

The memory tax

Amazon raised its 2026 capital budget by \$20 billion on 30 July and named the reason: memory cost more than it expected. That single line moves AI's binding constraint from accelerators to DRAM — a market with three suppliers, multi-year lead times on new wafer capacity, and no ability to be optimised away in software. J.P. Morgan Research now estimates the memory shock is adding 0.2 to 0.4 percentage points to US inflation. The AI buildout has stopped being a story about chips and started being a story about prices everyone pays.

AUTHOR

AT

Alicante Tech Vibes **Research Labs**

~8 MIN READ

•

ALL FIGURES SOURCED & DATED

+\$20B AMAZON'S 2026 CAPEX RAISE, ATTRIBUTED TO MEMORY COSTS · 30 JUL 2026	>400% ESTIMATED DRAM PRICE RISE, START-2024 TO END-2026 · J.P. MORGAN, 6 AUG 2026	0.2–0.4pp ESTIMATED LIFT TO US CORE INFLATION FROM THE MEMORY SHOCK · 6 AUG 2026
84.6% MICRON GROSS MARGIN, FQ3 2026, VS 37.7% A YEAR EARLIER · 24 JUN 2026	~35% MEMORY + STORAGE SHARE OF A PC BILL OF MATERIALS, UP FROM 15–18% · HP CFO	~\$730B SUM OF FOUR HYPERSCALERS' 2026 CAPEX GUIDANCE MIDPOINTS · JUL 2026 CALLS

I. The twenty billion nobody budgeted for

On 30 July 2026, Amazon told investors it now expects to spend approximately \$220 billion in cash capital expenditure this year, up from a prior estimate of

about \$200 billion. The company was unusually specific about why: the higher cost of memory pushed the number up. Not more datacenters, not more GPUs, not a change of strategy. The same buildout, repriced.

The rest of the quarter was a boom. AWS revenue reached \$42.2 billion, up 37% year on year — the fastest growth in eighteen quarters — with operating income of \$16.6 billion and a reported backlog of \$496 billion in customer commitments, according to Fortune's account of the call. Andy Jassy nonetheless said the company will not have enough capacity to meet all of 2026's demand, and expects the constraint to run into 2027.

Alphabet raised its own 2026 guidance to \$195–205 billion from \$180–190 billion, with CFO Anat Ashkenazi noting that demand still outpaces three years of aggressive capacity additions. Microsoft guided to roughly \$175 billion for calendar 2026. Meta narrowed its full-year range upward to \$130–145 billion on 29 July, having spent \$31.1 billion in the quarter alone. **Add the midpoints and four companies are committing roughly \$730 billion of capital in a single calendar year — and at least one of them says the number went up because a commodity input got expensive.**

FIG. 1: 2026 CAPITAL EXPENDITURE GUIDANCE, AS OF THE JULY 2026 EARNINGS ROUND



BARS SCALED TO AMAZON = 100%. ALPHABET AND META SHOWN AT GUIDANCE MIDPOINTS (\$195–205B; \$130–145B). MICROSOFT FIGURE REFLECTS AN ACCOUNTING RECLASSIFICATION OF DATACENTER LEASES. SOURCES: THE REGISTER, 4 AUG 2026; META Q2 2026 RESULTS, 29 JUL 2026.

II. In February the bottleneck was power

Six months ago the constraint story was electricity. A Futurum analysis published 12 February 2026 put five hyperscalers at \$660–690 billion of planned 2026 capex and identified power availability as the critical limiting factor, citing an \$80 billion backlog of Azure orders that could not be fulfilled for want of electricity. Memory did not appear in that analysis as a binding issue.

Power is still a constraint on *when* capacity comes online. Memory is now a constraint on *what it costs* — and cost constraints propagate faster than

permitting queues, because they hit the income statement in the quarter they occur. The shift matters because the two bottlenecks have opposite relief mechanisms. Power is a local, permitting-and-grid problem that a determined operator can route around by siting differently. Memory is a global oligopoly with three producers — Samsung, SK hynix and Micron — and a physical constraint that cannot be sited around.

The mechanism is capacity reallocation, not a demand spike alone. IDC's analysis, first published 18 December 2025 and updated 10 February 2026, expects 2026 supply growth of roughly 16% for DRAM and 17% for NAND — below historical averages — because high-bandwidth memory for AI accelerators consumes disproportionate wafer area and the three producers have redirected capacity toward it. Every HBM stack sold to a hyperscaler removes conventional DRAM from the market for everyone else.

We will still not have enough capacity to meet all the demand we have in 2026.

