

KI-Infrastruktur: Wo die nächsten Engpässe jetzt entstehen

Von Mike Glöckner, Analyst für den Technologiesektor bei der DJE Kapital AG

Pullach im Isartal, 05.10.2026 – An der Börse wird der KI-Boom oft auf Chips und Netzwerke reduziert. Doch selbst der leistungsfähigste Grafikprozessor (GPU) kann sein Potenzial nicht ausschöpfen, wenn die Daten nicht schnell genug bereitstehen. Genau hier entstehen die nächsten Engpässe: bei Speicher, Chip-Substraten und optischen Verbindungen. Für Anbieter dieser Komponenten ergeben sich daraus Wachstums- und Margenpotenziale. Gleichzeitig ziehen hohe Preise zusätzliche Investitionen an. Neue Kapazitäten können das Marktgleichgewicht jedoch wieder verändern. Entscheidend ist, welche Unternehmen die Knappheit in freien Cashflow und hohe Kapitalrenditen übersetzen können

Speicher: HBM wird zum Leistungsfaktor

Im Mittelpunkt steht High Bandwidth Memory (HBM). Dabei werden mehrere DRAM-Chips gestapelt und über breite Datenwege direkt mit dem KI-Prozessor verbunden. Dadurch erreicht HBM eine deutlich höhere Bandbreite als herkömmlicher DRAM-Arbeitsspeicher. Dies ist vor allem bei großen KI-Modellen entscheidend. Während der Verarbeitung einer Anfrage speichert das Modell bereits verarbeitete Wörter und Zusammenhänge im sogenannten KV-Cache. Dieser Zwischenspeicher verhindert, dass der bisherige Kontext bei jedem neuen Wort komplett neu berechnet werden muss. Je länger die Eingabe ist, je mehr Nutzer gleichzeitig arbeiten und je häufiger autonome KI-Agenten mehrere Schritte ausführen, desto größer wird der Speicherbedarf.

Die Deutsche Bank¹ erwartet für HBM bis 2030 ein jährliches Nachfragewachstum von rund 40 Prozent gegenüber 21 Prozent bei Standard-DRAM. Morgan Stanley² rechnet damit, dass Server im Jahr 2028 bereits 59 Prozent des DRAM-Bedarfs ausmachen werden. 2023 waren es 37 Prozent. Bei NAND soll der Anteil von Enterprise-SSDs von 18 auf 65 Prozent steigen.

Hohe Nachfrage trifft auf träges Angebot

Das Angebot kann mit dieser Entwicklung nur verzögert Schritt halten. Der Bau einer neuen Speicherfabrik dauert meist zwei bis drei Jahre. Zudem beansprucht ein Bit HBM ungefähr dreimal so viel Waferfläche, also Platz, wie ein Bit herkömmlicher DRAM. Bei HBM4 und HBM4E, die ab dem zweiten Halbjahr 2026 verfügbar sein sollen, könnte es sogar das Vierfache

¹ Deutsche Bank Research, Juni 2026

² Morgan Stanley, Juni 2026

sein. Dieser zusätzliche Flächenbedarf geht zulasten anderer Anwendungen, etwa in Computern, Smartphones oder Autos. Für das zweite Quartal 2026 wurden bei Standard-DRAM Preissteigerungen von 58 bis 63 Prozent gegenüber dem Vorquartal erwartet.

Bei NAND-Speichern lagen die Erwartungen sogar bei Preissteigerungen von 70 bis 75 Prozent gegenüber dem Vorquartal. Das spricht einerseits für eine hohe Preissetzungsmacht, zeigt andererseits aber auch die Fallhöhe, sobald das Angebot schneller wächst als die Nachfrage. Der Markt ist stark konzentriert. SK Hynix, Samsung und Micron kontrollieren zusammen mehr als 90 Prozent des DRAM-Marktes.

Milliardeninvestitionen erhöhen das Angebotsrisiko

Das 200-Milliarden-US-Dollar-Programm von Micron in den USA ist auf 20 Jahre angelegt. Im Durchschnitt entspricht dies rund zehn Milliarden US-Dollar pro Jahr. 150 Milliarden US-Dollar entfallen auf die Fertigung und 50 Milliarden US-Dollar auf die Forschung. Geplant sind Werke in Idaho und New York, die Modernisierung des Standorts in Virginia sowie Kapazitäten für das HBM-Packaging. Samsung gab 2025 rund 40 Milliarden US-Dollar für Anlagen aus und plant für 2026 mehr als 73 Milliarden US-Dollar für Anlagen und Forschung, auch in anderen Geschäftsbereichen. SK Hynix beschloss Investitionen von rund 38 Milliarden US-Dollar für zwei Speicherfabriken. Zudem kündigten Samsung, SK Hynix und Zulieferer ein südkoreanisches KI- und Halbleiterprogramm von rund 518 Milliarden US-Dollar an. Geplant sind vier Fabriken und ein HBM-Packaging-Cluster. Der zeitliche Rahmen ist bislang offen.

