Infineon zum Konsultationsprozess zur nationalen Rechenzentrumsstrategie

Einleitung

Infineon Technologies AG begrüßt die Beteiligungsmöglichkeit im Rahmen der Konsultation zur nationalen Rechenzentrumsstrategie. In unserer Rückmeldung möchten wir unsere Perspektive und Erfahrungswerte aus der Halbleiterbranche. Mit Blick auf die gestellten Leitfragen fokussieren wir uns auf die Herausforderungen, Chancen und politischen Handlungsempfehlungen für die Bundesregierung. Dabei stehen für uns insbesondere die gestiegenen Anforderungen durch Künstliche Intelligenz (KI) an Rechenzentren und deren Energieeffizienz im Vordergrund. Hier sehen wir entscheidende Hebel für einen wettbewerbsfähigen und zukunftsorientierten Rechenzentrumsstandort Deutschland.

Herausforderungen

1. Powering AI ist die zentrale Herausforderung für den Rechenzentrumsstandort Deutschland

Laut dem Gutachten „Stand und Entwicklung des Rechenzentrumsstandorts Deutschland“ wird der Strombedarf von Rechenzentren in den kommen 20 Jahren stark ansteigen. Ausgehend von 20 TWh Strombedarf im Jahr 2024, prognostiziert das Gutachten einen Anstieg auf 31 TWh bis 2030 und 80 TWh bis 2045.[1] Damit würde der Anteil des Stromverbrauchs von Rechenzentren bis 2045 auf 6% des Gesamtstromverbrauchs in Deutschland anwachsen.

Einer der stärksten Treiber dieses Anstiegs ist Künstliche Intelligenz (KI). Alle 3,4 Monate verdoppelt sich die Rechenleistung, die benötigt wird, um hochmoderne KI-Modelle zu trainieren.[2] Dementsprechend steigt auch der Stromverbrauch der Prozessoren rapide an. Die folgende Tabelle zeigt, wie KI die Rechenleistung und den Energiebedarf erhöht.

	Computing power	Energy demand
Standard Google query	>1 TFLOPs	>1 Wh
GPT-3 query (Inference)	350 TFLOPs	1-10 Wh
GPT-3 training	$3x10^{11}$ TFLOPs	~1,300 MWh

2. KI ist schnell. Das Netz ist langsam.

Während es etwa 6 Monate dauert, ein neues KI-Modell zu entwickeln und etwa 1-2 Jahre, um ein neues Rechenzentrum zu bauen, kann der Aufbau der entsprechenden Energieinfrastruktur 5 bis 15 Jahre dauern.[3] Daher ist die Energieversorgung eines bestehenden Rechenzentrums mittelfristig begrenzt. Um diese Herausforderung anzugehen, muss nicht nur die Rechenleistung pro Watt erhöht werden, sondern auch Energieeffizienz und ein erhöhter Stromverbrauch bereits in der Entwicklungsphase neuer KI-Rechenzentren und Serversysteme berücksichtigt werden.

[1] https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Publikationen/Technologie/stand-und-entwicklung-des-rechenzentrumsstandorts-deutschland.pdf?__blob=publicationFile&v=10

[2] https://cset.georgetown.edu/wp-content/uploads/AI-and-Compute-How-Much-Longer-Can-Computing-Power-Drive-Artificial-Intelligence-Progress.pdf

[3] https://www.iea.org/reports/electricity-grids-and-secure-energy-transitions/executive-summary

Copyright © Infineon Technologies AG 2023. All rights reserved.

Durch die immer weiter ansteigende Rechenleistung der KI werden wir gegen Ende der Dekade Rechenzentren sehen, die 1 Gigawatt Strom benötigen. Dies entspricht ca. dem Stromverbrauch einer Stadt wie München in einem Jahr. Im Idealfall wird dieser zusätzliche Strombedarf durch erneuerbare Energiequellen gedeckt. Eine weitere Neuerung wird sein, dass die MV-Netzspannungen für KI-Rechenzentren 20 bis 30 kV statt wie heute üblich 10 kV betragen werden.

3. **KI stellt die Betreiber von Rechenzentren vor große technologische Herausforderungen**

Total cost of ownership (TCO): KI-Rechenzentren verbrauchen wesentlich mehr Strom als Nicht-KI-Rechenzentren. Außerdem sind die Kosten für KI-Server bis zu 30-mal höher als die von herkömmlichen Servern.[4] Hinzu kommt eine stärkere Belastung von Hardware und Equipment, insbesondere für die Stromversorgung. Somit treiben Innovationen im Bereich KI und insbesondere generativer KI die Gesamtbetriebskosten von KI-Rechenzentren signifikant in die Höhe. Die geschätzten Kosten für das Training von GPT-4 belaufen sich auf ~$100 Millionen.[5]

Robustheit: Ein effizienter Betrieb, genaue Ergebnisse und eine effektive Skalierung sind entscheidend für den Erfolg von KI. Dafür bedarf es einer robusten Infrastruktur. In einer ITIC-Umfrage aus dem Jahr 2024 gaben 41 % der befragten Unternehmen an, dass ihre stündlichen Ausfallzeiten zwischen 1 und 5 Millionen US-Dollar kosten.[6] Ein reibungsloser Betrieb wird somit immer wichtiger, um Ausfallkosten zu reduzieren. Gleichzeitig werden KI-Serversysteme immer komplexer, was ihre Überwachung zu einer immer größeren Herausforderung macht.

