

June 2, 2026 08:00 PM GMT

Global Technology

Chipflation – Navigating A Memory Crisis

Surging memory prices and supply scarcity are becoming a cross-sector risk as AI reprices a critical input across the digital economy. What began as an AI infrastructure bottleneck is now spreading into hardware margins, device affordability, cloud costs, inflation and policy.

AI is turning memory into a structural bottleneck. AI demand is rising across HBM, DRAM and enterprise SSDs, pushing memory prices up six-fold in the past year. New supply takes years to build, qualify and ramp. That makes this cycle look less like a normal semiconductor upturn and more like a durable supply-demand reset.

Allocation is replacing commodity-market pricing. Hyperscalers and AI buyers are increasingly using long-term agreements (LTAs), prepayments and strategic commitments to secure capacity. That leaves a smaller, tighter and more volatile supply pool for traditional buyers.

HBM is crowding out conventional memory. HBM is essential for AI accelerators but consumes disproportionate advanced DRAM capacity. As suppliers prioritize HBM, server DRAM and enterprise SSDs, less supply remains for smartphones, PCs, autos, networking & industrial markets. Even as total DRAM wafer capacity expands ~30% by 2027, we see memory supply for smartphones/PCs falling 12-15% short.

Chipflation is spreading beyond Big Tech ... Large cloud buyers can secure supply and capitalize higher costs, while non-AI buyers face higher COGS and weaker allocation, price increases, spec cuts and delayed launches.

... and creates a clear divide between suppliers and OEMs. Memory producers benefit from stronger pricing, margins and visibility. Downstream hardware companies must absorb costs, pass them through, redesign products or risk demand destruction, especially in price-sensitive consumer markets.

The macro impact is bigger than CPI alone. Headline CPI effects may be modest given small basket weights, but pressure is visible in PPI, corporate margins, cloud bills, capex budgets and delayed technology deployment.

Policy could ease pressure, but would take years. Even if the US and China deploy tools such as subsidies, tax credits or permitting reform, supply responses will take years and near-term China capacity growth is insufficient. We assume US policy stays restrictive and is unlikely to ease pressures in the near or long term.

Stock implications. Pricing power sits with DRAM suppliers (Samsung, SK hynix, Micron), NAND (SanDisk, KIOXIA), HDD (Seagate, Western Digital) infrastructure (ASML, AMAT, KLA). OEMs over-indexed to the consumer, with less pricing power and elevated memory cost exposure face the sharpest margin headwinds.

VORGAN STANLEY & CO. INTERNATIONAL P.L.C.+

Shawn Kim

Equity Analyst

Shawn.Kim@morganstanley.com

+1 212 761 1018

VORGAN STANLEY & CO. LLC

Joseph Moore

Equity Analyst

Joseph.Moore@morganstanley.com

+1 212 761 1516

Erik W Woodring

Equity Analyst

Erik.Woodring@morganstanley.com

+1 212 761 8883

Diego Anzoategui

Economist

Diego.Anzoategui@morganstanley.com

+1 212 761 8573

VORGAN STANLEY & CO. INTERNATIONAL P.L.C. (DIFC BRANCH)+

Rajeev Sibal

Senior Global Economist

Rajeev.Sibal@morganstanley.com

+91 11 709 11 01

VORGAN STANLEY & CO. LLC

Ariana Salvatore

Equity Strategist

Ariana.Salvatore@morganstanley.com

VORGAN STANLEY & CO. INTERNATIONAL P.L.C.+

Lee Simpson

Equity Analyst

Lee.Simpson@morganstanley.com

+1 212 761 1318

Cindy Huang

Equity Analyst

Cindy.Huang@morganstanley.com

+1 212 761 1416

VORGAN STANLEY ASIA (PRIVATE)+

Duan Liu

Equity Analyst

Duan.Liu@morganstanley.com

+852 2239 7357

VORGAN STANLEY & CO. LLC

Mason Wayne

Research Associate

Mason.Wayne@morganstanley.com

+1 212 761 6012

Shane Brett

Equity Analyst

Shane.Brett@morganstanley.com

+1 212 761 1012

Maya C Neuman

Research Associate

Maya.Neuman@morganstanley.com

+1 212 761 1416

Dylan Liu

Research Associate

Dylan.Liu@morganstanley.com

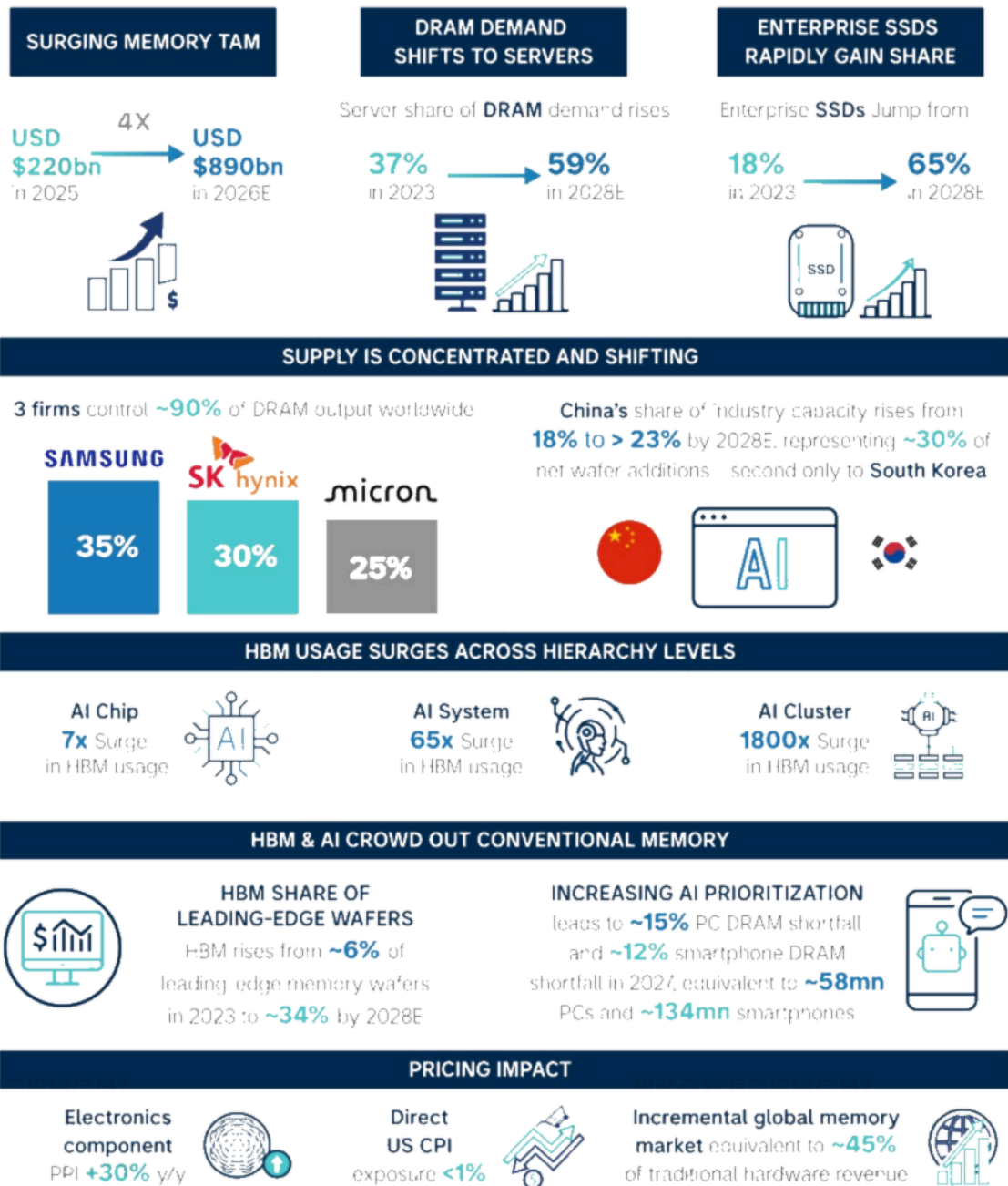
+1 212 761 4519

Morgan Stanley does and seeks to do business with companies covered in Morgan Stanley Research. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of Morgan Stanley Research. Investors should consider Morgan Stanley Research as only a single factor in making their investment decision.

For analyst certification and other important disclosures, refer to the Disclosure Section, located at the end of this report.

+ = Analysts employed by non-U.S. affiliates are not registered with FINRA, may not be associated persons of the member and may not be subject to FINRA restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account.

Chipflation – The Story in Numbers



Source: Trendforce projections (TAM and capacity), Morgan Stanley Research estimates

What is "Chipflation"?

The shift from historical, deflationary pricing for commoditized microelectronics toward sustained, structural price increases poses a continuous upside risk to overall consumer goods pricing. For memory, inflation is no longer just a component-price issue.

AI infrastructure is absorbing a growing share of memory (DRAM, HBM and enterprise SSD) supply, while suppliers prioritize higher-margin AI/server products over mainstream consumer and industrial demand. The potential impacts of higher memory prices include:

- constrained supply and increases in consumer hardware prices
- a squeeze in corporate profit margins
- a division in the hardware industry between the 'have' and 'have not'
- persistent upside risks to core inflation

Chipflation is demand-driven by the rapid rollout of AI infrastructure and datacenters, where tighter memory availability is feeding into higher device ASPs, higher cloud and enterprise IT costs, lower product specifications, delayed refresh cycles and margin pressure.

Memory shortages have become a macroeconomic concern. The direct CPI impact may be limited, but the effects are increasingly showing up in corporate COGS, cloud bills, capex budgets and delayed technology deployment – making this cycle broader than a traditional semiconductor upcycle. Insufficient chip supply can delay data center projects, slow cloud development and increase costs for businesses, ultimately affecting productivity growth.

Differentiated datasets

This report updates a range of proprietary Morgan Stanley models and introduces several new datasets.

- **Two-tier DRAM supply waterfall (Exhibit 7).** Our framework decomposes total DRAM supply down to the residual pool available to non-AI buyers. We estimate PCs could face a 15% memory shortfall (~58mn units) and smartphones a 12% gap (~134mn units) as AI crowds out non-server allocation.
- **Demand elasticity by hardware product (Exhibit 8).** We rank which end markets are most exposed to demand destruction from higher memory prices, from traditional servers (least elastic) to low-end PCs (most elastic).
- **CPI impact of higher memory costs (Exhibit 11).** Higher memory costs carry a measurable CPI pass-through: PCs and smartphones add 0.08pp to headline CPI, with total consumer electronics impact reaching 0.10pp.

Table of Contents

Section	Page Number
Executive Summary	5
Memory – A Multi-Year Bottleneck	14
HBM Cannibalization and the Two-Tier Market	20
Chipflation Passthrough and Sector Impact	26
Chips, Growth and Inflation – A Macro Perspective	37
US/China Policy Options Don't Offer Near-term Relief	40
Primer: Memory 101 – Types, Market Structure and the Cycle	47
Appendix: Memory Demand per Gigawatt of AI Data Center Deployment	53

Executive Summary



Tech Diffusion

A Morgan Stanley Research
Key Theme of 2026

A multi-year memory bottleneck

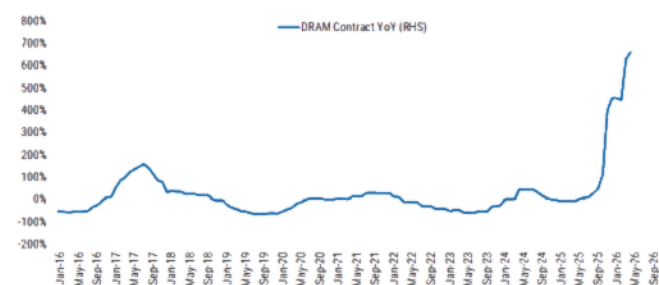
Memory chips are emerging as a critical input in the global AI ecosystem, underpinning the expansion of today's agentic AI architectures (*Technology: Rise of the AI Agent – Global Implications, 19 Apr 2026*). We believe that rising memory prices are rationally anticipating an AI-led price-inelastic leap in memory demand. This pricing strength is less a function of conventional supply discipline by producers, and more a reflection of the exponential growth in end-demand – evidenced in the annual capital expenditure commitments from hyperscalers eager not to be left behind in the race for superintelligence.

Rising memory prices create an increasingly challenging cost-push dynamic, where server build costs rise, cloud capex increases and hardware prices inflate. For non-hyperscaler corporates, this inflationary pressure could require choices on whether to pass on price increases, reduce device specifications, or accept reduced profits.

Structural bottleneck

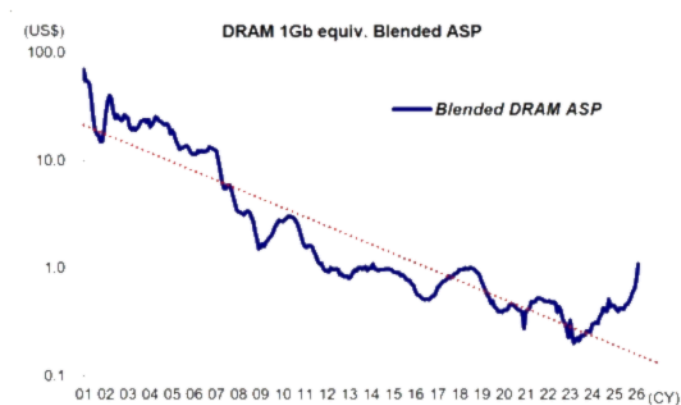
Prices for memory have risen more than 6-fold over the last year (*Exhibit 1*) – a sharp discontinuity from the multi-decade price declines. The price of a gigabyte of DRAM fell by around a factor of ten every 5 years from 1957 to 2020. Since 2010, the price has fallen more slowly as Moore's Law drove sustained price declines. However, this trend no longer applies in the AI economy, which has driven DRAM ASPs sharply higher (*Exhibit 2*). Building memory fabrication (fabs) plants takes years, so there is no quick response to address demand. This means sustained upside risks to consumer goods prices in the coming years.

Exhibit 1: DRAM prices YoY



Source: TrendForce, Morgan Stanley Research

Exhibit 2: Long-term DRAM price/Gb



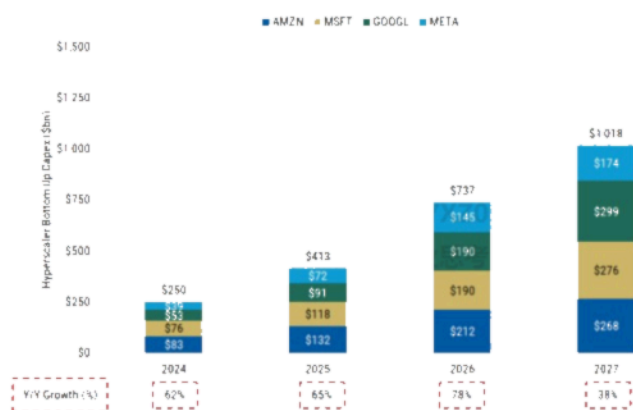
Source: VLSI

Exhibit 3: Memory supply relief depends on install, qualification, yield ramp and product allocation – a ~2 year process

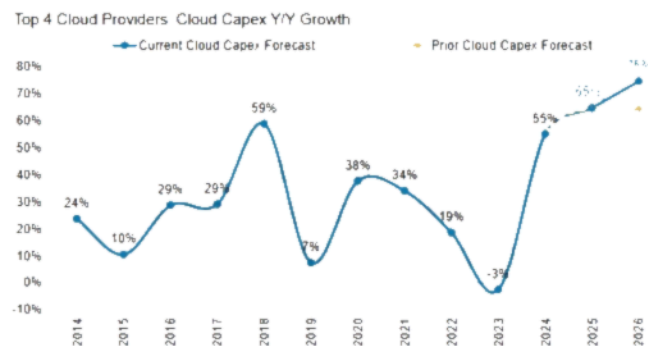
Source: Morgan Stanley Research

What's different this time? Previous chip shortages during COVID demonstrated how even minor supply disruptions can halt entire industries. However, this memory chip shortage looks more structural, and AI-related demand could amplify such effects on a much larger scale.

- **Mobile AI agents and physical AI agents – another very large additional surge in demand for memory chips is likely as these are rolled out.** As AI increasingly moves beyond data centers into edge devices and the physical world, the computational requirement for AI is set to grow exponentially. These components would no longer be seen as commodity memory chips but as critical pillars of the global digital infrastructure.
- **Long-Term Agreements (LTAs) are now absolute necessities to secure capacity** given long lead times and the persistence of the supply-demand mismatch. Memory producers are establishing strategic supply chain partnerships with all hyperscale customers via LTAs, through which memory producers are able to lock in long-term orders and prices, establish a stable high-margin model, and mitigate the industry's cyclical volatility.

Exhibit 4: We expect hyperscaler capex to surpass US\$1tr in 2027 and an aggregate ~US\$2tr invested since 2024

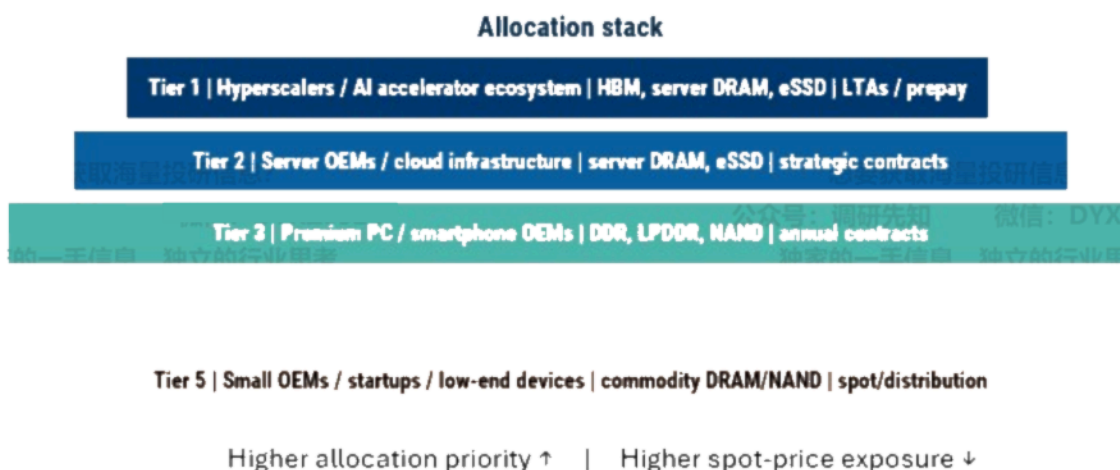
Source: Company data; Morgan Stanley Research estimates

Exhibit 5: The 2026 consensus cloud capex forecast increased to +75% Y/Y post results from 64% previously

Source: Company data; Factset consensus estimates; Morgan Stanley Research estimates

Exhibit 6:

Buyer Hierarchy: Who gets memory supply first?



Source: Morgan Stanley Research

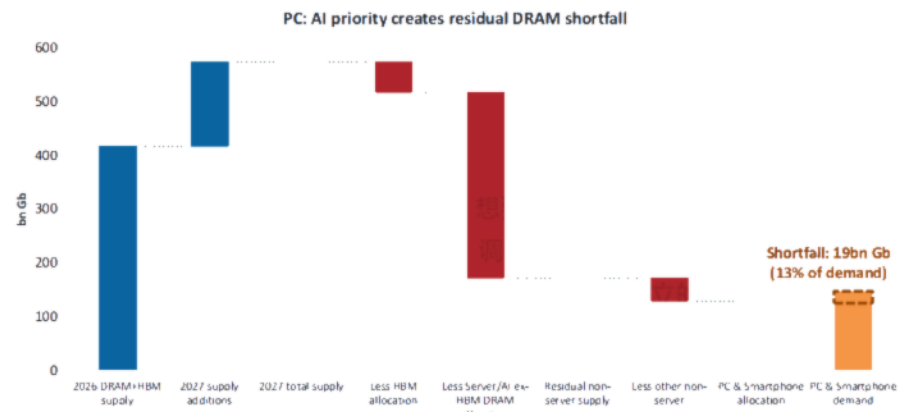
Margin risk for the broader hardware sector

Memory inflation is turning into a broad hardware-sector margin risk, extending beyond PCs and smartphones into gaming consoles, servers, cloud, networking, autos, medical devices, broadband equipment, industrial systems and low-cost education/embedded devices. The impact is not only higher DRAM/NAND pricing, but also allocation priority:

- Suppliers are prioritizing HBM, server DRAM and enterprise SSD demand tied to AI infrastructure, while traditional consumer and industrial memory users are increasingly coping with significantly higher prices, longer lead times and weaker bargaining power.
- The more impacted group are companies that are not directly monetizing AI but still have to pay AI-driven memory prices. There is accounting asymmetry, where hyperscalers can capitalize memory-heavy AI servers and depreciate them over time, while consumer hardware and industrial OEMs see memory inflation flow more directly through inventory and COGS. This will require choices on whether to pass on price increases to consumers, reduce device specifications, or accept reduced profits.
- Memory inflation is not just a financial transfer from non-AI sectors to memory producers. It is increasingly a constraint on what gets built and when for many non-tech industries.

