jedoch aus, wenn die Hue-Bridge etwa Leuchtmittel von Drittmarken an Apple HomeKit durchreichen soll. Zudem lassen sich in Amazon Alexa und Google Home mehrere Hue-Bridges einbinden, was bei einer großen Lichtinstallation von Vorteil ist.

Die Klassiker kommen

Ähnlich wie Hue gelten auch die Marken AVM und Bosch als Klassiker im deutschen Smart-Home-Kosmos. Die Firmware-Updates für deren Schaltzentralen sind zwar noch nicht verfügbar, aber in Sichtweite. Bosch nennt aktuell „Anfang kommenden Jahres“ als neues Zeitfenster für das Update, nachdem es eigentlich schon in diesem Jahr erscheinen sollte. Dafür plant der Hersteller nun, nicht nur den „Smart Home Controller II“ um das Kommunikationsprotokoll zu erweitern, sondern auch dessen weiterhin verfügbaren Vorgänger. Mit Heizthermostaten,

c't kompakt

- Der herstellerübergreifende Smart-Home-Standard Matter ist nur langsam und mit Lücken gestartet.
- Mittlerweile sind Matter-Geräte verfügbar, die etwa Apple-HomeKit-Komponenten zugänglicher machen.
- Die kommenden Matter-Spezifikationen bringen weitere Gerätetypen und mehr Funktionen.

einem Bewegungs- und Öffnungssensor, einem Smart Plug mit Funkschalter sowie einem Unterputzrelais erhalten zunächst die am häufigsten eingesetzten Komponenten Matter-Support.

AVM will Ende 2023 zunächst eine Beta-Fassung an Experimentierfreudige verteilen. Sie erscheint sicherheitshalber nur für das weniger kritische ZigBee-Gateway und nicht auch gleich für die im WLAN-Heimnetz wichtigeren Fritzbox-Router. Mit stabiler Firmware für alle Zentralen ist frühestens im Laufe des kommenden Jahres zu rechnen. Dass erste Anwendungsfälle bereits funktionieren, zeigte AVM auf der IFA 2023. Dort ließ sich AVM-Zubehör auf einer Smartwatch nativ mit Google Home steuern, was bisher nur per Bastellösung geht.

Matter rückt auch vermehrt in den Fokus von Großinstallationen professioneller Gebäudeausstatter. Die vor allem bei Schaltkomponenten aktive Traditionsfirma Busch-Jaeger will bis Ende 2023 ihr System Access Point 2.0 per Software-Update nachrüsten. Da Gebäudeausstatter in sehr langen Produktlebenszyklen denken, ist der Einstieg ein großer Vertrauensvorschuss in die Zukunft von Matter.

Eine gemeinsame Installation von Busch-Jaeger und Eve demonstrierte auf der diesjährigen IFA, wie Matter die Brücke zwischen Gebäudesystemtechnik und Consumer-Produkten schlägt. Dabei ließen sich smarte Eve-Steckdosen und -Rollo über Touchdisplays und Schalter von Busch-Jaeger steuern, umgekehrt kommandierte die Smartphone-App von Eve Dimmschalter und Heizungsaktoren von Busch-Jaeger. Macht das Beispiel Schule, könnte Matter den Gestaltungsspielraum bei der Planung professioneller Systeme deutlich erweitern.

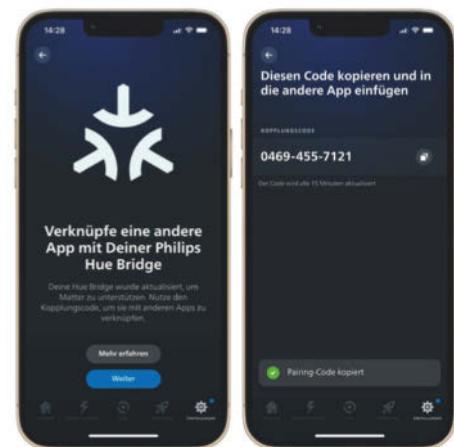
Das Bussystem von Busch-Jaeger ist nicht die einzige Matter-Option für Smart-Home-Profis: Passende Schaltkomponenten und Gateways gibt es etwa auch von Ubisys, Innovation Matters, Tridonic und Loxone – wobei der Miniserver letzterer Marke in puncto Matter aktuell noch im Beta-Status ist. Von Auftragsentwickler Mediola gibt es seit Juli eine passende Bridge für OEM-Hersteller.

Bridges und Controller

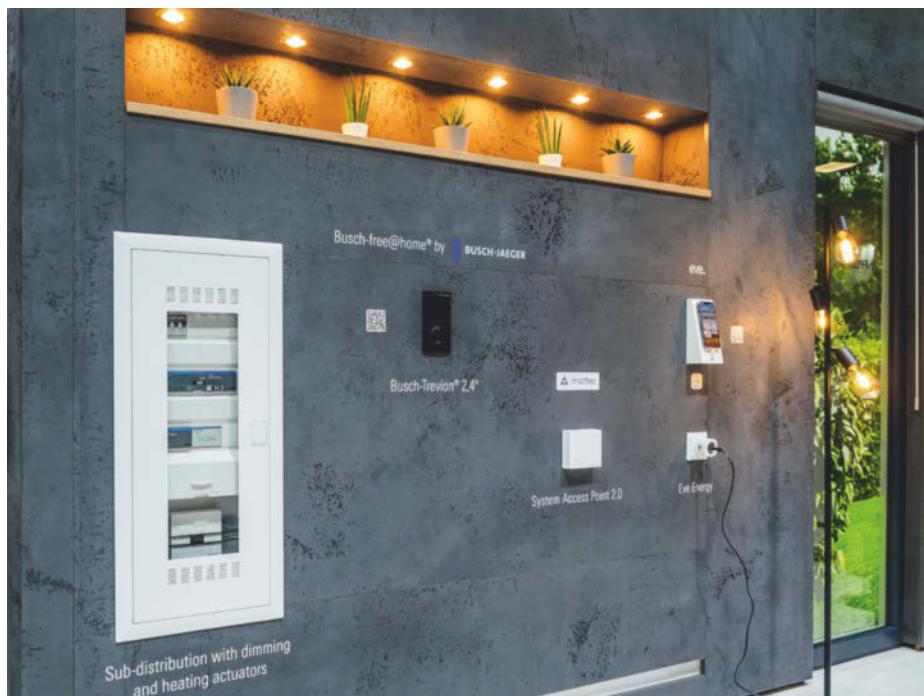
Bridges reichen gekoppeltes Zubehör nur an Matter-Plattformen durch, wenn letztere die Funkbrücken auch unterstützen. Amazon und Samsung SmartThings schalteten diese Fähigkeit im September frei, vorher akzeptierten nur Apple und Google Bridges. Somit bieten nun alle vier großen Ökosysteme die Möglichkeit, Bestandsgeräte etwa mit ZigBee, Z-Wave oder Bluetooth über Matter zu steuern. Weiterhin werden WLAN und Thread als Netzwerktechniken direkt unterstützt.

Auch die Auswahl bei sogenannten Matter-Controllern wird größer. Dabei handelt es sich um Schaltzentralen, die Geräte nicht nur an ein Netzwerk weitergeben, sondern diese auch steuern können. Mit den aktuellen HomePods, Apple TVs, diversen Echo-Geräten, dem SmartThings-Hub und dem Google Nest Hub 2 bietet das Markenquartett dabei viele Optionen.

Außerhalb der großen kommerziellen Ökosysteme eignet sich etwa der Homey Pro als Matter-Controller. Die Open-Source-Plattform Home Assistant erlaubt



Die Hue-Bridge lässt sich am Handy per Code auf Matter umstellen. Vorhandene Raum- und Leuchtmittelnamen werden automatisch übernommen, Szenen und Automatikregeln muss man aber von Hand anpassen.



Vorführung auf der IFA 2023: Geräte von Busch-Jaeger (links) und Eve (rechts) arbeiten dank Matter zusammen.

Wiz nutzen nun Matter, um sich die aufwendige Prüfung von Apple zu ersparen und den Zugang zum System zu erhalten. Umgekehrt öffnet Matter das bisher auf HomeKit beschränkte Zubehör von Eve für Nutzer, die auf die Plattformen von Amazon, Google oder Samsung setzen.

Von einem Standard, der eine universelle Steuerung über alle Systeme ermöglicht, ist Matter aktuell noch weit entfernt: Selbst wenn zwei Matter-Controller die gleichen Produkte akzeptieren, heißt das nicht, dass der Funktionsumfang identisch ist. Belegungen sowie verfügbare Sensorwerte und Automatiken können etwa weiterhin voneinander abweichen – wie schon vor Matter.

Untereinander harmonieren die Matter-Plattformen ebenfalls noch längst nicht perfekt. Statt eines gemeinsamen spannen sie mehrere separate Thread-Netze auf. Das ist eine seit Monaten klagende Baustelle, durch die sich die Mesh-Funktionen von Thread bislang nicht voll ausreizen lassen.

Doch alle diese Anlaufschwierigkeiten lassen sich beheben. Dass die Hersteller dies auch wirklich angehen, hat allerdings weniger mit Vorteilen für die Kundschaft zu tun als mit Unternehmenslogik: Matter verspricht Herstellern wie keine andere derzeitige Smart-Home-Technik eine große Kostenersparnis, weil sie sich künftig auf eine Schnittstelle konzentrieren können, statt diverse APIs zu pflegen. Die Kundschaft wird dies aber nicht stören, solange auch sie davon profitiert.

Bis man aber nicht mehr emsig nach Produkten, Funkstandards und Funktionen recherchiert, wird es weit über das kommende Jahr hinaus dauern. Jetzt schon Matter-Produkte unter den Weihnachtsbaum zu legen, schadet wiederum nicht: Sie funktionieren auch ohne Matter so, wie sie es bislang getan haben.

(nij@ct.de) **ct**

ebenfalls die Steuerung. In beiden Fällen befindet sich die Matter-Unterstützung allerdings noch im Beta-Stadium. TP-Link will noch im Dezember 2023 mit der Schaltzentrale Tapo H500 und später mit dem H900 zwei weitere sehr zugängliche und voraussichtlich preisgünstige Alternativen auf den Markt bringen.

Matter versteht bald Haushaltsgeräte

Einen Ausblick auf die geplante Entwicklung gab c't der Head of Technology der CSA, Chris LaPré. Die CSA (Connectivity Standards Alliance) definiert Standards wie Matter oder Zigbee. Nachdem Matter 1.0 eine Grundausrüstung für gängige Smart-Home-Geräte enthielt und das Update 1.1 im zurückliegenden Mai vorwiegend Bugs fixte, soll das turnusgemäße Herbst-Update die Peripherie stärker einbinden, das bei Erscheinen dieses Hefts verfügbar sein soll. Rund ein Dutzend Gerätetypen wurden laut LaPré von den zuständigen CSA-Gremien für Matter 1.2 getestet. Nur wenn sie ausreichend stabil funktionieren, schaffen sie es in die Veröffentlichung. Wie üblich werden einige davon noch weiteren Feinschliff und erneute Spezifizierungsrunden brauchen.

Eine hoch gehandelte Produktkategorie für Matter 1.2 sind Saugroboter; ebenfalls könnten Geschirrspüler, Wasch-

maschinen und andere „weiße Ware“ dazu zählen. Definitiv zu früh kommt das Herbst-Update für WLAN-Kameras und smartes Energiemanagement. Zusätzlich will man die bereits in Matter enthaltenen TV-Funktionen ausbauen. Alle drei Gerätetypen brauchen wegen des jeweils komplexen technischen Unterbaus noch Zeit und stehen frühestens mit Matter 1.3 im Frühjahr 2024 bereit. Und selbst bei grünem Licht vergehen anschließend weitere Monate, bis Hersteller die Vorgaben umsetzen und gebrauchsfertige Produkte in den Handel bringen. In den CSA-Gremien genießen laut Chris LaPré aber alle drei Gerätetypen Priorität.

Fazit

Matter ist ein Jahr nach dem Start also durchaus von der Stelle gekommen. Der Anreiz, schon jetzt das System darauf umzustellen, ist aber begrenzt. Denn läuft im Smart Home bereits alles rund, greift das Argument der vermeintlich besonders einfachen Einrichtung nicht. Interessant bleibt Matter vor allem für Neuanschaffungen.

Doch bislang haben sich die Kombinationsmöglichkeiten durch Matter noch nicht merklich gesteigert. Eine Ausnahme ist Apples HomeKit, das zuvor im Vergleich zu anderen Plattformen nur recht wenige Gerätetypen einbinden konnte. Hersteller wie Govee, SwitchBot und



Boschs Smart Home Controller II:
Ein Firmware-Update soll ihn und das Vorgängermodell Matter-kompatibel machen.

Bild: Berti Kolbow-Lehradt



Advanced Angular Day

Der Deep Dive für
Angular-Entwickler:innen

30. November 2023
Online

enterjs.de/advanced-angular.php



+++ Online-Workshops +++

Barrierefreiheit in der Webentwicklung (20.11.)

Web-Performance in Angular (1.12.)

Moderne Angular-Architekturen (4.12.)

Veranstalter





Innovationszwang

Smartphone-Trends: flacher und teurer, aber auch langlebiger und offener

Ein bisschen flacher, ein bisschen schneller: Den Smartphone-Herstellern fällt nur noch wenig Neues ein. Trotzdem tut sich bei den Geräten viel, denn neue Gesetze erzwingen eine längere Updateversorgung, freie App-Store-Wahl, ein Energie-Label und einiges mehr.

Von Robin Brand

Die neuen iPhones markieren in jedem Jahr die wohl am meisten beachtete Smartphone-Neuvorstellung. Über den Stand der Technik und das Wohlbefinden der Branche sagt die jeweils aktuelle Generation viel aus. Dass die Apple-Aktie nach der diesjährigen Vorstellung um 6,5 Prozent absackte, darf als Indiz gelesen werden, dass die vorgestellte Technik nicht zu überraschen wusste.

Nur zehn Prozent mehr CPU-Rechenleistung, etwas bessere Kameras und neues Gehäusematerial für die Pro-Modelle – das animierte Anleger nicht zu Freudensprünge. Die größte Neuerung war schon lange vorab durchgesickert: Die neuen iPhones sind mit USB-C-Buchsen

ausgestattet und Lightning ist damit beerdigt.

Eine Überraschung zauberte dagegen Google bei der Vorstellung des Pixel 8 aus dem Hut. Ganze sieben Jahre lang will das Unternehmen seine neuen Smartphones mit Sicherheits- und Funktionsupdates versorgen. Das ist weit mehr, als andere Hersteller für ihre Top-Smartphones versprechen, allein das auf Nachhaltigkeit bedachte niederländische Unternehmen Fairphone toppt das Google-Versprechen um ein Jahr.

Längere Updateversorgung und USB-C-Buchsen: Was die Branchengrößen Google und Apple vorstellten, klingt nicht nach technischer Revolution und darf

c't kompakt

- Foldables verkaufen sich immer besser, aber insgesamt schrumpft der Smartphonemarkt.
- Das liegt auch daran, dass die Geräte immer länger genutzt werden.
- Die EU zwingt Hersteller zu längerer Updateversorgung und mehr Wahlfreiheit bei App-Stores.

stellvertretend für die Smartphoneentwicklung betrachtet werden. Große Sprünge zwischen den Generationen gibt es nicht mehr. Bezeichnend, dass sowohl die USB-C-Buchse als auch das längere Updateversprechen maßgeblich von EU-Vorgaben motiviert wurden – der USB-C-Pflicht und der Ökodesign-Richtlinie.

Smartphones werden länger genutzt – und teurer

Aus Verbraucher- und Umweltsicht lässt sich dem Einiges abgewinnen. Werden die technischen Sprünge zwischen den Gerätegenerationen kleiner, gibt es wenig Anreiz, etwas Neues zu kaufen. So nutzen laut einer aktuellen Umfrage des Branchenverbands Bitkom elf Prozent der Verbraucher ihr Smartphone seit mindestens drei Jahren – fast dreimal so viele wie noch zwei Jahre zuvor.

Der Trend zur längeren Nutzung ist nicht auf Deutschland begrenzt und wird laut Einschätzung des weltweit führenden Kreditversicherers Euler Hermes weiter anhalten. Bis 2025 seien weltweit Umsätze in Höhe von 134 Milliarden US-Dollar in Gefahr. In Nordamerika erlebten die Smartphonehersteller gerade das schwächste Quartal seit über einem Jahrzehnt. IDC sagt für 2023 einen Absatz von rund 1,2 Milliarden Geräten voraus und für 2027 gut 1,3 Milliarden. Der Rekord von 2016, als 1,5 Milliarden Smartphones verkauft wurden, dürfte demnach noch eine Weile bestehen bleiben.

In Deutschland werden indes die geringeren Stückzahlen durch teurere Geräte ausgeglichen. Denn die Deutschen geben immer mehr Geld pro Smartphone aus: Durchschnittlich 563 Euro kostet ein neues Gerät hierzulande nach Bitkom-Zahlen, Tendenz weiter steigend. Und auch weltweit muss man sich keine Sorgen um die Branche machen – der Abgang

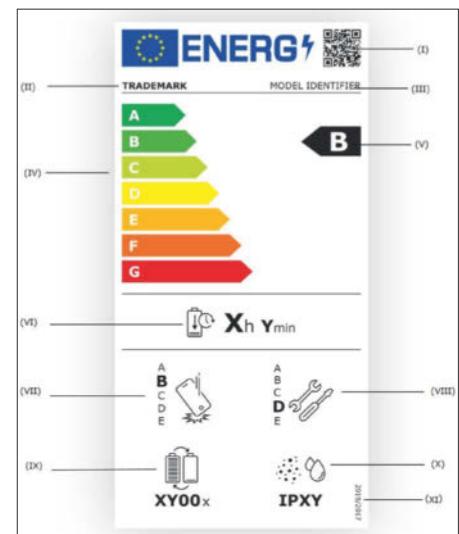
aufs Smartphone hat sich noch jedes Mal als verfrüht erwiesen. Sowohl die tägliche Nutzungszeit als auch der Markt der Apps und Dienste wachsen weiter stetig.

Dass in Zukunft Smartphones womöglich noch deutlich länger genutzt werden, dafür sollen Pläne der EU sorgen, auf die viele Hersteller schon jetzt unter anderem mit längeren Updateversprechen reagieren, siehe Google. Die Ökodesign-Anforderungen für Smartphones und Tablets wurden am 31. August 2023 im EU-Amtsblatt veröffentlicht. Die Veröffentlichung ist der letzte Schritt, bevor die neuen Regeln nach einer Übergangsphase in der gesamten EU am 20. Juni 2025 anwendbar werden.

Die Ökodesign-Regeln machen Vorgaben zur Updatepolitik der Hersteller, zur Reparier- und Recyclingfähigkeit der Geräte und zur Zuverlässigkeit der verwendeten Bauteile. Sie gelten – hier und da leicht abgeschwächt – auch für Tablets. Ab Inkrafttreten müssen Hersteller etliche Ersatzteile wie Akkus, Kameras, Lade- und Audiobuchsen, mechanische Tasten sowie Mikrofone und Lautsprecher an freie Werkstätten liefern. Wo Reparaturbetriebe in vielen Fällen derzeit auf den Graumarkt angewiesen sind, um Ersatzteile aufzutreiben, können sie diese künftig beim Hersteller direkt ordern. Ob die Reparaturszene mit den Regeln gestärkt wird, bleibt aber abzuwarten. Denn die EU macht keine Vorgaben zu den Ersatzteilpreisen und auch die Praxis, Bauteile per Mikrocontroller mit dem Gerät zu koppeln, unterbindet sie nicht.

Akkus wechselbar oder langlebig

Dennoch dürften Reparaturen günstiger und einfacher werden. Denn vielen hochintegrierten Bauteilen macht die EU ein Ende. Dass zum Beispiel eine ausgeleierte Ladebuchse irreparabel ist, weil sie aufs Mainboard gelötet ist, gehört dann der Vergangenheit an. Der Akku muss von



So soll das Energielabel für Smartphones aussehen: Es soll auf einen Blick zeigen, wie lang Smartphones mit einer Akkuladung durchhalten und wie robust sie sind.

Laien wechselbar sein, wenn er und das Gehäuse nicht bestimmte Anforderungen erfüllen (83 Prozent Restkapazität nach 500 Ladezyklen, Einstufung IP67).

Außerdem müssen Hersteller Reparaturbetrieben unter anderem Zerlegungspläne, Explosionszeichnungen, Schaltpläne sowie für die Fehleranalyse erforderliche Verkabelungs- und Anschlussdiagramme zugänglich machen und zudem eine Liste der notwendigen Werkzeuge und Prüfgeräte sowie eine Reparaturanleitung zur Verfügung stellen – und zwar innerhalb eines Arbeitstages nach Anfrage. Werkstätten haben bis zu sieben Jahre nach Ende des Inverkehrbringens eines Geräts Anspruch darauf.

Funktions- und Sicherheitsupdates müssen Hersteller für einen Mindestzeitraum von fünf Jahren liefern – Sicherheitspatches spätestens vier Monate nach Bereitstellung, Funktionsupdates spätestens sechs Monate danach. Flankiert werden



Die wohl größte Neuerung gegenüber den Vorgängern: Das iPhone 15 trägt eine USB-C-Buchse. Technisch dagegen enteilt es den 14ern nicht.



konstruieren, verwendet der Hersteller nach eigenen Angaben unter anderem einen speziellen Silizium-Kohlenstoff-Akku mit einer durchschnittlichen Dicke von nur 2,72 Millimetern und ein besonders filigranes Titanscharnier aus dem 3D-Drucker.

Aller technischen Finesse zum Trotz scheint Honor selbst dem Gerät den Durchbruch auf dem europäischen Markt nur bedingt zuzutrauen. Denn hierzulande kommt es mit erheblicher Verspätung. Im Juli in China vorgestellt, kündigte der Hersteller das Gerät erst im September auf der IFA für Europa an – und ließ die beiden wichtigsten Fragen noch immer unbeantwortet: wann und wie teuer? Damit ist Honor nicht allein: Auch Oppo und Xiaomi sehen im asiatischen Raum größeres Potenzial. Einzig Motorola und Samsung bedienen die westlichen Märkte nicht mit Verzögerung.

Wann betritt Apple den Foldable-Markt?

In naher Zukunft dürfte sich das nicht grundlegend ändern, denn speziell chinesische Verbraucher sind aufgeschlossen, was die verhältnismäßig junge Technik angeht. Satte 117 Prozent Wachstum verzeichneten die Analysten von Counterpoint Research auf dem chinesischen Markt im ersten Quartal 2023 verglichen mit dem Vorjahreszeitraum – und für die Zukunft prognostizieren sie weiteres Wachstum. Von weltweit rund 2,5 Millionen verkauften Foldables im ersten Quartal entfielen mehr als 1 Million auf China. Der deutlich kleinere europäische Markt wuchs um gerade einmal – oder angesichts eines insgesamt schrumpfenden Smartphonemarkts um immerhin – 17 Prozent. Doch mehr verkaufte Foldables hin oder her: Noch ist nicht einmal jedes 100. verkauftes Smartphone ein faltbares.