Verluste im Stromverteilungsnetz: Die durchschnittliche Stromflusseffizienz in KI-Rechenzentren liegt laut unseren Berechnungen derzeit bei lediglich 85 %. Bei jedem Leistungsumwandlungsschritt geht Energie verloren. Angesichts des steigenden Stromverbrauchs von KI-Rechenzentren ist es wichtig, diese Stromverluste zu verringern und die Energieeffizienz von KI-Rechenzentren in jedem Schritt der Umwandlungskette zu erhöhen.

Chancen

1. **Europäische Halbleiterunternehmen werden eine wichtige Rolle bei der Ausbildung, Entwicklung und Einführung von KI spielen.**

Europäische Halbleiterunternehmen sind bestens positioniert, um den skizzierten Herausforderungen entgegenzutreten. Sie werden eine wichtige Rolle bei der Ermöglichung der Entwicklung und Einführung von KI spielen. Die Verbesserung der Energieeffizienz von Rechenzentren erfordert spezifische Hardwarekomponenten von europäischen Halbleiterherstellern, darunter Leistungshalbleiter, Power Management ICs und Application Specific ICs (ASICs). Diese Lösungen kommen in allen Phasen der Energieumwandlung zum Einsatz. In einem typischen KI-Server-Rack sind sie Bestandteile z. B. von Top-of-Rack-Switches, Netzteilen, Batterie-Backup-Einheiten, 48-V-BUS-Konvertern, KI-Karten, intelligenten Netzwerkkarten und Server-Motherboards. Europäische Halbleiterunternehmen sind Weltmarktführer bei Halbleiterlösungen für Rechenzentren und sind Technologieführer bei der Integration der drei Halbleitermaterialien Silizium (Si), Siliziumkarbid (SiC) und Galliumnitrid (GaN). Um die Energieeffizienz, Leistungsdichte, thermische Leistungsfähigkeit und Robustheit von Rechenzentren zu verbessern und KI in Europa auf das nächste Level zu bringen, führt somit kein Weg an europäischen Halbleiterherstellern vorbei.

---

[4] https://omdia.tech.informa.com/om121647/server-market-analysis--1h24#:~:text=Covers%20the%20data%20center%20and,%3B%20and%20semiconductor%2Dlevel%20transitions

[5] https://www.forbes.com/sites/katharinabuchholz/2024/08/23/the-extreme-cost-of-training-ai-models/

[6] https://itic-corp.com/itic-2024-hourly-cost-of-downtime-report/

2. Edge AI bietet große Potentiale für das industrielle Ökosystem in Deutschland

Während große KI-Modelle wie Large Language Models (LLMs) auch in Zukunft auf Rechenzentren angewiesen sein werden, ermöglichen innovative EdgeAI-Designs für spezifische Anwendungen die Energiebelastung durch Rechenzentren zu reduzieren, indem ein Teil des Betriebs von KI-Modellen von der Cloud auf Geräte, Sensoren und Aktoren verlagert wird. Zudem ermöglicht dieser Ansatz eine Echtzeitverarbeitung, reduzierte Latenzzeiten und verbesserte Sicherheit und eignet sich daher besonders für Anwendungen in der industriellen Automatisierung, der Automobilindustrie und dem Gesundheitswesen. In diesen Anwendungsbereichen sind europäische Halbleiterunternehmen führend und vereinen notwendige Hardware-Sicherheitsmodule, vertrauenswürdige Ausführungsumgebungen sowie funktionale Produktsicherheit und -zuverlässigkeit.

Handlungsempfehlungen

Energieeffizienz in der Versorgung und im Betrieb von Rechenzentren, insbesondere KI-Rechenzentren, kann Deutschland als erstklassigen Standort für KI und Rechenzentren positionieren. Die Rechenzentrumsstrategie kann hierfür einen ehrgeizigen Rahmen schaffen. Dabei sollten insbesondere die technologischen Kapazitäten der europäischen Halbleiterindustrie genutzt werden. Vor dem Hintergrund möchten wir die folgenden Maßnahmen für die Rechenzentrumsstrategie der Bundesregierung vorschlagen:

1. Leitmarkt für Nachhaltigkeit schaffen

Der Rechenzentrumssektor hat das Potenzial, ein globaler Leitmarkt für innovative und nachhaltige Produkte zu werden. Deutschland sollte sich als führender Standort für energie- und ressourceneffiziente KI-Rechenzentren mit hoher Robustheit und Skalierbarkeit etablieren. Dafür sind Anreize für Investitionen und Wettbewerb auf der Grundlage von Nachhaltigkeits- und Energieeffizienzkriterien notwendig, z. B. bei der öffentlichen Beschaffung, bei staatlichen Beihilfen oder durch Priorisierung bei Genehmigungsverfahren. Öffentliche Vergaben für Rechenzentrums-, Cloud- und KI-Dienste sollen verbindliche Mindestanforderungen an Effizienz, Energieherkunft und Abwärmenutzung enthalten.