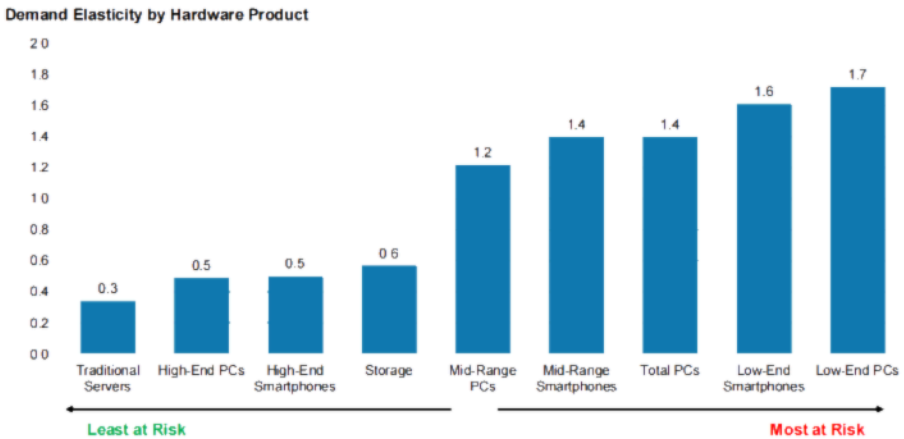
Applying our memory sufficiency framework separately to PC and smartphones, we find both segments face supply shortfalls in 2027. On our assumptions (PCs absorbing 17% of non-server DRAM and smartphones 58%) total addressable memory falls ~15% and ~12% short of demand, respectively. This implies unit downside of ~58mn PCs (-15% vs MS shipment forecast) and ~134mn smartphones (-12% vs MS forecast). These numbers are sensitive to assumptions on memory content per device, but our analysis indicates demand rationing risk if the supply build does not accelerate.

Exhibit 7: AI prioritization turns 2027 supply growth into a consumer memory shortfall



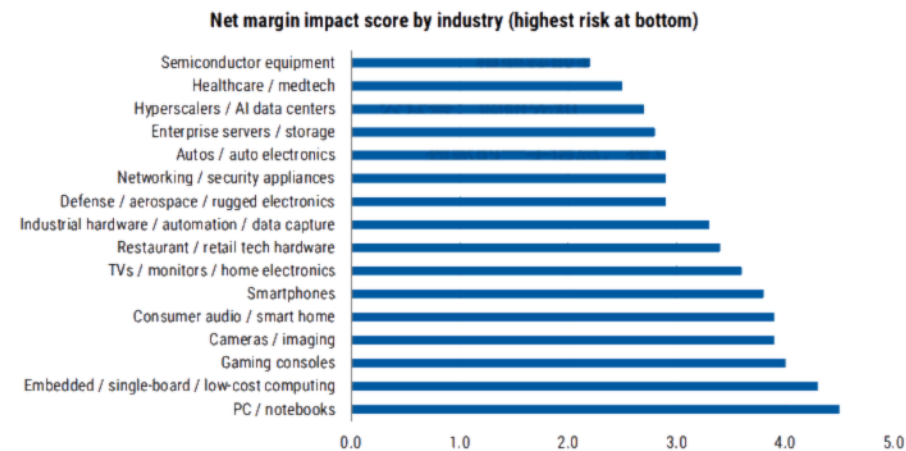
Source: Morgan Stanley Research estimates

Exhibit 8: Our quant analysis implies low-end smartphones and PCs are the most demand elastic, while servers, storage and high-end smartphones are least at risk of demand destruction from higher prices



Source: IDC Company data, AlphaWise, Morgan Stanley Research. Demand elasticity takes the absolute value. PC data goes back to 1995, for storage it starts in 2008, servers in 2003 and smartphones in 2007.

Exhibit 9: The Chipflation impact by industry (illustrative)

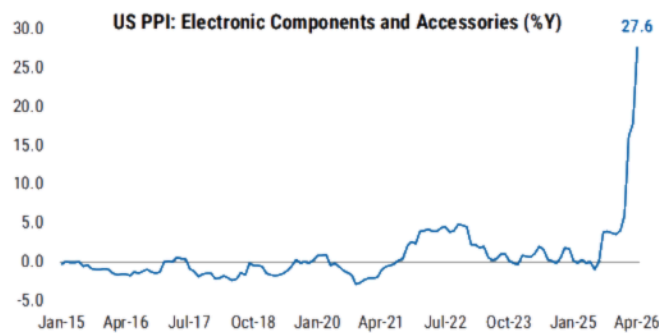


Note: Ranking reflects net margin pressure after passthrough, assuming: 45% gross cost impact + 25% demand/volume risk + 20% allocation risk + 10% pass-through burden. Score scale: 1 = low risk, 5 = highest risk. Source: Morgan Stanley Research estimates

Macro implications

Chipflation is translating into higher producer price inflation, notably in measures of computer and electronics linked equipment. How this translates into consumer inflation in these segments will define margin pressures – for now the effect is strong in US PPI and softer in CPI. The effect on headline CPI inflation is likely less because of the small weight of memory-sensitive goods in the basket, but the longer chip prices stay elevated, the more macro consequences we can expect.

Exhibit 10: The effect of rising costs is visible in exposed PPI, with some sub-components rising to all time highs



Source: BLS, Morgan Stanley Research

Exhibit 11: Higher memory costs could drive 15pp CPI for PCs and smartphones in the US this year – even if the headline CPI effect is more muted because of basket weights

CPI component	Memory costs effect on CPI in 2026 (pp)	CPI weight
PCs	15	0.30%
Smartphones	15	0.20%
TVs	10	0.10%
Cars	0	3.80%
Major household appliances	5	0.06%
Game consoles	125	0.01%
Impact on headline CPI: PCs and smartphones only	0.08	
Total impact on headline CPI	0.10	

Note: Smartphones and game consoles weights are estimates. Source: BLS, Morgan Stanley Research forecasts

Concentration strengthens the position of memory producers and boosts regional markets. For export-oriented economies dominated by semiconductor chip manufacturing (especially South Korea and Taiwan for logic chips), the surge in memory revenues and market demand has substantially bolstered national GDP and local stock indices. Samsung, SK hynix and Micron Technology together control roughly 90% of the global DRAM market and 100% of the HBM market, giving them significant pricing power. South Korea alone produces nearly 75% of the global DRAM chips.

Policy options

Memory chips are increasingly being viewed as a strategic resource, in particular where HBM and advanced DRAM directly enable AI end uses (and consequently, US policy options to support this end are [likely to persist](#) following the midterm elections). We therefore expect policy efforts in this area to focus less on near-term cost relief and more on protecting AI-strategic memory through supply-chain resilience, trusted capacity, and geopolitical de-risking. For less sophisticated (non-frontier) memory end uses (automotive, consumer, etc.) policymakers may seek to distinguish commodity or legacy memory from AI-strategic memory through differentiated licensing and targeted supply-side support for trusted capacity in the former category. Importantly, political obstacles, timing lags, and associated supply chain constraints still likely dilute these efforts' effectiveness. Forced allocation would likely be a last resort, in our view. It might protect favored sectors in the short term, but would risk turning memory into a rationed geopolitical commodity.

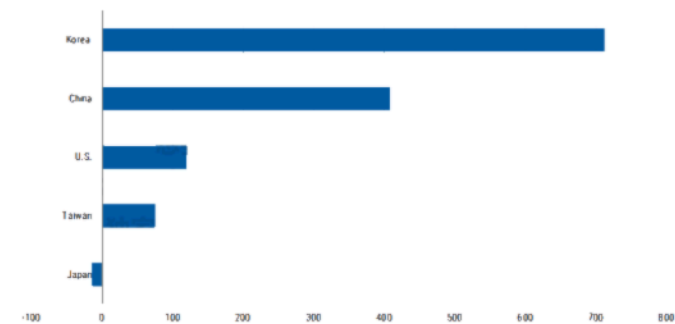
Across the board, the policy options may mitigate, but ultimately not solve, the problem.

- Direct subsidies can support new fabs, packaging and test capacity, but may not lower prices quickly.
- Tax credits can encourage investment but are less targeted. Procurement guarantees can de-risk capacity additions, but may distort allocation.
- Expedited permitting can improve time-to-capacity, but execution is difficult.
- Equipment-access coordination can support allied capacity expansion, but must be balanced against export-control objectives.
- Strategic stockpiles may help critical sectors, but are hard to manage in fast-moving technology markets.

China fabs are the largest swing factor for global memory supply, but the upside is policy-gated

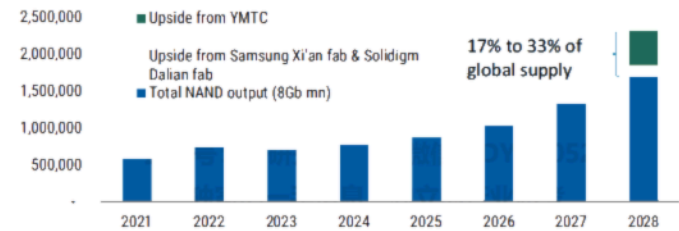
Mainland China accounts for ~30% of net wafer additions over 2023-28E, second only to South Korea ([Exhibit 12](#)). However, not all capacity is equally advanced, accessible, or trusted. On NAND, the potential uplift is considerable: combining YMTC capacity and yield expansion with node migration at Samsung's Xi'an and Solidigm's Dalian facilities, China-based incremental output could reach 17-33% of 2028E global NAND supply ([Exhibit 13](#)). This acceleration scenario assumes a meaningful relaxation of US export controls and is not our base case, as we do not assume the EUV ban on exports to China is lifted. The key gating factor is not Chinese ambition or investment appetite, but access to leading-edge equipment.

Exhibit 12: Mainland China accounts for 30% share of 2023–28E net DRAM wafer additions, behind South Korea



Source: TrendForce estimates, Morgan Stanley Research

Exhibit 13: China NAND acceleration scenario – YMTC and Koran suppliers' China fabs could add up to 17%-33% of 2028 global NAND supply



Source: Morgan Stanley Research estimates

Stock exposure

Companies selling into AI enjoy pricing power, while the buyers face margin pressure. In an agentic AI world, demand and pricing power gravitates towards memory and CPU plus the associated supply chains. This is evident in year-to-date stock performance: global consumer electronics stocks have fallen 1% on average, while a memory makers have risen by almost 300%, alongside unprecedented EPS revisions of 333% YTD. We continue to like DRAM suppliers (Samsung, SK hynix, Micron), NAND (SanDisk, KIOXIA), and HDD (Seagate, Western Digital).

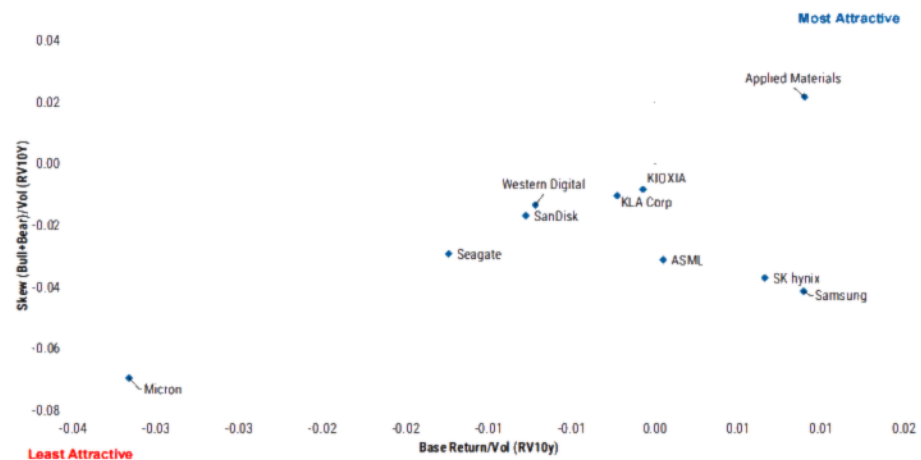
The opportunity extends beyond individual memory chips to the full AI system. The beneficiaries of this shift are global and full-stack. The surge in memory is mainly driven by the recent agentic AI shift broadening well beyond headline AI chips – we lay out in [Exhibit 14](#) which stocks offer exposure. Beneficiaries include CPU vendors, memory suppliers, storage companies, and advanced packaging and substrate providers, alongside foundries, equipment makers and server manufacturers. In short, agentic AI widens the AI investment landscape, shifting focus from owning the best accelerator to enabling the full system that makes intelligent agents work.

Exhibit 14: How to play the theme?

Global Exposure Across the Stack ? Names by exposure					
CPU	DRAM	NAND	HDD	FOUNDRY	IC-design
• NVDA • Intel	• Samsung	• Kioxia	• Seagate	• TSMC	• GUC
• AMD • Arm	• Hynix	• SanDisk	• WDC		• Egis
	• Micron		• TDK		
PCB/Substrate/CCL & Materials			BMC, CPU & Memory interface		
• SEMCO • Unimicron • NYPCB			• Aspeed • Renesas • Montage		
• Ibiden • Nittobo • MEC			• WPG • AP Memory		
MLCC & CPU socket		ODM	SPE		
• Murata • TDK • Yageo		• Wiwynn	• ASML • ASMi • AMAT • Besi • KLAC		
• FIT Hon Teng • Lotes		• Hon Hai	• Tokyo Electron • Ulvac • Wonik		

Source: Morgan Stanley Research

Exhibit 15: Risk/return positioning for key memory and SPE stocks



Source: Factset (historical share price data), Morgan Stanley Research estimates. Note: Methodology: X-axis (Base Return/Vol) represents our case case price target return divided by 10-year real zero volatility (RV10y). Y-axis (Skew) represents the sum of our bull and bear case returns divided by RV10y, capturing payoff asymmetry.

- **US Semiconductors:** Agentic AI-driven CPU demand structurally favors AMD in cloud share gains, but we prefer exposure via AI enablers in NVIDIA and Micron where token growth and capex translate more directly into earnings upside.
- **US Semiconductor Equipment:** Rising compute and CPU TAM drive incremental WFE demand, with DRAM and leading-edge logic (<5nm/2nm) capacity expansion supporting upside for equipment names such as AMAT and KLA.
- **US Hardware:** AI agent proliferation is a structural tailwind for HDD demand, as persistent storage is required to capture growing volumes of context, history, and system-level data, with ~80% of cloud data still residing on disks. We expect HDD demand to remain strong (Demand growing at 50% EB vs. 30%+ supply CAGR), with supply tightness supporting a "stronger for longer" pricing backdrop, benefiting Seagate and Western Digital. On the flip side, we see US Hardware OEMs and hyperscalers as most exposed to memory inflation/shortages. Among the US Hardware OEMs, vendors with (1) more consumer exposure, (2) smaller scale/TAMs, and (3) less pricing power/differentiation face the most significant

margin risk. While Server and Storage OEMs have greater exposure to memory inflation than this cohort, structural demand for AI supports greater pricing power/ability to pass through costs to the end customer (DELL, HPE). This puts small cap consumer electronics and PC vendors most at risk of seeing margin pressure. AAPL is our only Overweight-rated OEM in US Technology Hardware.

- **European Semiconductors:** ASML's EUV tools are a critical enabler of future DRAM supply expansion, and the company benefits from rising compute intensity and design complexity.
- **Korea Technology:** We continue to see Samsung Electronics and SK hynix as key beneficiaries of memory content growth from AI CPU penetration.
- **Greater China Semiconductors:** Montage is leveraged to the structural increase in memory content per server, with higher CPU and DRAM intensity driving demand for memory interconnect solutions.
- **Japan Semiconductors:** Increasing CPU complexity and memory scaling drive demand for advanced semi-cap (Tokyo Electron, Ulvac) and memory interface solutions, positioning Renesas as a beneficiary of higher DRAM bandwidth/capacity requirements. Agentic AI also supports NAND demand (KIOXIA) and broader data center exposure across Japan semi names.

What could challenge the chipflation theme

AI spend slows down. Reduced competition among frontier LLMs could result in lower AI capex spend and a shift back to ample memory supply. Alternatively, dramatic improvements in AI model architectures and KV cache/tokenization could shrink the amount of memory needed by data centers per token generation. The high costs of funding AI infrastructure could also lead major cloud service providers to scale back buildouts.

Electricity and power supply. The availability of reliable, large-scale electricity could be an obstacle, as scaling advanced AI models and running agentic ecosystems requires unprecedented electrical loads. A new phase of agentic systems designed to reason, plan and operate over extended periods with continuous, high-intensity computing is pushing data centers into sustained peak power usage, far beyond the short bursts typical of earlier chatbot-style AI. This poses a significant challenge to global grid infrastructures.

Demand destruction. Memory demand destruction is already here for traditional technology products and could take the form of a hardware margin squeeze, higher retail prices, and a structural market shift where smaller vendors and mid-tier OEMs are disproportionately affected. A material slowdown in demand would ease supply tightness and hence reduce prices.

Geopolitics and trade policy. Memory and advanced packaging sit in a global supply chain. Any new restriction or disruption could alter patterns of supply and demand.

Memory – A Multi-Year Bottleneck

Shawn Kim, Joseph Moore

The explosive demand for AI infrastructure has caused supply constraints in all types of memory (DRAM, HBM, NAND, HDD). As AI increasingly moves beyond data centers into edge devices and the physical world, it could trigger a further surge in demand for memory chips. In that scenario, these components would no longer be seen as commodity chips but as critical pillars of the global digital infrastructure.

AI and data centers are becoming the dominant buyers of memory – we expect the server share of DRAM demand to rise from 37% in 2023 to 59% in 2028e, while enterprise SSDs jump from 18% to 65% of NAND demand. The result is a structurally tighter market where non-AI buyers may face weaker access to supply for an extended period.

A quick primer – what is memory?

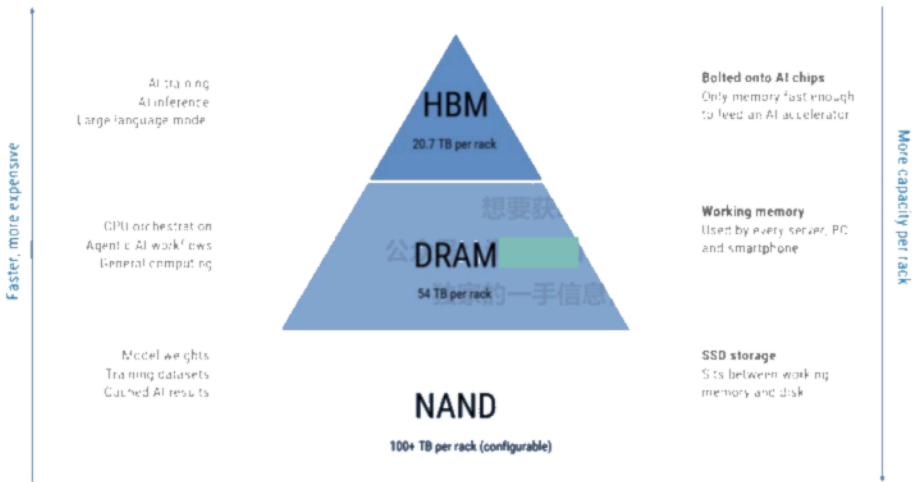
Memory is the working space a computer uses to hold data it is actively processing – the digital equivalent of the desk you are working on right now, as opposed to the filing cabinet behind you. Three types of memory matter for the AI economy:

- **DRAM (Dynamic Random Access Memory)** is the fast but expensive working memory used by every server, PC and smartphone. It is what the CPU uses to hold and fetch the data it is computing.
- **HBM (High Bandwidth Memory)** is a premium type of DRAM focused on providing maximum data bandwidth. HBM is DRAM stacked vertically and bolted directly onto AI chips – it is the memory layer that makes large AI models possible, feeding data fast enough to limit an AI accelerator downtime. Think of DRAM as a single lane road and HBM as a 12-lane highway for data transmission.
- **NAND** is the storage memory used in solid-state drives (SSDs), which sit between working memory and slower HDD disk storage. NAND holds the trained model weights, the training datasets, and the cached results that AI systems retrieve from. It is also used for picture, video and data storage.