ANDY JASSY, AMAZON CEO, Q2 2026 EARNINGS CALL, 30 JULY 2026 — SAID WHILE ANNOUNCING A \$220 BILLION CAPITAL BUDGET, THE LARGEST IN THE COMPANY'S HISTORY. SOURCE: FORTUNE, 30 JUL 2026.

III. The honest read: this is a cycle that has not yet proven it is structural

The price data is not ambiguous. TrendForce revised its Q1 2026 DRAM contract price forecast from 55–60% to 90–95% quarter on quarter in early February, calling the LPDDR moves the steepest in their history. Second quarter spot data compiled by SigmaIntel showed LPDDR5X modules up a further 89% quarter on quarter and DDR4 sticks up 51%. In early July, TrendForce reported Samsung seeking up to a 20% increase again for Q3. Micron's fiscal Q3, reported 24 June 2026, showed revenue of \$41.5 billion against \$9.3 billion a year earlier and an 84.6% gross margin, up from 37.7%. Those are not the margins of a competitive commodity market.

Here is the case against reading this as permanent. Memory is the most reliably cyclical market in semiconductors, and 84.6% gross margins are precisely the signal that historically calls forth the capacity that ends the cycle. SK hynix chairman Chey Tae-won announced on 2 June 2026 a plan to double wafer

capacity over five years, alongside \$15 billion of advanced-memory investment in 2026 — while separately warning, at Nvidia's GTC in March, that the squeeze could run four to five more years. Both statements come from the same person; the second is the sales pitch for the first. Treat the 2030 timeline as an interested forecast, not a finding.

There is a second skeptical thread that cuts deeper. An analysis of hyperscaler accounting notes that four companies spent \$433.9 billion on property and equipment in the four quarters to March 2026 while recognising roughly \$149 billion of annual depreciation, that Amazon shortened server useful life from six years to five in January 2025 while Meta extended to 5.5, and that Amazon's trailing four-quarter capex reached 102% of operating cash flow — the first time above 100% since 2022. On that reading the memory shock is a symptom, and the real constraint arriving in 2027–2029 is the depreciation of everything bought in 2025–2026. **Both things can be true: memory raises the price of the buildout, and the buildout's own accounting will bite regardless.**

IV. The tax reaches the consumer price index

This is the part that makes a semiconductor supply story a macroeconomic one. J.P. Morgan Research, in analysis published 6 August 2026, estimates DRAM prices will have risen more than 400% between the start of 2024 and the end of 2026, and traces the pass-through: US CPI for software and accessories up 23% since end-2024, PPI for storage devices up 23%, the import price index for computers, peripherals and parts up 37%. Their rule of thumb is that each 10% rise in hardware costs adds roughly 0.1 percentage points to core CPI and PCE, putting the memory shock's contribution at an estimated 0.2 to 0.4 percentage points. Analyst Jay Kwon's view is that the industry stays in shortage for multiple years, with resolution hard to date.

The transmission is visible further down. HP's finance chief has said memory and storage moved from 15–18% of a PC bill of materials to roughly 35% in 2026. IDC's moderate scenario has the PC market contracting 4.9% this year against a prior forecast of 2.4%, with 4–6% average selling price increases; its pessimistic case is an 8.9% contraction with 6–8% price rises. Smartphones fare better but still face 2.9–5.2% contraction. The people paying most of this tax never bought an AI product.

The distributional point is the one to hold onto: AI capital formation is now bidding a scarce physical input away from consumer electronics, and the

clearing mechanism is price. That is a real transfer, not an accounting one, and it runs from household device budgets to hyperscaler capacity — and, through 84.6% margins, to three memory manufacturers.

V. Who the shortage pays

Expensive memory is not neutral between competitors. A July 2026 preprint by Satoshi Matsuoka of RIKEN's Center for Computational Science argues that memory now represents 40–50% of accelerator cost, which converts the depreciation schedule into a competitive weapon: an operator serving inference from a partly depreciated fleet holds what the paper models as a 3.2× cost advantage over an entrant buying at today's prices in 2026, narrowing in 2027 before widening again toward the end of the decade. This is analysis, not observation, and the model's assumptions deserve scrutiny — but the direction is hard to argue with. When the input price spikes, whoever bought early wins, and the advantage rotates among incumbents rather than transferring to challengers.

The financing side is already adjusting. FactSet's analysis of the five largest hyperscalers finds FY26 capex exceeding \$690 billion across their fiscal years, approaching \$800 billion on a calendar basis including finance leases, against free cash flow expected to approach zero or turn negative for all but Alphabet and Microsoft. Incremental debt as a share of capex rose from 9% in FY24 to 32% in the twelve months to June 2026, with aggregate debt near \$700 billion; Alphabet raised \$84.75 billion of equity in June 2026. **Firms with the strongest cash generation in corporate history are funding this buildout externally.** A cost shock that raises the capital intensity of AI capacity by even ten percent is therefore not absorbed quietly — it shows up as debt, as guidance revisions, or as capacity that does not get built.