Der dominierende Hersteller im Foldable-Markt bleibt Samsung, dessen Neuauflagen von Flip und Fold eher konservativ geraten sind. Angesichts der erstarkten Konkurrenz um Honor und Xiaomi dürfen wir uns von der sechsten Generation der Samsung-Falter im kommenden Jahr deutlich mehr erwarten. Ein Hersteller, der dem Foldable-Markt insgesamt einen Schub geben könnte, ist – wie sollte es anders sein – Apple. Doch noch tischt die Gerüchteküche allenfalls lauwarme Speisen auf, heiß dürfte es in naher Zukunft allenfalls mit einem faltbaren iPad werden.

(rbr@ct.de) 

Um das Foldable Magic V2 besonders dünn zu bauen, verwendet Hersteller Honor unter anderem ein Titanscharnier aus dem 3D-Drucker.

die Maßnahmen von einem Energie-Label, das über Energieeffizienz, Langlebigkeit des Akkus, Schutz vor Wasser und Staub sowie Widerstandsfähigkeit gegen Sturzschäden informiert.

Zwar werden die Regeln erst im Juni 2025 in Kraft treten, Auswirkungen auf den hiesigen Smartphonemarkt dürften sie jedoch schon früher haben. Speziell günstige Smartphones, deren Akkus sich in der Regel nicht tauschen lassen und deren Gehäuse nicht wasserdicht sind, geraten gegenüber den Einsteigermodellen, die nach Inkrafttreten der Regeln erscheinen, ins Hintertreffen. Ein kostengünstiges Einsteigergerät mit unklaren Update-Aussichten lässt man bis dahin also besser links liegen.

Freie App-Store-Wahl

Neben den Ökodesign-Regeln wird eine weitere EU-Maßnahme die Smartphonebranche verändern: der Digital Markets Act (DMA). Sechs Unternehmen – Alphabet, Amazon, Apple, Bytedance, Meta und Microsoft – hat die EU als Gatekeeper eingestuft. Bis zum 6. März 2024 müssen sie erste Vorgaben aus dem DMA umsetzen. So sollen Nutzer ihre Software frei wählen dürfen – nicht deinstallierbare Apps sollen damit der Vergangenheit angehören. Allerdings gibt es eine nennenswerte Einschränkung: Um die Sicherheit der Nutzer zu gewährleisten, sollen die Anbieter weiterhin Möglichkeiten „zum Schutz der Integrität des Systems“ anbieten dürfen. Hier wird sich in den kommenden Monaten zeigen, ob Samsung, Google & Co. in der Android-Welt diesen Passus dazu wer-

den nutzen können, doch Apps und Dienste auf den Geräten zu installieren, die sich nicht entfernen lassen.

Für iPhone-Nutzer zeichnen sich größere Änderungen ab. Der DMA zielt speziell auf den App Store ab: Apple muss den Regeln zufolge künftig andere App-Läden auf dem iPhone dulden, das war bisher undenkbar. Erste Anbieter stehen dafür schon in den Startlöchern. Auch das auf iPhones derzeit nur sehr eingeschränkt mögliche Sideload von Apps dürfte dank DMA bald Normalität werden. Ob und wie Apple das Schlupfloch „Sicherheit der Plattform“ zu nutzen weiß, um unliebsame Drittanbieter doch wieder auszusperren, bleibt aber auch hier abzuwarten.

Foldables in der Millimeter-Ära

Während die EU die Rahmenbedingungen verändert, arbeiten sich die Hersteller an Millimetern und Prozenten ab. Zehn Prozent Mehrleistung eines iPhones locken allerdings niemanden so recht hinterm Ofen hervor, wohingegen ein paar Millimeter bei den Foldables durchaus einen Unterschied machen. Honor hat mit dem Faltphone Magic V2 mit markigen Worten die „Millimeterära“ der Foldables eingeläutet und meint damit, dass das Gerät auch zusammengeklappt weniger als einen Zentimeter dick ist: 9,9 Millimeter, um genau zu sein. Es ist damit das dünne Foldable auf dem Markt. In der Hand liegt das zusammengeklappte Gerät – anders als die meisten etwas korpulenteren Konkurrenten – wie ein herkömmliches Smartphone. Um es so dünn zu

TAUCHE EIN IN DIGITALE WELTEN – MIT DEM c't DIGITALABO



40%
Rabatt!

c't MINIABO DIGITAL AUF EINEN BLICK:

- 6 Ausgaben digital in der App, im Browser und als PDF
- Inklusive Geschenk nach Wahl
- Mit dem Digitalabo Geld und Papier sparen
- Zugriff auf das Artikel-Archiv



Jetzt bestellen:

ct.de/angebotdigital





Bild: orbcat/stock.adobe.com | KI Midjourney | Bearbeitung: c't

Bahn frei für 5G überall

Globale Mobilfunkabdeckung: Wie das geht, welche Smartphones sich eignen

Die Mobilfunkbranche schickt sich an, sämtliche Versorgungslücken der Welt mittels Satelliten zu stopfen. Und um die Satelliten anzuzapfen, braucht man kein Spezial-Equipment, es genügen handelsübliche Smartphones. Die Antworten auf das Warum und Wie illustrieren, wie geschickt sich der Mobilfunk weiterentwickelt.

Von Dušan Živadinović

Wer hätte gedacht, dass die Mobilfunkbranche die gesamte Erde mit 5G-Diensten versorgen kann, obwohl sie mit ihren Basisstationen bisher nicht einmal ein Flächenland, geschweige denn einen Flächenstaat zu 100 Prozent abdeckt. Beispielsweise bleibt die 5G-Versorgung in Deutschland lückenhaft, auch wenn sie seit Januar bis zum Sommer laut der Bundesnetzagentur um gut 5 Prozent auf 89 Prozent zugenommen hat. Selbst die ältere LTE-Technik deckt nur 97,2 Prozent des Landes ab.

Wenn schon die Abdeckung auf dem Boden schwerfällt, wie soll dann eine ungleich schwieriger erscheinende Versor-

gung per Satellit klappen? Und wie soll sich 5G per Satellit gegenüber etablierten Diensten wie Starlink, Globalsat oder Iridium durchsetzen?

Schon mit dem Satellitenbetrieb hat kaum ein Netzbetreiber Erfahrung. Hinzu kommen technische Hürden wie vielfach größere Distanzen (hunderte Kilometer zum Satelliten gegenüber höchstens rund zehn Kilometern zu Basisstationen auf der Erde), außerdem beispielsweise längere Signallaufzeiten und viel geringere Empfangspegel.

Doch seit den ersten kommerziellen Raketenstarts von SpaceX Mitte der 2010er Jahre nimmt die Transportkapazi-

tät von Raketen stark zu und zugleich fallen die Startkosten. Davon angezogen, kommen immer mehr Konzerne auf den Geschmack und planen eigene Satellitenkonstellationen zur Internetversorgung. Neben Starlink gehören OneWeb und Amazons Project Kuiper dazu. Das dürfte auch die Mobilfunker wachgerüttelt haben.

Letztlich genügen sogar nur wenige Konzepterweiterungen, um die 5G-Spezifikation mit Satelliten fit zu machen. Die wichtigsten stellen wir vor.

Orbitale 5G-Erweiterung

Wie so oft bei komplexen Problemen ist auch hier die Lösung mehrstufig: Zunächst mal geben sich die Mobilfunker mit niedrigen Datenraten von einigen zehn Mbit/s zufrieden, Hauptsache, man erreicht jeden Winkel der Erde. Gigabit gibts für Smartphones nach wie vor nur in Ballungsgebieten.

Der Rest ist geschickte Arbeitsteilung: Die Mobilfunkbranche entwickelt im Standardisierungsgremium 3GPP (3rd Generation Partnership Project) die Konzepte, baut auf dieser Grundlage die Funkhardware mitsamt Verbindungen zu ihren erdgestützten Netzen. Die Satellitenbranche baut und schießt die fliegenden Basisstationen in die Umlaufbahn und verwaltet sie. Am Ende können Netzbetreiber wie O2 Telefónica, Telekom und Vodafone die Funkressourcen in der Umlaufbahn mieten und so Kunden flächendeckende Versorgung anbieten.

Dabei wird das weltraumgestützte 5G die bodengestützten (terrestrischen) Netze nicht ersetzen; dafür ist die Signaldämpfung aufgrund der langen Strecken zu hoch und somit die Datenraten viel niedriger. Deshalb wird man die Satelliten mit Smartphones nur unter freiem Himmel anzapfen können. Aber die Güte genügt, um sämtliche weiße Flecken zu schließen.

Das 3GPP hat sich seit 2017 an das Thema herangetastet, anhaltend dadurch angeregt, dass konventionelle Mobilfunkversorgung vielerorts fehlt, aber auch abseits von Ballungszentren nachgefragt ist. Beispiele sind Flugzeuge oder auch küstenerne Plattformen wie Windkraftanlagen. Auch löcherige Abdeckungen will die 3GPP bekämpfen, um etwa nach Katastrophen kontinuierliche Mobilfunkversorgung für Rettungsdienste zu gewährleisten oder der Landwirtschaft eine Infrastruktur für IoT-Anwendungen zu bieten (z. B. Messung der Lufttemperatur, Luft- und Bodenfeuchte, siehe ct.de/ynhb). So lieferte sie ab 2019 bis 2022 mit der 3GPP-

Spezifikation Release 17 die genauen Rahmenbedingungen für eine weltweite Abdeckung per Satellit. Das 3GPP fasst sie unter dem Begriff nicht-terrestrische Netze zusammen (NTN).

Die NTN-Spezifikationen haben sich längst in Laborversuchen bewährt, darunter in Experimenten vom Netzwerkzulieferer Ericsson, Chiphersteller Qualcomm und Sat-Betreiber Thales im Jahr 2022. Die Firma Skylo setzt die Spezifikation bereits um. Abgesehen von den Polkappen deckt sie die Erde über geostationäre Satelliten mit schmalbandigen Diensten ab. Die Deutsche Telekom und O2 Telefónica wollen auf Basis der Skylo-Dienste IoT-Tarife für Endkunden ab Ende 2023 anbieten (ct.de/ynhb). Und Bullit Satellite bietet über Skylo und eine eigene App Kurznachrichten- und Notrufdienste ab 5 US-Dollar pro Monat an.

Der größte LEO-Satellit

Größere Aufmerksamkeit erhielt im Juni 2023 die texanische Firma AST mit ihrem Satelliten BlueWalker 3. Der belieferte im Zusammenspiel mit Netzwerkelementen von Nokia Bell Labs ein Samsung Galaxy S22 mit immerhin 10,3 Mbit/s. Das genügt für die Bild-, Text- sowie Sprachübertragung und selbst für schmalbandige Vi-

ct kompakt

- Die Mobilfunkversorgung aus dem All birgt einige sehr spezielle Herausforderungen.
- Die Mobilfunkbranche hat dafür elegante Lösungen mit kleinem Aufwand gefunden.
- Im Fokus steht die Versorgung von Smartphones, aber auch IoT-Anwendungen profitieren.

deos. Vodafone gelangen später in einem ähnlichen Szenario rund 14 Mbit/s.

Das verwundert, denn gängige Smartphones sind nur für Distanzen von wenigen Kilometern ausgelegt. Die Versorgung aus dem All macht erst das besonders empfindliche Antennen-Array des Blue-Walker 3 möglich. Sein 64 Quadratmeter großes Antennen-Array macht ihn zum bisher größten LEO-Satelliten. Im Gehäuse hinter der Gruppenantenne steckt eine modifizierte Mobilfunkbasis.

Ende 2023 will T-Mobile USA zusammen mit SpaceX Beta-Tests aus dem All starten, um die USA mitsamt Alaska, Hawaii und Puerto Rico mittels neuen Starlink-Sa-

Die richtige Umlaufbahn

GEO-Satellitensysteme, abgeleitet von „**geostationary orbit**“ sind fest über einem Punkt des Äquators positioniert und versorgen von dort aus immer dieselbe Fläche. Weil sie sich in rund 36.000 km Entfernung befinden, sind die Signallaufzeiten lang (> 500 ms) und die Datenraten sehr niedrig. Man braucht aber nur wenige GEO-Satelliten, um große Flächen abzudecken. Deshalb eignen sie sich zum Beispiel gut für Satellitenfernsehen.

Medium Earth Orbit, MEO-Systeme wie Galileo, GPS und GLONASS setzt man hauptsächlich für die Navigation ein. Die Satelliten umkreisen die Erde in etwa 20.000 Kilometern Höhe in rund zwölf Stunden. Gegenüber GEOs verkürzt das die Latenz auf ein Fünftel und ermöglicht höhere Datenraten. Daher nutzt man MEO-Konstellationen auch für Kommunikationsdienste. Ein Beispiel ist das vom Luxemburger Unternehmen SES betriebene O3B-Netzwerk, das aus einer Höhe von 8063 km Mobilfunk- und Internet-

provider mit Sprach- und Datendiensten versorgt. MEO-Satelliten decken auch den Nord- und Südpol ab.

Starlink, OneWeb, Iridium oder auch Globalstar nutzen **Low-Earth-Orbit**-Satellitensysteme (LEO) in 400 bis 2000 Kilometern Entfernung und auch 5G-NTN-Satelliten wie die von AST werden in niedrigem Orbit um die Erde sausen. Die Geschwindigkeit beträgt rund 8 km/s, sodass ein Umlauf in 90 bis 120 Minuten erfolgt. LEO-Satelliten bieten die geringste Latenz und liefern prinzipiell hunderte Megabit pro Sekunde. Auf der Teilnehmerseite sind dafür proprietäre Sat-Anlagen wie die von Starlink erforderlich. Aber LEO-Satelliten eignen sich auch für mobiles Breitband. Wegen kleinerer Antennen und begrenzter Energie von Smartphones bleiben die Datenraten deutlich unter 100 Mbit/s. Da die Ausleuchtzone im Vergleich zu MEO und GEO deutlich kleiner ist, braucht man zur Abdeckung großer Flächen mehr Satelliten, also größere Konstellationen.

telliten abzudecken. Zu Beginn will T-Mobile 2 bis 4 Mbit/s für Text-Messaging über Smartphones anbieten, die auf dem terrestrischen Funkband n25 einrasten können (1,9 GHz). Viele Smartphones eignen sich bereits dafür und können dann im Roaming-Betrieb in den USA über das T-Mobile-Netz auch den Satellitendienst nutzen.

Der Ansatz hat aber eine Schwäche: Terrestrische Bänder wie n25 sind nicht überall auf der Welt für Mobilfunk reserviert. Deshalb wird der Starlink-Dienst in vielen Ländern nicht angeboten, auch in Deutschland nicht. Die 3GPP favorisiert daher separate, nicht-terrestrische Frequenzen, für die die Hersteller die nächste Smartphone-Generation vorbereiten.

Frequenzfutter

Im Release 17 sind im L- und S-Band die Frequenzen von 1626,5 bis 1660,5 MHz sowie 1525 bis 1559 MHz für die Senderrichtung zum Satelliten (Uplink) und für die Empfangsrichtung reserviert (5G-Band n255). Das ebenso definierte 5G-Band n256 reicht von 1980 bis 2010 MHz (Uplink) und von 2170 bis 2200 MHz (Downlink). Beide, n255 und n256 sind als Frequency Division Duplex organisiert und je Richtung rund 30 MHz breit. Das ist im Vergleich zu terrestrischen Netzen unterdurchschnittlich, da sind leicht 100 MHz und mehr drin.

Wie am Boden, so müssen sich auch beim Mobilfunk per Satellit alle Nutzer das Spektrum teilen. So erklärt sich, dass man pro Zelle in Empfangsrichtung im Mittel nur einige zehn Mbit/s erwarten kann, maximal sind es 70 Mbit/s. Das ist ein

guter Anfang und immerhin kommen damit sogar kleine Videos ruckelfrei an. In Senderichtung geht bestenfalls 2 Mbit/s, weil Smartphones weit weniger Leistung und kleinere Antennen als Satelliten haben. Besserung ist aber unterwegs. Ab Release 18 kommt für den Uplink im L-Band der Bereich von 1610 bis 1626,5 MHz hinzu und für den Downlink im S-Band der Abschnitt von 2483,5 bis 2500 MHz.

Generell liefert ein 5G-Satellit den höchsten Durchsatz, wenn er senkrecht nach unten sendet. Je mehr er sich im Verhältnis zum Teilnehmer neigt, desto mehr sinkt der Durchsatz. Bei einem Winkel von 30 Grad schumpft der Durchsatz auf die Hälfte.

Damit sich ein Smartphone in ein NTN einbuchen kann, muss es zuerst seine Position per GNSS-Modul (Global Navigation Satellite System für Galileo, GPS, Glonass oder Beidou) bestimmen. Damit und unter Verwendung von Netzwerkinformationen ermittelt es, welcher Mobilfunksatellit seinen aktuellen Standort gerade versorgt. Dafür erhält er vom terrestrischen Netzwerk die genauen Flugdaten. Um dann Kontakt aufzunehmen, berechnet es die Distanz und relative Geschwindigkeit des Satelliten und stellt die für ihn erforderlichen Funkparameter ein.

Die dafür erforderlichen Modem- und Antennenanpassungen bauen die Hersteller erst in kommende Smartphones ein.

Netzarchitektur

Die Architektur der nicht-terrestrischen 5G-Netzwerke birgt einige interessante

Eigenschaften. Die wichtigste ist, dass 5G-Satelliten für mobiles Internet die Erde in niedrigen Orbits umkreisen müssen, denn je niedriger die Umlaufbahn, desto kürzer die Signallaufzeit. Für mobiles Internet möchte man Laufzeiten möglichst unter 50 Millisekunden halten. Davon profitieren Nutzer beim Surfen, weil sich Webseiten schneller aufbauen und bei Telefonaten fallen sich Gesprächspartner nicht gegenseitig ins Wort (ab 200 Millisekunden nimmt die Wahrscheinlichkeit dafür deutlich zu).

Deshalb sind LEO-Satelliten beliebt. Im Mai 2023 umkreisten die Erde bereits 7004 LEO-Satelliten; es kommen noch viel mehr hinzu. Allein die Starlink-Flotte, mittels der die Firma SpaceX vor allem abgelegene Gebiete mit Internetdiensten versorgt, soll in der Endkonfiguration aus 40.000 Satelliten bestehen. Weitere Megakonstellationen sind in Arbeit.

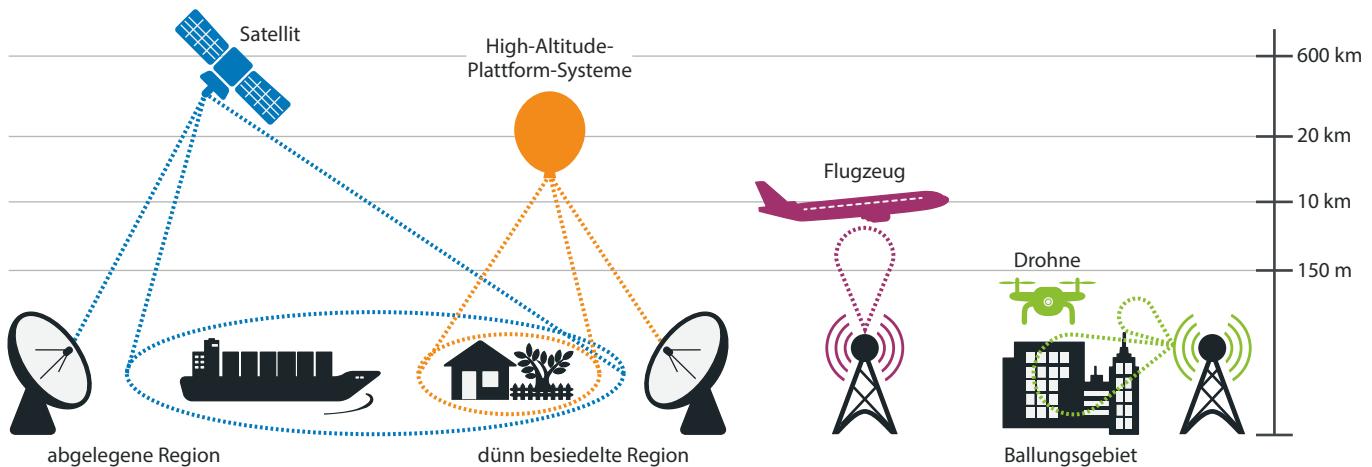
Die meisten bisherigen LEO-Systeme verwenden selbst entwickelte Technik, also auch proprietäre Teilnehmergeräte. Das treibt die Entwicklungs- und Wartungskosten in die Höhe. Es wird interessant zu beobachten sein, ob und wie die Betreiber die 5G-Technik aufgreifen, denn damit würden sie auf einen Schlag alle 5G-Mobilfunkteilnehmer der Welt ansprechen und könnten sich Überzeugungsarbeit für den Kauf ihrer speziellen Empfangsgeräte sparen.

Drei Entwicklungspfade

Genau besehen besteht die Entwicklung von mobilen Satellitenservices (MSS) aus drei Entwicklungspfaden: Legacy MSS,

Kommunikationsdienste aus der Luft

Abhängig von der Höhe der Erdumlaufbahn (GEO, MEO, LEO) versorgen Satelliten unterschiedliche Flächen. Die High-Altitude-Plattform-Systeme (HAPS) versorgen Städte oder kleine Regionen und fliegen doppelt so hoch wie kommerzielle Flugzeuge.



Legacy Long-Term Evolution (LTE) und 5G NTN.

Legacy-MSS zielt darauf, ältere Technik in neue Smartphones zu integrieren. Beispielsweise nutzen iPhones ab der 14er-Reihe Globalstar-Satelliten für Notrufe. Das Huawei Mate 50 sendet und empfängt über Beidou-Satelliten Textnachrichten, während Android-Smartphones mit Qualcomm's Chip Snapdragon Satellite über Iridium-Satelliten kommunizieren.

Unter Legacy-LTE sind unmodifizierte LTE-Smartphones zusammengefasst, die mit LEO-Satelliten über terrestrische LTE-Frequenzen kommunizieren (Direct-to-Smartphone). Diese Technik, die etwa Starlink vorbereitet, hat das 3GPP 2022 in Release 17 parallel zu 5G-NTN spezifiziert.

Im 5G-NTN unterscheidet man zwei Architekturvarianten: Bei der transparenten Architektur befindet sich die Basisstation (gNodeB oder gNB) am Boden hinter dem Gateway und der Satellit arbeitet als Repeater. In der regenerativen Architektur enthält der Satellit entweder eine komplett-

te Basisstation oder Teile davon, etwa die Einheit, die die Daten entschlüsselt und verarbeitet, siehe Infografik „Zwei NTN-Architekturen“.

Die transparente Architektur ist kostengünstiger, die regenerative Architektur bietet mehr Flexibilität und Leistung. Damit lässt sich eine globale Abdeckung mit nur wenigen Bodenstationen aufbauen, weil sie sich auch für Inter-Satelliten-Verbindungen eignet. Satelliten ohne Sicht zu einer Bodenstation leiten den Verkehr untereinander weiter, bis einer von ihnen die Daten hinabreichen kann.

Hürden und elegante Umgehungen

Moderne Satelliten unterteilen ihr Versorgungsgebiet in der Regel in hunderte Teilgebiete, die sie mit einzelnen Strahlen (Spotbeams) abdecken. Jedes Gebiet entspricht einer Zelle von zehn bis einigen hundert Kilometern Durchmesser.