All three types of memory matter as AI workloads use all three at once and in massive quantities. Next-generation rack-scale AI systems such as NVIDIA's Vera Rubin NVL72 contain 20.7TB of HBM attached to GPUs and 54TB of LPDDR5X serving the CPUs, before considering local or external SSD storage.

AI demand is scaling across three layers at once — more memory per chip, more chips per system, and more systems per cluster. Agentic AI adds another layer of pressure through longer context windows, larger conversation memory and repeated inference across multiple agents. The result is a structural increase in memory intensity across all aspects of the AI stack

Exhibit 16: AI servers are becoming memory systems



Source: NVIDIA, Morgan Stanley Research. Note: Rubin NVL72 memory capacity based on platform disclosures. NAND/SSD capacity varies by OEM and storage architecture.

For readers less familiar with the memory market structure, a full primer is included in [Primer: Memory 101 – Types, Market Structure and the Cycle](#)

The AI Memory Wall

The bottleneck in AI is increasingly memory, not just compute. As inference shifts from single-turn chat to agentic workflows, memory intensity rises sharply: large models need weights resident in memory, while longer context windows and KV caches increase memory usage per user, per session and per agent. Agent swarms multiply this effect because many agents maintain or reload context simultaneously, turning inference into a much larger memory-capacity and memory-bandwidth problem. Hyperscalers therefore need more HBM, server DRAM and enterprise SSD capacity to meet this surge in memory use now, not when new fabs ramp in a few years' time.

That urgency is pushing Big Tech toward multi-year LTAs, to secure future capacity via prepayments. The challenge for producers is that supply is unable to respond immediately: EUV tool availability, fab construction, process qualification, HBM stacking, packaging, testing and yield ramp are multi-year, multi-billion-dollar projects. This is why visible demand today can still translate into shortages through 2027 and beyond.

Exhibit 17: From tool capacity to qualified memory output – lead times are multiple quarters

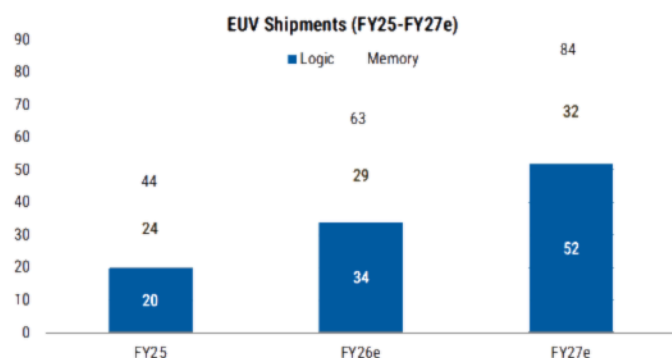


Source: Morgan Stanley Research

Supply response requires demand visibility. The supply response is coming, but it requires demand visibility and commercial commitment. Our Europe semis team expects ASML to expand capacity in line with sustained EUV demand, with shipments guided from 44 systems last year to 60+ systems this year and potential capacity moving higher as demand visibility improves. The issue is therefore not simply ASML lead time. The constraint is the full conversion chain: memory customers need to commit, tools need to ship, fabs need to be ready, processes need to qualify, yields need to ramp, and HBM packaging/test capacity needs to scale.

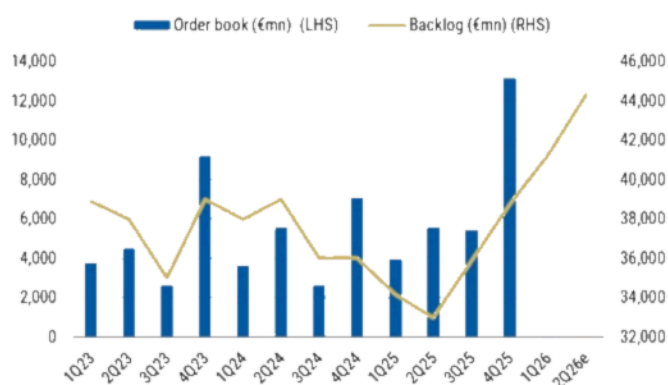
The post-Covid cycle was a lesson here. Intel and Samsung both deferred EUV orders after demand weakened and execution challenges delayed ramps, creating order volatility for the equipment supply chain. This cycle, memory makers and hyperscalers may need to provide firmer LTAs, prepayments and allocation commitments before capacity is built. That supports our view that supply relief is real, but phased – and that memory tightness could persist into 2027 even as EUV capacity expands.

Exhibit 18: ASML EUV shipments FY25-FY27e



Source: Company data; Morgan Stanley Research estimates

Exhibit 19: ASML order book and backlog

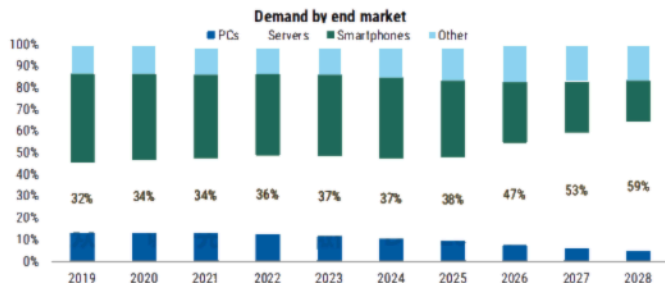


Source: Company data; Morgan Stanley Research estimates

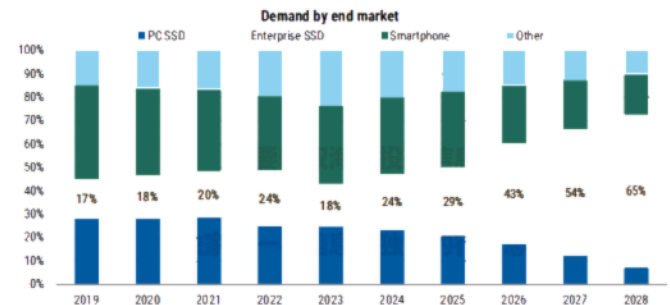
Demand is shifting from consumer electronics to AI/server infrastructure

DRAM server share of bit demand rises from 37% in 2023 to 59% in 2028e, while the share of shipments for PCs and smartphones declines sharply. This matters because the dominant buyer increasingly has stronger purchasing power, longer-duration demand visibility and greater willingness to sign LTAs or prepay for supply. As a result, pricing and shipment allocation continue to gravitate toward AI/server infrastructure requirements at the expense of traditional consumer end markets.

NAND is undergoing a rapid mix shift in enterprise SSDs. Enterprise SSD share of NAND bit demand rises from 18% in 2023 to 65% in 2028e, making AI/server storage the key swing factor for NAND. This reflects the storage intensity of AI workloads, including training data, model checkpoints, retrieval-augmented generation, inference infrastructure and enterprise data lakes.

Exhibit 20: DRAM – AI server demand becomes dominant ...

Source: Trend force estimates, Morgan Stanley Research

Exhibit 21: ... as well as NAND enterprise SSDs as key growth drivers

Source: Trend force estimates, Morgan Stanley Research

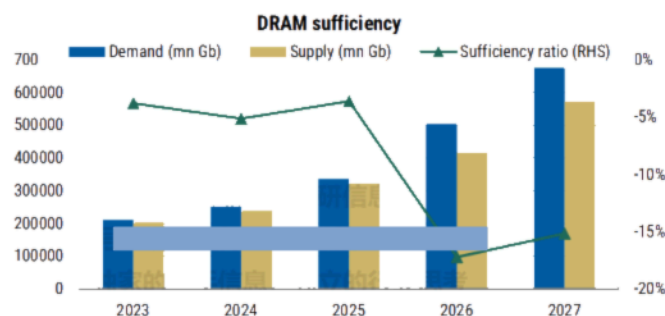
Supply slow to respond

New memory capacity will come online from late 2027 but is still unable to meet demand for all markets. Strategically, most of the valuable capacity is likely to be dedicated to where pricing, margins and customer commitments are strongest – namely to HBM customers, server DRAM and enterprise SSDs. While AI semiconductors are forecast to grow at a ~50% CAGR by 2030 (as per [TSMC forecasts](#) at its annual technology symposium on May 14, 2026), we expect DRAM production to grow at a pace that will meet only about 60% of demand, given annual bit shipments are unlikely to rise much above the 30% level due to EUV constraints.

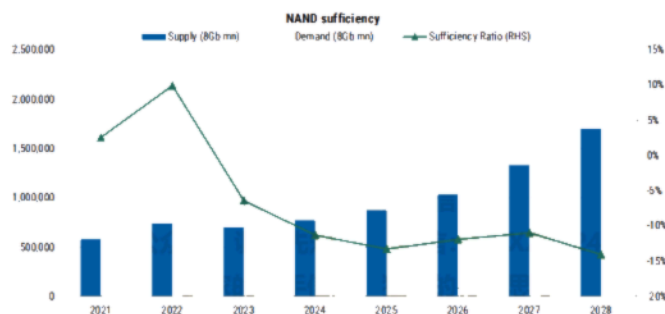
We model DRAM supply remaining tight even as capacity additions accelerate. DRAM undersupply widens from ~4–5% in 2023–24 to ~17% in 2025–26, before easing but still remaining tight at ~15% in 2027e. The key issue is timing.

- Annual DRAM wafer capacity additions rise from ~105kwp in 2025e to ~390kwp in 2027e and ~460kwp in 2028e, led by Samsung, SK hynix, Micron and CXMT. This represents meaningful supply growth by historical standards, but arrives after the period of maximum tightness.
- NAND remains under-supplied through 2028e, with the sufficiency ratio deteriorating to roughly ~14% undersupply in 2028e. Capacity additions recover only after net cuts in 2024–25, rising to roughly ~170kwp in 2028e, led by Samsung, KIOXIA/SanDisk, SK hynix/Solidigm, Micron and YMTC.
- Of these, CXMT and YMTC are Chinese memory producers whose recent acceleration represents a meaningful shift in the supply-side landscape. From 2023 to 2028e, we expect Mainland China to account for ~31% of net global DRAM capacity additions, behind South Korea but ahead of the US, Taiwan and Japan. We discuss China's role in the global memory landscape and the geopolitical questions it raises [here](#).

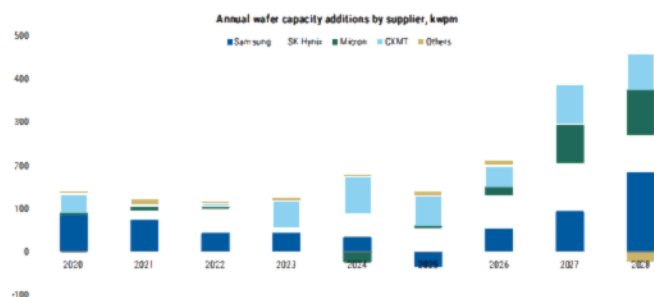
The timing issue in bringing on new capacity is compounded by execution risk. New memory capacity requires fab construction, tool delivery, process qualification, customer validation and yield ramp. Even when capacity is announced, usable output can lag by several quarters or years.

Exhibit 22: DRAM supply sufficiency ratio

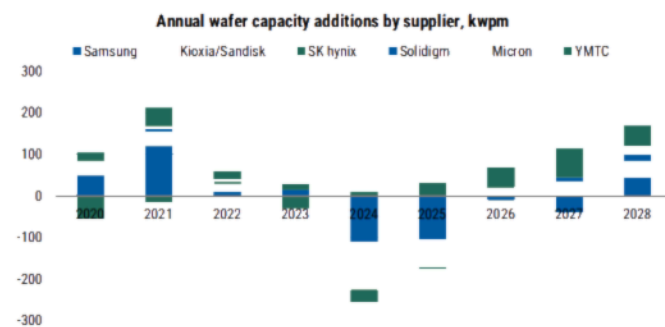
Source: TrendForce estimates, Morgan Stanley Research

Exhibit 23: NAND supply sufficiency ratio

Source: TrendForce estimates, Morgan Stanley Research

Exhibit 24: DRAM annual wafer capacity additions are moving to all-time high

Source: TrendForce estimates, Morgan Stanley Research

Exhibit 25: NAND annual wafer capacity additions are held back by DRAM producers

Source: TrendForce estimates, Morgan Stanley Research

LTAs: strategies to guarantee supply allocation and priority

Shifting from price-driven cycles to supply-security focused LTAs. The memory industry is undergoing a structural transformation in how supply is contracted. Historically, memory customers often procured opportunistically, with pricing largely linked to short-term supply-demand conditions. During upcycles, customers sometimes signed LTAs to secure allocation, but these agreements often had limited enforceability and could be renegotiated when the market softened.

This cycle looks different as memory availability is now tied directly to AI deployment roadmaps. Leading Cloud Service Providers (CSPs) are no longer managing just price risk; they are managing supply risk as well. As a result, multi-year agreements, production capacity guarantees and prepayments are becoming more important.

- Samsung has discussed moving quarterly and annual contracts into 3-5 year agreements.
- SK hynix has highlighted increasing customer requests for medium- to long-term supply commitments.
- Micron has announced a 5-year Strategic Customer Agreement.
- SanDisk has described multi-year partnerships backed by firm financial guarantees.
- KIOXIA has indicated that long-term supply agreements with hyperscale customers are extending into FY28-29.

This is where we see a two-tier market starting to emerge. AI/cloud buyers lock in supply first, while non-LTA buyers compete for a smaller and more volatile residual pool.

HBM Cannibalization and the Two-Tier Market

Shawn Kim, Joseph Moore

HBM is increasingly becoming the mechanism through which AI demand crowds out conventional memory supply, by consuming advanced DRAM wafer capacity and locking in future supply through LTAs. This is sharply reducing available DRAM supply – our analysis implies a ~15% PC DRAM shortfall and ~12% smartphone DRAM shortfall, equivalent to ~58mn PCs and ~134mn smartphones. This could drive a two-tier market, where large AI buyers secure supply first, while non-AI buyers face higher prices, weaker allocation and higher volatility.

HBM demand is scaling non-linearly

The HBM demand curve is compounding at every layer of the AI stack. With growing complexity in AI models and larger systems architectures, memory requires higher capacity, lower latency, higher bandwidth, and improved energy efficiency. All leading AI chips deployed for AI training and inference use HBM to scale memory capacity and bandwidth per chip by adding more stacks, higher layer counts, with faster generations of HBM. HBM's critical role in AI chip architecture is scaling across three layers at once:

- HBM content rises 7.2x at the chip level, from 40GB per A100 GPU to 288GB per Rubin GPU.
- At the system level, memory content rises even faster: an 8-GPU A100 server had roughly 320GB of HBM, while a Rubin NVL72 rack-scale system carries 20.7TB, or around 65x more.
- At cluster scale, the memory consumption becomes exponential. A 2020 training cluster with ~256 A100 GPUs contained roughly ~10TB of HBM. A 2026 frontier cluster with ~200,000+ GPUs could contain roughly ~18PB of HBM, or around ~1,800x more.

Exhibit 26: AI memory intensity compounds across chip, server, rack and cluster



Source: NVIDIA disclosures, Morgan Stanley Research estimates. Note: HBM capacities are illustrative platform-level comparisons; cluster estimate assumes frontier-scale training clusters

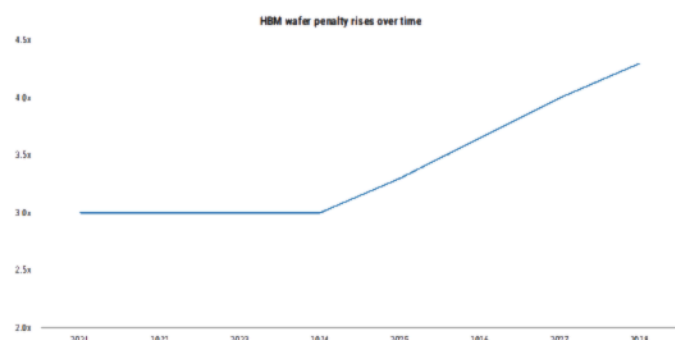
HBM consumes disproportionate DRAM capacity

HBM is not just incremental DRAM demand. It causes a meaningful throughput penalty due to significantly larger die sizes and yield challenges from 3D stacking, TSVs, advanced packaging, intensive testing and customer qualification. These steps significantly reduce effective bit output versus conventional DRAM and increase the burden on wafer, packaging and test capacity.

Our model shows the HBM bit-output penalty rising from roughly 3.0x conventional DRAM wafers in 2021–24 to ~4.3x by 2028e. In other words, each unit of HBM output consumes materially more wafer capacity than conventional DRAM output.

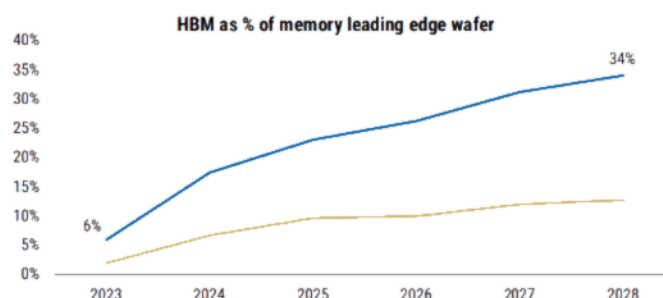
At the same time, HBM is absorbing a growing share of the most constrained wafer pool. HBM rises from ~6% of leading-edge memory wafers in 2023 to ~34% by 2028e. AI customers are not only adding new demand; they are taking a larger share of leading-edge capacity that might otherwise support DDR, LPDDR and conventional server DRAM.

Exhibit 27: HBM die penalty – bit-output is reduced due to much larger DRAM dies



Source: Morgan Stanley Research estimates

Exhibit 28: HBM cannibalization: share of advanced DRAM wafer is rising sharply



Source: Morgan Stanley Research estimates

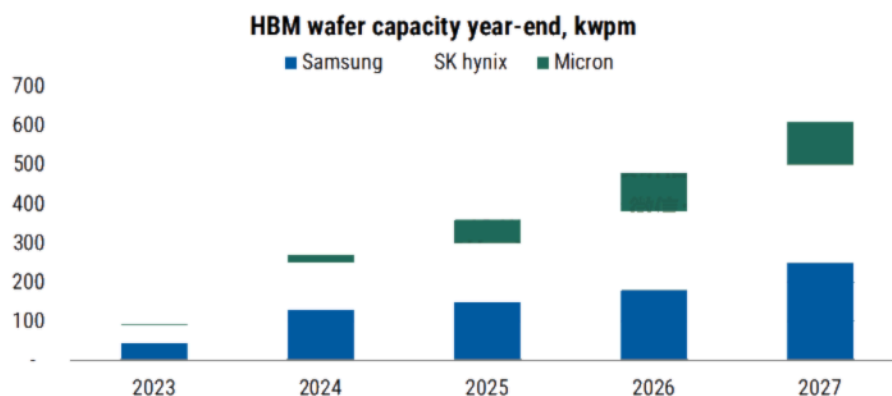
HBM supply is ramping, but remains concentrated

Significant growth in HBM bit demand is driven by AI chip demand. HBM capacity is ramping quickly, but the market remains concentrated across Samsung, SK hynix and Micron. We forecast year-end HBM capacity to rise from ~480kwpm in 2026e to ~610kwpm in 2027e.