Kurzfristig bleibt Speicher knapp. Mittelfristig steigt jedoch das Risiko eines neuen Angebotszyklus. Auch im NAND-Markt verschärft sich der Wettbewerb. Ein japanischer NAND-Spezialist profitiert zwar von der steigenden Nachfrage nach Enterprise-SSDs, ist vom HBM-Boom jedoch kaum betroffen. Gleichzeitig gewinnen chinesische Anbieter Marktanteile. So erreichte ein chinesischer Wettbewerber im zweiten Quartal 2026 rund 14 Prozent der weltweiten NAND-Auslieferungen nach Bit und überholte damit beim Volumen einen etablierten Anbieter. Bei hochwertigen Enterprise-Produkten besteht dagegen weiterhin ein Rückstand. Auch im DRAM-Markt nimmt der Wettbewerb durch chinesische Anbieter zu. Kurzfristig bietet die KI-Nachfrage also Rückenwind für die NAND. Langfristig könnte der wachsende Wettbewerb jedoch den Preisdruck erhöhen.

Effizienzgewinne können den Speicherbedarf verändern

Gleichzeitig könnte technischer Fortschritt den Speicherbedarf je KI-Anfrage senken. Googles Kompressionsverfahren TurboQuant kann den KV-Cache laut veröffentlichten Tests ohne messbaren Genauigkeitsverlust verkleinern. Dadurch reduziert sich der rechnerische Speicherbedarf je Anfrage, was Speicheraktien 2026 belastete. Dem steht der sogenannte Jevons-Effekt gegenüber. Sinken die Kosten der KI-Nutzung, kann die Nutzung selbst zunehmen. Ein geringerer Speicherbedarf je Anfrage muss daher nicht zwangsläufig zu einer geringeren Gesamtnachfrage führen.

Substrate: Die unsichtbare Brücke

Bei leistungsstarken KI-Chips bildet das IC-Substrat die Verbindung zwischen Silizium und Leiterplatte. Damit verbindet es das Prozessor-Speicher-Paket mit dem System. Ohne Substrat funktioniert kein modernes Chipgehäuse. Lösungen mit HBM und mehreren Chiplets zählen zum Advanced Packaging. Hochleistungsprozessoren nutzen vor allem ABF-Substrate. Das Isoliermaterial ermöglicht besonders feine Leiterbahnen. Größere Chip-Pakete und mehr HBM erhöhen jedoch den Flächenbedarf und den Fertigungsaufwand.

Goldman Sachs³ schätzt, dass NVIDIAs kommende Rubin-GPU 123 Prozent mehr ABF-Fläche benötigt als Blackwell. Bei der neuen Vera-CPU gegenüber Grace sind es 47 Prozent. Neue Serverprozessoren könnten mehr als 30 Prozent zusätzliche Substratfläche verlangen. Der ABF-Markt könnte Goldman Sachs zufolge von 7,7 Milliarden US-Dollar im Jahr 2025 auf 33 Milliarden US-Dollar im Jahr 2028 wachsen. Gleichzeitig modelliert Goldman Sachs eine Unterversorgung von 51 Prozent. Anfang 2026 stiegen die Lieferzeiten von drei bis vier auf mehr als zwölf Monate. Neue Kapazitäten dürften erst ab 2028 oder 2029 spürbar entlasten. Lange Bau- und Qualifikationszeiten verschaffen dabei den etablierten Anbietern einen Vorteil.

Langfristige Verträge stabilisieren das Geschäft

Zwei Anbieter führen den Markt mit Kapazitätsanteilen von rund 21 beziehungsweise 20 Prozent an. Einer davon ist bei KI-Chips und Serverprozessoren breit aufgestellt. Mehr als 70 Prozent seiner ABF-Lieferungen sind langfristig gebunden. Das stabilisiert die Auslastung und die Erlöse, begrenzt zugleich aber die Wirkung höherer Spotpreise auf die Erlöse. Ein weiterer Anbieter ist stärker den Marktpreisen ausgesetzt. Das erhöht das Margenpotenzial, birgt aber auch ein höheres Rückschlagrisiko.