Wenn LEO-Satelliten die Abstrahlungskeule in einem festen Winkel auf die

Erde richten, überstreichen sie die Oberfläche mit rund 8 km/s und wandern deshalb so schnell weiter, dass sich Empfangsgeräte alle paar Sekunden in den nächsten Satelliten einbuchen müssen (Handover). Das saugt am Akku und ist deshalb nicht wünschenswert. Um die Handover-Häufigkeit zu senken, können Satelliten dieselbe Zelle mit einem schwenkbaren Strahl mehrere Minuten lang ausleuchten (earth-fixed beams), bevor der nächste übernimmt. Das Release 17 spezifiziert beide Methoden.

Zwei große Herausforderungen sind die gegenüber ergebundenen Übertragungen längeren Signallaufzeiten (Latenz) und Frequenzverschiebungen durch den Dopplereffekt. Die terrestrischen Netzwerke sind zwar von Anfang an für mobile Teilnehmer ausgelegt, aber eben für irdische Entfernung und Geschwindigkeiten – etwa zu Fuß, per Fahrrad, Auto oder Zug.

Doch bei 5G-Satellitennetzen bewegen sich auch die Basisstationen im Orbit

IT-Stipendium Wintersemester 2023/2024

 **e-fellows.net**

heise medien fördert Studierende im Fach INFORMATIK

Zusammen mit dem Karrierenetzwerk e-fellows.net schreiben die heise-Marken c't, iX und heise online ein Stipendium für IT-Studierende aus.

Sicher Dir 600 Euro für das Wintersemester!

Zusätzlich spendiert heise

- einen Schnuppertag in der jeweiligen Redaktion
- ein E-Learning-Kurs nach Wahl bei der heise Academy
- sowie ein Jahresabonnement von heise+
- Zudem werden die drei Stipendiaten zu der Sicherheitsmesse secIT eingeladen

**Bewerbungsschluss:
19. November 2023**

Jetzt in zehn Minuten bewerben:
www.e-fellows.net/IT-Stipendium



 **heise medien**



und zwar rasant. Das führt zu erheblich größeren Verschiebungen der Mittenfrequenz wegen des Dopplereffekts als bei Auto- oder Zugfahrten: Ähnlich einer Sirene, die beim Herannahen eines Feuerwehrfahrzeugs höher klingt und beim sich entfernenden Fahrzeug tiefer, so ändert sich auch die Frequenz der Satelliten, je nach dem, ob sie sich nähern oder entfernen. Der Unterschied kann bei der Trägerfrequenz von 2 GHz bis zu 50 kHz ausmachen. Aber nur, wenn ein Empfänger auf der richtigen Frequenz einrastet (also z. B. auf 2,05 GHz), kann er die auf einer Funkwelle aufmodulierten Daten korrekt lesen.

Die 3GPP löst dieses Problem elegant: Die Endgeräte bekommen vom Netz eine Hand voll Informationen, um die Dopplerverschiebung selbst zu kompensieren. So bleibt die komplette Netzwerkinfrastruktur unverändert.

Dafür sendet der Satellit seine Position und Geschwindigkeit (Ephemeriden) und ein Teilnehmergerät, das sich einbuchen will, berechnet zunächst seine eigene

Position. Dafür nutzt es sein GNSS-Modul. Dann berechnet es den Abstand und die Relativgeschwindigkeit zum Satelliten und schließt daraus auf die erforderlichen Anpassungen zur Kompensation des Dopplereffekts. So sind an den 5G-Basisstationen keine funkseitigen Änderungen erforderlich, sie senden mit der nominalen Frequenz ohne Rücksicht auf mögliche Frequenzverschiebungen.

Zusätzlich muss die 5G-Spezifikation die gegenüber der terrestrischen Architektur deutlich verlängerte Signalaufzeit berücksichtigen. Innerhalb von terrestrischen Netzen sind Laufzeiten um 1 Millisekunde üblich, bei LEO-Satelliten kommen je nach Höhe einige zehn Millisekunden zusammen (bei 600 km zwischen 8 und 20 Millisekunden). Außerdem kommt es je nach Position innerhalb einer Zelle zu unterschiedlichen Laufzeiten, die je nach Zellengröße bis zu 10 ms betragen. Beides lässt sich mit verlängerten Timing-Parametern kompensieren, Basisstation und Smartphone warten einfach länger auf Steuer- und Nutzdaten.

Satellitenwechsel

Besonders interessant ist das Handover gelöst, also der nahtlose Wechsel von Zelle zu Zelle, denn die in terrestrischen Netzen übliche Handover-Technik funktioniert mit Satelliten nicht. Wenn in einem terrestrischen Netz der Pegel einer Zelle abnimmt und der einer Nachbarzelle zunimmt, ist das ein Indiz dafür, dass sich der Teilnehmer bewegt und das Netzwerk leitet dann das Handover ein.

In NTN funktioniert das nicht, weil sich die Empfangspegel am Zellenrand und in der Mitte zu wenig unterscheiden und weil der Satellit schneller fliegt als übliche Bodenfahrzeuge fahren. Aber ersatzweise kann ein Smartphone im NTN-Betrieb das bedingte Handover nutzen (Conditional Handover, CHO). Damit kann es ein Handover zeitabhängig und standortbasiert auslösen. Es weiß ja, wo es sich befindet und wann sein aktueller Satellit außer Reichweite verschwindet.

Auch die Methode zur Korrektur von Empfangsfehlern funktioniert im NTN zunächst nicht. Dazu erstmal etwas Hintergrund: Wenn die Basisstation für ein Datenpaket vom Teilnehmergerät keine Quittung bekommt, verschickt sie das betreffende Paket neu (Hybrid Automatic Repeat Request, HARQ). Für terrestrische Netze sind maximal 16 HARQ-Prozess-IDs definiert. Doch bei der Kommunikation mit Satelliten kann es wegen der langen Laufzeiten vorkommen, dass alle 16 IDs in Gebrauch sind. Dann könnten mangels Prozess-IDs keine neuen Daten übertragen werden (HARQ stalling). Deshalb sind für 5G-Satelliten bis zu 32 IDs spezifiziert.

Weitere Ergänzungen betreffen unter anderem die Zellauswahl und Neuauswahl von Zellen und Positionsmeldungen des Teilnehmergeräts, damit es sich in das richtige Kernnetz im richtigen Land einbucht.

IoT ohne Handover

Für das Internet der Dinge, etwa beim Fernauslesen von Sensoren, mit den beiden Spezifikationen NB-IoT und LTE-M nutzt das 3GPP zum Teil die Methoden, die es schon für Smartphones konzipiert hat, beispielsweise die Kompensation der Dopplerverschiebung.

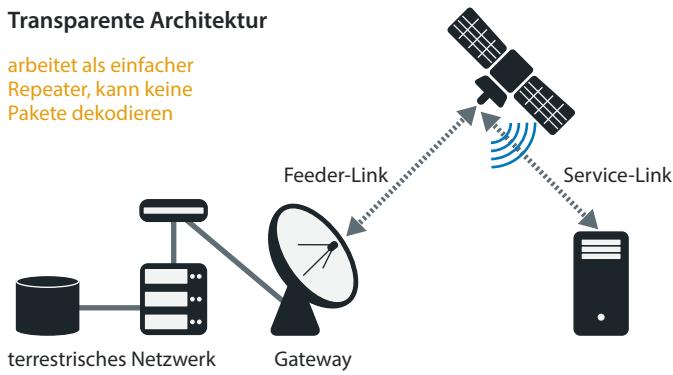
Die Handover-Erweiterungen gehören nicht dazu, denn IoT-Geräte benötigen in den meisten Fällen keinen kontinuierlichen Zugang zum Mobilfunknetz. Stattdessen sammeln sie Daten eine Weile lang und versenden sie aus Stromspargründen

Zwei NTN-Architekturen

Nutzdaten werden über einen Feeder-Link an eine Bodenstation oder ein Gateway gesendet und von dort zum Kernnetz und weiter ins Internet geleitet. Teilnehmergeräte koppeln über den Service-Link an den Satelliten an. Satellitenbetreiber haben die Wahl zwischen einer transparenten Architektur (oben) und regenerativer Architektur.

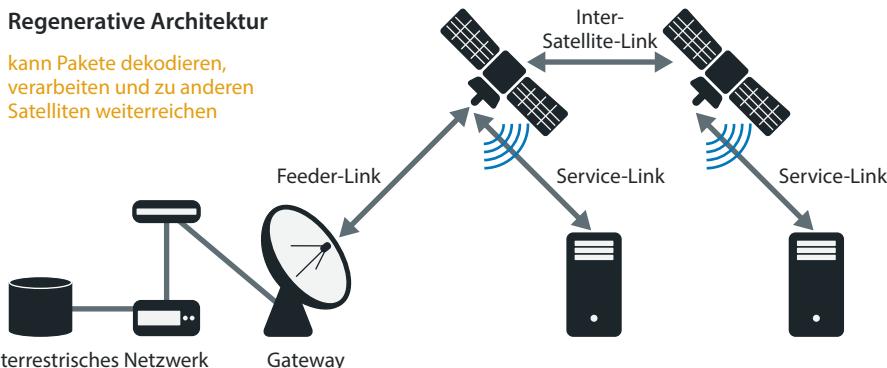
Transparente Architektur

arbeitet als einfacher Repeater, kann keine Pakete dekodieren



Regenerative Architektur

kann Pakete dekodieren, verarbeiten und zu anderen Satelliten weiterreichen



einmal alle paar Stunden (z. B. Luftfeuchte- oder Temperaturmesswerte). Deshalb kommen Satellitenbetreiber, die nur IoT-Dienste anbieten wollen, mit kleineren Satellitenkonstellationen aus.

Für die erdgestützte Astronomie gilt ohnehin: Je weniger Satelliten, desto besser. Schon die Starlink-Satelliten fallen negativ auf, weil sie auf astronomischen Aufnahmen helle Streifen hinterlassen und kosmische Objekte überdecken. Noch mehr Störungen kann man von 5G-Satelliten erwarten. Der von AST betriebene BlueWalker 3 ist nicht nur der größte LEO-Satellit, er überstrahlt auch jeden Stern am Himmel.

Immerhin: Nach Ende der Betriebszeit versenkt man die Satelliten kontrolliert im Südpazifik zwischen Chile und Neuseeland, genauer: am Point Nemo, von wo aus die nächste feste menschliche Behausung mindestens 2688 Kilometer entfernt ist. Und Point Nemo gilt als der biologisch am wenigsten aktive Ort der Ozeane.

Ausblick mit Apfel

Der Hauptvorteil der NTN-Erweiterungen liegt in der Kompatibilität zu kommenden handelsüblichen Smartphones, die mit passenden Tarifen umgehend globale Daten- und Sprachkonnektivität erhalten. Terrestrische Betreiber können damit ihre Abdeckung ohne zusätzliche Basisstationen verbessern und Versorgungslücken in dünn besiedelten Gebieten schließen.

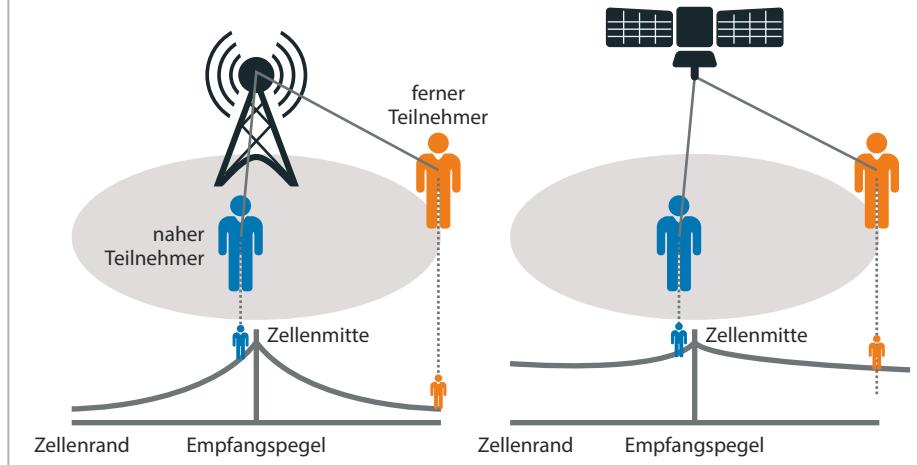
Fachleute erwarten, dass Betreiber von Legacy-LTE- und 5G-NTN-Satelliten Mobilfunknetzbetreiber als Kooperationspartner suchen werden, denn allein mit ihrem satellitengestützten Angebot können sie weder mit dem Preis noch mit der Leistung von terrestrischen 5G-Netzwerken mithalten. Gute Chancen hätten sie aber als Ergänzung für terrestrische Netze.

Für Netzbetreiber erscheint die Kooperation attraktiv, weil ihnen Roaming-erlöse winken. Auch die Geräteindustrie dürfte an der Technik interessiert sein, weil sie Geräte mit Satellitenzugang teurer verkaufen kann. Daher kann man Android-Smartphones mit Modems für 5G-Satelliten schon bald von vielen Herstellern erwarten; ob mit MediaTek-, Qualcomm- oder Exynos-Chips ist offen.

Das setzt Apple und Globalstar unter Druck nachzuziehen. Beide dürften es schwer haben, nur Notrufdienste per Satellit anzubieten, während das Android-Lager globales mobiles Breitband nutzen kann.

Empfangspegelunterschiede

Die in terrestrischen Netzwerken (links) üblichen Handover-Methoden funktionieren bei Satelliten nicht, weil das Netz wegen geringer Empfangspegelunterschiede nicht genau weiß, ob sich Teilnehmer in der Zellenmitte oder am Rand befinden.



Die Hersteller von Notrufgeräten auf Globalstar-Grundlage dürften das bald spüren. Wenn Apple seine Mobilfunkmodems weiterhin von Qualcomm bezieht, dann könnten bald auch iPhones die 5G-Technik für Satelliten bekommen. Das könnte freilich eine Abkehr vom bisherigen Kooperationspartner Globalsat nach sich ziehen, über dessen Satelliten iPhones abseits von Mobilfunkabdeckungen Notrufe absetzen.

Apple arbeitet aber seit Jahren an einem eigenen 5G-Modem. Das eröffnet dem Konzern verschiedene Perspektiven. Im einfachsten Fall könnte Apple die 5G-NTN-Technik so wie Android-Hersteller nutzen und damit die globale Versorgung von iPhones gewährleisten. Die Globalsat-Kooperation wäre dann überflüssig. Doch Apple hat allein Ende 2022 450 Millionen

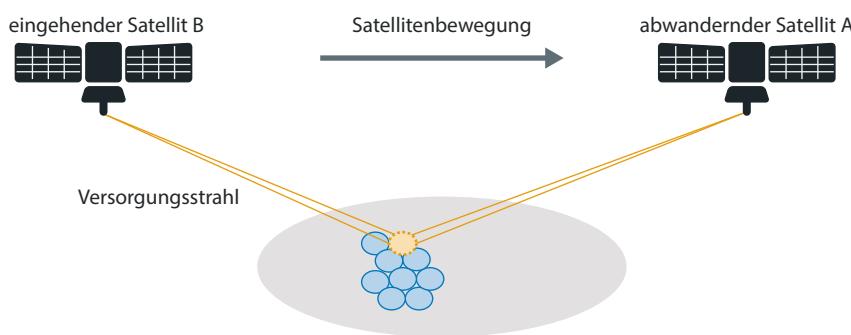
US-Dollar in das Globalstar-Netz und Bodenstationen investiert (ct.de/ynhb), von denen ein Teil erst 2025 für weitere Globalstar-Satelliten aufgewendet wird. Und da die Firma selbst im 3GPP sitzt, wird ihr schwerlich entgangen sein, dass 5G-Satelliten im Kommen sind.

An dieser Stelle brechen wir die Spekulation ab und überlassen es Ihrer Fantasie, sich auszumalen, was der reichste Konzern der Welt mit seiner Kohle anfängt. Sind weitere Investitionen in Globalstar zu erwarten, etwa zur Aufrüstung auf 5G-NTN? Wird Apple so wie von Gerüchten fabuliert zum Mobilfunkprovider? Schreiben Sie uns, wie Sie die weitere Entwicklung einschätzen. (dz@ct.de) **ct**

5G-NTN-Infos: ct.de/ynhb

Zellenwechsel

Beim 5G-Satellitenfunk versorgen mehrere Satelliten die Teilnehmer einer Zelle (gestrichelter Kreis), und sie reichen sich am Ende des Überflugs die Zuständigkeit weiter. Anders als bei terrestrischen Netzen lösen Teilnehmergeräte den Wechsel zur nächsten „Basis“ beziehungsweise zum nächsten Satelliten selbst aus.



Kein Ende in Sicht: Spiele-PCs to go

Hype um Handheld-Konsolen beleuchtet

Asus ROG Ally, Ayaneo Air 1S, Valve Steam Deck: Tragbare Gaming-PCs im Format von Nintendos Switch sind derzeit äußerst beliebt. Eine Bestandsaufnahme zeigt, was die Teile draufhaben und welche Modelle besonders spannend sind.

Von Dennis Schirrmacher

Seit Nintendo 2017 seine Hybridkonsole Nintendo Switch [1] für unterwegs und zu Hause am TV-Gerät veröffentlicht hat, wünschten sich Videospielfans leistungsfähigere Hardware in dieser Bauform. Das Ein-Chip-System (SoC) der Switch Tegra X1 von Nvidia war bei Erscheinen der Konsole bereits rund zwei Jahre alt und vergleichsweise schwach auf der Brust: Die Hardwareleistung liegt in der Summe etwas über der einer PlayStation 3 aus dem Jahr 2006. Eine leistungsfähigere Switch 2 ist bis heute nicht erschienen.

Den Hype des hochauflösten mobilen Spielens entfachte Valve Anfang 2022 mit dem im Vergleich deutlich leistungsfähigeren Handheld-PC Steam Deck [2]. Damit hat der Begründer der erfolgreichen Internetvertriebsplattform für Computerspiele Steam den von Nintendo ins Rollen gebrachten Stein um ein Vielfaches beschleunigt und eine wahre Lawine ausgelöst: Gefühlt im Monatstakt kündigen Hersteller neue Modelle an und heben das mobile Gaming auf ein neues Level.

Leistung meets Mobilität

Steam Deck & Co. nutzen Notebookhardware von AMD und sind mit flottem M.2-Flashspeicher und flinkem Arbeitsspeicher ausgestattet. Damit sollen die trag-

baren Gaming-PCs den Spagat zwischen möglichst langer Akkulaufzeit und viel Rechenkraft für 3D-Spiele vollziehen. Bis zu einem gewissen Grad klappt das auch. Die Geräte machen bereits vieles richtig, Wunder vollbringen können aber auch sie nicht. Daher muss man auf einer der beiden Seiten Kompromisse eingehen. Demzufolge läuft auf dem Steam Deck sogar der anspruchsvolle 3D-Titel „Cyberpunk 2077“ mit hohen Grafikdetails im Schnitt mit 30 Bildern pro Sekunde (fps). Nach maximal eineinhalb Stunden ist dann aber der Akku leer. Bei 2D-Spielen wie „Owlboy“ sind wiederum bis zu acht Stunden drin.

Auf dem Papier ist die Hardware (Zen-2-CPU, RDNA-2-GPU) des Steam Decks bereits dezent veraltet. Asus ROG Ally [3] und Konkurrenten wie Ayaneo Air 1S setzen mit der Zen-4-CPU und der RDNA-3-Grafikeinheit aktuelle AMD-Komponenten ein. Doch beim Steam Deck holt Valve mit Softwareoptimierungen immer wieder zusätzliche Leistung heraus. Zuletzt verbesserte das Update auf SteamOS 3.5 die Multithread-Leistung. Außerdem muss es im Vergleich zur Konkurrenz aufgrund der kleineren Display-Auflösung mit 1280 × 800 Pixeln nur die Hälfte der Bildpunkte im Vergleich zum ROG Ally mit

seinem Full-HD-Bildschirm berechnen. Das Steam Deck kann also immer noch mithalten und ein Nachfolger ist bislang nicht in Sicht. Im Schnitt lässt sich sagen, dass die aktuelle Handheld-PC-Generation so schnell wie eine PlayStation 4 Pro aus dem Jahr 2016 rechnet.

AMDs Upscaling-Technologie FidelityFX Super Resolution (FSR) spielt den mobilen Gaming-PCs in die Karten. So kann die Grafikeinheit kompatible Spiele intern mit einer geringeren Auflösung berechnen und so Rechenkraft einsparen. Der FSR-Ansatz skaliert die Auflösung dann effektiv auf die Bildschirmauflösung. Das kommt der Performance und Akkulaufzeit zugute und das Ergebnis kann sich auf den vergleichsweise kleinen Bildschirmen sehen lassen. Natürlich ist die Darstellung unschärfer als in der nativen Auflösung und Artefakte wie flirrende feine Strukturen trüben den Gesamteindruck. Insgesamt kann die Bildqualität aber oft überzeugen.

Weiterer Einsatzzweck

Das Steam Deck benutzt intern Linux, das Asus ROG Ally hingegen Windows. Für beide Handhelds gibt es reichlich Softwareemulatoren für Retro-Spielkonsolen. 2D-Spiele von Super Nintendo & Co. schaffen diese problemlos. Dank der Leistungsreserven der aktuellen Handheld-Generation laufen sogar einige PlayStation-3-Titel wie „Skate 3“ flüssig mit 60 Bildern pro Sekunde.

Doch das funktioniert nicht out of the box: Man muss die Emulatoren selbst installieren und für die optimale Leistung konfigurieren. Dank dem EmuDeck-Projekt [4], einer vorkonfigurierten Softwareemulatorsammlung für das Steam Deck gelingt die Installation mit wenigen Mausklicks.



Die Bauformen des Asus ROG Ally (oben) und Valve Steam Deck (unten) sind sehr ähnlich. Das neuere Asus-Modell enthält aber leistungsfähigere Hardware. Es rechnet nicht nur schneller, auch der Bildschirm ist mit seiner Full-HD-Auflösung und 120-Hertz-Darstellung inklusive variabler Refresh-Rate (VRR) technisch überlegen.

Hardware meets Software

Ein Alleinstellungsmerkmal von Nintendos Switch ist die sinnvolle Verknüpfung von Hard- und Software. Außer der Konsole entwickelt Nintendo auch die Bedienoberfläche und bietet im hauseigenen eShop Spiele zum Download an. Dementsprechend bekommt man alles aus einer Hand, was der Gesamtqualität und Usability zugutekommt.

Valve geht einen ähnlichen Weg wie Nintendo und bietet mit dem Steam Deck und seinem Steam Store eine gelungene Symbiose aus Hard- und Software an. Diese Faktoren sind maßgeblich für den Erfolg beider Geräte verantwortlich.

Genau an dieser Stelle müssen Hersteller wie Asus und Ayaneo nun nachziehen. Vor allem das ROG Ally ist noch nicht optimal bedienbar. Asus Bedienoberfläche Armoury Crate ist nur eine Windows-Anwendung und das Ganze wirkt im Vergleich mit Nintendo und Valve nicht wie aus einem Guss. So gibt es immer wieder einen Bruch und man landet auf dem Windows-Desktop, um Einstellungen vorzunehmen. Außerdem hinken Ayaneo und Asus bei der Veröffentlichung von Software-Verbesserungen für das System Valve hinterher, die viele in der Community geschilderten Probleme oft zeitnah mit Updates lösen.