Even with this ramp, supply remains constrained because HBM capacity is not just a wafer-start question. It also depends on yield, utilization, TSV capability, advanced packaging, test capacity and customer qualification. Qualification is especially important: not all HBM capacity is immediately usable by all accelerator customers.

What about China HBM? Current export restrictions ban all raw HBM stacks into China; however, chips with HBM can still be shipped as long as they do not exceed the FLOPS regulations. CXMT is rapidly expanding HBM capacity, targeting HBM3 this year and HBM3e in 2027 to support domestic AI chips. Other new entrants, such as XMC (Wuhan Xinxin) are producing the HBM wafers with new DRAM dies set for production by YMTC later this year. Current capacity is at R&D scale but is planned to ramp in coming years.

Exhibit 29: HBM wafer capacity is still highly concentrated in the Big 3 DRAM companies



Source: Company data, Morgan Stanley Research

HBM Model Update

We update our HBM model with the following key changes:

- Capacity:** We expect all three players to continue their aggressive back-end capacity expansion. We expect total capacity to reach 610k/wpm by the end of 2027e with the potential upside if pricing negotiation goes smoothly.
- Yield and UTR:** We increase the UTR and yield assumptions for Samsung Electronics, as the current qualification progress for HBM4 at key customers has been going well and company management has been confident on the market share gain moving into 2H26.
- Sufficiency:** On the HBM side, we believe the supply and demand balance will be tight this year, but with Rubin Ultra ramping up in late 2027 and under current capacity expansion plans, HBM might be in shortage moving into 2027. On the overall sufficiency of the DRAM industry, we model a -7-14% decline in smartphone, PC and consumer DRAM demand in 2026 but nearly 100% growth in server DRAM demand, based on our channel checks. As a result, our 2026 forecast commodity DRAM growth increases to 48% versus supply growth at 29%, and our total DRAM & HBM sufficiency ratio decreases to -17% from -4% in 2025e. Our 2027e forecasts for YoY growth in commodity DRAM supply (37%) and demand (32%) are based on our US Semiconductor team's WFE model forecast, together with our HBM forecast. For 2027e, we model sufficiency slightly improving to -15%

Exhibit 30: HBM supply and demand model

Sufficiency estimates	2023	2024	2025	2026e	2027e
HBM TSV Capacity (K wpm)					
Samsung	45	130	150	180	250
SK Hynix	45	120	150	200	250
Micron	3	20	60	100	110
Total	93	270	360	480	610
Yield rate assumptions					
Samsung		50%	60%	55%	70%
SK Hynix		60%	75%	65%	70%
Micron		50%	70%	65%	70%
UTR assumptions					
Samsung		80%	60%	70%	100%
SK Hynix		100%	100%	100%	100%
Micron		100%	100%	100%	100%
Implied HBM production (mn Gb)					
Samsung	1,500	4,435	6,387	9,696	21,239
SK Hynix	1,500	6,273	12,830	17,363	22,226
Micron	150	729	3,548	7,937	10,372
Total	3,150	11,436	22,765	34,997	53,837
Sufficiency Ratio					
HBM Demand (mn Gb)	1,866	9,059	19,516	34,205	56,085
Commodity DRAM market demand (mn Gb)	209,527	243,423	316,640	469,675	619,970
Total HBM+DRAM demand (mn Gb)	211,393	252,481	336,156	503,879	676,055
Commodity DRAM market supply (mn Gb)	200,214	228,086	301,257	382,107	519,666
Total HBM+DRAM supply (mn Gb)	203,364	239,523	324,022	417,104	573,503
Total DRAM sufficiency	-4%	-5%	-4%	-17%	-15%

Source: Company data; Morgan Stanley Research estimates

NVIDIA accounts for the largest share of HBM demand, and we expect this to continue in 2027 driven by its Rubin Ultra roadmap, which increases per GPU capacity to 1TB. Broadcom follows as TPU and MTIA volumes surge, while other incremental projects add smaller increases. Amazon also emerges as one of the top HBM customers, with an aggressive Trainium AI chip roadmap.

Exhibit 31: HBM – AI chip consumption in 2026

AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	HBM chip density (GB)	HBM chip units	Total HBM size (GB)	HBM generation	HBM vendor	Total HBM demand (k GB)
AI GPU (2026e)										
NVIDIA	B300	390	14	5,460	36	8	288	HBM3e 12hi	Hynix/Micron/Samsung	1,572,480
	Vera CPU	90	23	2,070						
	Spectrum/CPX	60		-						
	Rubin R200	260	8	2,080	36	8	288	HBM4	Hynix/Micron/Samsung?	599,040
AMD	H200	75	27	2,025	24	6	141	HBM3e 8hi	Hynix	285,525
	Mi300	3	12	36	24	8	192	HBM3	Samsung	6,912
	Mi350/375	7	12	84	36	8	288	HBM3e 12hi	Samsung/Micron	24,192
	Mi400	65	10	650	36	12	432	HBM4	Samsung/Micron	280,800
AI ASIC (2026e)										
Google	TPU v7p (Ironwood; AVGO)	145	16	2,320	24	8	192	HBM3e 8hi	Hynix/Samsung	445,440
	TPU v8i (Sunfish; AVGO)	80	12	960	36	8	288	HBM3e 12hi	Hynix/Samsung/Micron	276,480
	TPU v8t (Zebrafish; MediaTek)	40	20	800	36	6	216	HBM3e 12hi	Hynix/Micron	172,800
AWS	Trainium 3	100	17	1,700	36	4	144	HBM3e 12hi	Hynix/Samsung/Micron	244,800
Microsoft	Maia 200	4	29	116	16	4	64	HBM3	Samsung	7,424
	Maia 300	5	11	55	36	8	288	HBM4	Samsung	15,840
Total		1,479								4,034,885
Total HBM demand (mn Gb)										32,279

Source: Company data; Morgan Stanley Research estimates. Note: Estimates are compiled using our Asian supply chain checks.

LTAs could create a two-tier memory market

As top-tier CSPs secure multi-year supply through LTAs, the residual memory market could become structurally tighter. Customers without LTAs may need to procure from a smaller pool of uncommitted supply, potentially at higher and more volatile prices. We view this as a potential structural cost borne by every buyer who is not at the front of the LTA queue.

Historically, weak PC or smartphone demand would help loosen DRAM markets and lower prices for everyone. In this cycle, AI demand keeps allocation tight even if consumer electronics demand softens, because HBM and server applications now set the priority order, and supplier incentives reinforce the allocation. After the last memory downcycle, suppliers have limited incentive to flood the market with commodity supply when AI customers offer stronger pricing, better margins and more durable contracts.

We run DRAM sufficiency analysis on the two largest non-server end markets: PCs and smartphones

Starting from total 2027 DRAM + HBM supply of 573,503mn Gb, we allocate 70% to servers and 30% to non-server applications. Within non-server, we assign 17% to PCs and 58% to smartphones, consistent with historical patterns. On the PC side, bottom-up demand of 34,386mn Gb implies a shortfall of 5,138mn Gb (-15%). The picture is comparably challenging for smartphones: total smartphone DRAM demand of 113,125mn Gb against an addressable supply of 99,790mn Gb produces a 13,335mn Gb gap (-12%).

The shortfall proves highly sensitive to content assumptions. On the PC side, our base case of 89Gb per device implies a -15% shortfall; holding content flat year-on-year at 83Gb narrows this to -9%, and de-spec'ing to 77Gb reduces it to just -2%. Smartphones follow the same pattern: against a 99Gb base case (-12%), flat content at 95Gb narrows the gap to -7%, and de-spec'ing to 90Gb brings it to -2%.

The read-through is twofold. First, the unit-volume risk is a function of memory content assumptions, not an absolute supply wall — OEMs have a clear lever to protect shipment volumes. Second, that lever is not free: every gigabyte of de-spec'ing that protects unit volumes is a gigabyte of forgone content growth, so the adjustment ultimately resolves the unit shortfall by suppressing memory bit demand. The market clears either through fewer devices or lower memory intensity per device — and the more likely outcome is a blend of the two.

This reinforces supplier pricing power and changes the cycle. The more supply is committed to strategic AI customers, the less flexible supply remains for opportunistic or non-LTA buyers. This could support ASP resilience even if parts of consumer demand weaken.

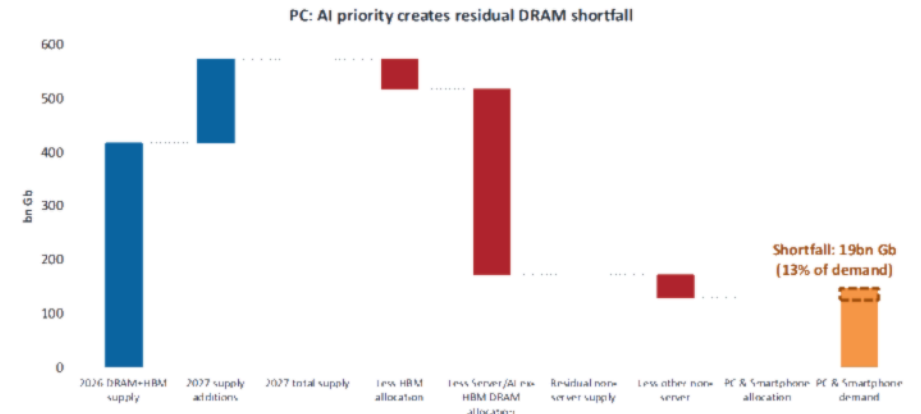
Exhibit 32: Our memory sufficiency framework suggests PC and smartphone could see 12-15% memory shortfall

PC memory sufficiency in 2027			Smartphone memory sufficiency in 2027		
Total DRAM + HBM supply (mn Gb)	573,503		Total DRAM + HBM supply (2027)	573,503	
Server gets 70% of allocation	401,452		Server gets 70% of allocation	401,452	
Non-Server gets 30% of allocation	172,051		Non-Server gets 30% of allocation	172,051	
PC gets 17% of non-server allocation	29,249		Smartphone gets 58% of non-server allocation	99,790	
Total PC memory demand (mn Gb)	34,386		Total smartphone demand (2027)	113,125	
Shortfall (mn Gb)	(5,138)		Shortfall (mn Gb)	(13,335)	
	-15%			-12%	
Memory content per device (Gb)	89		Memory content per device (Gb)	99	
Shortfall (mn units)	(58)		Shortfall (mn units)	(134)	
vs. MS PC shipment forecast	388		vs. MS smartphone forecast	1,137	
Downside vs. MSE	-15%		Downside vs. Mse	-12%	

Sensitivity analysis			Sensitivity analysis		
Memory content per device (Gb)	Shortfall		Memory content per device (Gb)	Shortfall	
Current assumption	89	-15%	Current assumption	99	-12%
Fiat content YoY	83	-9%	Fiat content YoY	95	-7%
De-specing	77	-2%	De-specing	90	-2%

Source: Morgan Stanley Research estimates. Note: PC includes Laptops, desktops and tablets

Exhibit 33: AI prioritization turns 2027 supply growth into a consumer memory shortfall



Source: Morgan Stanley Research estimates

Chipflation Passthrough and Sector Impact

Erik Woodring, Shawn Kim

The current memory cycle is being driven by a structural reallocation of wafer capacity and supplier attention toward AI-related demand. This memory shortage is reshaping PC and smartphone markets as suppliers prioritize higher-value AI/server demand over lower-margin consumer applications. This is already translating into end-demand risk.

We now expect surging memory costs to lower 2026 PC shipments by 10%+ and smartphone shipments by 13% versus 2025, with memory and SSD price inflation driving higher device prices and demand increasingly skewing toward premium devices.

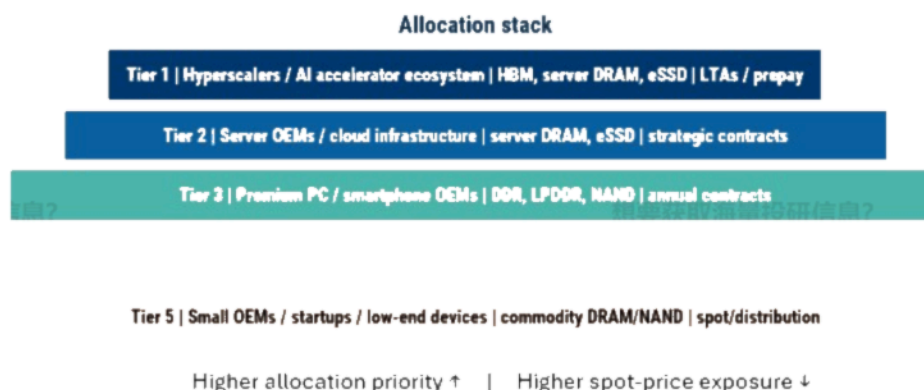
There are long-term ramifications too – device redesigns, a structural shift in unit economics, industry consolidation, and a re-evaluation of the workload mix on-prem versus in the cloud.

Buyer-seller matrix: who gets priority?

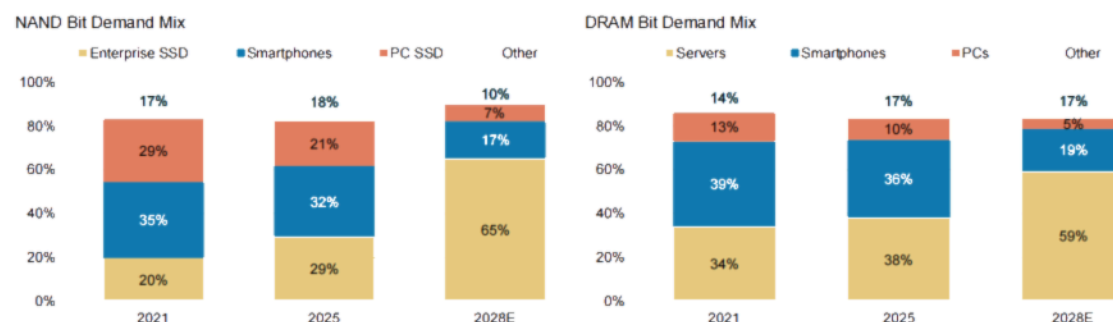
The buyer-seller map is shifting from a commodity market to an allocation market. On the supply side, Samsung, SK hynix and Micron dominate advanced DRAM and HBM, with CXMT increasingly relevant in DRAM capacity additions. In NAND and enterprise SSDs, the key suppliers include Samsung, KIOXIA/SanDisk, SK hynix/Solidigm, Micron and YMTC.

On the buyer side, hyperscalers, AI infrastructure providers and GPU/accelerator ecosystems are the most strategic customers for HBM, high-density server DRAM and enterprise SSDs. These buyers are more likely to receive priority allocation because they offer scale, roadmap visibility and willingness to commit through LTAs – and as a result, they sit at the front of the allocation queue. Traditional consumer and industrial buyers are more exposed: PC OEMs, smartphone OEMs, gaming console makers, networking companies, automakers, industrial firms, medical device makers and smaller hardware companies all rely on DRAM, LPDDR and NAND, but have weaker bargaining power than AI/cloud customers.

The strategic tension is that upstream supply is concentrated while downstream demand is broad. When AI customers move to the front of the allocation queue, the remaining markets face tighter access and higher price volatility.

Exhibit 34: Buyer Hierarchy: Who gets memory supply first?

Source: Morgan Stanley Research

Exhibit 35: Memory demand is overwhelmingly shifting towards data center technologies and away from more consumer-facing Smartphone and PC markets

Source: TrendForce estimates, Morgan Stanley Research

Sector impact and passthrough – where the pain lands

Memory inflation is becoming a broad hardware-sector margin risk, extending beyond PCs and smartphones into gaming consoles, servers, cloud, networking, autos, medical devices, broadband equipment, industrial systems and low-cost education/embedded devices. The issue is not only higher DRAM/NAND pricing, but also allocation priority: suppliers are prioritizing HBM, server DRAM and enterprise SSD demand tied to AI infrastructure, while conventional consumer and industrial memory users are increasingly exposed to higher prices, longer lead times and weaker bargaining power.

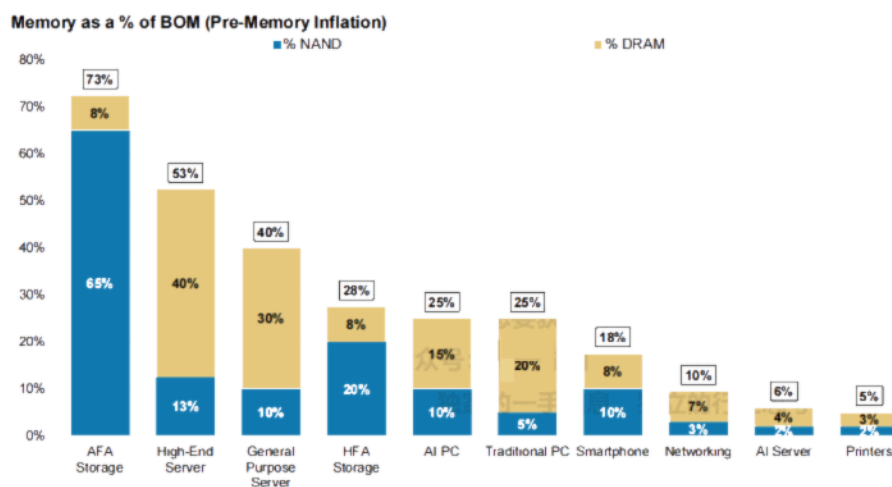
On a gross cost basis, servers and cloud and PCs are among the most exposed. But on net margin impact after passthrough, the most impacted sectors are different: midrange and low-end PCs, low-end smartphones, consumer electronics, gaming consoles, embedded and education devices, and smaller industrial OEMs ([Exhibit 36](#)). While it is difficult to predict whether historical demand elasticity will repeat this cycle given such significant price hikes, our quant analysis would paint a similar picture: low/mid-range PCs and smartphones face relatively elastic demand responses and thus face more significant net margin risk than enterprise infrastructure, where demand is much more inelastic with relatively more margin protection ([Exhibit 38](#)). These more consumer-led end markets combine higher memory BOM exposure with weaker pricing power and higher demand elasticity.