Die Substratproduktion hängt nicht allein von Fertigungslinien ab. Mit steigender Leistungsdichte wird T-Glass wichtiger. Dabei handelt es sich um ein hitzebeständiges Glasgewebe, das feine Substratstrukturen stabilisiert. Engpässe bei diesem Material könnten den Ausbau der ABF-Kapazitäten bremsen. Ein europäischer Anbieter hat für ein KI-Projekt Investitionen von 1,5 bis 2 Milliarden Euro in Aussicht gestellt. Hohe Investitionen, kritische Materialien und Verzögerungen bergen jedoch weiterhin Risiken.

Optik: Kupfer stößt an physikalische Grenzen

Moderne KI-Modelle verteilen ihre Arbeit auf Tausende Beschleuniger. Wenn das Netzwerk stockt, können auch die Grafikprozessoren (GPUs) ihre Leistung nicht voll ausschöpfen. Bei 800 Gigabit, 1,6 Terabit und später 3,2 Terabit pro Sekunde stößt Kupfer zunehmend an seine physikalischen Grenzen. Elektrische Signale verlieren über längere Strecken an Qualität, erzeugen Wärme und verbrauchen mehr Energie. Glasfasern hingegen übertragen Daten als Licht mit hoher Bandbreite über größere Distanzen und verbrauchen je Bit weniger Energie. Für kurze

³ Goldman Sachs, September 2026

Wege bleibt Kupfer günstiger. Bislang galt daher das Motto: „Kupfer, wenn möglich. Optik, wenn nötig.“ Mit steigender Leistungsdichte verschiebt sich dieses Verhältnis jedoch zunehmend zugunsten optischer Verbindungen.

So kann ein KI-Rack 16- bis 36-mal mehr Glasfaser benötigen als ein traditionelles Rack. Bei Co-Packaged Optics (CPO) sitzen die optischen Komponenten direkt neben dem Netzwerkchip. Dadurch werden die elektrischen Wege verkürzt, was bei 1,6 Terabit pro Sekunde den Energieverbrauch von etwa 30 auf neun Watt senken kann. Weltweit erzielen CPO-Lösungen heute weniger als eine Milliarde US-Dollar Umsatz. Eine Studie erwartet jedoch bis 2030 einen Umsatz von rund 15 Milliarden US-Dollar. Dabei bleiben Kundenfreigaben, Qualität und Wirtschaftlichkeit entscheidend.

Hohe Nachfrage trifft auf Konzentrationsrisiken

Ein führender Anbieter optischer Komponenten schätzt seinen Marktanteil bei sogenannten EML-Lasern auf 50 bis 60 Prozent. Diese wandeln elektrische Datensignale in Lichtsignale um, die über Glasfasern übertragen werden. Im Juni-Quartal 2026 stieg der Umsatz dieses Anbieters um 109 Prozent auf 1,006 Milliarden US-Dollar. Die Bruttomarge erreichte 50,4 Prozent und die operative Marge 36,6 Prozent. Laut Management lag die Nachfrage nach EML-Lasern mehr als 30 Prozent über dem Angebot. Ein weiterer Anbieter, der ein breiteres Spektrum von Lasern bis hin zu optischen Schaltern abdeckt, erwartet breitere CPO-Auslieferungen erst ab 2028.

Auch chinesische Anbieter profitieren von der steigenden Nachfrage nach optischen Modulen. Ein führender Hersteller erzielte 2025 einen Umsatz von 5,3 Milliarden US-Dollar. Im ersten Quartal 2026 stieg der Umsatz um 192 Prozent auf 2,7 Milliarden US-Dollar. Der Gewinn nahm um 274 Prozent auf 0,9 Milliarden US-Dollar zu. Dies unterstreicht die hohe Nachfrage nach 800G- und 1,6T-Verbindungen. Gleichzeitig wird die Abhängigkeit von einzelnen Märkten und Großkunden sichtbar: 61,7 Prozent des Umsatzes entfielen auf die USA und Google zählt zu den größten Kunden.

Hinzu kommen Bewertungs- und geopolitische Risiken. So prüfte die US-Regierung zuletzt Einfuhrbeschränkungen für optische Komponenten aus China. Zudem erweitern die Anbieter ihre Kapazitäten. Entscheidend ist daher, ob die Nachfrage mit dem steigenden Angebot Schritt hält und sich die hohen Margen behaupten lassen.