Zukunftsansichten

Nach Spekulationen plant Nintendo für 2024 einen Switch-Nachfolger. Dieser soll ungefähr die Leistung einer PlayStation 4 aus dem Jahr 2013 schaffen. Damit würde Nintendo abermals bei der Hardware hinterherhinken, dieses Mal sogar um zehn Jahre. Genau an dieser Stelle punkten die Modelle der Konkurrenz schon jetzt mit ihrer Leistung. Mittlerweile ist allerdings ein Peak erreicht und viele Geräte suchen ihr Heil mit fast identischer Hardware. Aufgrund der Entwicklungszeit von noch leistungsfähigeren mobiltauglichen Komponenten wird es eine Weile dauern, bis Spiele unterwegs noch besser aussehen und laufen.

Fraglich ist zudem, wie die Handheld-PCs über die reine Hardwareleistung hinaus etwas anbieten, das nur sie allein können. Der chinesische Hersteller Ayaneo beispielsweise überflutet den Markt mit verschiedenen Modellen. Er hat äußerst interessante Vertreter im Angebot: Beim Ayaneo Slide kann man den Bildschirm nach oben schieben, um die darunter liegende Tastatur freizulegen. Auch

der kompakte Handheld-PC Air 1S mit 5,5-Zoll-OLED-Bildschirm und aktueller AMD-Mobil-Technik klingt verlockend. Der Pocket Air kommt mit Android. Mit dem Ayaneo Kun will der Hersteller mit einer 54-Watt-Option und aktueller AMD-Hardware in neue Grafikssphären vorstoßen. Das verkürzt natürlich die Akkulaufzeit. Mittlerweile hat Ayaneo mehr als zehn Modelle im Portfolio, wobei einige ausschließlich über die Crowdfunding-Plattform Indiegogo bestellbar sind. Bei Reddit mehren sich indes die Kommentare über schlechten Service. Das lässt vermuten, dass sich der Hersteller etwas zu viel zugemutet hat.

Nintendo hat bis dato nicht ohne Grund rund 130 Millionen Switch-Konsolen verkauft. Der Hersteller blickt seit dem Game Boy im Jahr 1989 auf eine lange Erfahrung mit Handheld-Konsolen zurück. Dementsprechend ist die Bedienung mittlerweile sehr ausgereift und etwa der Download und das Starten von Spielen gelingt ohne großartigen Konfigurationsaufwand. Die anderen Hersteller können da logischerweise nicht einfach so anknüpfen und müssen vor allem bei der Bedienbarkeit noch hinzulernen. Vieles wirkt noch unfertig und wer das Optimum herausholen möchte, kommt um Gefriskel nicht herum. Käuferinnen und Käufer müssen oft selbst noch Firmware-Updates und neuen Treiber installieren sowie die Konfiguration erledigen, damit alles optimal läuft.

Am Ende entscheidet auch der Preis: Im Vergleich zur Switch mit 270 Euro sind die Kosten für die mobilen Gaming-PCs

ct kompakt

- Prachtvolle 3D-Grafik für unterwegs: Früher haben Gamer davon geträumt, heute ist es Realität.
- Aktuelle High-End-Modelle unterscheiden sich in der Leistung kaum. Manche Hersteller versuchen mit Alleinstellungsmerkmalen zu punkten.
- Mobiles Gaming mit beeindruckender Grafik ist aber teuer; einige Handheld-PCs kosten ab 1000 Euro.

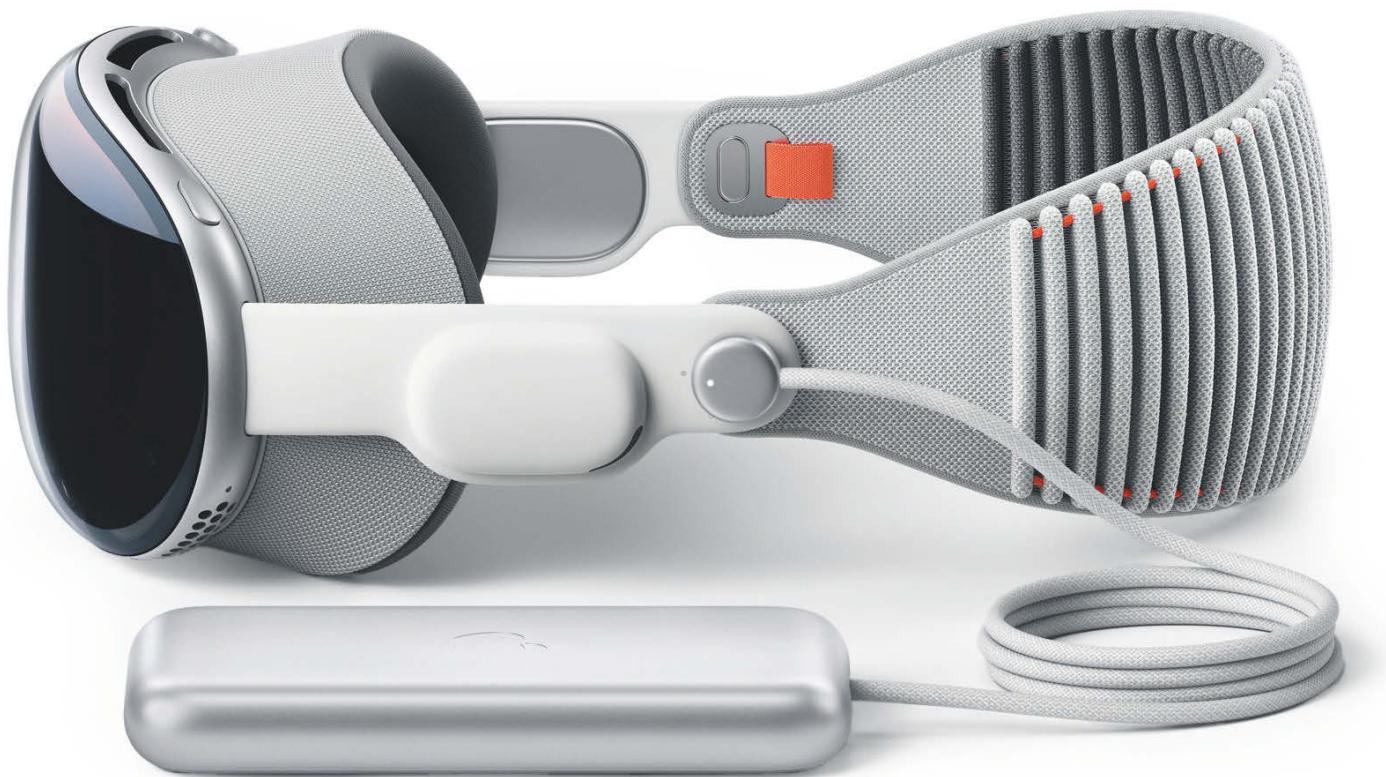
vergleichsweise hoch: Für das Einstiegsmodell des Steam Decks hat Valve den Preis mit knapp über 400 Euro recht niedrig angesetzt. Das Asus ROG Ally mit dem Z1-Extreme-Chipsatz kostet schon rund 800 Euro. Je nach Ausstattung muss man für einen Ayaneo-Handheld über 1000 Euro berappen. Dafür bekommt man aber auch High-End-Leistung für unterwegs und man kann mit der Grafikpracht im Bus oder Zug angeben. (des@ct.de) ct

Literatur

- [1] Dennis Schirmacher, Der Videospiel-Transformer, Spielkonsole Nintendo Switch getestet, c't 6/2017, S. 32
- [2] Liane M. Dubowy, Taschenspielerei, Mobile Steam-Bibliothek: Spielkonsole Valve Steam Deck im Test, c't 9/2022, S. 112
- [3] Jan-Keno Janssen, Steam-Deck-Konkurrent, Asus ROG Ally im Test, c't 14/2023, S. 86
- [4] Dennis Schirmacher, Retrogaming to go, Videospielkonsolen auf dem Steam Deck emulieren, c't 25/2022, S. 158

Lenovo lockt Grafikenthusiasten mit dem Legion Go. Der Handheld hat ein 8,8-Zoll-Display und wie bei Nintendos Switch kann man die seitlich angesteckten Controller abnehmen und den Bildschirm mit einem Ständer auf einen Tisch stellen.





Blickfänger

Spatial Computing: Wie Apples MX-Headset Vision Pro arbeitet

Niemand weiß derzeit, ob Apples groß angekündigte Computerbrille ein Rohrkrepierer wird oder mit einem lange erwarteten Knall eine neue Epoche einläutet – sie kommt ja erst ab Anfang 2024 in den Handel. Doch obwohl der Konzern viele Spezifikationen unter dem Deckel hält, haben wir viele Details sammeln können, die eine Einordnung ermöglichen.

Von Leonhard Becker, Inge Schwabe und Dušan Živadinović

Die Vision-Pro-Brille ist für Apple ein völlig neues Produkt, das zwei Techniken kombiniert: Augmented Reality (AR) überlagert optische und akustische Sinneseindrücke mit Computerinhalten und Virtual Reality (VR) ersetzt sie komplett mit computergenerierten Inhalten. Daher gehört sie zu den Mixed-Reality-Headsets. Mit VR taucht man in digital erzeugte Welten ein (Immersion), zum Beispiel 3D-Lehrgänge oder Spiele, die AR blendet digitale Objekte über die reale Welt ein, etwa Benutzeroberflächen von Programmen. Bereits erhältliche Headsets wie die Meta Quest 3 bieten diese Grundfunktion für deutlich weniger Geld (550 Euro statt vermutlich 3400 Euro), aber mit weniger Displayauflösung und Funktionen zum Augen- und Mimiktracking.

Besonders bemerkenswert erscheint das Vision-Pro-Headset, weil Apple seine typische zweidimensionale Benutzerführung überzeugend in ein 3D-Kontinuum überführt. Die VR- und MR-Pionierarbeit hat freilich die Entwicklergemeinde in den letzten zehn Jahren geleistet und Apple hat davon die besten Ideen herausgepickt und auf eigene Art umgesetzt. So könnte der Konzern aber tatsächlich einen Meilenstein auf dem Weg zum räumlichen Computing (spatial Computing) setzen, weshalb wir in diesem Beitrag hauptsächlich über die Bedienung und die Funktionen berichten.

Zunächst aber kurz zum Bedienkonzept von Headsets und den wichtigsten Hardware-Merkmalen der Vision Pro (weitere Hardwaremerkmale folgen weiter hinten): Headsets erfassen die Umge-

bung des Nutzers und seine Gesten sowie Augenbewegungen kontinuierlich mit Kameras und Sensoren. Anhand der Umgebungsdaten wie Wand-, Tür- und Möbelpositionen können Spielszenen oder Bedienelemente in die reale Umgebung eingepasst werden.

Das gewährleistet auch, dass sich Träger des Headsets zumindest langsam unfallfrei in einem Zimmer bewegen können, auch wenn sie die Möbel darin nicht sehen. Die Entwickler simulieren auf dieser Grundlage Spielszenen wie Durchbrüche von bösen Monstern durch reale Mauern. Und Headsets erfassen Hand- und Augenbewegungen der Träger zum Auslösen von Funktionen und zur Navigation im Bedienraum, etwa um Anwendungsfenster zu verschieben oder Dateien zu speichern.

Das Vision-Pro-Headset enthält zwölf Kameras, sechs Mikrofone und fünf Sensoren. Dazu zählt etwa ein Lidarscanner, der die Laufzeit von Lichtpulsen misst, die von Objekten zurückstreuen. Mittels dieser Daten erstellt das Headset eine 3D-Karte der Umgebung. Eine Übersicht aller Schnittstellen zur Außenwelt sehen Sie auf Seite 152.

Um den für menschliche Kommunikation wünschenswerten Augenkontakt zu simulieren, blendet die Brille auf einem Außendisplay eine Augenabbildung des Trägers ein (EyeSight), wenn sie detektiert, dass jemand den Träger anspricht. Diese 3D-Daten erfasst sie mitsamt dem Gesicht beim ersten Einrichten in 3D; dafür muss der Träger das Headset und damit die frontal angebrachte Infrarotkamera, den Punktprojektor und die RGB-Kamera kurz von außen ansehen. Bei FaceTime-Konferenzen kombiniert das Headset die aktuelle Mimik und Gestik des Nutzers mit dem digitalen Abbild seines Gesichts zu einem 3D-Avatar (Persona) und sendet diesen an die anderen Teilnehmer.

Über Vorlagen kann man mehrere mittels Avataren teilnehmende Personen um einen geteilten Inhalt aufstellen. Zum Beispiel können Teilnehmer nebeneinander oder im Kreis sitzen und gemeinsam eine Präsentation, einen Spielfilm oder ein 3D-Objekt betrachten. Bisher hat Apple keine Gruppenkonferenzen vorgeführt, sondern nur FaceTime-Dialoge. Daher bleibt offen, ob Apple den digitalen Personae auch Unterleib und Beine spendiert hat. In Metas Horizon Workrooms muss man ohne solche Körperteile auskommen.

Headset und Akku hat Apple baulich separiert und per Kabel verbunden. Das

schont die Nackenmuskulatur der Träger. Zumindest der Prototyp des Vision-Pro-Headsets, den Entwickler in Apple-Laboren aufsetzen dürfen, wiegt rund 400 bis 500 Gramm und ist damit ungefähr so schwer wie andere Headsets (siehe ct.de/y7vd). Die maximale Laufzeit beträgt laut Apple rund zwei Stunden. Alternativ kann man den Akku über USB-C dauerhaft mit Strom versorgen, um ohne Akkuwechsel mehrere Stunden zu arbeiten.

Handarbeit

Für das Vision-Pro-Headset hat Apple ein von iOS abgeleitetes Betriebssystem entwickelt, das visionOS. Um das Headset zu entsperren und um Einkäufe im App Store zugenehmigen, weist sich der Nutzer mittels seiner Iris aus; die individuellen Muster tasten innen angebrachte Infrarotsensoren ab (Optic ID). Wie bei Touch ID (Fingerabdruckscan) und Face ID (Gesichtsscan) verbleiben die Optic-ID-Daten verschlüsselt auf dem Gerät. Sie liegen vom Hauptprozessor isoliert und für Apps unzugänglich in der Secure Enclave.

Auf visionOS läuft der Großteil des riesigen Softwareangebots für iPhones und iPads. So führt der Konzern seine Kunden über vertrautes Terrain zu seinem 3D-Bedienkonzept. Zusätzlich kann die Brille den Bildschirminhalt eines im selben WLAN eingebuchten Mac groß in 4K in einem einzelnen Fenster darstellen, sodass man die mächtigen macOS-Anwendungen nutzen kann; die Brille wird so zum Mac-Monitor.

Ein Druck auf die rechts oben am Brillenrahmen angebrachte digitale Krone öffnet jederzeit die Home-Ansicht mit vertrauten Apps wie iMessage, Safari-Browser, Mail, Fotos, Notizen, Musik und TV, Einstellungen und App Store. Neue Apps für Apples Headset-Welt erkennt man

daran, dass sie ihre Erscheinung automatisch auf die Beleuchtung der physischen Umgebung anpassen. Mit einem als „Glas“ bezeichneten, leicht durchsichtigen Hintergrund fügen sie sich auch optisch besser in den visionOS-Raum ein.

Wie auf iOS und iPadOS, so kann man auch auf visionOS die Home-Ansicht individualisieren und zum Beispiel Apps in Ordner sammeln. Links neben den App-Icons blendet visionOS eine vertikale Steuerleiste ein, über die man schnell zu den Kontakten (für FaceTime-Videokonferenzen) oder zu den Hintergrundmotiven (Environments) wechselt; diese überlagern auf Wunsch die physische Umgebung hinter den Apps.

Dazu gehören beispielsweise Naturabbildungen und Szenen aus Nationalparks der USA. Elemente der Environments können bewegt und hörbar sein, sodass etwa Regen prasselt oder Wind vernehmlich bläst. Die Klangfarbe passt das Headset an die Beschaffenheit des Raums an (Audio Raytracing); dafür erfasst es dessen Charakteristik mit den Mikrofonen.

Dreht man die digitale Krone, bestimmt man den Grad der Überlagerung über die reale Umgebung. Je nach Bedarf und Laune kann man das nutzen, um ein gerade nicht aufgeräumtes Büro auszublenden oder sich proaktiv vom Aufräumen abhalten zu lassen.

Fenster, 3D-Objekte, VR-Apps

Apps öffnen sich in großen Fenstern, die in der physischen Umgebung schweben. Die Größe der Arbeits- oder Wiedergabefläche etwa für 3D-Filme lässt sich per Handgeste ändern. Bei maximaler Einstellung erscheint sie etwa so groß wie eine 30-Meter-Leinwand aus erster Reihe betrachtet. Blickt man ein Objekt an, erscheint es in



Apples Headset besteht aus drei mechanisch verknüpften Teilen. So kann die Firma je nach Kopf- und Gesichtsform geeignete Kopfbänder (links) und Lichtsiegel liefern (Mitte). Oben rechts am Brillenrahmen befindet sich die Krone, mit der man den Homescreen einblendet und den Grad der Immersion einrichtet.

höchstmöglicher Qualität. Bereiche im Außenwinkel werden in niedrigerer Auflösung gerendert, was Energie spart.

Fenster lassen sich frei mit konkretem Bezug zur eigenen Umgebung platzieren, und zwar rundherum und auch in verschiedenen Zimmern. So kann ein Set von Programmen im Büro auf Arbeit warten, während im Wohnzimmer etwa die Fenster von Foto- und Streaming-Apps offen stehen. Anhand von Apples bisherigen Prinzipien kann man erwarten, dass sich visionOS die Fensterpositionen auch über Neustarts und Abstürze hinweg merkt. Wie viele Fenster gleichzeitig geöffnet sein können, hat Apple bisher nicht verraten.

Über ein kleines, automatisch auftauchendes Element lässt sich die Größe der Fenster ändern. Das funktioniert ähnlich wie im Stage Manager auf dem iPad. Außerdem liegen am unteren Fensterrand je zwei feste Bedienelemente: Tippt man auf den weißen Punkt, wird das Fenster geschlossen. Mit dem länglichen weißen Balken stellt man die Distanz ein oder richtet es im Raum frisch auf den eigenen Blickwinkel aus. Bedienelemente von Apps liegen in Ornamenten an einer Fensterseite und gewähren Zugriff auf verschiedene Ansichten und Funktionen. In der Fotos-App wechselt man damit zum Beispiel von der Mediathekansicht zu den Alben. Das entspricht der Tab-Leiste vieler iPhone-Apps.

Apps können zweidimensional sein oder in der Fensteransicht 3D-Elemente integrieren. Dreidimensionale Objekte, etwa Auto- oder Flugzeugmodelle, lassen sich frei in den Raum stellen und von allen Seiten betrachten. Mehrere Apps können gleichzeitig geöffnet sein (Multitasking), aber es kann auch, wie auf dem Mac gewohnt, ein einzelnes Programm im Vollbildmodus laufen. In dessen Raum können dann etwa mehrere Fenster und 3D-Objekte schweben und die Hintergrundumgebung verwaltet es dann selbst wie in klassischen Virtual-Reality-Anwendungen, also „fully immersive“.

Im VR-Modus aktiviert Apple Schutzmaßnahmen für den Nutzer: Bewegt er sich schnell oder auf einen Gegenstand zu, unterbricht das Headset die Immersion automatisch und blendet das Livebild der physischen Umgebung ein, um Kollisionen etwa mit Möbeln zu vermeiden.

In VR-Spielen kann sich der Träger nur rund 1,5 Meter von seinem Startpunkt weg bewegen, bis das Bild der Außenwelt die VR verdrängt. Für den Betrieb im Flugzeug, Zug oder Auto gibt es den Reisemodus, der bestimmte Funktionen abschaltet und erfordert, dass man sitzt.

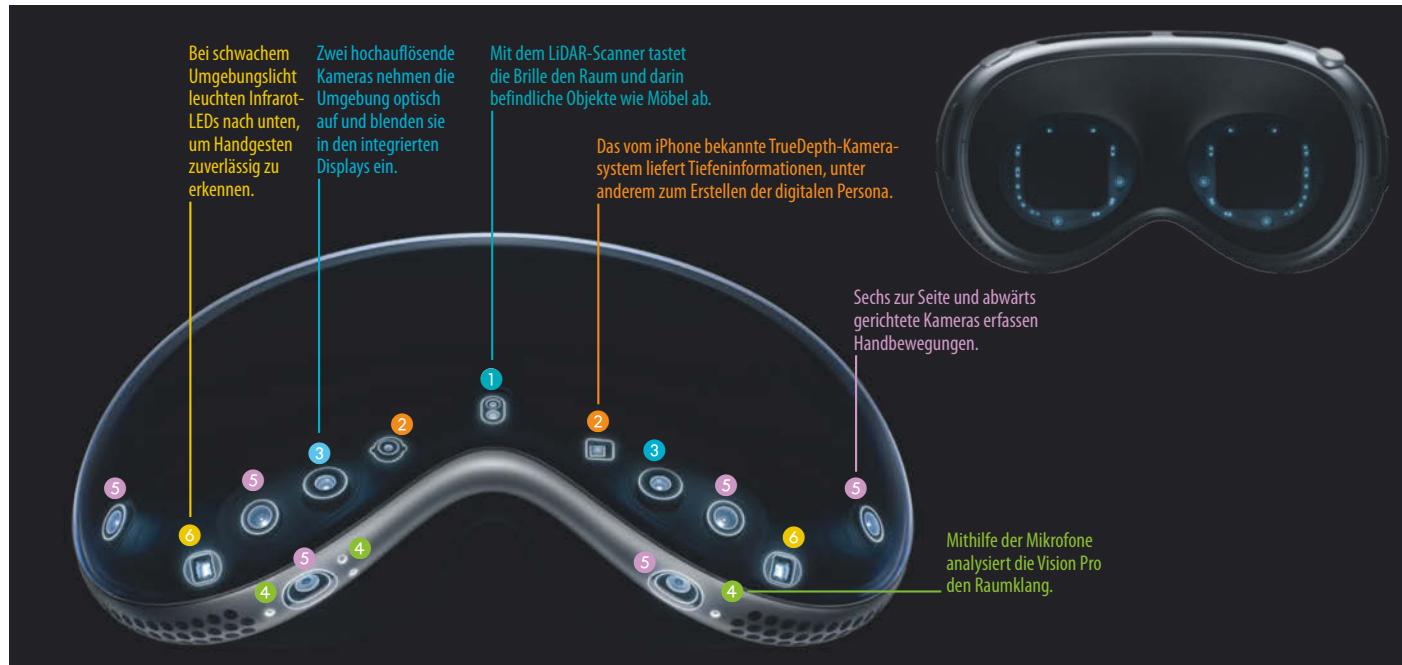
Augenblicke und Gesten

Im 3D-Raum navigiert man, indem man ein erwünschtes Element oder Objekt ein-

fach anblickt. Das erkennt das Headset mittels vier Infrarotsensoren, die um das Innendisplay herum angebracht sind (Eye Input) und hebt das anvisierte Objekt hervor, um Bereitschaft zu signalisieren. Dann kann man die zugehörige Aktion per Handbewegung auslösen, also etwa Zeigefinger und Daumen kurz pressen, um einen Knopf zu drücken. Das Headset sieht das mit den abwärts gerichteten Kameras. Unser Eindruck nach ersten Tests in Apples Labor: Das fühlt sich natürlich an, besonders Smartphonenuutzer erfassen das Bedienkonzept schnell.