Exhibit 36: Consumer electronics companies continue to call out margin pressures relating to memory cost inflation

Consumer Electronics Commentary on Memory Cost Inflation Pressures		
Source	Commentary	Date
HP Inc	"Looking ahead, we expect the memory and storage environment to remain constrained. In addition, we also anticipate broader inflationary pressures beyond memory and storage."	27-May-26
Xiaomi	"I think given the cost increase, pressure will continue . That is an objective fact. So we need to solve this problem. We need to balance scale, profit and also gross margin... We cannot simply pass the cost increase to consumers . Given the new cost structure, we need to make sure that users are happy with the value for money."	26-May-26
ASUSTek Computer	"Due to the AI server boom, supply for both memory and CPUs have become very tight. And furthermore, we're seeing that memory shortages, in particular, and those price hikes have been extreme for the past few quarters ... So I think that it's important to note that both memory and CPU are undergoing both supply shortages and price hikes. "	12-May-26
GoPro	"Macro challenges in the consumer electronics sector, including rising memory costs , supply chain constraints and fluctuating tariffs prompted us to take some discrete actions in the quarter that impacted gross margins and earnings per share for the period. "	11-May-26
Sony	"Of course, the memory prices going up would increase the cost of the BOM . So the cost of manufacturing will go up. And if that leads to passing on cost to prices, there would be a big impact on the gaming console prices. "	6-May-26
Fractal Gaming Group	"We are seeing a significant market decline across regions... driven by a global imbalance in the memory market, where increased investments in AI and data centers have redirected capacity from the end consumer market and thereby driven up prices for DRAM and NAND memory. This has significantly increased the cost of building a PC with memory now accounting for a much larger share of the total system cost , at times even exceeding the CPU. As a result, we are seeing more cautious consumer behavior with many customers delaying or scaling back upgrades, which is directly impacting demand across the do-it-yourself and broader consumer electronics market. "	7-May-26
Logitech	"We are working to make sure that we get as much memory as possible to protect our video conference portfolio. So whatever we can get, we get it, and that may impact the inventory turns. So don't expect— don't model greater than 100% [cash conversion] every quarter in fiscal year '27. "	5-May-26
Sonos	"Looking to the second half and beyond, we're managing the headwind of higher memory costs, which are putting downward pressure on our gross margin . As you know, the semiconductor industry is in the middle of a transition from DDR4 to DDR5 and high-bandwidth memory, driven by AI and data center demand. That is tightening the supply for the DDR4 chips we use and increasing costs across consumer electronics. "	4-May-26
Apple	"We said it would be a bit more in the March quarter, and we did see higher memory costs in the March quarter , and they were partially offset by benefits from carry-in inventory that we had. For the June quarter and what's embedded in the guidance that Kevin went through earlier, we expect significantly higher memory costs . They are also partly offset by the benefit of carry-in inventory. And then where we don't give color beyond June, I can tell you that beyond the June quarter, we believe memory costs will drive an increasing impact on our business. "	30-Apr-26
Samsung Electronics	"With the growing demand for AI server memory, memory supply shortages for mobile and upward trend in prices have persisted. In 1Q 2026, memory prices surged, weakening profitability year-on-year. Also in 2Q, prices are expected to rise further, adding to cost pressures. We will leverage our stable supply to expand sales of S26 and new A-Series. At the same time, across development, procurement and sales, we will enhance cost efficiency to mitigate the impact of rising memory prices on profitability."	30-Apr-26
LG Electronics	"For monitors, apart from certain smart monitor models, the impact from memory-driven price increases is minimal. For PCs, which have relatively high memory content, the industry is facing significant cost pressure. As a result, price increases of approximately 15% to 20% have already been implemented. Should the sharp rise in memory prices persist, additional price adjustments may become unavoidable. "	29-Apr-26
Best Buy	"For next year, our guide for gross profit is about 30 basis points increase year-over-year... there could be some categories, some pressure on margins because of memory costs ... there could be unique areas within computing that might have some impact. "	3-Mar-26

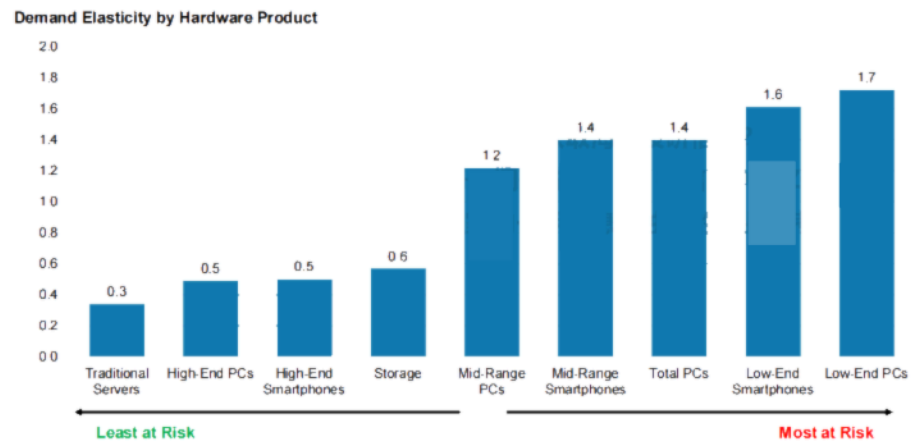
Source: Company Data / Morgan Stanley Research (emphasis added)

Exhibit 37: We estimate that memory accounts for 5-70% of tech hardware products, with servers and PCs more exposed to DRAM while storage is over-indexed to NAND



Source: Company Data / Morgan Stanley Research estimates

Exhibit 38: Our quant analysis implies low-end smartphones and PCs are the most demand elastic, while servers, storage and high-end smartphones are least at risk of demand destruction from higher prices



Source: IDC, Company data, AlphaWise, Morgan Stanley Research. Demand elasticity takes the absolute value. PC data goes back to 1995, for storage it starts in 2008, servers in 2003 and smartphones in 2007.

Accounting asymmetry matters

A less obvious transmission mechanism for chipflation is accounting treatment, disproportionately impacting companies that are not directly monetizing AI but still have to pay AI-driven memory prices. This creates an accounting asymmetry: hyperscalers can capitalize memory-heavy AI servers and depreciate them over time, while consumer hardware and industrial OEMs see memory inflation flow more directly through inventory and COGS.

The same memory price shock can therefore look like strategic capex for one buyer and immediate gross-margin pressure for another. This is one of the most under-discussed transmission mechanisms of chipflation in the current cycle.

The capability cost: what doesn't get built

Memory inflation is not just a financial transfer from non-AI sectors to memory producers. It is increasingly a constraint on what gets built and when.

For automakers, memory costs are pushing back ADAS rollouts and compressing EV program economics. GM has raised its 2026 commodity and logistics inflation guide by \$1bn-2bn, citing higher DRAM costs among the drivers.

- Magna expects more pricing tension at the vehicle level if DRAM costs remain elevated.
- Visteon's memory supply is constrained through 2027, and the company is qualifying emerging suppliers for 2026+ demand.

For healthcare and medical devices, the impact is more diffused but the challenge is real.

- GE HealthCare flagged ~\$100m of a \$250m inflation headwind tied to chip and cost actions.
- Siemens Healthineers said the EPS impact could reach €0.05 due to supply-chain

inflation including memory chips.

- Intuitive Surgical noted gross-margin pressure including roughly 100bps of impact from tariffs, freight, and semiconductor memory costs.

For the indirect channel – the largest and least visible – every enterprise running workloads on AWS, Azure or GCP is exposed to chipflation through its cloud bill. A bank running fraud detection, a logistics firm running route optimization, a retailer running inventory analytics: all are paying memory inflation indirectly without recognizing it as a memory cost. This is the channel with the broadest economic footprint, and the channel that is hardest to see in any individual company's filings.

Exhibit 39: Memory impact heat-map by end market

End market	Memory availability / allocation priority	Cost impact	Pass-through ability	Demand destruction / pull-forward risk	Major company commentary / impact one-liners
Hyperscalers / AI data centers	Highest priority — suppliers are optimizing for AI/HBM, server DRAM, enterprise SSDs	Very high, but strategic	High	Low	Meta: raised 2026 capex to \$125–145bn, partly due to higher component costs, especially memory. Amazon: said memory/storage component costs have “skyrocketed,” but large cloud buyers are better positioned to secure supply.
Enterprise servers / storage	High priority — large server OEMs get allocation, but at sharply higher prices	Very high	High	Low-medium	HPE: DRAM/NAND are now over half of traditional server BOM, with elevated pricing expected into 2027; using shorter quote cycles, repricing rights, LTAs and alternative configs. Dell: using shorter quote validity and dynamic pricing; flagged some pull-forward risk.
Semiconductor equipment	Medium — supply likely manageable, but not immune to electronics BOM inflation	Medium	Medium-high	Low	KLA: expects elevated memory pricing through at least calendar 2026 and about 100 bps negative gross-margin impact over the next several quarters.
Networking / security appliances	Medium-high for large vendors; weaker for smaller appliance makers	Medium-high	Medium-high	Low-medium	Cisco: immediate price increases across memory-heavy compute/networking products. Check Point: enough inventory for H1 but H2 margin headwind. Radware: 5–8% list-price increases on memory-intensive hardware; no buying-behavior change seen yet. HP: memory/storage moved to ~35% of PC BOM from 15–18%, with LTAs, supplier diversification and price increases.
PCs / notebooks	Medium — tier-1 OEMs have some allocation, but PC DRAM is not top priority versus AI/server	High	Medium-high	High in entry/mainstream PCs	Lenovo: said DRAM rose 40–50% in one quarter and almost doubled again, calling the imbalance structural. Acer/ASUS: passing higher memory costs into PC pricing; potential spec cuts such as 16GB to 8GB if pressure persists. LG PC: already seeing 15–20% price increases. Apple: memory becomes a more visible gross-margin headwind from the June quarter and increases beyond that.
Smartphones	Medium — Apple/Samsung can secure supply, but low-end Android is lower priority	Medium-high	Medium at premium; low at low end	High at low end; lower at premium end	Xiaomi: secured 2026 memory supply but warned shortages could push smartphone prices higher. Samsung: memory division benefits from higher ASPs, while handset-side memory cost pressure is less separately quantified in official commentary.
Gaming consoles	Medium — large platform owners can secure supply, but consoles compete with higher-margin AI/server demand	High	Medium	Medium-high	Nintendo: built around ¥100bn of FY27 COGS impact from rising component prices, especially memory, plus tariffs; also raised Switch 2 prices in Japan, the U.S. and Europe. Sony: secured minimum memory supply for the next holiday season, but did not fully quantify the cost impact.
TVs / monitors / home electronics	Medium-low — lower memory content, less priority than servers/PCs/phones	Low-medium	Medium	Low-medium	LG Electronics: TV memory content is relatively low, monitor impact is limited except smart monitors, but PCs are meaningfully exposed; mitigation includes supply MOUs, multi-sourcing, spec optimization and premium mix. Canon: mostly secured memory supply but faced an estimated ~¥50bn negative memory-cost impact and guidance pressure.
Cameras / imaging	Medium-low — discretionary hardware, less supplier priority than AI/server/PC/phone	Medium-high	Low-medium	Medium-high	Nikon: secured volumes, but DRAM costs rose severalfold and could create a high-single-digit-billion-yen operating profit hit. Fujifilm: flagged semiconductor memory and silver as cost pressures, with pricing and offsets planned.
Autos / auto electronics	Medium — automotive volumes are smaller, specs are specialized, and 2027 supply security is a focus	Medium	Medium, but lagged	Low near term; medium if prolonged	GM: raised commodity/logistics inflation guide, including higher DRAM costs, to \$1.5–2.0bn. Ford: expects sufficient DRAM supply but sees pricing pressure that may affect vehicle pricing. Magna: more a price issue than availability issue; annual DRAM buy under \$100m. Visteon: memory remains constrained through 2027; qualifying emerging suppliers for ~10% of 2026 demand and negotiating customer recoveries. Renesas: automotive DRAM volumes not large, but DDR4/DDR5 compatibility matters.
Healthcare / medtech	Medium — mission-critical products and larger OEM scale help, but memory is still a cost shock	Medium	Medium-high, but lagged	Low	GE HealthCare: memory chips were about \$100m of a \$250m inflation headwind, partly offset by price and cost actions. Siemens Healthineers: cut EPS midpoint by around €0.05 due to supply-chain inflation including memory chips. Intuitive Surgical: gross-margin guide includes about 100 bps of impact from tariffs, freight and semiconductor memory costs. AeroVironment: disclosed lead-time and price uncertainty for memory-related microprocessors.
Defense / aerospace	Low-medium — lower-volume, long-life, rugged/legacy memory can be vulnerable	Medium-high where DDR4/legacy-heavy	Medium-high structurally, but lagged	Low	JLT Mobile: DDR4 prices up more than 10x and lead times longer; secured 2026 supply through early purchases. Nexteq / Ependion: DDR4/memory availability is a key 2026 risk, mitigated by LTAs and price increases.

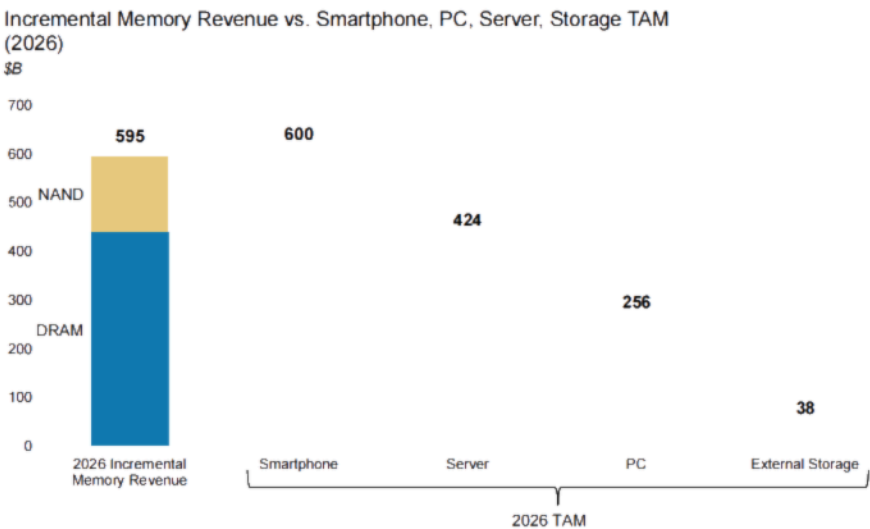
Source: Company data; Morgan Stanley Research

The chipflation impact in dollars

Memory cost inflation has become large enough to matter at the industry level. While it varies greatly by product, nearly every hardware product – a PC, server, storage array, smartphone, etc. – has some combination of NAND and DRAM content, which we estimate makes up anywhere between 5% of a device bill of materials (printers) to 70% (high-end servers). Simplistically, higher-end, more performance-based products such as All-Flash storage arrays, high-end & traditional servers, higher-end PCs, and AI servers all have 25-70%+ NAND/DRAM mix within their product bill of materials. Conversely, products such as printers or networking hardware are generally lighter on memory content. Nevertheless, this means that most hardware OEMs and ODMs are over-exposed to memory markets, especially those with high-end device exposure ([Exhibit 37](#)).

The memory market is expected to add more Y/Y revenue in 2026 (~\$600B) than the entire standalone smartphone, PC, or server TAMs. To put the overall impact of memory inflation into perspective, in just the last 3 months, expectations for aggregate memory market revenue in 2026 have increased by 71%, from \$520B to \$890B, relative to a \$220B market in 2025. Not only are memory buyers facing an incremental \$300B of standalone memory costs this year – equivalent to the annual GDP of Finland or New Zealand – but the memory market is expected to add more Y/Y revenue in 2026 (~\$600B) than the entire standalone smartphone, PC, or server TAMs ([Exhibit 40](#)). This will show up in 2026 as COGS or Capex for memory buyers. In aggregate, the global hardware industry must now contend with incremental memory costs that equate to over half of the \$1.3T aggregate hardware TAM, with our estimation at the industry level suggesting the highest dollar burden is on Servers at \$246B, followed by Smartphones at \$175B, PCs at \$132B, and External Storage at \$13B.

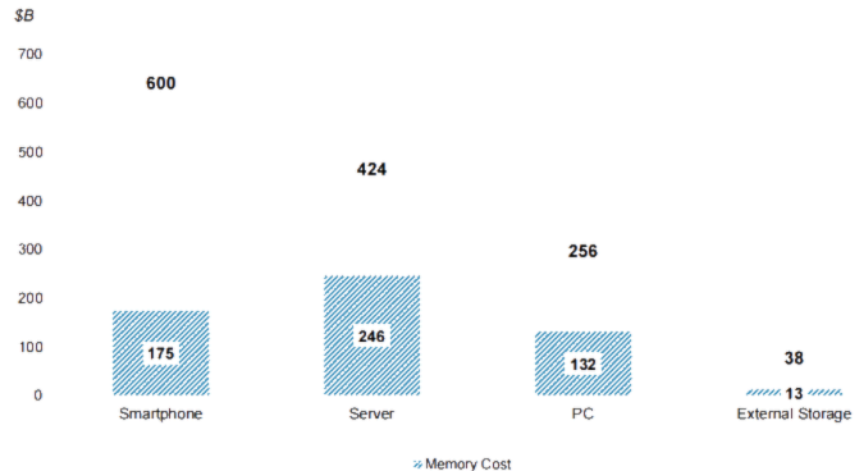
Exhibit 40: Based on latest industry estimates, memory revenue will increase by \$600B in 2026 ...



Source: Trendforce estimates, IDC data, Morgan Stanley Research

Exhibit 41: ... which we estimate translates into ~\$600B of incremental memory cost across Smartphone, Server, PC, and External Storage markets

Incremental Memory Cost & TAM by Market (2026)



Source: TrendForce, IDC, Morgan Stanley Research estimates

This does not mean that every hardware company faces a direct margin hit of the same magnitude. The chipflation burden will be distributed across hyperscalers, OEMs, component suppliers, distributors, enterprises and end-customers. But the order of magnitude matters: the memory supercycle is large enough to force some combination of price increases, margin pressure, specification cuts, product delays, budget reallocation and demand destruction. For hyperscalers, higher memory costs is being absorbed as strategic AI capex. For many non-AI hardware buyers, it is showing up as higher prices, COGS pressure, lower product specs, and/or delayed product launches/refresh cycles.

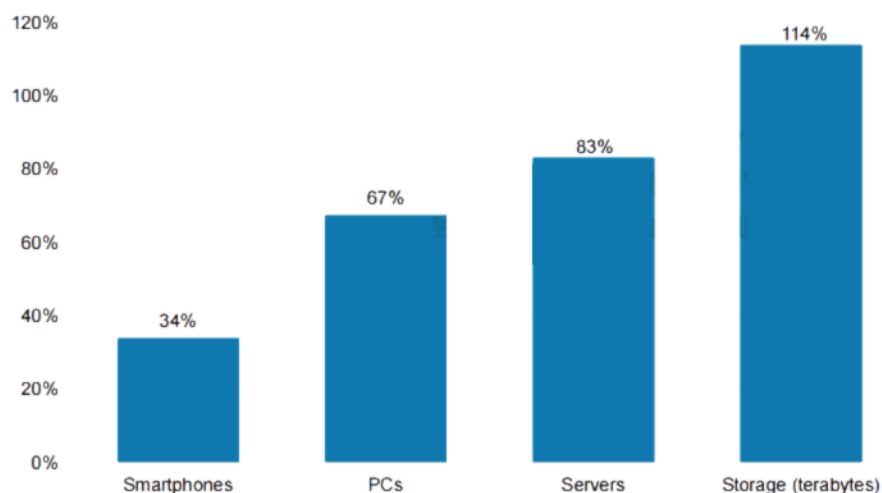
And in non-AI hardware markets scale matters. A vendor's ability to aggregate demand across multiple products (i.e. PC, server, storage) or come to market with exceptionally strong buying power (Apple, Samsung) makes a meaningful difference in getting memory allocations.

But even sufficient memory allocations come with a significant cost – higher prices across the hardware ecosystem. Using the latest TrendForce memory price forecasts, and making the assumption that hardware vendors will prioritize maintaining margin rates, we estimate like-for-like unit price hikes would have to be ~35% for Smartphones, ~70%+ for PCs, 80%+ for Servers and 110%+ for Storage arrays. This is the average though. Anecdotally, we are already hearing instances of significant price hikes – 100%+ in some cases in high-end Storage/Server markets, due to memory plus other semiconductor inflation (CPU, PCB, MLCC, etc.).

For context, these required pricing increases sit significantly above historical ASP behavior across virtually every major hardware category. PCs are the clearest example: over the past ~30 years, annual PC ASP growth has typically fluctuated within a +/-10% band, with even the strongest inflationary periods – including COVID-era supply constraints – driving aggregate pricing increases in the high-single digits Y/Y. In fact, if PC prices increase 67% Y/Y this year, it would represent the strongest Y/Y ASP growth in PC market history by a magnitude of 8-9x ([Exhibit 43](#)).