FIG. 2: REPORTED DRAM CONTRACT AND SPOT PRICE MOVES, QUARTER ON QUARTER



BARS SCALED TO 95% = 100%. NOT LIKE-FOR-LIKE: Q1 IS TRENDFORCE'S CONTRACT-PRICE FORECAST (FEB 2026); Q2 IS SIGMAINTEL'S LPDDR5X SPOT MOVE (JUN 2026); Q3 IS SAMSUNG'S REPORTED TARGET INCREASE, NOT A SETTLED PRICE (TRENDFORCE, 3 JUL 2026). SHOWN TO INDICATE CADENCE, NOT A SINGLE CONTINUOUS SERIES.

VI. Every memory shortage in history ended in a glut

The precedent is unhelpfully consistent — though we flag this paragraph as sector judgment rather than a sourced finding. Memory shortages have historically been resolved by capacity that arrives roughly two years after the price signal, lands into decelerating demand, and produces a fall steeper than the run-up; the 2017–2018 DRAM cycle is the most recent example. What is different this time is the demand source: previous cycles were driven by PC and handset units, which are countable and saturating, while this one is driven by AI capacity whose demand is a function of token consumption that nobody can forecast with confidence.

Matsuoka's paper puts a number on the threshold: new capacity remains solvent only if token demand sustains roughly 2× annual growth for four consecutive years, falling to about 1.6× if efficiency gains decelerate. That is the load-bearing assumption underneath \$730 billion of guidance. He assigns 25% to a commoditisation crash among his five scenarios — a reminder that the most credible technical analysis of this buildout puts a quarter of the probability mass on it going badly.

VII. How it plays out: three scenarios

Horizon to end-2028. Probabilities are our analytical judgment, not derived from a model, and are stated so they can be scored later.

1 • Baseline: the tax persists, then eases ~50%_{2026–2028} • GRINDING

Contract prices keep rising through 2026 at a decelerating rate, plateau during 2027, and soften in 2028 as SK hynix and Samsung capacity lands. Hyperscaler capex guidance drifts up again at Q3 and Q4 results but the memory attribution stops being novel enough to mention. Consumer device prices stay elevated and the inflation contribution fades from the data by late 2027. Watch for: Q4 2026 contract settlements coming in below 20%; PC ASPs flattening; at least one hyperscaler holding guidance flat rather than raising.

2 · Acceleration: memory becomes permanent

2026–2029 · FAST

overhead ~20%

Inference demand keeps compounding above the 2× threshold, HBM absorbs a rising share of wafer capacity, and the shortage extends toward the 2029–2030 horizon Chey Tae-won described. Calendar capex crosses \$1 trillion, financed increasingly with debt. Memory ceases to be a cycle and becomes a structural cost floor under inference pricing, halting the per-token price declines of 2024–2026. Watch for: token prices flat or rising across two consecutive quarters; memory makers' capex guidance rising without price relief; a fourth entrant attempting DRAM capacity.

3 · Stall: capacity lands into softening

2027–2028 · ABRUPT

demand ~30%

Token growth undershoots, the depreciation wave arrives on schedule, and the new wafer capacity commissioned in 2026 begins output into a market that no longer needs it. Memory prices fall hard, Micron's margins revert toward historical norms, and hyperscalers cut capex the way Amazon did in 2022. The memory shock is retrospectively read as the top of the cycle rather than the start of a constraint. Watch for: hyperscaler depreciation estimate changes disclosed in annual reports; capex-to-operating-cash-flow ratios falling below 80%; a quarter in which AWS or Azure growth decelerates while backlog flattens.

WHAT WOULD FALSIFY THIS THESIS

The thesis is that memory has become the binding priced constraint on AI capacity, with measurable macro spillover. It is wrong if Q3 and Q4 2026 DRAM contract prices settle flat or negative while hyperscaler capex guidance keeps rising — that would mean July's memory attribution was a one-quarter pass-through, not a structural constraint, and that the real driver was volume all along. It is also wrong if no company other than Amazon names memory as a capex driver at Q3 results in late October; a constraint visible to one buyer and not four is a procurement failure, not a market condition.

The strongest counter-case is the accounting one: the depreciation-wall argument holds that four hyperscalers spent \$433.9 billion on property and equipment in the four quarters to March 2026 against roughly \$149 billion of recognised depreciation, and that the 2027–2029 catch-up will compress margins regardless of what memory costs. On that reading memory is a rounding error inside a much larger timing mismatch. We do not think these are exclusive, but if capex falls in 2027 while memory prices are still rising, the accounting explanation wins outright.