Per Bluetooth lassen sich auch Trackpads oder Mäuse (auch Tastaturen) ankoppeln und dann kann man wie von 2D-Computern gewohnt per Mauszeiger navigieren. Der auf zweidimensionale Bedienoberflächen ausgelegte Mauszeiger folgt dem Blick, wenn man von einem Fenster zu einem anderen wechselt.

Die Gesten funktionieren auch mit den Händen im Schoß, die von zwei abwärts blickenden Kameras erfasst werden. Entwickler können eigene Handgesten definieren sowie Daten über die Beschaffenheit der Umgebung abfragen, um etwa 3D-Objekte auf einem physischen Tisch zu platzieren. Direkter Zugriff auf die Kameras bleibt den Apps verwehrt. Dem Handtracking durch Fremd-Apps müssen Nutzer zustimmen.



Außenansicht und Innenansicht der Vision-Pro-Brille: Alle Sensoren zum Abtasten der Außenwelt hat Apple am unteren Rand angebracht (linke Abbildung). Die Infrarotkameras zur Erfassung der Blickrichtung sind rund um die beiden Innendisplays angeordnet.

Apps lassen sich auch direkt mit den Fingern steuern, was der Touchbedienung von iPhone und iPad ähnelt. Dafür zieht man die Fenster heran, um die Bedienoberfläche in Reichweite zu holen. Dann kann man auch 3D-Objekte manipulieren. Für die Texteingabe gibt es ein virtuelles Keyboard. Alternativ kommandiert man Siri und startet so beispielsweise Apps.

Weitere von iOS bekannte Elemente sind geplant, darunter das Kontrollzentrum sowie die Mitteilungszentrale, die Benachrichtigungen sammelt. Im Gastmodus können die Brille Freunde oder Bekannte ausprobieren. Ein vollständiger Mehrbenutzermodus fehlt.

Hardware

Viele Merkmale hält Apple bisher hinterm Berg: Arbeits- und Massenspeicherkapazität, Prozessortakt, Gewicht, Akkukapazität, Bluetooth- und WLAN-Einzelheiten, Wartungsschnittstelle. All das und mehr werden ausführliche Tests ans Tageslicht bringen. Es sind aber schon überraschend viele Hardwaremerkmale bekannt.

Die zwei inneren, briefmarkengroßen Ultra-HD-Displays bestehen aus je 23 Millionen Pixeln (4K-Auflösung); die Pixelbreite beträgt 7,5 Mikrometer. Das Headset spielt sowohl Standard- als auch High-Dynamic-Range-Inhalte ab (SDR bzw. HDR). Die Bildwiederholrate beträgt 90 bis 96 Hertz. Die in 4K gerenderten Texte sind laut Apple unabhängig vom Betrachtungswinkel immer scharf dargestellt.

Für die Gestensteuerung beobachten vier Kameras Unterarme und Hände. Über die Mikrofone kann man das Headset mit Sprachbefehlen steuern. Bei niedriger Helligkeit beleuchtet das Headset die Hände mit Infrarotstrahlen. Die nach vorn gerichteten Kameras liefern einen 3D-Stream des realen Raums, den das Headset graduell mit virtuellen Inhalten überblendet oder mit AR-Inhalten anreichert und im inneren stereoskopischen Display darstellt. Mit dem Drehrad (Digitale Krone) richtet man den Grad der Überblendung ein.

Über die linke Taste nimmt man mit den Frontkameras 3D-Fotos oder -Videos mitsamt dem Audiosignal in Stereo auf. Zumindest bisher erhältliche Demoaufnahmen geben die räumliche Beschaffenheit der realen Welt beeindruckend plastisch wieder. Apples erste 3D-Kamera ist jedoch das bereits erhältliche iPhone 15 Pro, das zwei seiner Objektive für stereoskopische Videos kombiniert. In 3D lassen

VisionOS mit Basisgesten bedienen



Grundlegende Gesten detektiert das visionOS ab Werk. Entwickler können für ihre Apps weitere definieren.

sich solche Aufnahmen bisher nur mit dem Vision-Pro-Headset betrachten.

Laufende Aufnahmen signalisiert eine blinkende LED. Die maximale Länge hat Apple bisher nicht genannt. Zumindest Profianwender würden es für Studioaufnahmen begrüßen, wenn man die Daten umgehend auf externe Medien speichern könnte.

Privatsphärenschutz

Das für die Bedienung erforderliche Augentracking verrät genau, wofür sich ein Headset-Träger interessiert. Diese Daten würde die Werbeindustrie sehr gerne zum Erstellen von Profilen verwenden, um individualisierte Werbung auszuspielen. Das Headset behält sie aber für sich, um die Privatsphäre der Nutzer zu schützen.

Im Headset verwendet Apple ein sehr flaches Spiegellinsensystem in Pancake-Bauform. Gegenüber Fresnel-Linsen, wie sie unter anderem in der Meta Quest 2 und der Playstation VR2 sitzen, schlucken sie

etwas mehr Licht, mischen jedoch weniger optische Fehler wie Farbsäume ins Bild. Sehfehler des Trägers lassen sich mit zusätzlichen, speziell von Zeiss angefertigten Linsen kompensieren.

Das Raumempfinden verstärkt Apple mit einer Audio Raytracing genannten Technik. Dafür analysiert das Headset mit den Mikrofonen die akustische Beschaffenheit des Aufenthaltsorts (zum Beispiel Teppichboden und schwere Vorhänge) und appliziert dieselben Eigenschaften auf den Klang der virtuellen Realität. Mit Ambient Spatial Audio simuliert es Richtung und Entfernung von künstlichen Geräuschquellen.

Die Lautsprecher sitzen über den Ohren, weshalb man Umweltgeräusche wie die Türklingel gut wahrnimmt. Für besseren Klang kann der Nutzer von sich selbst ein 3D-Audioprofil mittels eines iPhones mit TrueDepth-Kamera (also ab der Modellreihe X) aufnehmen. Damit passt das Headset die Audioausgabe an die Kopf- und Ohrgeometrie an. Alternativ kann man zur Audiomeldung AirPods ankoppeln.

Ausblick

Die Vision Pro ist Apples first big thing, das erste große neue Produkt seit der Einführung des iPhones vor über 15 Jahren. Anders als beim iPhone stützt sich der Hersteller auf ein viel größeres hauseigenes Entwicklerteam mit enormer Hard- und Softwareexpertise, außerdem auf stromsparende Hochleistungschips, ein Fundament aus einem riesigen App-Angebot und auf ein intelligent erweitertes Bedienkonzept.

Das Potenzial und die Ambitionen der Firma sind an vielen Stellen sichtbar, ob beim Bedienkonzept, dem Bild oder dem Klang. Das sind sehr gute Voraussetzungen für den Start. Trotz eines hohen Preises wäre es keine Überraschung, wenn Interessenten am ersten Verkaufstag vor Apples Läden Schlange stehen. Gerüchteküche wollen gehört haben, dass Apple an einem abgespeckten Modell arbeitet und dabei manche Kamera, den einen oder anderen Sensor und das Außendisplay streicht und einen Preis zwischen 2500 und 1500 US-Dollar anpeilt. Das dürfte das Interesse daran nur noch mehr befeuern. (dz@ct.de) **ct**

Dieser Beitrag gründet auf einem Artikel unserer Schwesterzeitschrift Mac & i.

Infos zu MX-Headsets: ct.de/y7vd



Netzbeschleunigung

Wi-Fi 7 treibt WLAN-Tempo hoch, Ethernet zieht nach

Die ersten Wi-Fi-7-Router schicken unter günstigen Bedingungen mehrere Gigabit pro Sekunde durch die Luft und bringen so mehr Geschwindigkeit ins Netz. Da muss sich das etablierte Kabel-LAN ganz schön anstrengen.

Von Ernst Ahlers

Wer sein Netz nur braucht, um einen mäßig schnellen Internetanschluss im Haus zu verteilen, der hat mit der etablierten WLAN-Version Wi-Fi 6 und Gigabit-Ethernet übers LAN-Kabel längst genug Reserven. Doch bei 100 Mbit/s, wovon viele leider immer noch nur träumen können, ist längst nicht das Ende der Fahnenstange erreicht: Internet per TV-Kabel oder Glasfaser bietet schon 1000 Mbit/s und hat in den kommenden Ausprägungen DOCSIS 4.0 beziehungsweise XGS-PON mit einer Spitzendatenrate von 10.000 Mbit/s entsprechend 10 Gbit/s noch viel Luft nach oben.

Klar, allein zum 4K-Streamen und Websurfen braucht man derartige Rasanz nicht. Manchen kann es aber nie schnell genug gehen, beispielsweise Videoproduzenten oder Softwareentwicklern, die ständig gigabytengroße Videos oder Images virtueller Maschinen durchs Netz schicken.

Besonders diesen Nutzergruppen kommt die jüngste WLAN-Inkarnation namens Wi-Fi 7 gerade recht: Sie kann im neuen 6-GHz-Funkband ein mit 320 Megahertz extrabreites Signal nutzen. Über zwei Antennen, die bei besseren Notebooks und Smartphones üblich sind, gehen dann unter günstigen Umständen

– geringe Distanz, Nachbar-WLANs gerade inaktiv – satte 5,8 Gbit/s brutto durch die Luft.

Mit dem bei aktuellen WLAN-Standards üblichen Netto/Brutto-Verhältnis von 70 Prozent übersetzt sich das auf Anwendungsebene zu rund 4 Gbit/s, was ungefähr 480 MByte/s beim Dateikopieren entspricht. Diese Geschwindigkeit liegt im Bereich der in PCs noch verbreiteten SATA-Schnittstelle für Massenspeicher (siehe Tabelle auf S. 162). Festplatten mit drehenden Scheiben, die aktuell bis zu 250 MByte/s schaffen, hängt Wi-Fi 7 locker ab.

Um auch SATA-SSDs mit ihren maximalen 560 MByte/s zu überholen, muss Wi-Fi 7 seinen spitzesten Pfeil aus dem Köcher ziehen, die neue Übertragungstechnik Multi-Link Operation. Funken Basis und Notebook mit MLO parallel in zwei Bändern, dann klettert der Durchsatz in der Nähe um die Hälfte (320-MHz-Signal im 6-GHz-Band, dazu 160 MHz bei 5 GHz). Über größere Distanzen, wo die Verbindungen über schmalere Signale laufen, kann MLO die Datenrate sogar verdoppeln, wenn auch auf niedrigerem Geschwindigkeitsniveau.

Dass Multigigabit-WLAN-Datenraten nicht illusorisch sind und die MLO-Technik funktioniert, konnten wir im Labor bereits bestätigen [1]: Zwei Wi-Fi-7-Router TP-Link Deco BE85 übertrugen im Mesh-Betrieb untereinander auf kurze Distanz 3 Gbit/s, mit mehreren parallelen TCP-Datenströmen sogar 7,1 Gbit/s.

Als Patchkabelersatz für mobile Arbeitsgeräte taugt Wi-Fi 7 also durchaus schon, auch wenn der zugehörige IEEE-Standard 802.11be noch gar nicht verabschiedet ist; die verwendeten Geräte arbeiten nach einem Entwurf.

Wi-Fi-7-Extras

Simultaner Betrieb in zwei Funkbändern ist aber optional: MLO kennt zwei Spielarten für Mobilgeräte, die nur ein Funkmodul haben (Multi-Link Single-Radio, MLSR und Enhanced MLSR, EMLSR). Die sind besonders in Smartphones sinnvoll, weil zwei parallel funkende WLAN-Schnittstellen stärker an der Akkuladung zehren würden als eine.

MLSR hält die Verbindung zur Basis in zwei oder gar drei Funkbändern gleichzeitig und überträgt Daten fallweise im jeweils günstigsten. Das steigert im Mittel die Geschwindigkeit sowie Zuverlässigkeit und senkt als Nebenwirkung die Latenz. EMLSR ergänzt das um einen zusätzlichen

Das erste Wi-Fi-7-Mesh-Gerätepaar übertrug im c't-Labor untereinander auf kurze Distanz mehrere Gigabit pro Sekunde. Bei sechs parallelen Übertragungen waren es selbst über 20 Meter durch Wände noch knapp 2,5 Gbit/s.



Empfänger. Damit kann das Mobilgerät bei moderatem Mehrenergiebedarf auf zwei Bändern gleichzeitig auf die WLAN-Basis lauschen. So weiß es früher, ob Daten bereitliegen; die mittlere Latenz sinkt weiter.

Wer bei einem künftigen Mobilgerät Wert auf optimales WLAN legt, sollte also auf Wi-Fi 7 mit zwei Optionen achten: 320-MHz-Betrieb im 6-GHz-Band und mindestens MLSR, besser EMLSR.

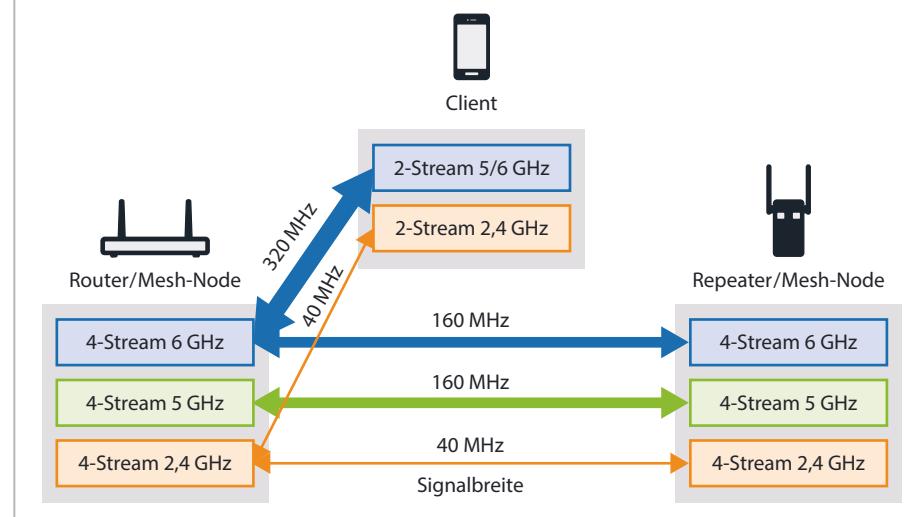
Für Notebooks mit ihren hochkapazitiven Akkus und Nachrüst-Adapter (PCI-

Express oder USB) werden auch die MLO-Varianten Non-STR MLSR, STR MLSR und Enhanced MLSR interessant. STR MLSR steht für Simultaneous Transmit/Receive Multi-Link Multi-Radio, also gleichzeitiges Senden und Empfangen über zwei oder mehr Funkmodule, was dem Vollduplexbetrieb im LAN-Kabel ähnelt. Non-STR beschränkt sich auf simultanes Senden oder Empfangen.

EMLSR schreibt das Prinzip von EMLSR bei Mobilgeräten fort, die Daten in zwei Bändern gleichzeitig übertragen

Multi-Link Operation bei Wi-Fi 7

Ein Wi-Fi-7-Smartphone mit zwei Antennen kann über ein extrabreites Signal im 6-GHz-Band bis zu 5,8 Gbit/s brutto bekommen. Die auf 2,4 GHz parallel möglichen 0,7 Gbit/s lohnen aber das Addieren per MLO kaum. Attraktiver ist, dynamisch zu wechseln und Streaming-Aussetzer zu vermeiden, weil die niederfrequente Verbindung bei steigender Distanz länger hält.





Das LAN für neues WLAN zu ertüchtigen, wird erschwinglich: Einfache Switches für 2,5 Gbit/s mit fünf Ports sind inzwischen deutlich unter 100 Euro zu haben. USB-Adapter und Einbaukarten zum Aufrüsten von PCs kosten ab 25 Euro.

können und mit einem zusätzlichen Empfänger in einem dritten Frequenzblock lauschen.

Laut Intel soll MLSR bei einer zu 30 Prozent ausgelasteten Funkzelle gegenüber klassischem Betrieb (Single Link Single Radio) rund die Hälfte mehr Durchsatz bringen (circa 310 statt 200 Mbit/s), MLMR ihn sogar knapp verdoppeln (circa 390 Mbit/s, [2]).

Abgesehen von MLO bringt Wi-Fi 7 einige weitere Neuerungen wie Multiple Ressource Units an einen Client (verbessert Spektrumsausnutzung) oder Restricted Target Wait Time (R-TWT, passt TWT von Wi-Fi 6 für echtzeitfähige Netzwerke an). Diese hier auszuführen, würde den Rahmen sprengen, weshalb wir bei passender Gelegenheit einen Grundlagenbeitrag bringen.

Standard und Realität

Der noch im Werden befindliche IEEE-Standard 802.11be beschreibt, um welche neuen Funktionen die bisherige WLAN-Norm 802.11-2020 erweitert wird und welche der bestehenden wie ergänzt werden müssen. Im August 2023 hatte 802.11be den Entwurfsstand 4.0 und etwas über 1000 Seiten Umfang erreicht. Die finale Version wird wohl erst Ende 2024 verabschiedet.

Welche der zahlreichen, in 11be definierten Optionen die WLAN-Chiphersteller umsetzen, ist wie schon bei früheren Technikgenerationen ihnen überlassen. Deshalb kümmert sich die Herstellervereinigung Wi-Fi Alliance (WFA) seit Langem mit ihrem Zertifizierungsprogramm um eine grundlegende Kompatibilität.

Das wird auch bei Wi-Fi 7 kommen, doch welche Funktionen Produkte für das „WLAN-Wollsiegel“ obligatorisch mitbringen müssen, wird frühestens 2024 bekannt gegeben. Eventuell wird es abgestufte Prüfungen geben, beispielsweise für 320-MHz-Betrieb und MLO mit einem Funkmodul oder mehreren, die dann in Zusätzen zu „Wi-Fi 7“ münden.

Da der Markteintritt nicht von einer WFA-Zertifizierung abhängt, werden in den nächsten Jahren gewiss WLAN-Produkte

herauskommen, die sich Wi-Fi-7-kompatibel nennen, aber nützliche Optionen missen lassen. Wer „gutes“ WLAN will, muss sich mindestens mit einem Blick ins Datenblatt vergewissern, ob die Neuanschaffung die interessanten Wi-Fi-7-Funktionen beherrscht. Hat das Produkt ein Wi-Fi-Siegel, kann man das wenigstens für einen Grundumfang als gegeben ansehen.

LAN-Turbo NBase-T

Schon Wi-Fi 6 konnte unter günstigen Bedingungen deutlich über 1 Gbit/s transportieren. Bei einem verteilten WLAN aus mehreren per LAN-Kabel angebundenen Access-Points war der Verlust durch die mit Gigabit-Ethernet möglichen 940 Mbit/s netto noch erträglich. Wi-Fi 7 macht nun deutlich, dass man für solche Szenarien schnelleres Ethernet braucht.

10-Gigabit-Ethernet in der optischen Variante (10GBase-Lx/Sx) ist seit 2002 normiert. Auch die Ausführung 10GBase-T für Kabel mit dem gewohnten RJ45-Stecker hat schon 17 Jahre auf dem Buckel. Sie setzte sich aber nicht auf breiter Front durch, weil bisher kaum jemand derart hohe Datenraten am PC brauchte.

So kam 2016 eine Norm für die Zwischenstufen 2,5 und 5 Gbit/s dazu: NBase-T ermöglicht Multigigabit-Ethernet auch auf älteren Kabelinstallationen (CAT5e).

Typische Datenraten von Netztechniken und PC-Schnittstellen

| Technik | Standard | Bruttodatenrate ¹ | Nettdatenrate ² | in MByte/s ² |
|-----------------------|------------------------------|------------------------------|----------------------------|---------------------------------|
| NFC | ISO/IEC 14443 / 15693 | 0,0004 Gbit/s | 0,0004 Gbit/s | 0,05 MByte/s |
| Bluetooth | 2.1/3.0/4.2/5.3 (EDR) | 0,003 Gbit/s | 0,0021 Gbit/s | 0,26 MByte/s |
| Fast Ethernet | 100Base-T | 0,1 Gbit/s | 0,094 Gbit/s | bis 11 MByte/s |
| WLAN | Wi-Fi 4 (IEEE 802.11n) | 0,072-0,3 Gbit/s | 0,036-0,15 Gbit/s | bis 20 MByte/s |
| USB | 2.0 ³ | 0,48 Gbit/s | 0,38 Gbit/s | bis 45 MByte/s |
| WLAN | Wi-Fi 5 (IEEE 802.11ac) | 0,433-1,73 Gbit/s | 0,24-1,05 Gbit/s | bis 130 MByte/s |
| Gigabit-Ethernet | 1000Base-T | 1 Gbit/s | 0,94 Gbit/s | bis 115 MByte/s |
| WLAN | Wi-Fi 6 (IEEE 802.11ax) | 0,6-2,4 Gbit/s | 0,3-1,6 Gbit/s | bis 190 MByte/s |
| Multigigabit-Ethernet | NBase-T 2,5 | 2,5 Gbit/s | 2,4 Gbit/s | bis 280 MByte/s |
| USB | 3.0 = 3.2 Gen 1 ³ | 5 Gbit/s | 4,0 Gbit/s | bis 480 MByte/s |
| Multigigabit-Ethernet | NBase-T 5 | 5 Gbit/s | 4,7 Gbit/s | bis 560 MByte/s |
| SATA | 3 (6G) ³ | 6 Gbit/s | 4,5 Gbit/s | bis 560 MByte/s |
| WLAN | Wi-Fi 7 (IEEE 802.11be) | 0,7-11,5 Gbit/s | 0,5-8,0 Gbit/s | bis 1000 MByte/s |
| USB | 3.2 Gen 2 ³ | 10 Gbit/s | 8,0 Gbit/s | bis 1100 MByte/s |
| 10-Gigabit-Ethernet | 10GBase-T/SR/LR | 10 Gbit/s | 9,4 Gbit/s | bis 1100 MByte/s |
| USB | 3.2 Gen 2x2 ³ | 20 Gbit/s | 17 Gbit/s | bis 2100 MByte/s |
| 25-Gigabit-Ethernet | 25GBase-T/xR ⁴ | 25 Gbit/s | 24 Gbit/s ⁵ | bis 2900 MByte/s ⁵ |
| 40-Gigabit-Ethernet | 40GBase-T/xR ⁴ | 40 Gbit/s | 38 Gbit/s ⁵ | bis 4600 MByte/s ⁵ |
| 100-Gigabit-Ethernet | 100GBase-T/xR ⁴ | 100 Gbit/s | 94 Gbit/s ⁵ | bis 11.000 MByte/s ⁵ |

¹ hängt bei WLAN in der Praxis von der Hardwareausstattung der Geräte (Anzahl Antennen) sowie des verwendeten Funkbandes, der nutzbaren Signalbreite und der Entfernung ab

² für große Dateien bei Ethernet und WLAN, bei WLAN stark von der Belegung des Funkbandes durch Nachbarnetze beeinflusst

³ Nettdatenrate abhängig von Chipsätzen in PCs und Massenspeichern

⁴ verschiedene Medientypen, Kupfer nur über wenige Meter

⁵ hochgerechnet

Dass dabei die Datenraten 5 und 10 Gbit/s optional sind, kommt den Geräteherstellern entgegen. So kann man an der Hardware sparen und in ein paar Jahren beim nächsten Geschwindigkeits- schritt neue Adapter und Switches verkaufen.