Exhibit 42: We estimate that in order to protect gross margins, hardware vendors would need to raise prices by 35-115% Y/Y in CY26

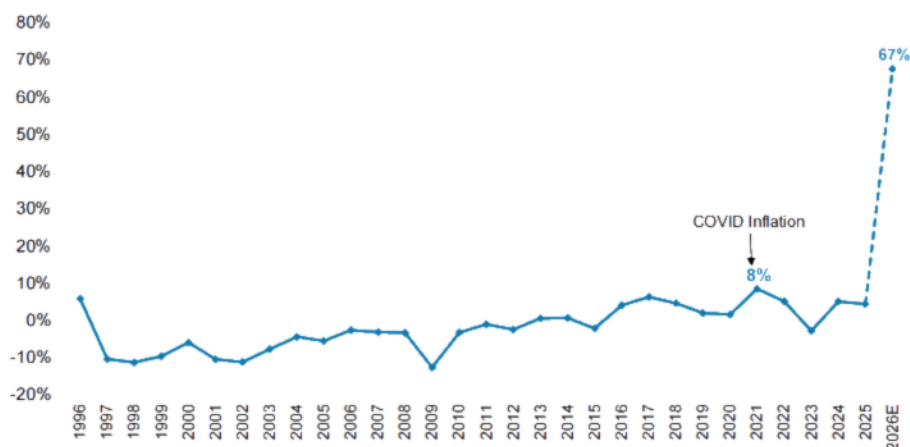
Y/Y ASP Growth Needed to Hold Gross Margins Steady



Source: TrendForce, IDC, Company Data, Morgan Stanley Research estimates

Exhibit 43: For context, if PC ASPs were to rise 67% Y/Y, it would mark the strongest Y/Y ASP growth in PC market history, far surpassing COVID-era inflation

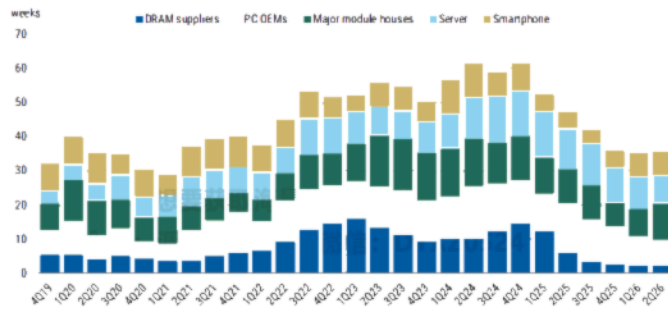
PC ASP Growth Y/Y



Source: IDC, Morgan Stanley Research

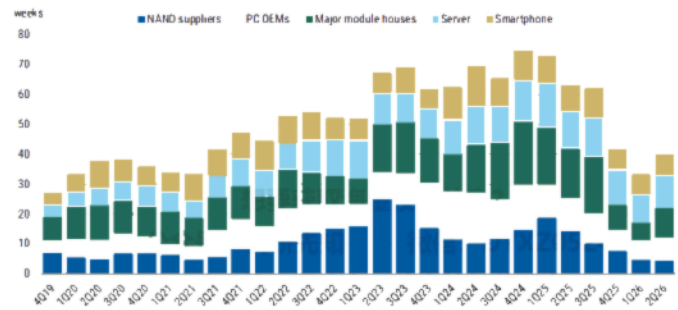
The timing of the impact is likely lagged. Many hardware companies are still working through lower-cost inventory, which can mask the near-term earnings impact. This explains why some memory-exposed companies can still beat current-quarter expectations while investors look ahead to margin risk in calendar 2H26 and beyond. However, memory inventory has depleted quickly across channels. As higher-priced memory flows through inventory and COGS, companies will need to choose whether to raise prices, cut specifications, delay launches or absorb margin pressure.

Exhibit 44: DRAM inventory



Source: TrendForce estimates, Morgan Stanley Research

Exhibit 45: NAND inventory



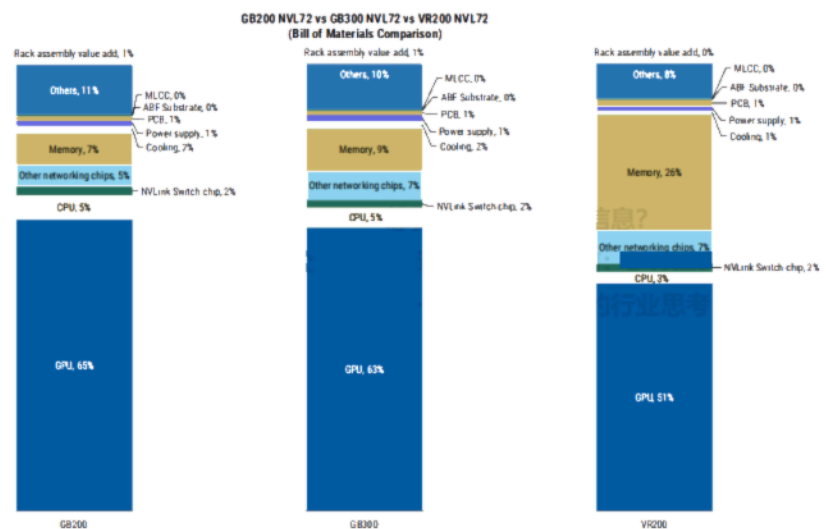
Source: TrendForce estimates, Morgan Stanley Research

Memory vs GPU cost convergence

Memory is now 25%+ of a Rubin rack BOM. Our bottom-up Rubin rack BOM analysis shows how large the memory cost has become inside AI infrastructure. Memory was only 5-10% of the GB200 NVL72 rack BOM, but with higher memory content and materially higher pricing, memory rises to roughly 25-30% of the VR200 / Rubin rack BOM. In the detailed BOM comparison, memory increases from \$374k per GB300 rack to \$2.0mn per VR200 rack, or +435%, making it the largest source of non-GPU cost inflation.

This is the clearest dollar-based evidence of chipflation. For hyperscalers, memory inflation becomes part of strategic AI capex and can be capitalized over time. For non-AI buyers, the same supply shock shows up more directly through COGS, gross margins, product pricing and availability. The issue is therefore not only that memory prices are rising, but that AI infrastructure is absorbing a much larger share of total component spend and supplier allocation.

Exhibit 46: Owing to recent increases in memory prices, memory will become 25%+ of the rack BOM for Rubin



Source: Morgan Stanley Research estimates

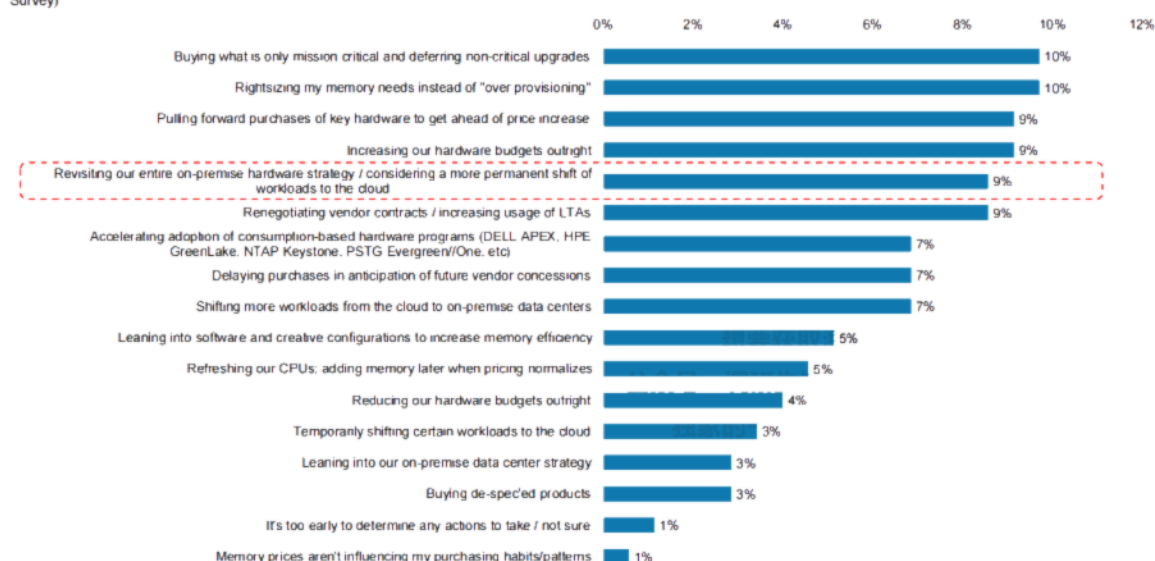
Identifying the cyclical and secular risks of chipflation

The cyclical risks. Prices across nearly every hardware end-market will move higher throughout 2026, and devices are already being de-sped'd. Despite these efforts, hardware margins are likely to face pressure in 2H26 and beyond, with the impact likely greatest in the markets where demand is most elastic: PCs, smartphones, and lower-end consumer electronics. The biggest cyclical variable remains the demand elasticity response, especially once accelerated purchasing (i.e. pull-forward) ceases. In PCs, this already appears to be happening, with our expectations for double-digit Y/Y unit declines in 2H26. For Servers and Storage, demand is more resilient due to workload growth and growing AI use cases, with industry analysts still expecting unit growth throughout 2026 and into 2027, despite record price hikes.

The bigger question in hardware markets remains: what is the secular risk of memory inflation? In consumer markets, we could see a forced re-design of products, the elimination of lower-end products, and industry consolidation as Tier 2 and 3 vendors struggle to survive a prolonged memory supercycle. In enterprise markets, we are seeing early evidence that chipflation could accelerate workload shift to the cloud, which is a product of accounting asymmetry – hardware prices are rising more rapidly than cloud CPU instance prices. It is too early to assess all the possible impacts, but it is increasingly clear to us that there will be long-term ramifications of chipflation across hardware markets.

Exhibit 47: 9% of CIOs say they are revisiting their on-prem hardware strategy and considering a more permanent shift of workloads to the cloud given the risk of significant price increases caused by memory price inflation.

Actions Taken Given the Risk of Significant Price Increases Caused by Memory Price Inflation and Risk of Insufficient Memory Supply (1Q26 Survey)



Source: AlphaWise #61 (US and EU data), Morgan Stanley Research

Exhibit 48: A combination of our own checks and public commentary from management teams suggests some organizations are migrating to the cloud due to increasing memory costs

Commentary on Potential Migration to Cloud Due to Memory Cost Inflation		
Source	Commentary	Date
Radware	"There is a question that we don't know the answer yet. Will it push customers to shift from on-prem to cloud?"	14-May-26
VAR Checks	"The part that's different is that companies that have traditionally said no to cloud are maybe more interested in taking a look and understanding the price delta."	11-May-26
Cloudflare	"I think that as you see that there are vulnerabilities in hardware... as you see, again, supply chain shortages, especially around memory right now, I do think all of those things are pushing more and more people to evaluate that they need at least some part of the cloud as part of their infrastructure."	7-May-26
Ingram Micro	"We've been through shortages. We've been through macroeconomic headwinds. A couple of thoughts here ... We are starting to see some movement to from on-prem solutions to actually cloud. And we're starting to see that, and we expect that to happen going into Q2 too."	30-Apr-26
Amazon	"One of the interesting things that we see right now with the change in price and in supply on things like memory is that it is a further impetus pushing companies who have on-premises infrastructure into the cloud."	29-Apr-26
Morgan Stanley Procurement	"In terms of how we're reacting ... Next I would say is the consideration of offloading things to cloud."	10-Apr-26

Source: Company data; Morgan Stanley Research



Chips, Growth and Inflation – A Macro Perspective

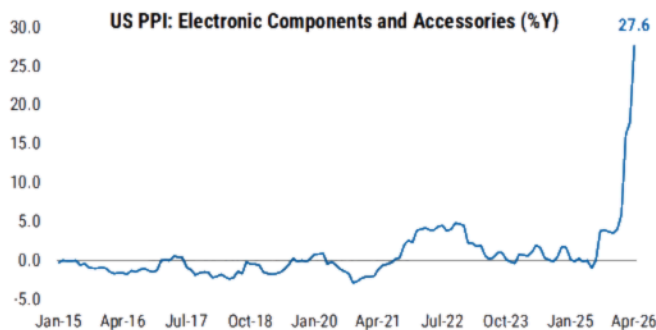
Diego Anzotegui, Rajeev Sibal

Chip Inflation is already visible in producer price inflation (PPI) segments tied to computing and electronics. The effect on CPI is not as strong, perhaps suggesting difficult margin decisions being made by consumer-facing producers. We expect that pressure to continue to be visible in the data as the upswing in electronics components inflation has yet to abate. The translation of the inflationary forces to headline macro data, however, will be more muted because chips account for less than 1% of both PPI and CPI indices. Goods are less than one-third of CPI indices and goods inflation is currently heavily distorted by the fading effects of tariffs ([see discussion here](#)).

Memory Costs in Producer Price Inflation (PPI)

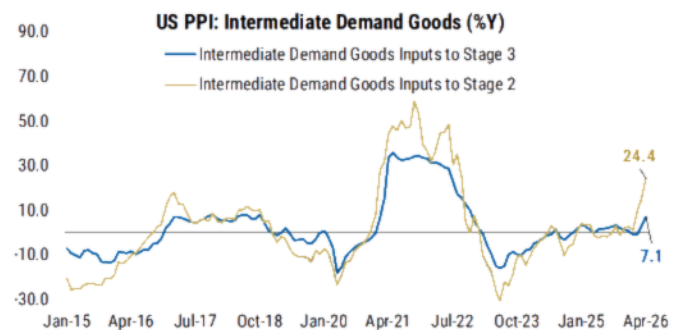
The effects of memory costs on inflation are most visible in PPI subcomponents linked to electronics and related equipment. Across the various categories with memory cost risks, the most notable is the electronic components and accessories sub-category moving near ~30% y/y without a clear top as of yet ([Exhibit 49](#)). The rise in electronics PPI is also spreading across the supply chain, specifically in Stage 2 and Stage 3 intermediate goods where the headline indices are affected ([Exhibit 50](#)). Stages 2 and 3 of the PPI measures are arguably best placed to capture the rise because of the effect in the supply chain as it moves to the consumer. The contribution of electronics and components is less than ~1% of these measures, but we can see the effect. The inflationary push reduces at the headline PPI level, however, because the goods contribution within PPI is far smaller than that of services.

Exhibit 49: Sharp rise in PPI electronics ...



Source: BLS, Morgan Stanley Research

Exhibit 50: ... boosting upstream input costs



Source: BLS, Morgan Stanley Research

The translation from PPI to CPI is important because it gives an indication of the extent of margin pressure for producers. The equivalent products in CPI are not showing as strong signs of inflation, with materially lower inflation readings (~5% y/y in CPI measures of computers relative to the ~30% in components). The implication of course is that margin pressures will force difficult decisions in the supply chain. While we expect further pressure in consumer indices, the level seems to be structurally lower than what we are seeing in PPI because of a relatively low chips share in the overall cost structure and given how much consumers can absorb.

We do, however, expect the rise in PPI to have material downstream effects in US supply chains, which could have a read-across to growth. Given the importance of AI capex to US growth ([see discussion here](#)), it is important to monitor the continued translation from PPI to CPI and to investment sentiment. One of the key drivers of the bull case for investment in the US is continued AI capital expenditure. So far AI capex has been orthogonal to macroeconomic conditions, but if it slows because chip costs become too high, it has the double impact of potentially slowing investment as well as delaying the translation to AI productivity in the US. The contours of how the costs are absorbed will have important ramifications for the bullish narrative in US investment.

Memory costs in Consumer Price Inflation (CPI)

We think that the impact on CPI will be more micro than macro – visible in individual categories, but with a modest effect on the overall index, in the order of ~10bps to headline inflation in 2026. This is because memory-sensitive goods carry a relatively small weight in the CPI basket (see [Exhibit 51](#)).

Exhibit 51: CPI impact of higher memory costs

CPI component	Memory costs effect on CPI in 2026 (pp)	CPI weight
PCs	15	0.30%
Smartphones	15	0.20%
TVs	10	0.10%
Cars	0	3.80%
Major household appliances	5	0.06%
Game consoles	125	0.01%
Impact on headline CPI: PCs and smartphones only	0.08	
Total impact on headline CPI	0.10	

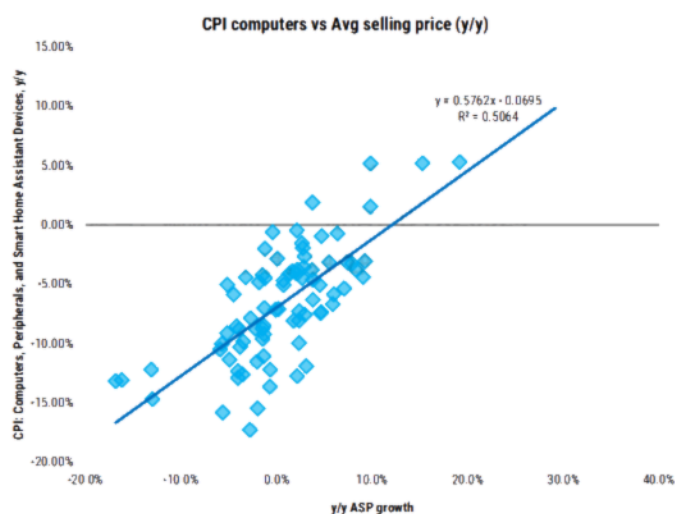
Source: BLS, Morgan Stanley Research forecasts. Note: Smartphone and game console weights are estimates.

That said, the transmission is likely to be uneven across categories. PCs and smartphones are the most important components to monitor, given their relatively high memory intensity in the bill of materials and their comparatively larger weights within memory-sensitive core goods CPI. Industry analysts suggest that average selling prices (ASPs) for PCs could rise by 20% this year, with smartphone prices increasing by around 15% ([see here](#)). However, the sharp rise in memory costs in recent months might imply upside risks to those estimates.

The relationship between ASPs and CPI is not one-to-one. CPI measures for these goods are quality-adjusted, whereas ASP series are not. And this adjustment seems to make CPI series smoother compared to ASPs. Historical experience illustrates this point. For computers, a 30% y/y increase in ASPs is broadly consistent with roughly a 10% y/y increase in CPI prices ([see Exhibit 52](#)), implying a significantly lower magnitude to measured inflation. The relationship is even weaker for smartphones ([see Exhibit 53](#)), suggesting downside risk to CPI inflation relative to ASP-based estimates.

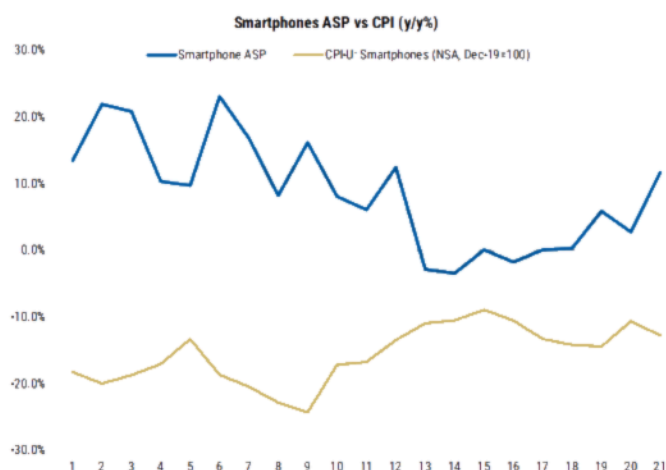
Taking into account the link between ASP and CPI measures and possible upside risks to current analyst estimates, we forecast CPI computers at 10% y/y in 2026 (consistent with 30% y/y ASP) – implying an acceleration of 15pp vs 2025. The outlook for CPI smartphones is less certain, but conservatively we assume an acceleration similar to PCs for this category.

Exhibit 52: The link between ASP and CPI computers is not 1-to-1 ...



Source: BLS, Morgan Stanley Research

Exhibit 53: ... and the relationship for smartphones is even weaker



Source: R/S, Morgan Stanley Research

For other goods categories, the impact is more limited. TVs and household appliances have lower memory content in their cost structure, implying more muted price pressure. By contrast, game consoles are highly exposed, with memory accounting for roughly 30% of the bill of materials. In a full passthrough scenario, price increases could be substantial – potentially exceeding 100%. However, the very small weight of game consoles in the CPI basket means that their contribution to aggregate inflation remains negligible.

For autos, the impact is smaller still. Our analysts estimate that higher memory costs are raising production costs by [roughly 1–3% this year](#). However, OEM pricing guidance has remained broadly unchanged, suggesting that these costs are being absorbed in margins rather than passed through to consumers.