Was Anleger im Blick behalten sollten

Bei Speicherchips sind HBM-Verträge, Waferstarts, Ausbeuten und die Verteilung der Kapazitäten entscheidend. Waferstarts geben früh Hinweise auf das spätere Angebot. Bei Substraten sind Lieferzeiten, Preise und Kundenverträge wichtig. Bei optischen Verbindungen sind der Übergang von 800G zu 1,6T, CPO-Kundenfreigaben und die Kundenkonzentration entscheidend. Warnsignale sind doppelte Bestellungen, fallende Lieferzeiten, verschobene

Rechenzentren oder neue Fabriken ohne entsprechende Abnahmeverträge. Entscheidend bleibt, ob die Unternehmen die Knappheit dauerhaft in freien Cashflow und hohe Kapitalrenditen übersetzen können.

Fazit: Knappheit trifft auf steigende Investitionen

Speicher, Substrate und optische Komponenten sind zentrale Bausteine der KI-Infrastruktur und zugleich potenzielle Engpassfaktoren. Speicher halten Daten bereit, Substrate verbinden Prozessor und weitere Komponenten und Optik transportiert die Daten innerhalb der Rechenzentren. Die Anforderungen steigen derzeit schneller als die verfügbaren Kapazitäten. Das stützt bei führenden Anbietern Wachstum, Auslastung und Preissetzungsmacht.

Im Speicherbereich nimmt HBM eine Schlüsselrolle ein. Umfangreiche Investitionsprogramme der Branche zeigen, dass das Angebot mittelfristig deutlich wachsen dürfte. Das könnte die Kräfteverhältnisse und die Ertragslage beeinflussen.

Bei Substraten profitieren etablierte Anbieter von langen Bau- und Qualifikationszeiten sowie der zunehmenden Komplexität moderner Chip-Pakete. Langfristige Lieferverträge können die Auslastung und die Erlöse stabilisieren, während eine stärkere Abhängigkeit von Spotpreisen zwar größere Margenchancen, aber auch höhere Schwankungen mit sich bringt. Kritische Materialien und der hohe Kapitalbedarf begrenzen zugleich die Geschwindigkeit des Kapazitätsausbaus.

Bei optischen Verbindungen treiben höhere Datenraten und der steigende Vernetzungsbedarf in Rechenzentren die Nachfrage. Gleichzeitig bleiben Kundenkonzentration, geopolitische Risiken und der Ausbau neuer Kapazitäten wichtige Unsicherheitsfaktoren. Mit zunehmendem Angebot wird es deshalb entscheidend sein, ob die Nachfrage hoch genug bleibt, um Auslastung und Margen zu stützen.

Damit ist die Knappheit in allen drei Bereichen ein wichtiger Treiber, jedoch kein dauerhafter Zustand. Je stärker die Anbieter investieren, desto größer wird mittelfristig das Risiko eines neuen Angebotszyklus. Entscheidend ist daher nicht allein das Wachstum der KI-Infrastruktur, sondern auch, welche Unternehmen ihre Marktposition in nachhaltigen freien Cashflow und hohe Kapitalrenditen übersetzen können.

Über DJE

Die DJE Kapital AG gehört zur DJE-Gruppe und ist seit 1974 als unabhängige Vermögensverwaltung am Kapitalmarkt aktiv. Das Unternehmen aus Pullach bei München verwaltet mit rund 200 Mitarbeitern (davon rund 25 Fondsmanager und Analysten) aktuell über 19,2 Milliarden Euro (Stand: 31.08.2026) in den Bereichen individuelle Vermögensverwaltung, institutionelles Asset Management sowie Publikumsfonds. Zudem bietet DJE seit 2017



Marketing-Anzeige

eine einzeltitelbasierte Online-Vermögensverwaltung an – als digitale Lösung im Rahmen aktiv gemanagter Depots. DJE hat den Anspruch, ihren Kunden weitsichtige Kapitalmarktexpertise in allen Marktphasen zu bieten.

www.dje.de

Pressekontakt

DJE Kapital AG

Konstantin Krenz von Essen

Konstantin.KrenzvonEssen@dje.de

+49 89 790453 – 674

PB3C GmbH

Paul Oberhofer

+49 151 72 51 42 46

oberhofer@pb3c.com

Rechtliche Hinweise

Marketing-Anzeige: Alle hier veröffentlichten Angaben dienen ausschließlich Ihrer Information und stellen keine Anlageberatung oder sonstige Empfehlung dar. Die in diesem Dokument enthaltenen Aussagen geben die aktuelle Einschätzung der DJE Kapital AG wieder. Diese können sich jederzeit, ohne vorherige Ankündigung, ändern. Alle getroffenen Angaben sind mit Sorgfalt entsprechend dem Kenntnisstand zum Zeitpunkt der Erstellung gemacht worden. Für die Richtigkeit und Vollständigkeit kann jedoch keine Gewähr und keine Haftung übernommen werden.