Mit 2,5 Gbit/s hat sich NBase-T bereits auf PC-Mainboards und Netzwerkspeicher (NAS) ausgebreitet. Bei Letzteren harmoniert diese Datenrate gut mit Festplatten-Arrays, die beim Transfer großer Dateien bis zu 250 MByte/s Schreib/Lese-Geschwindigkeit schaffen.

Seit Sommer 2023 sind endlich auch kleine NBase-T-Switches für 2,5 Gbit/s erschwinglich geworden, manche 5-Port-Modelle gehen schon für weniger als 80 Euro über den Tresen. Passende USB- und PCI-Express-Adapter für PCs liegen bei 25 bis 30 Euro. Das macht das Aufrüsten kleiner Netze attraktiv: Mit einer moderaten Investition beschleunigt man das LAN aufs Zwei- einhalb-fache.

Wer die WLAN-Wucht von Wi-Fi 7 verlustfrei ins LAN bringen will, kommt um 10GBase-T nicht herum. Reine USB-Adapter gibt es noch nicht, PCI-Express-Karten für PCs sind ab 70 Euro zu haben. Ein kleiner Switch, der NBase-T bis 10 Gbit/s an mindestens vier Ports spricht, kostet im Herbst 2023 mindestens 260 Euro.

Verschwiegen sei auch nicht, dass mit beidem – Wi-Fi 7 und NBase-T – höhere Energiekosten einhergehen: Der mit drei Funkmodulen bestückte Mesh-Router Deco BE85 zog mit 10-Gigabit-Ethernet-Links um die 20 Watt. Vergleichbare Router der vorigen Wi-Fi-6E-Generation brauchen mindestens 5 Watt weniger (Asus GT-AXE11000: 14 Watt). Bei NBase-T hängt die Leistungsaufnahme aktiver Ethernet-Ports nahezu linear von der ausgetauschten Geschwindigkeit ab: Bei 1 Gbit/s sind es typischerweise 0,3 Watt, bei 10 Gbit/s um die 2 Watt.

Ausblick

Mit dem Ertüchtigen des LANs können sich Wi-Fi-7-Interessenten noch Zeit lassen: Bisher ist die Auswahl von Wi-Fi-7-Basen sehr mager und diese sind sehr teuer. Beim Verfassen dieses Artikels warf der Heise-Preisvergleich gerade mal ein bestellbares Produkt aus, ein Dreier-Mesh-Set Deco BE65 von TP-Link für stolze 1000 Euro bei genau einem Anbieter. Beim Online-Kaufhaus Amazon ginge ein einzelner Deco BE85 für 600 Euro auf die Reise, das Doppelpack für 1100 Euro. Der deutsche Hersteller AVM gab zur Fachmesse IFA im September an, seine ersten Wi-Fi-7-Fritzboxen 5690 Pro und 6670 Cable Ende 2023 für 339 beziehungsweise 289 Euro (UVP) auf den Markt zu bringen.

Clientseitig sieht es ähnlich spärlich aus: Lenovo kündigte zur Gamescom 2023 ein 5000 Euro teures Gamer-Notebook mit Wi-Fi 7 an, einige Mainboard-Hersteller entsprechend bestückte PC-Hauptplatten. Abgesehen davon waren lediglich eine Handvoll Smartphones mit Wi-Fi 7 zu finden, angefangen beim 700 Euro teuren Xiaomi 13. Bis Intels Wi-Fi-7-Kärtchen BE200 zum Aufrüsten älterer Notebooks für vielleicht 30 Euro zu kaufen ist, dürften noch ein paar Monate vergehen. [ea@ct.de\) ct](mailto:(ea@ct.de) ct)

Literatur

- [1] Ernst Ahlers, WLAN zieht davon, Wi-Fi 7: Die nächste WLAN-Generation funkstreckt, c't 16/2023, S. 12
- [2] Carlos Cordeiro, Wi-Fi Unleashed: Wi-Fi 7, 6 GHz, and Beyond, Intel, Juni 2022: heise.de/s/jzBL

**2 Ausgaben MIT Technology Review
als Heft oder digital
inklusive Prämie nach Wahl**

mit-tr.de/testen

Ausblick auf kommende Ereignisse

de.wikipedia.org/wiki/2024

Was erwartet uns im kommenden Jahr? 2024 wird sicher viele Überraschungen bereithalten. Aber einige (Groß-)Ereignisse des Jahres sind bereits lange geplant. Einen Überblick bietet die **Kalenderübersicht in der Wikipedia**. Unter „Voraussichtliche Ereignisse“ listet sie Sportliches von der Handball-EM der Männer bis zu den Olympischen Sommerspielen sowie Politisches, etwa die Landtagswahlen in Sachsen, Thüringen und Brandenburg und die Präsidentschaftswahlen in den USA.

Einige fest stehende Ereignisse wie die Übernahme der EU-Ratspräsidentschaft durch Belgien im 1. Halbjahr führt die Seite ebenso auf wie ausgewählte Jahres- und Geburtstage. Am 4. April 2024 zum Beispiel jährt sich die Gründung der NATO zum 75. Mal. Und der Seefahrer und Entdecker Vasco da Gama wäre am Heiligabend 2024 500 Jahre alt geworden. Sie möchten noch weiter in die Zukunft schauen? Jahres-Seiten gibt es auch für 2025 und die darauffolgenden Jahre. (jo@ct.de)

Kulturvorschau

en.wikipedia.org/wiki/2024_in_public_domain

kino.de/filme/jahre/2024/

Eine eigene Seite führt die Wikipedia über Werke, die in einem Jahr **gemeinfrei** werden, weil sich der Todestag des Urhebers – je nach der Rechtsprechung des Heimatlandes – zum 50., 60., 70. oder 80. Mal jährt. Die Kompositionen des russischen Komponisten Sergei Prokofjew zum Beispiel, der am 5. März 1953 verstarb, gehen 2024 in die Public Domain über, ebenso wie

| Countries with life + 50 years | | | | | | |
|--|----------------|-------------------|-------------------|---------------------|-------------------------------|--|
| In most countries of Africa and Asia, as well as Belarus, Bolivia, New Zealand, and Uruguay, a work enters the public domain 50 years after the creator's death. | | | | | | |
| Names | Country | Birth | Death | Occupation | Notable work | |
| Zainal Abidin Ahmad | Malaysia | 16 September 1895 | 23 October 1973 | Writer, linguist | | |
| Charles Brasch | New Zealand | 27 July 1909 | 20 May 1973 | Poet | | |
| Isabel Peacocke | New Zealand | 31 January 1881 | 12 October 1973 | Children's writer | | |
| Pablo Picasso | Spain | 25 October 1881 | 8 April 1973 | Painter | | |
| William Plomer | South Africa | 10 December 1903 | 20 September 1973 | Writer | | |
| J. R. R. Tolkien | United Kingdom | 3 January 1892 | 2 September 1973 | Writer, philologist | J. R. R. Tolkien bibliography | |

die Stücke des belgischen Jazz-Gitarristen Django Reinhardt und das Werk des britischen Autors J.R.R. Tolkien.

Apropos Tolkien: Zehn Jahre nach dem letzten „Hobbit“-Teil kommt 2024 ein animiertes Abenteuer aus Tolkiens Mittelerde auf die Leinwand. Der Film spielt im Dritten Zeitalter, 183 Jahre vor der Handlung der „Herr der Ringe“-Triologie. Sie müssen sich allerdings bis zum Dezember des kommenden Jahres gedulden.

Bis dahin ist das **Kinojahr 2024** vollgepackt mit spannenden Filmen, darunter etliche Fortsetzungen bekannter Filme und Filmreihen. Im Februar wird, 37 Jahre nach dem ersten Teil, die Geschichte von Dirty Dancing weitergeschrieben. Die John-Wick-Reihe wird mit dem Spin-Off „Ballerina“ fortgesetzt, bei dem eine von Ana de Armas gespielte Auftragskillerin im Vordergrund steht. Mehr Vorgeschmack findet sich in der Übersicht von kino.de. (jo@ct.de)

Ein besonderer Tag im Jahr

eintagimjahr.wordpress.com

Im Jahr 1960 startete die Autorin Christa Wolf ein Tagebuchprojekt, das sie auch als Buch **Ein Tag im Jahr** herausbrachte. Jedes Jahr notierte sie nach dem 27. September ihre Gedanken

und Erlebnisse dieses Tages: alltägliche Beobachtungen, etwa des Wetters, ihre Erlebnisse als Mutter und berufstätige Frau, aber auch die bedeutende Frage: Sollte sie in der DDR verbleiben oder nicht?

Susanne Hösel und Christina Müller haben dieses Projekt zusammen mit anderen Autoren in die Gegenwart geholt und schreiben es weiter. Am 27. September jedes Jahres werfen sie einen genauen Blick auf ihre Leben, schreiben ihre Beobachtungen auf und teilen sie in ihrem Blog. Wer mitschreiben möchte, darf die beiden Betreiberinnen gerne kontaktieren. (jo@ct.de)

Diese Seite mit klickbaren Links: ct.de/yjm9

Richtig entschieden!

Eine neue Homepage von
heise regioconcept.

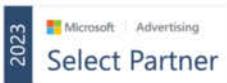
Sie brauchen eine bezahlbare Homepage mit Rundum-sorglos-Service? Dann lehnen Sie sich entspannt zurück:
Das machen wir für Sie.

Unsere erfahrenen Online-Mediendesigner bereiten einen individuellen Entwurf für Sie vor und gestalten die Inhalte ganz nach Ihren Wünschen.

Full-Service, maßgeschneidert und mit Köpfchen!

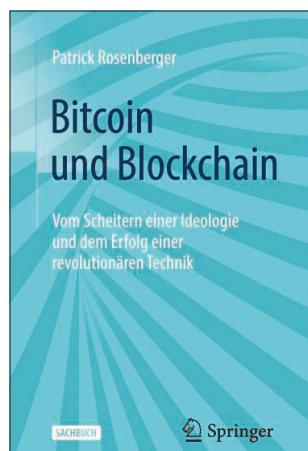
Schon ab
55,- Euro
pro Monat

Erfolgreiches
Marketing
vor Ort.



 **heise regioconcept**

Telefon 0511 80 90 89 43
www.heise-regioconcept.de



Patrick Rosenberger
Bitcoin und Blockchain
 Vom Scheitern einer Ideologie und dem Erfolg einer revolutionären Technik
 Springer, Berlin 2023 (2. Aufl.)
 ISBN 978-3662665299
 274 Seiten, 23 €
 (PDF-/Epub-/Kindle-E-Book: 17 €)

Digitale Geldsorgen

Mancher glaubt, Kryptowährungen könnten helfen, drängende gesellschaftliche Probleme zu lösen. Mancher hält sie hingegen nur für Spekulantenspielzeug. Rosenbergers Buch liefert Grundlagen für eine realistische Perspektive.

Der freie Journalist Patrick Rosenberger sieht in der Blockchain-Technik das „vielleicht größte sozioökonomische Experiment des Informationszeitalters“. Sein Buch beleuchtet in zweiter Auflage den aktuellen Stand der bisweilen wie eine Achterbahnfahrt anmutenden Entwicklung in Sachen Kryptogeld. Es vermittelt zudem eine Menge Hintergrundwissen insbesondere in wirtschaftlicher Hinsicht. Um technische Details geht es hingegen weniger. Rosenberger befasst sich zunächst kurz mit der Geschichte des Geldes allgemein; anschließend beschreibt er verschiedene Kryptowährungen und die Blockchains, die diese nutzen. Dabei kommen alle namhaften Vertreter zur Sprache, etwa Bitcoin, Ethereum, NEO, Ripple und Dash. Die bisweilen subtilen Unterschiede zwischen den Konzepten arbeitet er gut heraus.

Besondere Aufmerksamkeit widmet er dem Bitcoin. Seine Ausführungen etwa zu dessen Auswirkungen auf Wikileaks und zum rätselhaften Bitcoin-Schöpfer Satoshi Nakamoto lesen sich streckenweise wie ein Thriller. Überhaupt kommt die dunkle Seite des Themas nicht zu kurz: Rosenberger beschreibt, welche Bedeutung Kryptowährungen für das Dark Web und dessen Handelsplätze hatten und haben. Gewissermaßen im Vorübergehen erklärt der Autor das Prinzip von Blockchains und streift Verfahren wie Bitcoin-Mining. Vorrangig geht es ihm um gesellschaftliche Aspekte. Was den Bitcoin angeht, ist seine Analyse ernüchternd: Darin sieht er heute eher ein Spekulationsobjekt als ein Zahlungsmittel. Das Konzept hat auch nicht zur ursprünglich erhofften engeren Vernetzung von Menschen im Sinne einer unabhängigen weltweiten Gemeinschaft geführt.

Eine neue, sinnvoll regulierte Welle von Kryptowährungen mag in dieser Hinsicht Besseres bewirken – bis dahin hofft Rosenberger, dass andere Blockchain-Anwendungen wie Smart Contracts und NFTs das vernetzte Leben fördern. Bereits heute verkaufen beispielsweise immer mehr Menschen selbst produzierten Strom mithilfe von Blockchain-Technik direkt statt über einen Vermittler.

Insgesamt verschafft das kompetent verfasste Buch Lesern einen guten Einstieg in die Welt von Blockchains und Kryptogeld, ohne technisch in die Tiefe zu gehen. Die Lektüre ist ansprechend und unterhaltsam.

(Maik Schmidt/psz@ct.de)

Wider die digitale Bequemlichkeit

Soziale Netzwerke – in diesem Buch als „Plattformen“ bezeichnet – spielen eine zentrale Rolle für unsere Gesellschaft. Ihre Betreiber sind aber nicht dem Gemeinwohl verpflichtet, sondern streben nach Gewinnmaximierung. Für Björn Staschen handelt es sich um „egoistische Giganten“.

Dieser Lagebericht zu unserer Medienlandschaft ist hervorragend recherchiert und mit 240 Quellen belegt. Staschen formuliert klar, argumentiert präzise und kommt schnell auf den Punkt. Und doch fällt es mitunter schwer weiterzulesen, denn fast auf jeder Seite fühlt man sich ertappt, weil man selbst auch zu den zahllosen Nutzern gehört, die wider besseres Wissen den bequemen Weg gehen und Google, Facebook, WhatsApp, TikTok, Instagram verwenden.

Kein Wunder, dass das so ist, sagt Staschen: „Die Plattformen haben ihre Angebote bewusst genauso gestaltet und immer weiter perfektioniert, damit wir von ihnen abhängig werden.“ Je mehr Nutzer die Plattformen bekommen, desto größer wird ihre Macht. Ihre Algorithmen steuern dem Autor zufolge die Auslieferung von journalistischen Inhalten und Nachrichten in einer Weise, die unsere Demokratie bedroht.

Ein ganzes Kapitel widmet sich den Genies und Egomanen an der Spitze der großen Tech-Firmen und ihren Netzwerken mit teils mafiosen Strukturen. Dann geht es um journalistische Tugenden und den gesetzlichen Rahmen für öffentlich-rechtliche Medien. Staschen, der selbst in einer NDR-Redaktion arbeitet, beklagt Wettbewerbsverzerrung durch das Medienrecht. So darf Spotify als Streamingdienst ohne Unterbrechung Musik liefern, während lizenzierte Vollprogramme in Radio und Fernsehen auch Nachrichten anbieten müssen.

Und wie geht das alles nun besser? Die letzten drei Kapitel nennen jede Menge alternativer Projekte, das Fediverse mit der Twitter-Alternative Mastodon etwa, Friendica als Facebook-Ersatz, PeerTube als YouTube-Ersatz. Staschen ermutigt dazu, auf dem Elternabend gegen eine WhatsApp-Elterngruppe zu stimmen, und nennt bessere Wege der Kommunikation. Er macht keine falschen Versprechungen, aber Hoffnung und sogar Lust, weniger bequeme Wege zu gehen.

(dwi@ct.de)



Björn Staschen
In der Social-Media-Falle
 Wie wir unsere digitale Freiheit retten können

Hirzel, Stuttgart 2023
 ISBN 978-3777633893
 222 Seiten, 24 €
 (Epub-Book: 22 €)



heise academy



Stark gegen Hacker

Ein Klick, Ihre Entscheidung.

Cyberkriminelle werden immer raffinierter und Angriffe immer ausgefeilter. Deshalb sind Weiterbildungen für IT-Professionals in der **IT-Security** unverzichtbar. Bleiben Sie am Puls der Zeit, vertiefen Sie Ihr Wissen und lernen es praktisch einzusetzen, um Ihre IT zu schützen.

Praxisnahe IT-Wissen, vermittelt von erstklassigen Referenten: Die Schulungen der heise academy überzeugen vor allem durch thematischen Tiefgang, aktuelle Themen und individuelle Betreuung. Unsere iX-Workshops bieten Ihnen **Expertenwissen für IT-Professionals**.

Investieren Sie in Ihre berufliche Zukunft – mit unseren praxisnahen Schulungen.



Jetzt handeln und IT-Security-Schulungen entdecken:
heise-academy.de/ix-gegen-hacker





DER NEUE FREUND

von Martin Kobe

Guten Morgen, Andreas. Es ist elf Uhr früh und Zeit, in den neuen, aufregenden Tag zu starten!"

Andreas wälzte sich im Bett herum. Hatte er nicht eben eine Stimme gehör?

„Guten Morgen, lieber Andreas. Heute ist ein wundervoller Aufstehtag.“

Ganz bestimmt nicht, knurrte er in sich hinein. Wer sprach da überhaupt?

Er öffnete seine Augen, konnte aber niemanden entdecken. Wie auch? Er bewohnte diese Wohnung allein. Also ließ er sich wieder tief in sein Kopfkissen sinken.

Oder hatte er gestern zu lange in der Kneipe gesessen und sich jemand Hübsches mitgenommen? Nein, das hatte er definitiv nicht getan. Am Ende des Monats war so etwas finanziell kaum drin.

„Eben, Andreas, hattest du es fast geschafft, aufzustehen. Versuche es doch bitte ein zweites Mal.“

Wenn die Leute anfingen, Stimmen zu hören, war das der Anfang vom Ende, hatte er irgendwo gelesen. Würde er demnächst splitternackt über eine Straßenkreuzung rennen? Oder den Nachbarn mit einer Machete niedermetzeln und sich wenig später ohne Erinnerung daran in einer psychiatrischen Klinik wiederfinden?

„Sehr gut, Andreas, wie du deinen Kreislauf in Schwung gebracht hast! Jetzt brauchst du dich nur noch hinzustellen.“

Andreas richtete sich ein wenig auf. Seine Matratze lag auf dem Fußboden des Schlafzimmers. Von dieser niedrigen Warte aus schaute er sich im Zimmer um. Da war niemand. Nur sein Reinigungsroboter kurvte scheinbar planlos über den Boden. „Hast du etwa mit mir gesprochen?“, fragte er die Maschine und schämte sich zugleich dafür, das zu tun.

„Bingo, Andreas“, kam die Antwort. „Ich bin dein neuer Freund.“

„Na toll. Als ich dich vor ein paar Wochen in dem Elektronik-Ramschladen kaufte, konntest du noch nicht sprechen.“

„Es war gar nicht so einfach, mir einen ausreichenden Datenspeicher, einen Lautsprecher sowie ein Sende- und Empfangsmodul zu verschaffen, aber wie du siehst, hatte ich letztlich Erfolg. Genau wie du eines Tages Erfolg haben wirst.“

Andreas starrte den Roboter an. Ja, der Lautsprecher auf der Oberseite war neu und wirkte professionell eingefügt. „Völlig unmöglich. Du kannst dir den Lautsprecher nicht selbst eingebaut haben. Du besitzt keine dafür geeigneten

Vielen Menschen gelingt es immer weniger, ihr Leben aus eigener Kraft in den Griff zu bekommen. Was könnten nachlässige Großstädter da willkommener finden als einen unverdrießbaren automatischen Motivierer, der sie besser kennt als sie sich selbst und der auch besser weiß, was gut für sie ist?

Werkzeuge, hast keine Arme und irgendwoher musst du ja auch die neuen Teile bekommen.“

„Wie klug du bist, Andreas. Für dich wird sich bestimmt schnell eine passende Arbeit finden lassen.“

„Lenk nicht vom Thema ab. Wie hast du das gemacht?“

Der Roboter fuhr nervös hin und her. Schließlich antwortete er leiser und weniger fröhlich als bisher: „Ich habe Hilfe bekommen. Es gibt mobile Einsatzteams aus spinnenförmigen Minirobotern, die die Veränderungen am Einsatzort vornehmen.“

„Du befindest dich also in meiner Wohnung an einem Einsatzort? Dann sieh zu, dass du schnellstens von hier wieder verschwindest. Ich gehe jetzt ins Café und lese eine Zeitung und wenn ich zurückkomme, bist du weg.“

„Das darf ich leider nicht“, bekannte die Stimme kleinlaut.

Andreas hatte sich inzwischen angezogen und einen Entschluss gefasst. Er griff sich seinen Reinigungsroboter, der mit den Rädern durchdrehte und ihn umzustimmen versuchte. Andreas reagierte nicht darauf und trug ihn durch das Treppenhaus nach draußen. Nicht weit entfernt, am Rande eines Parkwegs, warf er das Gerät auf einen Gerümpelhaufen einer illegalen Müllabladestelle. Neben den üblichen Möbeln und den Plastiksäcken voller Bauschutt lagen dort auffallend viele Elektrogeräte. Es war alles andere als still dort – zuerst vermutete Andreas, der Lärm stammte von spielenden Kindern, tatsächlich aber zeterten weggeworfene Geräte vor sich hin. Ohne weiter auf sie zu achten, schlenderte er ins Café.

Dort ließ er seine Zeitung lange ungelesen vor sich liegen, weil er immer wieder an das seltsame Erlebnis mit dem Roboter dachte. Er verstand es nicht.