US/China Policy Options Don't Offer Near-term Relief

Ariana Salvatore, Shawn Kim

Key Takeaways

- Even if the US and China make policy choices aimed at alleviating price pressure, actually doing so will take years, for three key reasons:
 - US policy efforts (tax credits, subsidies, expedited permitting) come with constraints and lags
 - China's currently available incremental supply won't close the gap with demand
 - China's policy incentives to boost supply would take a couple years to show results
- Moreover, our expectation is that US policy choices are actually skewed toward a more, not less, restrictive direction: we expect little in terms of export control removal or loosening of the current regime.

US Could Loosen Demand Restrictions, But Supply Wouldn't Be Enough to Alleviate Strain

Given the focus on memory as a strategic resource, we expect policy discussions in particular on HBM and advanced DRAM to focus on global supply chain vulnerabilities as well as supplementing domestic availability. To that end, we expect these discussions [to persist](#) regardless of the outcome of the 2026 midterm elections and the configuration of power in Washington.

The key policy distinction is between AI-strategic memory – primarily HBM and advanced DRAM that directly enable frontier AI systems – and commodity or legacy memory used in automotive, industrial, consumer and other non-frontier applications. The starting point is concentration: advanced DRAM and HBM are dominated by Samsung, SK hynix and Micron, while NAND supply is broader but still concentrated. China is becoming more relevant through CXMT in DRAM and YMTC in NAND, adding a new policy dimension to an already concentrated market.

To that end, policy efforts likely distinguish between HBM and legacy (non-frontier) memory, but even the latter would be limited in scope and effectiveness. We expect policy efforts in HBM and advanced DRAM to prioritize protecting AI-strategic memory through supply-chain resilience, trusted capacity, and geopolitical de-risking over nearer-term cost pressures (i.e. not alleviating existing export controls). For less sophisticated (non-frontier) memory end uses (automotive, consumer, etc.) policymakers may seek to distinguish commodity or legacy memory from AI-strategic memory through differentiated licensing and targeted supply-side support for trusted capacity in the former category

(more options detailed below). Importantly, political obstacles, timing lags, and associated supply chain constraints still likely dilute these efforts' effectiveness. Forced allocation would likely be a last resort, in our view. It might protect favored sectors in the short term, but would risk turning memory into a rationed geopolitical commodity. **Across the board, the policy options may mitigate, but ultimately not solve, the problem.** Specifically, loosening demand restrictions could allow more Chinese supply into the US market, but we don't think the supply would be enough to alleviate near-term price pressures (as we [detail below](#)).

Policy Options: federal support is possible ...

We often frame policy options in the context of supply chain resiliency as either 'defensive' or 'offensive' tactics. 'Defensive' refers to the more traditional trade restrictions (i.e. tariffs, non-tariff barriers, licensing requirements, export controls, etc.) while 'offensive' refers to domestic investment initiatives (direct federal spending, loans, grants, tax incentives, etc.) to on-shore supply chains. In the memory context, we view the most likely options as supply-side measures aimed at expanding trusted HBM, advanced DRAM, packaging and test capacity.

In that context, we see the below policy options as the ones most likely to be discussed to mitigate (although ultimately not solve) the problem:

- Direct subsidies can support new fabs, packaging and test capacity, but may not lower prices quickly.
- Tax credits can encourage investment but are less targeted.
- Procurement guarantees can de-risk capacity additions, but may distort allocation.
- Expedited permitting can improve time-to-capacity, but execution is difficult.
- Equipment-access coordination can support allied capacity expansion, but must be balanced against export-control objectives.
- Strategic stockpiles may help critical sectors, but are hard to manage in fast-moving technology markets.

Subsidy accountability will become more important. If memory producers receive public support while enjoying high pricing and strong margins, policymakers may ask whether subsidies are translating into actual capacity expansion, broader access, domestic resilience or lower downstream costs. Forced allocation would likely be a last resort. It might protect favored sectors in the short term, but would risk turning memory into a rationed geopolitical commodity.

More export controls on top of the existing regime look less likely than supply-side initiatives in the near term. Aside from the geopolitical risks associated with a loosening export control regime, this type of policy cannot relieve the bottleneck or facilitate additional supply if not paired with domestic supply-side initiatives. To put a finer point on it, four reasons underpin our view: 1) the existing regime already places HBM as an explicit advanced-computing control item under the EAR, albeit with structural gaps; 2) allied coordination is likely difficult to implement and enforce; 3) directionally, the environment appears to be moving away from broader-diffusion style controls and more toward targeted controls, in particular for more advanced tech; and 4) the likely implication of further controls is more stockpiling and limited global availability, which could drive up costs versus alleviating the bottleneck.

The existing export controls already placed HBM into the advanced-computing regime, but with certain nuances, as mentioned above: the administration maintains HBM inside the AI export control umbrella, but gaps exist given the geographical dispersion of other nodes of the supply chain (packaging, in particular). Hence, a more comprehensive export control regime would require coordination with allies – the bipartisan MATCH Act has been the subject of many questions from investors on this topic, but the road to passage is challenging, and we view it more as a control alignment tool than a means to expand capacity. Aside from the procedural constraints (i.e. needing a legislative vehicle), coordination with allies may be difficult to implement and enforce (in particular because the relevant chokepoints are spread across a variety of geographies, not just one country). Further, as we note above, denying or delaying access to HBM beyond the existing controls does not seem imminent given the administration's recent actions to facilitate more cross-border selling and the bilateral statements following the US-China summit meeting two weeks ago (see [here](#)). Restrictions of this type could also risk escalating tensions outside of the boundaries of 'tactical escalation' between the two countries (which is likely not in either sides' interest following last year's détente, as we note [here](#)), and could increase scarcity or accelerate stockpiling, in particular if not paired with parallel supply-side investment.

For HBM and advanced DRAM, policy is likely to combine de-risking through existing export controls and allied alignment with supply-side efforts to expand trusted capacity.

For commodity or legacy memory, we think policy is less likely to rely on restrictive controls and more likely to focus on differentiated licensing and targeted supply-side support, though timing lags, permitting, workforce constraints, tool availability, and qualification cycles will limit near-term impact.

Memory Category	Policy Objective	Policy Tools
HBM/Advanced DRAM	Protect strategic AI inputs & frontier capabilities; delay adversary access	Maintain de-risking (export controls) strategy; expand supply-side efforts to expand trusted capacity
Commodity/legacy memory	Avoid overly restrictive policy; facilitate trusted supply & alleviate cost pressures	Differentiated licensing regimes; targeted domestic/allied capacity support vis-à-vis supply-side initiatives; procurement for critical sectors

Source: Morgan Stanley Research

... but key constraints exist: procedure, timing & public perception

As a result, we view the most feasible near-term response as increasing HBM and packaging supply in the US and allied countries. The options listed above are all a means toward that end, but each comes with its own constraints. The principal constraint is time: building domestic memory capacity comes with long physical lead times (likely years). Aside from the bottlenecks such as permitting, site selection, etc., the HBM-related constraints also expand to other nodes of the supply chain. Advanced packaging needs mean that government incentives cannot focus on HBM capacity alone. Workforce is also a constraint: a skilled labor force is a key bottleneck as the US has less experience relative to other countries in building & maintaining leading-edge memory and packaging facilities

at the scale needed. **Hence, policy efforts here are more likely to be aimed at building more resilient supply chains from a geopolitical perspective for HBM and advanced DRAM (which takes time and is likely more expensive) versus solving a nearer-term availability or pricing issue.**

Against a backdrop of the public increasingly skeptical of AI (see [here](#)), we also take into consideration the public perception angle and its associated political hurdles. While we ultimately think that the bipartisan consensus in DC is toward accelerating the US position vis-à-vis China and therefore the policy options supportive of that ambition, there is still [public skepticism](#) toward AI and [industrial subsidies](#) more broadly that may make meaningful or deficit-additive policies challenging in the near term.

China Could Boost Supply, But That Could Take Years

Could China be the swing factor? In the consumer segment, integrating Chinese consumer-grade memory into PCs and smartphones could be a solution. We note [recent reports](#) that major PC manufacturers have contacted CXMT about potential DRAM collaboration, and this shift is already seen in retail where US module brands, such as Corsair Vengeance DDR5, are using CXMT DRAM chips. US technology companies can use Chinese memory capabilities but have been increasingly reluctant to do so, in light of geopolitical tensions and national security concerns. While the US government restricts the sale of advanced chip tech to China, importing standard Chinese memory chips for consumer electronics remains largely permissible. One of the factors preventing companies such as CXMT from supplying DRAM chips to mainstream manufacturers is the possibility of changes to US legislation. For example, a new proposed rule issued by the Federal Acquisition Regulatory Council (FAR) is intended to restrict the use of chips from CXMT, YMTC, and SMIC in commercial products.

China is emerging as a conventional-memory swing factor, but not a near-term HBM solution. China-located DRAM capacity rises from ~18% of industry capacity in 2023 to ~23% in 2028, on TrendForce estimates, while domestic Chinese suppliers rise from ~7% to ~17%. CXMT is the key driver, adding 388kwpm and accounting for roughly 30% of global net DRAM additions between 2023 and 2028. This could make China an important relief valve for commodity DRAM, especially for non-AI buyers facing tighter allocation from the Big 3.

However, the effective output of Chinese producers is significantly discounted by yield and technology gaps. On NAND, YMTC has crossed the 128-layer threshold and is broadly competitive on headline technology metrics, but manufacturability and enterprise-grade yield remain the key differentiators, with bits/wafer running approximately 20% below western peers. On DRAM, CXMT is currently on 12nm (~16nm), roughly three generations behind the state of the art, having transitioned from 17-19nm through 2025. While 30% of production is now on DDR5, clock speeds remain insufficient for mainstream PC, server, smartphone or AI workloads. The roadmap targets HBM3e in 2027 using 1nm DRAM – achievable without EUV access per the Micron precedent – but bits/wafer runs at approximately one-third of western levels, with the node disadvantage accounting for only around one-third of that gap. The remainder reflects structural manufacturing inefficiency, suggesting the HBM supply contribution from China should be modelled conservatively, if at all.

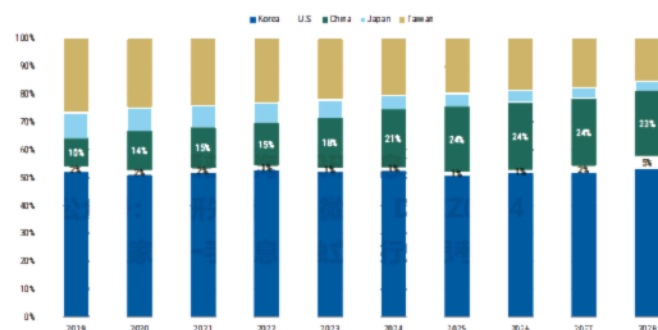
The regulatory backdrop reinforces these constraints. Both CXMT and YMTC remain on the Pentagon's Section 1260H "Chinese Military Companies" list as of today – CXMT was added in January 2025, YMTC in 2024. The list carries no statutory trade ban, but the FY2024 NDAA bars the DoD from contracting with listed entities directly from June 30, 2026. [Press reports](#) indicate that a February 13, 2026 update briefly showed both memory makers removed (while adding Alibaba, Baidu and BYD), but the Defense Department withdrew the document from the Federal Register within hours without explanation, leaving the January 2025 list operative. The episode suggests there could be internal debate over relief, though delivered none in practice. More bindingly, YMTC remains on the US Commerce Department's Entity List, where it has sat since late 2022, restricting US technology exports to the company. CXMT is not on the Entity List but operates under heightened scrutiny. Critically, neither designation prevents Chinese firms from selling memory into Western markets – the barriers to adoption are reputational and compliance-driven rather than a statutory import ban.

On the supply side, lithography access remains the hard ceiling. EUV tools have been off-limits to China since 2023 (ASML has never shipped an EUV system there), which is the primary reason CXMT's advanced-DRAM capacity appears to have plateaued. The proposed MATCH Act would tighten this further, naming CXMT, YMTC, SMIC, Hua Hong and Huawei as "covered facilities" and prohibiting not only DUV immersion tool sales but also the servicing of already-installed equipment – DUV fleets require regular maintenance to sustain yield. The bill remains in the legislative pipeline rather than enacted, but it underscores that the trajectory of policy is toward, not away from, constraining Chinese memory scaling.

China does not solve the strategic AI-memory (HBM) bottleneck. Our HBM model remains concentrated entirely in Samsung, SK hynix and Micron, with no modelled Chinese HBM contribution. That distinction matters for policy: Chinese DRAM/NAND capacity may help relieve commodity tightness, but it does not address the frontier HBM constraint tied to AI accelerators.

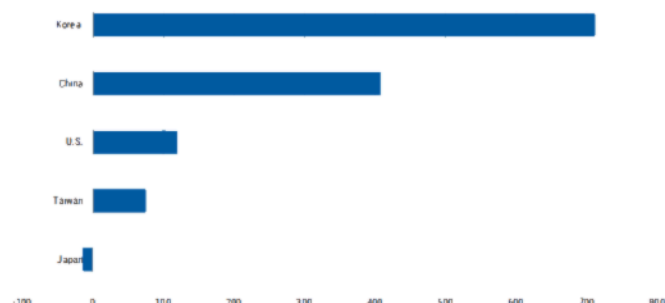
The key resilience question is not total global capacity, but trusted and accessible capacity. In our DRAM model, South Korea is the largest capacity pool and Mainland China is the second largest, followed by Taiwan, Japan and the US. South Korea plus Mainland China account for roughly 77% of 2028e DRAM wafer capacity. However, not all capacity is equally advanced, qualified or accessible under different geopolitical scenarios. Policymakers need to know where capacity sits, who controls it, what type of memory it can produce, and whether buyers can access it during a supply shock.

Exhibit 54: Capacity located in Mainland China rises from ~18% of industry capacity in 2023 to ~23% in 2028E



Source: TrendForce estimates, Morgan Stanley Research

Exhibit 55: Mainland China accounts for 30% share of 2023–28E net DRAM wafer additions, behind South Korea



Source: TrendForce estimates, Morgan Stanley Research

The potential for higher capacity in China

The equipment-access constraint now extends to the Big 3's own China fabs – making their upgrade path a key swing variable. Samsung and SK hynix previously operated their Chinese plants (Samsung's Xi'an NAND fab, SK hynix's Wuxi DRAM and Dalian NAND facilities) under Validated End User status, which allowed US-controlled tools without per-shipment licensing. That status expired on 31 December 2025 and has been replaced by an annual licensing regime requiring yearly US approval of equipment plans. The 2026 licenses are *understood* to cover maintenance and continued operation rather than capacity-expanding upgrades, effectively freezing this capacity at current nodes. Since these fabs supply a meaningful share of global mature-node DRAM and NAND, the question of whether upgrade licenses are eventually granted becomes a direct swing factor for global supply. Below, we size that potential upside on a fab-by-fab basis.

SK hynix Wuxi DRAM – migration scenario analysis

SK hynix has a DRAM fab in China (Wuxi C2) with 200k/wpm running mostly on 1a nm node. Against the backdrop of geopolitical tensions and the US semi equipment export restrictions, SK hynix has stopped the node migration process in its Wuxi fab, opting to let it run on current products until these become legacy products.

In a scenario in which management decides to migrate the Wuxi fabs to the advanced node, we would expect 1.5-3.4% incremental global DRAM supply from the Wuxi fab alone from purely brown-field node migration.

Exhibit 56: SK hynix Wuxi Fab node migration scenario – potential impact

Moving from 1a to...	Incremental bit addition per wafer (2Gb mn)	Annual bit output addition (2Gb mn)	As % of 2028 Global Supply
1beta nm	2.4	5,873	1.5%
1gamma/C	5.0	11,959	3.0%
1D nm	5.7	13,747	3.4%

Note: Underlying node migration and yield assumptions are from US WLE model. Source: Company data, Morgan Stanley Research estimates

Samsung Xi'an NAND – migration scenario analysis

Samsung has a Xi'an fab with 160k/wpm running mainly on 192L products. Solidigm (SK hynix's NAND subsidiary) runs 95-120k/wpm total eSSD capacity in Dalian, China producing 144L-192L products. Although the company's decision to slow down the NAND

node migration might reflect a number of factors (it is not migrating much capacity in the Korea fabs for NAND, owing to supply discipline), a more stable policy environment could to some extent incentivize the companies to migrate more aggressively in China amid current shortages.

Applying the same exercise, we estimate that migrating these China-based NAND fabs to more advanced layer counts could add between 3.0% and 8.9% to 2028 global supply, depending on the migration target – ranging from ~51,301mn (8Gb) of incremental annual bit output at 232-layer to ~150,863mn at 412-layer.

Exhibit 57: Korean suppliers' China fabs – node migration scenario

Moving from 1YY L to...	Incremental bit addition per wafer (8Gb mn)	Annual bit output addition (8Gb mn)	As % of 2028 Global Supply
2XX L (232-layer)	15.3	51,301	3.0%
2YY L (286-layer)	34.7	116,662	6.8%
3XX L (328-layer)	43.8	147,063	8.6%
4XX L (412-layer)	44.9	150,863	8.9%

Note: Underlying node migration and yield assumptions are from US WFE model. Source: Company data, Morgan Stanley Research estimates

China acceleration scenario

On the supply side, both CXMT and YMTC are expanding aggressively and collaborate closely with local WFE vendors.

On the DRAM side, EUV is still the bottleneck and we do not assume the EUV ban on exports to China is lifted in our base case scenario. We assume CXMT will continue to run on 1 alpha node with bit supply as a percentage of global DRAM supply increasing from 14% in 2026 to 16% in 2028 on greenfield capacity expansion.

On the NAND side, YMTC is currently on track to ramp up the yield on local equipment vendors. Our industry checks suggests that a lifting of the equipment ban from US vendors could help improve the yield by around 10%. Direct orders from US customers could also provide an incentive for Chinese players. Currently, YMTC remains on the US Commerce Department's Entity List, where it has sat since late 2022, restricting US technology exports to the company.

In a scenario in which YMTC accelerates the construction of Fab 4 & 5 and dedicates all the capacity to NAND at 3XXL mainstream products, together with yield improvement from US equipment, it would will increase the output by 2-10% from our current base case scenario.

Exhibit 58: YMTC – Potential supply acceleration analysis

	Base case bit supply (310k/wpm, 65% yield at 3XXL in 2028)	Uplift from yield improvement (65% to 75%)	Uplift from capacity expansion acceleration (310k to 500k)
Bit Supply (8Gb mn)	246,122	283,987	458,044
As % of Global Supply	14%	16%	24%

Note: Underlying node migration and yield assumptions are from US WFE model. Source: Company data, Morgan Stanley Research estimates

Primer: Memory 101 – Types, Market Structure and the Cycle

What is memory?

Memory is not a single commodity but a family of distinct products – each serving different applications, made by different suppliers, bought by different customers, and subject to its own supply and demand dynamics. Understanding these differences is essential to understanding who stands to benefit from the current market dynamics and who is disadvantaged.