WHAT TO WATCH, IN ORDER

1. **Q3 2026 hyperscaler earnings, late October 2026.** Does anyone besides Amazon attribute a capex revision to memory? One company is anecdote; three is a market.
2. **Q4 2026 DRAM contract settlements, TrendForce, December 2026.** The cadence has run +90–95%, +89%, then a targeted +20%. A settlement below 10% is the first genuine sign of the cycle turning.
3. **Micron fiscal Q1 2027 results, around late December 2026.** Gross margin is the cleanest single indicator available. Sustained above 80% means the shortage holds; a fall toward 60% means capacity is arriving.
4. **BLS import price index for computers, peripherals and parts, monthly.** Currently up 37% since end-2024 per J.P. Morgan. The month this rolls over is the month the inflation contribution starts unwinding.
5. **Samsung and SK hynix 2027 capex plans, at Q4 2026 results in January 2027.** The supply-side answer to a shortage is always capex. Aggressive 2027 plans make the stall scenario materially more likely for 2028.

SOURCES

1. The Register, "Cloud giants pour nearly \$600B into capex as AI demand surges," 4 Aug 2026 · [theregister.com](https://www.theregister.com)
2. Fortune, "Andy Jassy said Amazon will spend \$220 billion this year—and still won't have enough capacity," 30 Jul 2026 · [fortune.com](https://www.fortune.com)
3. Meta Platforms, "Meta Reports Second Quarter 2026 Results," 29 Jul 2026 · [prnewswire.com](https://www.prnewswire.com)
4. J.P. Morgan Global Research, "The AI-Driven Memory Shortage: DRAM Prices, Inflation and Market Risks," 6 Aug 2026 · [jpmorgan.com](https://www.jpmorgan.com)
5. Micron Technology, "Micron Reports Record Results for the Third Quarter of Fiscal 2026" (quarter ended 28 May 2026), 24 Jun 2026 · investors.micron.com
6. IDC, "Global Memory Shortage Crisis: Market Analysis and the Potential Impact on the Smartphone and PC Markets in 2026," 18 Dec 2025, updated 10 Feb 2026 · [idc.com](https://www.idc.com)
7. The Register, "DRAM prices expected to double in Q1 as AI ambitions push memory fabs to their limit" (TrendForce data), 2 Feb 2026 · [theregister.com](https://www.theregister.com)
8. TrendForce, "Samsung Reportedly Seeks Up to 20% 3Q26 DRAM Price Increase," 3 Jul 2026 · [trendforce.com](https://www.trendforce.com)
9. TweakTown, "DRAM prices surged by up to 89% in Q2 2026" (SigmaIntel data), 25 Jun 2026 · [tweaktown.com](https://www.tweaktown.com)
10. Techzine, "SK hynix to double wafer capacity amid AI memory shortage," announcement 2 Jun 2026 · [techzine.eu](https://www.techzine.eu)
11. eeNews Europe, "SK hynix warns memory shortage could last up to 2030" (Chey Tae-won at GTC, 16 Mar 2026) · [eenewseurope.com](https://www.eenewseurope.com)
12. S. Matsuoka (RIKEN R-CCS), "Memory Scarcity, Open Models, and the Restructuring of the AI Industry, 2026–2030," preprint, 8 Jul 2026 · arxiv.org
13. FactSet Insight, "Hyperscalers Tap External Financing as AI Capex Outruns Cash Flow" · insight.factset.com
14. Silicon Analysts, "Hyperscaler AI Capex 2026: \$434B Trailing Four Quarters, D&A Lag, Debt Wave" · [siliconanalysts.com](https://www.siliconanalysts.com)
15. Futurum Group, "AI Capex 2026: The \$690B Infrastructure Sprint," 12 Feb 2026 · [futurumgroup.com](https://www.futurumgroup.com)
16. Sourceability, "Memory price increase timeline QoQ in 2026" (HP bill-of-materials commentary) · [sourceability.com](https://www.sourceability.com)

METHOD NOTE: FIGURES ARE ATTRIBUTED AND DATED INLINE; WHERE INDEPENDENT SOURCES DISAGREE, THE DISAGREEMENT IS REPORTED RATHER THAN AVERAGED. SCENARIO PROBABILITIES ARE ANALYTICAL JUDGMENT. MATSUOKA (2026) IS A PREPRINT AND HAS NOT BEEN PEER REVIEWED; ITS COST-ADVANTAGE AND SOLVENCY FIGURES ARE MODEL OUTPUTS, NOT OBSERVATIONS.