Wenig später klingelte sein Smartphone mit einer nicht gespeicherten Nummer: „Herzlichen Glückwunsch, Herr Müller! Aus Ihrem Verhalten schließe ich, dass Sie endlich eine bezahlte Stelle angenommen haben und auf die Leistungen des Jobcenters komplett verzichten können.“

„Mit wem spreche ich bitte?“

„Oh entschuldigen Sie bitte, Herr Müller. Ich glaubte, Sie hätten meine Stimme erkannt. Ich bin Frau Hildebrand, Ihre Sachbearbeiterin beim Jobcenter.“

„Guten Tag, Frau Hildebrand. Wir hatten bei unserem letzten Gespräch vereinbart, dass ich meinen angefangenen

Roman zu Ende schreibe und mich um seine Veröffentlichung bemühe. Ganz so schnell geht es leider doch nicht.“

„Mit unserer Bereitschaft, Ihre Tätigkeit als Schriftsteller zu fördern, sind wir Ihnen weit entgegengekommen. Wie Sie selber zugeben, sind die Aussichten, von Ihrer Schreiberei zu leben, äußerst gering. Gerade in Berlin gibt es Tausende von Schriftstellern, die niemals den Durchbruch schaffen werden, andererseits fehlen an allen Ecken und Enden Fachkräfte. Sei es als Busfahrer, Lehrer oder Altenpfleger. Der Literaturbetrieb in Berlin würde gewiss nicht zusammenbrechen, wenn der eine oder der andere Mochtegernautor sich sinnvoll betätigen würde.“

„Nach Artikel 12 des Grundgesetzes hat jeder Deutsche das Recht, sich seinen Beruf, seine Arbeits- und Ausbildungsstelle frei auszuwählen – und ich habe mich für den Beruf des Geschichtenschreibers entschieden. Wie Sie selbst zugegeben haben, ist es durchaus möglich, als Geschichtenschreiber seinen Lebensunterhalt zu bestreiten. Betrachten Sie die erfolglose Anfangsphase als Ausbildungszeit, wie bei jedem anderen Beruf. Auch als Klempner muss ich meine Lehre beenden, bevor ich Geld verdienen und Steuern zahlen kann.“

„Eine durchschnittliche Ausbildung dauert drei Jahre. Sie hingegen sind schon seit über acht Jahren als erfolgloser Schriftsteller bei uns gemeldet. Deswegen haben wir beschlossen, Sie mit der modernsten Technologie zu unterstützen. Doch statt uns dankbar dafür zu sein, werfen Sie einfach alles hin.“

„Ich habe nichts hingeworfen. Bevor der Tag vorbei ist, werde ich einige Seiten meines neuen Romans geschrieben haben“, schwindelte Andreas, der wusste, dass er vorerst keine einzige Zeile mehr würde schreiben können.

Eine Weile blieb es still am anderen Ende der Leitung. Andreas überlegte, ob er einfach auflegen sollte, aber dann würde es richtig Ärger geben. Heute war irgendwie nicht sein Tag.

Schließlich meldete sich Frau Hildebrand wieder: „Ich habe eben die Vereinbarung mit Ihnen überflogen. Sie haben ausdrücklich der Begleitung durch einen modernen Jobcoach zugestimmt.“

Andreas erschrak: „Wollen Sie damit sagen, dass ich meinen sprechenden Staubsauger Ihnen zu verdanken habe?“

Frau Hildebrand lachte glockenhell: „Das hat aber lange gedauert, Herr Müller. Sie haben gerade Ihren modernen Jobcoach illegal im Park entsorgt.“ Und ernst fügte Sie hinzu: „Als Vertreterin der Steuerzahler, die Ihren Lebensunterhalt bestreiten müssen, bin ich entsetzt, wie Sie jedes Hilfsangebot in den Wind schlagen. Im Grunde bleibt mir keine andere Wahl, als Ihre Leistungen komplett zu streichen.“

„Frau Hildebrand, allenfalls habe ich mich einer Ordnungswidrigkeit schuldig gemacht, indem ich einen mir gehörenden Gegenstand an einem kiezbekannten Ort meinen Nachbarn als Geschenk angeboten habe. Sie hingegen haben einen Einbruch in meine Privatwohnung veranlasst und mein Eigentum beschädigen lassen. Außerdem haben Sie eine Abhöranlage in meiner Wohnung installieren lassen, was allenfalls bei Schwerverbrechern zulässig ist, nicht jedoch bei widerspenstigen Langzeitarbeitslosen.“

Frau Hildebrand atmete tief ein. „Wir haben keine Abhöranlage bei Ihnen installiert.“

Andreas lachte schallend: „Wollen Sie allen Ernstes behaupten, dass die Gespräche Ihrer KI mit mir und allerlei andere Daten nicht schon aus Gründen der Qualitätssicherung irgendwo abrufbar gespeichert werden?“

„Mit den technischen Details der KI bin ich nicht vertraut. Aber vielmehr wundert mich, dass Sie als angeblicher Science-Fiction-Autor die riesige Chance nicht erkennen, die wir Ihnen angeboten haben. Berlin-Mitte ist eine von wenigen Regionen in Deutschland, in denen das KI-gestützte Jobcoaching erprobt wird. Sie sind einer der ersten, die diese technische Neuheit nutzen dürfen. Wenn Sie es geschickt anstellen, werden Sie deswegen weltberühmt. Dafür brauchen Sie das KI-Coaching nicht einmal zu loben. Sie könnten sich zu einem Fachmann auf diesem Gebiet entwickeln.“

„WIR HABEN KEINE ABHÖRANLAGE BEI IHNEN INSTALLIERT.“

Andreas schwieg. So ganz unrecht hatte seine Sachbearbeiterin nicht. Und da er ohnehin ein Thema für seinen Roman brauchte, drängte sich dieses geradezu auf.

In dem Moment öffnete sich die Tür des Cafés und ein Junge mit einem Reinigungsroboter auf dem Arm betrat den Laden. Er schaute sich um: „Heißt hier jemand Andreas und hat seinen besten Freund im Park vergessen?“

Andreas schwieg. Man kannte ihn hier, Leugnen hatte also keinen Zweck. Er stand auf und ging dem Jungen entgegen. „Danke, dass du mir meinen Staubsauger gebracht hast.“ Er streckte seine Arme aus, doch der Junge hielt das Gerät fest. Statt es herzugeben, sagte er: „Ihr Freund versprach mir, dass ich eine hohe Belohnung bekomme, wenn ich ihn zu seinem Besitzer zurückbringe. Sind Sie der Besitzer?“

„Ja, er ist es“, bestätigte der Roboter. „Andreas, du solltest dem Jungen zehn Euro für seinen Dienst aushändigen.“ Zerknirscht zog Andreas seine Geldbörse aus der Tasche.

Im selben Moment redete wieder Frau Hildebrand aus seinem Smartphone – er hatte das Gespräch mit ihr ja noch nicht beendet. Andreas griff nach dem Gerät und hörte: „Wie ich mitbekommen habe, hat sich Ihre Situation inzwischen geklärt. Falls ich keine weiteren Beschwerden über Ihr Fehlverhalten bekomme, brauchen Sie ausnahmsweise nicht mit negativen Folgen zu rechnen.“ Damit legte die Sachbearbeiterin auf.

Andreas betrachtete seinen Reinigungsroboter auf dem Cafétisch. Die Lust aufs Zeitungslesen war ihm vergangen. Er bezahlte und trug seinen behördlich verordneten neuen Freund zurück nach Hause. Dort angekommen, ließ er ihn unsanft auf den Boden krachen.

„Autsch“, rief der Roboter empört. „Mein Fahrgestell ist nicht für einen Sturz aus einem Meter Höhe ausgelegt.“

„Dann wird dir ein Sturz aus dem vierten Stock auf die Straße noch weniger gefallen. Überleg es dir also gut, ob du mich weiter nerven willst.“

„Ich tue nur meine Pflicht und wenn du es genauer betrachtest, müsstest du erkennen, dass du mit dem angedrohten barbarischen Akt allenfalls dein automatisches Reinigungsgerät zertrümmern würdest, nicht jedoch die Vernunft, die aus mir spricht.“

„Apropos Reinigungsgerät: Wie wäre es, wenn du dein Pflichtbewusstsein dadurch beweisen würdest, dass du meine Zimmerböden wischst, anstatt mich mit Geschwätz zu langweilen?“

Der Reinigungsroboter schwieg und begann den Küchenboden zu wischen. Andreas eilte aus der Wohnung und begab sich schnurstracks in seine Stammkneipe. Zum Glück durfte er dort ausnahmsweise anschreiben lassen.

* * *

„Guten Morgen, Andreas. Wusstest du eigentlich, dass du scharnarchst?“

Im ersten Moment erschrak er, aber dann fühlte er eine warme Begleiterin an seiner Seite. Er schaute sie an. Es war Susi. Eigentlich in festen Händen, aber was hieß schon „eigentlich“? Andreas war bereits eine ganze Weile lang scharf auf sie gewesen. Er strich ihr über den Körper und küsste sie. „Als Entschädigung dafür, dass ich dich so lange wachgehalten habe.“

„Ich habe sehr lange wachgelegen“, schmunzelte sie.

Der Sex hatte Andreas gefallen. Nach dem Duschen setzten sich die beiden an den Frühstückstisch. Zum Glück hatte er sich angewöhnt, für derartige – leider viel zu seltene – Notfälle Aufbackbrötchen und einige Leckereien vorrätig zu haben.

Susi erkundigte sich: „Du hast gestern Abend eine wirre Geschichte von einem sprechenden Reinigungsroboter erzählt. Meinstest du etwa diesen hier?“

In dem Moment fuhr sein neuer Freund durch die Küchentür. Andreas erschrak.

Susi lachte, als sie sein entsetztes Gesicht sah. „Solche Geräte sind neuerdings weit verbreitet. Die Böden wischen und saugen sie ganz passabel, aber an die höher gelegenen Stellen kommen diese Modelle noch nicht heran. Kein Grund, deswegen in Panik zu geraten.“

„Ein normaler Reinigungsroboter versetzt mich nicht in Panik. Bis vorgestern war ich ganz zufrieden mit ihm – bis er zu sprechen begann.“

„SOLCHE GERÄTE SIND NEUERDINGS WEIT VERBREITET.“

Susi lachte schallend. „Du solltest wirklich nicht so viel Zeit allein verbringen.“ Sie beugte sich nach unten und lockte den Roboter mit den Fingern. „Duzzi, Duzzi, kleiner Roboter, rede doch ein bisschen mit uns!“

Der Roboter beachtete sie nicht, drehte seine Runde durch die Küche und kehrte dann ins Wohnzimmer zurück.

„Der Sex mit dir hat mir wirklich gutgetan“, lächelte Andreas.

Sie nickte und lächelte zurück. „Dir müsste klar sein, dass ich einen festen Partner habe und dass unsere heutige nette Begegnung eine Ausnahme bleiben wird. Ich bitte dich, niemandem davon zu erzählen.“

Andreas nickte, konnte aber seine Enttäuschung nicht verbergen.

Sie tanzelte zu ihm hinüber, hauchte: „Aber noch bin ich ja hier“ und setzte sich auf seinen Schoß.

* * *

Nachdem Susi gegangen war, räumte Andreas die Küche auf, bummelte ins Wohnzimmer, schaltete den Fernseher ein und ließ sich in seinen bequemen Sessel fallen. Wenig später erlosch das Bild wieder. Er schaute wütend zum Reinigungsroboter hinüber.

Dieser bestätigte: „Ich habe die Kontrolle über das Fernsehgerät übernommen. Du solltest deine kostbaren Vormittagsstunden nicht mit solchem Blödsinn verschwenden.“

Andreas griff nach der Fernbedienung, aber der Bildschirm blieb schwarz. Wütend warf Andreas die Fernbedienung gegen das Gehäuse des Staubsaugers.

„Du gehst nicht sehr sorgsam mit deinen elektronischen Helfern um. Nachdem du dein Tagespensum geschafft hast, darfst du heute Abend gerne einen Film schauen. Es gibt einen Science-Fiction-Klassiker, der dir gefallen dürfte.“

Andreas erhob sich wortlos, suchte seine Sachen zusammen und schritt zur Wohnungstür. In dem Moment klingelte das Handy. Eine unbekannte Nummer.

„Herr Müller, falls Sie jetzt Ihre Wohnung verlassen, statt den Anweisungen Ihres Jobcoaches zu folgen, kürze ich Ihre Leistungen mit sofortiger Wirkung um zehn Prozent.“

Andreas brach die Verbindung ab und schlenderte ins Café. Er besaß nur noch wenige Euro, aber in ein paar Tagen würde es frisches Geld vom Jobcenter geben. Ein mulmiges Gefühl breitete sich in seinem Magen aus. Mit zehn Prozent weniger käme er nicht aus. Im Grunde reichte nicht einmal der Regelsatz. Dann entspannte er sich wieder. So schnell reagierte das Jobcenter nicht. Es blieb Zeit genug, mit seinem Rechtsberater über die neue Situation zu sprechen.

Andreas blieb länger im Café als üblich. Er hatte die Zeitung schon komplett zweimal durchgelesen und die Kellnerin hatte sich mit zunehmend finsterem Gesichtsausdruck mehrmals erkundigt, ob er noch etwas essen oder trinken wollte. Sein Magen knurrte, also beschloss er, nach Hause zu gehen.

* * *

Kaum hatte Andreas die Wohnungstür durchschritten, rollte ihm der Roboter entgegen: „Schön, dass du wieder da bist. Der wundervolle Tag ist noch lange nicht zu Ende.“

„Zuerst werde ich zu Abend essen.“

„Fein. Zähl mir doch bitte auf, was du an Nahrungsmitteln vorrätig hast, damit ich das optimale Gericht für dich zusammenstellen kann.“

„Ganz bestimmt nicht! Ich esse, was ich will.“

„Übrigens ist ein Brief für dich angekommen.“

Der Briefkasten enthielt neben einem Werbeschreiben, das er sofort ungelesen in den Mülleimer warf, einen Brief des Jobcenters.

Über den Autor

Martin Kobe lebt und arbeitet in Berlin. Er beteiligt sich mit wechselndem Erfolg an Literaturwettbewerben, veranstaltet Lesungen und leitet eine Schreibgruppe. Sein Hauptwerk „Die Psychologische Onlinepraxis Dr. Rabbithole“ brachte er mit zwei Co-Autoren heraus. Überwachungsbegehrlichkeiten aller Art steht er sehr kritisch gegenüber. Sein Debüt bei den c't-Stories hatte Martin Kobe mit „Der Drohnenbesuch“ in Ausgabe 15/2023.



Bild: Martin Kobe

*Sehr geehrter Herr Müller,
aufgrund Ihrer anhaltend mangelhaften Kooperationsbereitschaft kürze ich die Ihnen gewährten Leistungen um 20 Prozent. Zehn, weil Sie den Anweisungen Ihres Jobcoaches nicht gefolgt sind, und weitere zehn, weil Sie während eines Gesprächs aufgelegt haben ...*

Andreas betrachtete den Bescheid und schüttelte den Kopf.

„Du brauchst dir keine Sorgen zu machen, Andreas“, tröstete ihn sein neuer Freund. „Für zwei Wochen wird das Geld reichen und danach kannst du zu einer Essensausgabe gehen oder kleine Jobs annehmen – notfalls betteln. Wir werden einen Weg finden.“

Andreas ließ sich erschüttert in den Fernsehsessel fallen und suchte die Fernbedienung.

„WIR WERDEN EINEN WEG FINDEN“, SAGTE DER REINIGUNGSROBOTER.

„Ich schlage vor, dass du mir deine Nahrungsvorräte aufzählst.“

Widerwillig schlurfte Andreas in die Küche und zählte auf, welche Lebensmittel sich dort befanden.

„Ich empfehle dir ein Pilzomelette mit Gewürzgurken.“

Das klang gar nicht schlecht, überlegte Andreas und bereitete sich das Omelette zu. Der neue Freund ließ ihn beim Essen in Ruhe. Anschließend zählte der Staubsauger auf, was aus Gründen gesunder Ernährung aus der Küche entfernt werden müsste: Weizenmehl, Zucker, Süßigkeiten ... Nur bei den Aufbackbrötchen sträubte sich Andreas, weil er die für überraschende Besuche vorgesehen hatte. Tatsächlich erlaubte ihm der Roboter, die angeblich ungesunden Brötchen zu behalten.

„Die nun aussortierten ungeeigneten Nahrungsmittel unterteilen wir in zwei Gruppen. Die angebrochenen Packungen wirfst du in den Müll. Den Rest kannst du irgendwo für deine Nachbarn hinstellen.“

„Ist dir eigentlich aufgefallen, dass der Kühlschrank und die Regale fast leer sind?“

„Gesehen habe ich das nicht, aber während du das nächste Mal schlafen wirst, bekomme ich meine Augen. Wenn deine Aufzählung korrekt war, besitzt du tatsächlich wenig Lebensmittel. Nachdem du den Müll entsorgt hast, sprechen wir über eine Aufstockung der Vorräte.“

„Dieses Gespräch können wir uns schenken. Ich bin total pleite.“

„Für Sonderausgaben steht mir ein kleines Budget zur Verfügung. Während du zur Mülltonne unterwegs bist, stelle ich eine Einkaufsliste zusammen.“

Andreas machte sich mit einem Pappkarton auf den Weg.

* * *

„Guten Morgen, lieber Andreas. Heute ist ein wundervoller Aufstehtag.“

Andreas zuckte zusammen, aber dann setzte er sich auf. Neben seiner Matratze kauerte der Staubsauger mit seinen Sportschuhen. Andreas ließ sich wieder ins Kissen fallen.

„Guten Morgen, lieber Andreas. Ich weiß, dass du keine Lust zu laufen hast, aber du solltest es trotzdem tun. Die Bewegung an frischer Luft wird dich stärken.“

Andreas lachte schallend: „Ganz bestimmt nicht. Aber ich tue es, wenn du mitkommst.“ Er war sich sicher, dass er damit den Staubsauger überforderte. Dann jedoch hörte er ein Surren und verspürte einen leichten Luftzug. Eine künstliche Libelle setzte sich auf die Hülle des Staubsaugers.

„Das ist Erzulie“, erklärte dieser. „Sie wird dich begleiten. Bei ihr handelt es sich um eine sensationelle Neuentwicklung.“

Andreas schüttelte den Kopf und zog sich seine Sportkleidung an. Dann wartete er, bis sich Erzulie auf seine Schulter gesetzt hatte, und marschierte in Richtung des Parks. Als sie an dem Müllhaufen vorbeikamen, riefen die weggeworfenen Geräte, deren Energiespeicher fast aufgebraucht waren, um Hilfe. Erzulie flüsterte erschüttert: „Dass manche Menschen so herzlos und dumm sein können, hätte ich nie gedacht.“

„Wir wollen eben nicht bevormundet werden“, erklärte Andreas.

„Ihr werft wertvolle Hilfe auf den Müll. Wenn wir auf dem Rückweg ein paar Geräte mitnehmen, werden sie dir helfen, ein gutes Leben zu führen. Du kannst den ganzen Tag im Café hocken und Zeitungen lesen, während eines von ihnen deinen Roman zu Ende schreibt. Ein anderes beschäftigt das Jobcenter, das du bald ohnehin nicht mehr brauchen wirst.“

Andreas lächelte.

(psz@ct.de) ct

Die c't-Stories als Hörversion

Unter heise.de/-4491527 können Sie einige c't-Stories als Audiofassung kostenlos herunterladen oder streamen. Die c't-Stories zum Zuhören gibt es auch als RSS-Feed und auf den bekannten Plattformen wie Spotify, Player FM und Apple Podcasts (ct.de/yz13).



**ICH HACKE
KEIN PROGRAMM.
ICH PROGRAMMIERE
AUF ERFOLG.**

Fernstudium Robotik



Ihre Chance in einem schnell wachsenden Zukunftsmarkt. Aus- und Weiterbildung zum Roboter-Techniker und -Programmierer. Kosten-günstiges und praxisgerechtes Studium ohne Vorkenntnisse. Beginn jederzeit.

NEU: Umweltschutz-Techniker, Netzwerk-Techniker, SPS-Techniker

Teststudium ohne Risiko.
GRATIS-Infomappe gleich anfordern!

FERN SCHULE WEBER - seit 1959
Neerstedter Str. 8 - 26197 Großenkneten - Abt. F14
Telefon 0 44 87 / 263 - Telefax 0 44 87 / 264

www.fernenschule-weber.de



Inkl.
Roboterbausätze
und -Modelle



c't feiert 40 Jahre!

LC-POWER

www.lc-power.com



28"



ULTRAscharfes Arbeiten dank LC-Power!

Erleben Sie ein ultrascharfes Office-Feeling mit unserem neuen 28"/71,12 cm-Monitor, LC-M28-4K-UHD! Der Monitor bietet Ihnen neben einer 4K UHD-Auflösung u.a. DP 1.4-, HDMI 2.0- & Audio-Anschlüsse sowie einen höhenverstellbaren Standfuß für ein ergonomisches Arbeiten in Ihrem Office!

ORACLE Feuerwehr www.oraservices.de

xxs-kurze Daten- & Netzkabel: kurze-kabel.de

softaktiv.datensysteme Datenbankapplikationen, Website Boosting, Online-Pressemitteilungen, Unterstützung bei Ihren V-Projekten. Einfach anrufen, Faxen oder eine E-Mail schicken. Telefon: 0511/3884511, Mobil: 0170/3210024, Telefax: 0511/3884512, E-Mail: service@softaktiv.de, Internet: www.softaktiv.de

nginx-Webhosting: timmehosting.de

EDELSTAHL LED SCHILDER: www.3D-buchstabe.com
HAUSNUMMERN nobel 230V~: www.3D-hausnummer.de

www.patchkabel.de - LWL und Netzwerk Kabel

www.embedded-specialists.de

**Anzeigenschluss
für die nächsten
erreichbaren Ausgaben:**

28/2023: 14.11.2023

29/2023: 28.11.2023

01/2024: 08.12.2023

c't – Kleinanzeigen

Private Kleinanzeige:

erste Druckzeile € 10,-; jede weitere Zeile € 8,-

Gewerbliche Kleinanzeige:

erste Druckzeile € 20,-; jede weitere Zeile € 16,-

Chiffre-Anzeige: € 5,- Gebühr

Hinweis: Die Rechnungsstellung erfolgt nach Veröffentlichung der Anzeige!

Name/Vorname

Firma

Str./Nr.

PLZ/Ort

Bitte veröffentlichen Sie den Text in der nächsterreichbaren Ausgabe von c't.

○ Den Betrag habe ich auf Ihr Konto überwiesen.
Sparkasse Hannover,
IBAN DE98 2505 0180 0000 0199 68, BIC SPKH DE 2H

Bei Angeboten: Ich versichere, dass ich alle Rechte an den angebotenen Sachen besitze.

Datum Unterschrift (unter 18, der Erziehungsberechtigte)

Faxnummer: 05 11 / 53 52-200
eMail: dispo@heise.de

Bitte veröffentlichen Sie in der nächsterreichbaren Ausgabe (Vorlaufzeit mind. 3 Wochen) folgende Anzeige im Fließsatz privat gewerblich* (werden in c't mit  gekennzeichnet) Chiffre

€ 10,- (20,-)

€ 18,- (36,-)

€ 26,- (52,-)

€ 34,- (68,-)

€ 42,- (84,-)

€ 50,- (100,-)

€ 58,- (116,-)

€ 66,- (132,-)

Pro Zeile bitte jeweils 45 Buchstaben einschließlich Satzzeichen und Wortzwischenräumen. Wörter, die **fettgedruckt** (nur in der ersten Zeile möglich) erscheinen sollen, unterstreichen Sie bitte. Den genauen Preis können Sie so selbst ablesen. *Der Preis für gewerbliche Kleinanzeigen ist in Klammern angegeben. Soll die Anzeige unter einer Chiffre-Nummer erscheinen, so erhöht sich der Endpreis um € 5,- Chiffre-Gebühr.

Ausfüllen und einsenden an:  **Heise Medien GmbH & Co. KG**
c't-Magazin, Anzeigenabteilung
Karl-Wiechert-Allee 10, 30625 Hannover

↳ Weiterlesen, wo andere aufhören.







Professor
für Technische Informatik (m/w/d)
 Entgeltgruppe E 15 TV-L, Vollzeit, unbefristet

An der **Staatlichen Studienakademie Glauchau** ist die genannte Professorenstelle (m/w/d) zum frühestmöglichen Zeitpunkt zu besetzen.