2. Mindeststandards für Energieeffizienz

Mindeststandards für die Energieeffizienz von KI-Rechenzentren sind ein effektiver Hebel, um die Herausforderung des Betriebs von KI-Rechenzentren sowie der Energieversorgung von KI anzugehen. Vor diesem Hintergrund ist das Ambitionsniveau des Energieeffizienzgesetzes zu begrüßen. Dieses sollte künftig auch europaweit einheitlich festgeschrieben werden. Daher empfehlen wir der Bundesregierung sich auf EU-Ebene für die Einführung europaweit einheitlicher Mindeststandards für die Energieeffizienz von Rechenzentren einzusetzen.

3. Ganzheitliche Netzplanung

Planungs- und Anschlussverfahren für Umspannwerke, Leitungen und Trafostationen müssen beschleunigt werden, mit standardisierten, modularen Anschlusslösungen und transparentem Queue-Management. Standorte sollten gemeinsam mit Übertragungsnetzbetreibern, Kommunen und Wärmeversorgern geplant werden. Dabei sollten zur Reduktion von Verlusten nahe Erzeugungsschwerpunkte und Netzknoten bevorzugt werden sowie höhere Spannungsebenen bis nahe ans Gebäude, ermöglicht durch moderne Leistungshalbleiter und 48-V-Rack-Verteilung, minimiert werden.

4. EU Chips Act

Die Bundesregierung sollte die Prioritäten der Rechenzentrumsstrategie in den Prozess zur Überarbeitung des EU Chips Act einbringen. Die wachsende Nachfrage nach Leistungshalbleitern, Power Management ICs und ASICs im KI-Rechenzentrumsmarkt sollten stärker berücksichtigt werden, insbesondere im Pillar I für die Förderung von Forschung und Entwicklung.

Unternehmensprofil: Infineon Technologies AG

Halbleiter sind essenziell, um die energiebezogenen Herausforderungen unserer Zeit zu meistern und die digitale Transformation mitzugestalten. Daher setzen wir bei Infineon alles daran, die Dekarbonisierung und Digitalisierung aktiv voranzutreiben. Als ein weltweit führender Anbieter von Halbleiterlösungen für Power-Systems und IoT ermöglichen wir wegweisende Lösungen für grüne und effiziente Energie, saubere und sichere Mobilität sowie ein intelligentes und sicheres IoT. Wir machen das Leben einfacher, sicherer und umweltfreundlicher. Gemeinsam mit unseren Kunden und Partnern. Für eine bessere Zukunft.

Infineon entwickelt, fertigt und vertreibt eine Vielzahl an Halbleitern und halbleiterbasierten Lösungen. Dabei liegt der Fokus auf den wesentlichen Märkten: von Automobil- über Industrie- bis hin zu konsumentennahen Sektoren. Das Produktangebot reicht von Standardkomponenten über spezielle Komponenten für digitale, analoge sowie Mixed-Signal-Anwendungen bis hin zu Rechenzentren für künstliche Intelligenz und einigen spezifischen konsumentennahen Sektoren. Etwa 55 Prozent des Umsatzes von Infineon stammen aus dem Bereich Leistungshalbleiter, rund 33 Prozent aus Embedded-Control- und Konnektivitätslösungen (Mikrocontroller für Automotive- und Industrieanwendungen, Wi-Fi- und Bluetooth®-/Bluetooth®-Low-Energy-Technologien, Combos sowie Sicherheitslösungen). Zudem kommen circa neun Prozent des Umsatzes aus dem Bereich Hochfrequenzkomponenten und Sensoren, während rund vier Prozent auf Speicherchips für spezifische Anwendungen entfallen. Infineon erzielte im Geschäftsjahr 2024 26 Prozent seines Umsatzes in Europa, 61 Prozent in Asien und 13 Prozent in Amerika.

Gründung:	1999
Mitarbeiter:	Rund 58.060 weltweit (Stand 30. September 2024)
Umsatz:	€14.955 Mio. im Geschäftsjahr 2024
Präsenz:	Weltweit 71 Forschungs- und Entwicklungsstandorte sowie 15 Fertigungsstandorte

Weitere Informationen sind auf unserer Website zu finden: https://www.infineon.com/cms/en/about-infineon/company/