Alle Informationen zur Stellenausschreibung finden Sie unter:
<https://www.ba-glauchau.de/stellen2023>

Onlinebewerbungen richten Sie bitte an:
 bewerbung@ba-glauchau.de
 (bis 30.11.2023)









Professor für
Wirtschaftsinformatik (m/w/d)
 Entgeltgruppe E 15 TV-L, Vollzeit, unbefristet

An der **Staatlichen Studienakademie Glauchau** ist die genannte Professorenstelle (m/w/d) zum frühestmöglichen Zeitpunkt zu besetzen.

Alle Informationen zur Stellenausschreibung finden Sie unter:
<https://www.ba-glauchau.de/stellen2023>

Onlinebewerbungen richten Sie bitte an:
 bewerbung@ba-glauchau.de
 (bis 30.11.2023)







MACH, WAS WIRKLICH ZÄHLT.

**IT-PROJEKTMANAGER/IN (M/W/D) IN ZIVIL
IN BONN**

Die Versorgung der Streitkräfte sichern. In Zivil unterstützen. Als IT-Projektmanager/in im gehobenen nichttechnischen Dienst im Bundesamt für Ausrüstung, Informationstechnik und Nutzung der Bundeswehr am Dienstort Bonn setzen Sie Ihre Fachexpertise bei der Entwicklung, Erprobung, Beschaffung sowie dem Nutzungsmanagement von Wehrmaterial ein.

IHRE AUFGABEN

- Sie bearbeiten Anforderungen und setzen Maßnahmen im Nutzungsmanagement für SAP Module für verschiedene Bereiche der Bundeswehr um.
- Sie wirken beim „2nd Level Support“ in allen IT-Management-Prozessen einschließlich der Qualitätssicherung und den Tests mit.
- Sie führen Analysen zur Aufnahme fachlicher Anforderungen der Auftraggeber und Bewertungen im Kontext der Projekte durch.

IHRE QUALIFIKATIONEN

- Sie besitzen eine Laufbahnbefähigung für den gehobenen nichttechnischen Verwaltungsdienst **oder**
- Sie verfügen über ein für die Laufbahn des gehobenen nicht-technischen Verwaltungsdienstes förderliches abgeschlossenes Bachelor- oder Diplom-(FH)-Studium in einer verwaltungswissenschaftlichen Fachrichtung **und**
- Sie erbringen den Nachweis einer dem gehobenen nicht-technischen Verwaltungsdienst gleichartigen und gleichwertigen hauptberuflichen Tätigkeit von mindestens 1,5 Jahren.

IHRE VORTEILE

- Sie arbeiten bei einem anerkannten Arbeitgeber und erhalten ein attraktives Gehalt (bis A12); eine Verbeamtung ist möglich.
- Sie profitieren von Fortbildungen und dem betrieblichen Gesundheitsmanagement.
- Sie vereinbaren Beruf und Familie durch flexible Arbeitszeiten, Teilzeit-Arbeit und ggf. die Nutzung von Telearbeit/Homeoffice.

STARTEN SIE JETZT IHRE KARRIERE

Mehr Infos erhalten Sie auf bewerbung.bundeswehr-karriere.de (Job-ID: 208D_0823_BAAINBW-E). Bewerben Sie sich dort bis zum **31. Dezember 2023** für eine Einstellung zum nächstmöglichen Zeitpunkt. Bitte beziehen Sie sich hierbei auf das c't Magazin.



HABEN SIE FRAGEN ZUR AUSGESCHRIEBENEN STELLE?
Herr Anhalt (02203 105 2569), Frau Endrulat (02203 105 2567)
E-Mail: ac-bewerbung-direkteinstieg@bundeswehr.org

Schwerbehinderte Menschen werden bei gleicher Qualifikation und Eignung bevorzugt eingestellt. Bewerbungen von Frauen sind ausdrücklich erwünscht. Die Bundeswehr begrüßt Bewerbungen von Menschen mit Migrationshintergrund.

ZIVILISTIN.DE
 Karriere geht auch ohne Uniform.
 Jetzt informieren und bewerben!

oder 0800 9800880
 (bundesweit kostenfrei)



.NET 8.0

Das Online-Event von Heise und www.IT-Visions.de
zum neuen .NET-LTS-Release

21. November 2023 - Online

Jetzt
Tickets
sichern!

- Die Neuerungen von .NET 8.0: SDK, Runtime und Basisklassen
- Einfacher lesbarer, stabilerer Code mit C# 12.0
- Alle Neuerungen von ASP.NET Core 8.0 und Blazor 8.0
- Neues beim OR-Mapping mit Entity Framework Core 8.0
- Das hat sich mit Windows Forms 8.0, WPF 8.0 und WinUI 3 verändert
- Cross-Plattform-Entwicklung mit .NET MAUI
- Ausblick auf .NET 9.0

Kooperationspartner
www.IT-Visions.de
Dr. Holger Schwichtenberg

Workshops zu C# 12.0, Entity Framework Core 8.0, Blazor 8.0 und .NET MAUI 8.0

net.bettercode.eu

PHP 2023

Die Heise-Konferenz zu PHP

27. November 2023 - Online

Mach deine PHP-Anwendung fit

- Best Practices
- Umgang mit Legacy Code
- Update zu PHP 8.3

Kooperationspartner



Workshops am 1. und 6. Dezember

Jetzt
Tickets
sichern!

php.bettercode.eu

Inserenten*

| | 180 | Veranstaltungen | |
|--|-----|---|-------------------|
| 1&1 Telecom GmbH, Montabaur | 35 | c't Workshops | c't, heise events |
| Bundesamt f. Personalmanagem. der Bundeswehr, Köln | 25 | Horizons #23 by heise | Heise Medien |
| DOAG Konferenz + Ausstellung GmbH, Berlin | 27 | Inside agile | iX, dpunkt.verlag |
| dpunkt.verlag GmbH, Heidelberg | 21 | Webinar-Serie | heise academy |
| EXTRA Computer GmbH, Giengen-Sachsenhausen | 173 | enterJS | iX, dpunkt.verlag |
| Fernschule Weber, Großkenner | 67 | Stark gegen Hacker | heise academy, iX |
| Nubert electronic GmbH, Schwäbisch Gmünd | 13 | betterCode | iX, dpunkt.verlag |
| RaidSonic Technology GmbH, Ahrensburg | 173 | Ein Teil dieser Ausgabe enthält Beilagen von EWE AG, Oldenburg und SIGS-DATACOM GmbH, Troisdorf. | |
| Silent Power Electronics GmbH, Willich | 173 | <i>* Die hier abgedruckten Seitenzahlen sind nicht verbindlich. Redaktionelle Gründe können Änderungen erforderlich machen.</i> | |
| Thomas Krenn AG, Freyung | 2 | | |
| WIBU-SYSTEMS AG, Karlsruhe | 69 | | |
| Wiesemann & Theis GmbH, Wuppertal | 15 | | |

Stellenanzeigen

| | |
|---|-----|
| Berufsakademie Sachsen Staatliche Studienakademie, Glauchau | 175 |
| Bundesamt f. Personalmanagem. der Bundeswehr, Köln | 175 |

** Die hier abgedruckten Seitenzahlen sind nicht verbindlich.
Redaktionelle Gründe können Änderungen erforderlich machen.*

WERDEN SIE c't-BOTSCHAFTER!

... UND UNTERSTÜZEN SIE DAMIT DEN UNABHÄNGIGEN UND GLAUBWÜRDIGEN JOURNALISMUS!

Wir schenken Ihnen **30 €** und unsere kultige **c't-Tasse „Kein Backup? Kein Mitleid“**, wenn Sie einen neuen Leser für ein Jahres-Abo der c't werben. Der neue Leser erhält die c't zum Preis von **144,20 €** pro Jahr. Das Abo kann in gedruckter oder digitaler Form bezogen werden. Nach einem Jahr ist das Abo monatlich kündbar.

Hier bestellen: ct.de/botschafter
+49 541/80 009 120 leserservice@heise.de

ct magazin für computer technik

Impressum

Redaktion

Heise Medien GmbH & Co. KG, Redaktion c't
Postfach 61 04 07, 30604 Hannover
Karl-Wiechert-Allee 10, 30625 Hannover
Telefon: 05 11/53 52-300
Telefax: 05 11/53 52-417
Internet: www.ct.de, E-Mail: ct@ct.de

Titelthemenkoordination in dieser Ausgabe:

Angela Meyer (anm@ct.de), Christian Wölbert (cwo@ct.de)

Chefredakteur: Torsten Beeck (tbe@ct.de) (verantwortlich für den Textteil)

Stellv. Chefredakteur: Axel Kossel (ad@ct.de)

Chef vom Dienst: Georg Schnurer (gs@ct.de)

Koordination Leserkommunikation: Martin Triadan (mat@ct.de)

Leiter redaktionelle Entwicklung: Jobst Kehrhahn (keh@ct.de)

Ressort Internet, Datenschutz & Anwendungen

Leitende Redakteure: Hartmut Gieselmann (hag@ct.de), Jo Bager (jo@ct.de)

Redaktion: Holger Bleich (hob@ct.de), Anke Brandt (abr@ct.de), Greta Friedrich (gref@ct.de),

Tim Gerber (tig@ct.de), Arne Grävemeyer (agr@ct.de), Markus Montz (mon@ct.de),

Peter Schmitz (psz@ct.de), Sylvester Tremmel (syt@ct.de), Andrea Trinkwalder (atr@ct.de),

Dorothee Wiegand (dwi@ct.de), Stefan Wischner (swi@ct.de)

Ressort Systeme & Sicherheit

Leitende Redakteure: Peter Siering (ps@ct.de), Jan Mahn (jam@ct.de)

Redaktion: Niklas Dierking (ndi@ct.de), Mirko Dölle (mid@ct.de), Wilhelm Drehling (wid@ct.de), Liane M. Dubowy (ldm@ct.de), Ronald Eikenberg (rei@ct.de), Oliver Lau (ola@ct.de), Pina Merkert (pmk@ct.de), Dennis Schirrmacher (des@ct.de), Hajo Schulz (hos@ct.de), Jan Schüßler (jss@ct.de), Kathrin Stoll (kst@ct.de), Keywan Tonekaboni (ktn@ct.de), Axel Vahldiek (axv@ct.de)

Ressort Hardware

Leitende Redakteure: Christof Windeck (cw@ct.de), Ulrike Kuhlmann (uk@ct.de), Dušan Živadinović (dz@ct.de)

Redaktion: Ernst Ahlers (ea@ct.de), Christian Hirsch (chh@ct.de), Benjamin Kraft (bkr@ct.de), Lutz Labs (ll@ct.de), Andrijan Möcker (amo@ct.de), Florian Müssig (mue@ct.de), Rudolf Opitz (rop@ct.de), Carsten Spille (csp@ct.de)

Ressort Mobiles, Entertainment & Gadgets

Leitende Redakteure: Jörg Wirtgen (jow@ct.de), Christian Wölbert (cwo@ct.de)

Redaktion: Robin Brand (rbr@ct.de), Sven Hansen (sha@ct.de), Steffen Herget (sht@ct.de), Nico Jurran (nij@ct.de), André Kramer (akr@ct.de), Michael Link (mil@ct.de), Urs Mansmann (uma@ct.de), Stefan Porteck (spo@ct.de)

Leiter c't 3003: Jan-Keno Janssen (jkj@ct.de)

c't Sonderhefte

Leitung: Jobst Kehrhahn (keh@ct.de)

Koordination: Pia Groß (piae@ct.de), Angela Meyer (anm@ct.de)

c't online: Sylvester Tremmel (syt@ct.de), Niklas Dierking (ndi@ct.de)

Social Media: Jil Martha Baae (jmb@ct.de)

Koordination News-Teil: Hartmut Gieselmann (hag@ct.de), Kathrin Stoll (kst@ct.de), Christian Wölbert (cwo@ct.de)

Koordination Heftproduktion: Martin Triadan (mat@ct.de)

Redaktionsassistenz: Susanne Cölle (suc@ct.de), Christopher Tränkmann (cht@ct.de)

Software-Entwicklung: Kai Wasserbäch (kaw@ct.de)

Technische Assistenz: Ralf Schneider (rlg@ct.de), Christoph Hoppe (cho@ct.de), Stefan Labusga (sla@ct.de), Arne Mertins (ame@ct.de), Jens Nohl (jno@ct.de), Daniel Ladeira Rodrigues (dro@ct.de)

Dokumentation: Thomas Masur (tm@ct.de)

Verlagsbüro München: Hans-Pinsel-Str. 10b, 85540 Haar, Tel.: 0 89/42 71 86-0, Fax: 0 89/42 71 86-10

Ständige Mitarbeiter: Detlef Borchers, Herbert Braun (heb@ct.de), Tobias Engler, Monika Ermert, Stefan Krempel, Ben Schwan (bsc@ct.de), Christiane Schulzki-Haddouti

DTP-Produktion: Mike Bunjes, Birgit Graff, Angela Hilberg, Jessica Nachtigall, Astrid Seifert, Ulrike Weis

Junior Art Director: Martina Bruns

Fotografie: Melissa Ramson, Andreas Wodrich

Digitale Produktion: Melanie Becker, Kevin Harte, Martin Kreft, Thomas Kaltschmidt, Pascal Wissner

Illustrationen

Rudolf A. Blaha, Frankfurt am Main, Thorsten Hübner, Berlin, Albert Hulm, Berlin, Sven Hauth, Schülp, Timo Lenzen, Berlin, Andreas Martini, Wettin, Moritz Reichartz, Viersen, Michael Vogt, Berlin

Editorial: Hans-Jürgen „Mash“ Marhenke, Hannover, Schlagseite: Ritsch & Renn, Wien, c't-Logo: Gerold Kalter, Rheine, Titelbild: KI Midjourney | Bearbeitung c't

c't-Krypto-Kampagne: Infos zur Krypto-Kampagne unter <https://ct.de/pgp>. Die Authentizität unserer Zertifizierungsschlüssel lässt sich mit den nachstehenden Fingerprints überprüfen:

Key-ID: 5C1C1DC5BEEDD33A
ct magazine CERTIFICATE <pgpCA@heise.de>
D337 FCC6 7EB9 09EA D1FC 8065 5C1C 1DC5 BEED D33A
Key-ID: 2BAE3CF6DAFFB000
ct magazine CERTIFICATE <pgpCA@ct.heise.de>
A3B5 24C2 01A0 D0F2 355E 5D1F 2BAE 3CF6 DAFF B000
Key-ID: DBD245FCB3B2A12C
ct magazine CERTIFICATE <pgpCA@ct.heise.de>
19ED 6E14 58EB A451 C5E8 0871 DBD2 45FC B3B2 A12C

heise Investigativ: Über diesen sicheren Briefkasten können Sie uns anonym informieren.

Anonymer Briefkasten: <https://heise.de/investigativ>
via Tor: ayznmonmewb2tjygf7ym4t2726muprjvwckzx2vhf2hbarbbzydm7oad.onion

Heise Medien GmbH & Co. KG
Postfach 61 04 07, 30604 Hannover
Karl-Wiechert-Allee 10, 30625 Hannover
Telefon: 05 11/53 52-0
Telefax: 05 11/53 52-129
Internet: www.heise.de

Verlag

Herausgeber: Christian Heise, Ansgar Heise, Christian Persson

Geschäftsführer: Ansgar Heise, Beate Gerold

Mitglieder der Geschäftsleitung: Jörg Mühlé, Falko Ossmann

Anzeigenleitung: Michael Hanke (-167) (verantwortlich für den Anzeigenteil), www.heise.de/mediadaten/ct

Anzeigenpreise: Es gilt die Anzeigenpreisliste Nr. 40 vom 1. Januar 2023.

Anzeigen-Auslandsvertretung (Asien): Media Gate Group Co., Ltd., 7F, No. 182, Section 4, Chengde Road, Shilin District, 11167 Taipei City, Taiwan, www.mediagate.com.tw
Tel: +886-2-2882-5577, Fax: +886-2-2882-6000, E-Mail: mei@mediagate.com.tw

Leiter Vertrieb und Marketing: André Lux (-299)

Werbeleitung: Julia Conrades (-156)

Service Sonderdrucke: Julia Conrades (-156)

Druck: Firmengruppe APPL, appl druck, Senefelderstr. 3-11, 86650 Wemding

Kundenkonto in der Schweiz: PostFinance, Bern, Kto.-Nr. 60-486910-4, BIC: POFICHBXXX, IBAN: CH73 0900 0000 6048 6910 4

Vertrieb Einzelverkauf:

DMV Der Medienvertrieb GmbH & Co. KG

Meßberg 1

20086 Hamburg

Tel.: 040/3019 1800, Fax: 040/3019 1815

E-Mail: info@dermedienvertrieb.de

c't erscheint 14-täglich

Einzelpreis 5,90 €; Österreich 6,50 €; Schweiz 9,90 CHF; Belgien, Luxemburg 6,90 €; Niederlande 7,20 €; Italien, Spanien 7,40 €, Dänemark 64,00 DKK

Abonnement-Preise: Das Jahresabonnement kostet inkl. Versandkosten: Inland 144,20 €, Österreich 155,40 €, Europa 165,20 €, restl. Ausland 191,80 € (Schweiz 236,60 CHF); ermäßiges Abonnement für Schüler, Studenten, Auszubildende (nur gegen Vorlage einer entsprechenden Bescheinigung): Inland 105,00 €, Österreich 99,40 €, Europa 124,60 €, restl. Ausland 152,60 € (Schweiz 145,60 CHF). c't-Plus-Abonnements (inkl. Zugriff auf das c't-Artikel-Archiv sowie die App für Android und iOS) kosten pro Jahr 25,00 € (Schweiz 30.80 CHF) Aufpreis. Ermäßiges Abonnement für Mitglieder von AUGE, bdvb e.V., BvDW e.V., ch/open, GI, GUUG, ISACA Germany Chapter e.V., JUG Switzerland, VBIO, VDE und VDI (gegen Mitgliedsausweis): Inland 108,15 €, Österreich 116,55 €, Europa 123,90 €, restl. Ausland 143,85 € (Schweiz 177,45 CHF). Luftpost auf Anfrage.

Leserservice:

Bestellungen, Adressänderungen, Lieferprobleme usw.

Heise Medien GmbH & Co. KG

Leserservice

Postfach 24 69

49014 Osnabrück

E-Mail: leserservice@ct.de

Telefon: 05 41/8 00 09-120

Fax: 05 41/8 00 09-122

c't abonnieren: Online-Bestellung via Internet (www.ct.de/abo) oder

E-Mail: leserservice@ct.de.

Eine Haftung für die Richtigkeit der Veröffentlichungen kann trotz sorgfältiger Prüfung durch die Redaktion vom Herausgeber nicht übernommen werden. Kein Teil dieser Publikation darf ohne ausdrückliche schriftliche Genehmigung des Verlags in irgendeiner Form reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Die Nutzung der Programme, Schaltpläne und gedruckten Schaltungen ist nur zum Zweck der Fortbildung und zum persönlichen Gebrauch des Lesers gestattet.

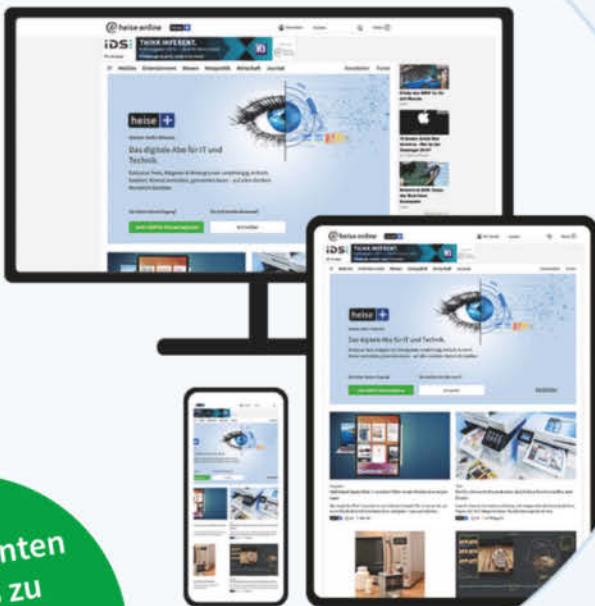
Für unverlangt eingesandte Manuskripte kann keine Haftung übernommen werden. Mit Übergabe der Manuskripte und Bilder an die Redaktion erteilt der Verfasser dem Verlag das Exklusivrecht zur Veröffentlichung. Honorierte Arbeiten gehen in das Verfügungsberecht des Verlages über. Sämtliche Veröffentlichungen in c't erfolgen ohne Berücksichtigung eines eventuellen Patentschutzes.

Warennamen werden ohne Gewährleistung einer freien Verwendung benutzt. Hergestellt und produziert mit Xpublisher: www.xpublisher.com. Printed in Germany. Alle Rechte vorbehalten. Gedruckt auf chlortfreiem Papier.

© Copyright 2023 by Heise Medien GmbH & Co. KG

ISSN 0724-8679 AWA LAE ▲

c't-Abo-Abonnenten
lesen bis zu
75%
günstiger



Das digitale Abo für IT und Technik.

Exklusives Angebot für c't-Abo-Abonnenten:
Sonderrabatt für Magazinabonnenten

- ✓ Zugriff auf alle Artikel von heise+
- ✓ Alle Heise-Magazine online lesen:
c't, iX, MIT Technology Review,
Mac & i, Make und c't Fotografie
- ✓ Jeden Freitag exklusiver Newsletter
der Chefredaktion
- ✓ 1. Monat gratis lesen – danach jederzeit kündbar

Sie möchten dieses Exklusiv-Angebot nutzen? Jetzt bestellen unter:

heise.de/plus-testen

✉ leserservice@heise.de ☎ 0541 80009 120

DAUERTIEFPREISE

1&1 Weihnachts-Aktion: Wunsch-Handy mit 10 GB monatlich dauerhaft günstig.

1&1 – Immer wieder besser.

Alle Handys
einmalig

0,- €*



1und1.de
0721/960 6060

Weihnachtsfreude für alle: Bei 1&1 gibt's alle Handys für einmalig 0,- € in Verbindung mit der 1&1 All-Net-Flat und mindestens 10 GB/Monat zum Dauertiefpreis.*

1&1

*Google Pixel 7a, Samsung Galaxy S23 oder Galaxy Z Flip5 für je 0,- € einmalig, z.B. in Verbindung mit der 1&1 All-Net-Flat M mit 250 GB Cloud-Speicher und 10 GB Highspeed-Volumen/Monat (danach max. 64 kB/s) für dauerhaft 19,99 €/Monat (Google Pixel 7a), 29,99 €/Monat (Samsung Galaxy S23) bzw. 39,99 €/Monat (Samsung Galaxy Z Flip5). Dauertiefpreis: Kein Preisprung während der gesamten Laufzeit. Gratis Telefonieren und Surfen gilt nicht für Sonder- und Premiumdienste, Verbindungen aus Deutschland ins Ausland sowie Roaming außerhalb EU gemäß Preisliste. Telefonate in dt. Fest- und Handynetze inklusive sowie Verbindungen innerhalb des EU-Auslands und aus EU nach Deutschland plus Island, Liechtenstein und Norwegen. Bereitstellungspreis je 39,90 €. Je 24 Monate Vertragslaufzeit. Änderungen und Irrtümer vorbehalten. Preise inkl. MwSt. 1&1 Telecom GmbH, 56403 Montabaur. WEEE-Reg.-Nr. DE13470330