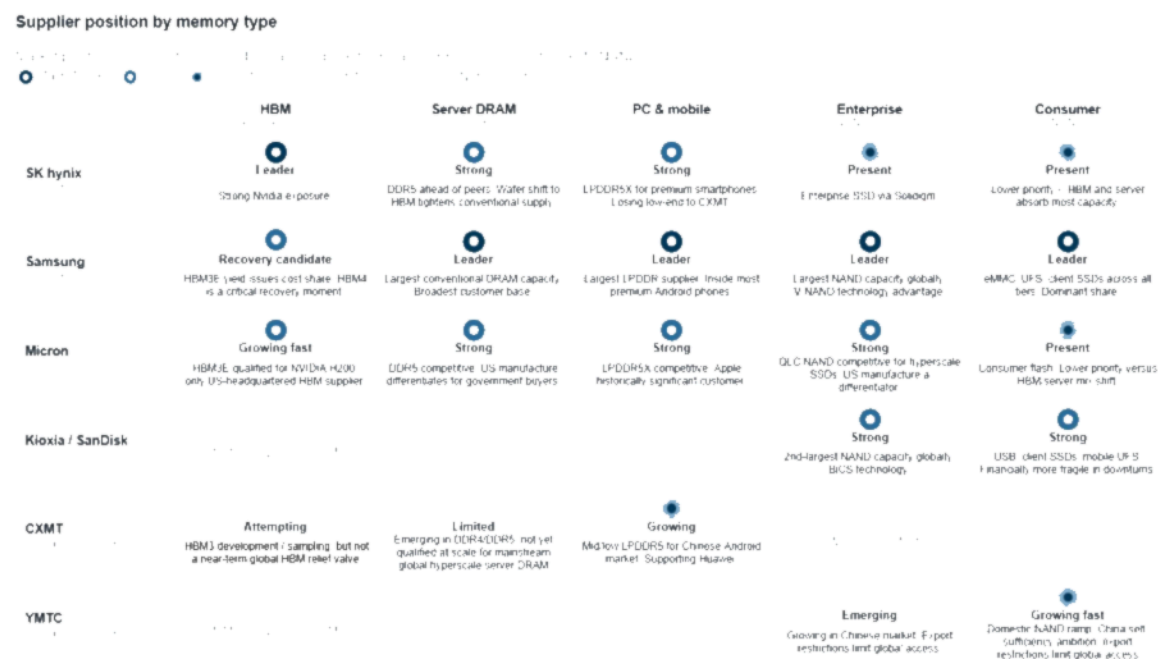
The 6 main types of memory:

- **HBM (High Bandwidth Memory)** sits directly on the AI accelerator package, stacked in layers connected through microscopic vertical pillars called through-silicon vias. It delivers extraordinary data throughput – far beyond conventional DRAM – and is the reason a single Nvidia H100 can process vast amounts of data per second. Because it is physically integrated into the chip package, HBM is not interchangeable with any other memory type; it must be designed in at the silicon level. It is also the most technically complex and expensive memory in the world to make.
- **Server DRAM** is the working memory of data centers – the DDR modules that sit in the server motherboard and hold active data for CPU and GPU processing. Every AI training server requires large quantities, and agentic AI workloads – where many simultaneous reasoning chains run in parallel – are driving server DRAM content per box to levels that were unimaginable two years ago.
- **PC and mobile DRAM** is the same fundamental technology as server DRAM but in lower-power form factors — the LPDDR chips inside laptops, tablets, and smartphones. It is a far more price-sensitive, consumer-facing market, and the one most exposed to demand destruction when prices rise.
- **Enterprise NAND** is flash storage used in data center SSDs and storage arrays. It holds data that does not need to be accessed in real time, but which must persist. NAND is structurally different from DRAM — it retains data without power — and is manufactured in a completely separate fab process.
- **Consumer NAND** covers the flash storage inside smartphones, PCs, gaming devices, and USB drives. It is high volume, low margin, and the first to suffer when budgets tighten.
- **HDD (Hard Disk Drives)** are used to store digital content and data on computers or other electronic devices. These non-volatile storage devices employ magnetic disks to write and read data from the disk. SSDs are generally faster and more reliable than HDDs, but they are also far more expensive.

Supplier map: strategic memory remains Big 3-led; China scales in commodity pools

The supplier landscape is highly segmented by product. HBM and advanced/server DRAM remain concentrated among Samsung, SK hynix and Micron, with SK hynix leading HBM today, Samsung positioned as the broadest DRAM/NAND supplier and potential HBM recovery candidate, and Micron emerging as the key US-headquartered strategic supplier. NAND supply is broader, with Samsung, Kioxia/SanDisk, SK hynix/Solidigm, Micron and YMTC all relevant. China is becoming more important through CXMT in commodity DRAM/LPDDR and YMTC in NAND, but neither is a near-term global HBM relief valve. The key implication is that supply relief is uneven: China can help in selected commodity memory markets, while frontier AI memory remains concentrated in the Big 3.

Exhibit 59: Supplier position by memory type



Source: Morgan Stanley Research estimates, company commentary. Note: Market positions are illustrative and intended to frame supplier participation by memory type.

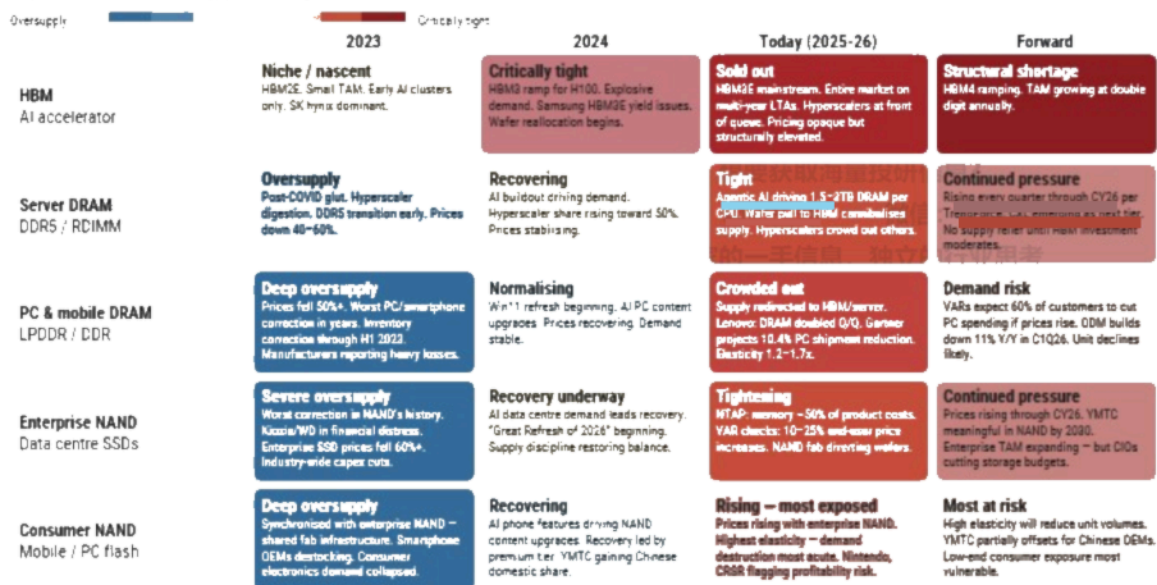
Memory cycle monitor

The memory cycle is no longer uniform. AI-facing products are structurally tight, while consumer-facing products are tight mainly because they are being crowded out. HBM is the most constrained pool; server DRAM is being pulled by AI/cloud demand; and enterprise NAND is tightening as AI storage demand grows. By contrast, PC/mobile DRAM and consumer NAND are not tight because end demand is booming; they are exposed because suppliers are prioritizing higher-value HBM, server DRAM and enterprise SSDs. This divergence means that AI buyers get priority allocation, while consumer and industrial buyers face higher prices, lower specs, delayed refresh cycles and margin pressure.

Exhibit 60: Memory cycle conditions by type

Memory cycle conditions by type

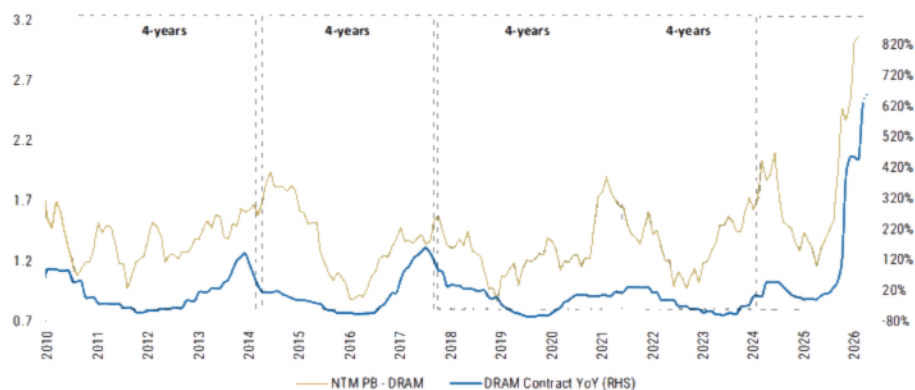
Supply/demand balance from oversupply to critically tight



Source: Morgan Stanley Research estimates, company commentary, TrendForce, Gartner, VAP checks. Note: qualitative supply/demand conditions shown by product type; price increases refer to dated market checks where available.

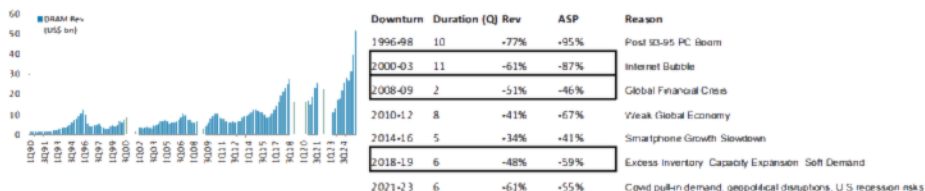
The rate of change in DRAM contract prices has been a consistent indicator of inflections in the memory cycle and stock prices: Investors have not been rewarded in past cycles for buying stocks that are cheap on peak earnings once meaningful second-derivative weakness has begun. Second derivative weakness at cycle peaks is such a consistent leading indicator because there are positive feedback loops with customer inventory behavior that are far more durable than week-to-week demand variances. When expectations build for future quarterly prices to plateau or fall, customer focus turns to reducing inventory to minimize the negative impact of depreciating inventory.

Share prices normally lead DRAM and are concurrent with the NAND cycle at trough: In most cases the trough in stock prices was concurrent with (or lagging by a month) the inflection point (second derivative), measured by the rate of revenue growth for DRAM. In the 2021-23 cycle, DRAM YoY contract pricing peaked in December 2021 and had troughed by June 2023, but stock prices bottomed two quarters ahead of this in December 2022.

Exhibit 61: DRAM Contract price YoY vs. NTM P/B

Source: Factset, TrendForce, Morgan Stanley Research

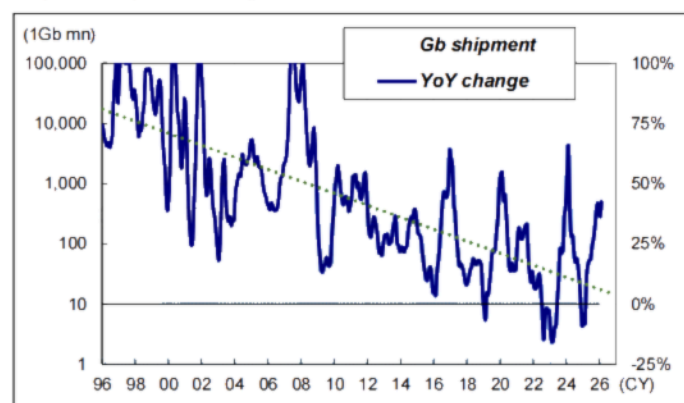
Duration is getting shorter now. Historically, the memory segment tended to be 4-8 quarters in a downturn and 4-9 quarters in an upturn. The slowdown tends to be shorter than the upturn in the memory cycle because it is easy to cancel orders and rationalize inventory but harder to bring back capacity. The 2018 share-price correction lasted only four quarters, from 1Q18 to 1Q19. The most recent downcycle in 2024 lasted for only 8 months before stocks found a bottom, and with tariffs and AI demand supporting an increase in pricing again, share prices also rebounded sharply. We are now 13 months into this upcycle, which began in April 2025.

Exhibit 62: DRAM cycles in perspective over the last 30 years

Source: Factset, Morgan Stanley Research

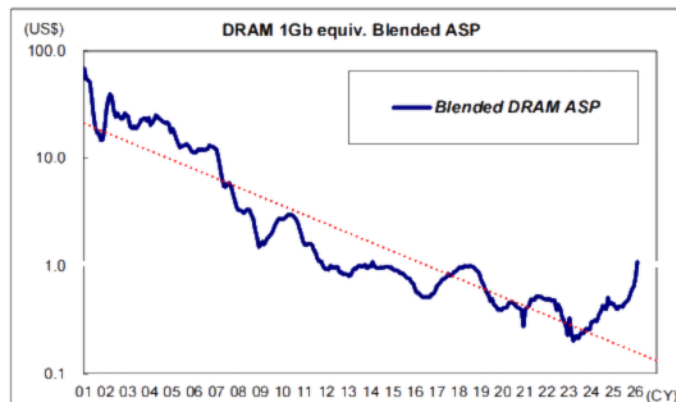
DRAM share price performance is driven by valuation in the early stages of a downturn: In the 2021 cycle, EPS estimate revisions lagged the peak in share prices by three months, while de-rating drove the initial downward leg of share price performance, anticipating the inflection of the DRAM cycle. Multiples tend to bottom around the inflection in the second derivative (either pricing or monthly revenues), which is often the catalyst for share prices to reverse performance. In terms of EPS estimate revisions breadth, peak to trough magnitude is almost 2x (+102% to -88%) on average, and the duration of EPS estimate revisions averages 12 months but only lasted 2 quarters in the last downcycle.

Exhibit 63: DRAM demand grows every year in terms of bit shipments (1mn Gb eq)



Source: WSTS

Exhibit 64: DRAM ASP has rebounded quickly since 2024



Source: WSTS

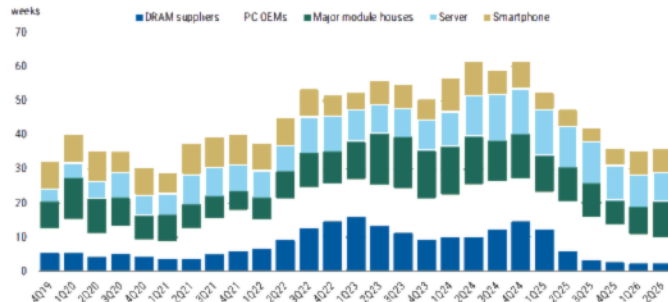
Memory inflections

We monitor four signposts as early inflection signals during a cycle:

1. **DRAM prices extending strength.** We expect LTAs to change the high volatility nature of memory pricing for the next 2-3 years. After significant hikes from 4Q25-2Q26, we expect pricing increases on a QoQ basis to narrow to around 8-13% QoQ for both DRAM and NAND. Instead of viewing this as a sign of inflection, we believe LTAs can instill stability, improve earnings predictability, and reduce the risk of overbuilding capacity with better supply management for memory companies by locking in pricing at an elevated level with prepayments. ([LTAs – How They Are Affecting Memory](#))
2. **Inventory adjustment cycle.** Supplier inventory remains at historical lows of only 2-3 weeks for DRAM and 4-5 weeks for NAND at the beginning of 2Q26. Server customers, despite aggressive orders, continue to work down their inventory, with remaining end customers retaining healthy inventory levels. During an upturn, lead times for the delivery of chips to manufacturers of PCs, servers and smartphones tend to grow longer. This prompts those customers to order more than they need and hoard that inventory in case longer lead times grow into a real shortage. We see this as an ongoing issue for consumer customers, who face a trade-off between soaring prices and limited supply allocation.
3. **Capex budgets pared down while capital intensity moves higher.** There is typically either too much or too little supply in memory. The industry is competitive but also rational, and should therefore be responsive to periods of price disequilibrium. In response to the increase in inventories and the prospect of lower profitability, chip manufacturers reduce their capacity growth to manufacture chips to meet that demand. Capital expenditure for DRAM was cut by 23% in 2019 and stayed flat in 2020, while the 2015 downturn saw only 9% cuts. However, AI demand is likely to keep absolute capex levels elevated in upcoming downturns and also prompt companies to require higher returns to justify those investments.
4. **Continued upward earnings revisions and valuation expansion.** The market often runs ahead of the actual revisions in earnings, and we would look for further

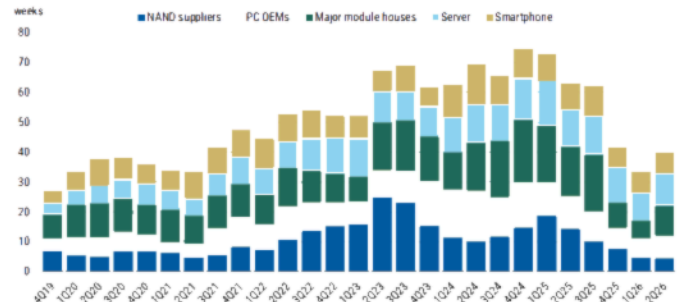
earnings upside through HBM pricing negotiations. Our sensitivity analysis ([here](#)) suggests that the market is not yet assigning any meaningful P/E premium to memory suppliers' LTA-backed earnings and FCF. For illustration, if we were to apply a 10x P/E multiple to LTA-backed earnings and assume 70% commodity LTA coverage, the 2027 P/E for Samsung/hynix could expand to ~8.5x, all else equal (from 5x currently).

Exhibit 65: DRAM inventory



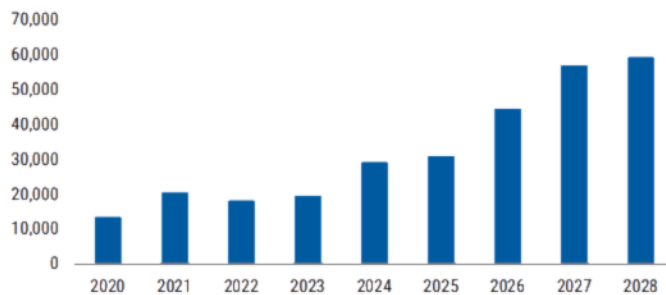
Source: Trend force estimates, Morgan Stanley Research

Exhibit 66: NAND inventory



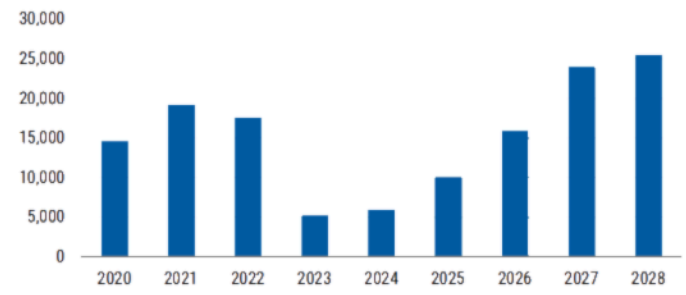
Source: Trend force estimates, Morgan Stanley Research

Exhibit 67: DRAM Capex (US\$ mn)



Source: Trend force estimates, Morgan Stanley Research

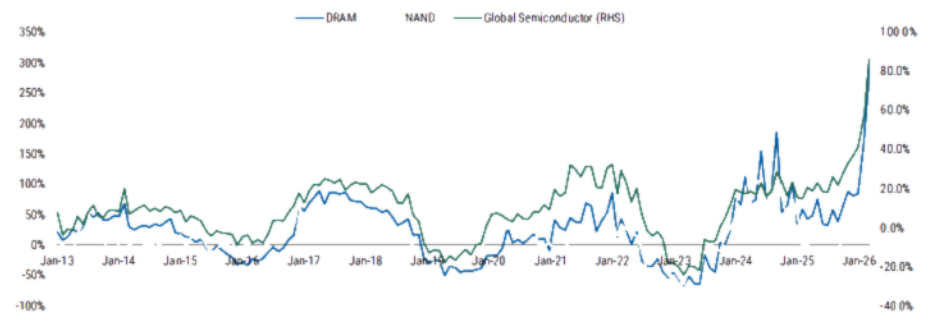
Exhibit 68: NAND Capex (US\$ mn)



Source: Trend force estimates, Morgan Stanley Research

Exhibit 69:

DRAM in the context of the semiconductor cycle – revenue YoY is highly correlated for memory with DRAM at a more advanced stage relative to NAND



Source: SIA

Appendix: Memory Demand per Gigawatt of AI Data Center Deployment

Framework overview

Quantifying memory demand per GW of AI compute requires disaggregating by hardware platform, as configurations differ materially between GPU-based and ASIC-based deployments. For GPU platforms (NVIDIA), our analysis anchors on the GB200 NVL72 and Vera Rubin NVL72 systems. ASIC configurations are non-standard and harder to model – LPDDR is not yet widely adopted for server main memory in ASIC deployments, making direct comparison unreliable – though HBM intensity should be broadly similar across platforms.

DRAM wafer intensity per 1GW of compute

Based on our SPE team's bottom-up analysis of wafer requirements to produce 1GW of compute annually, DRAM is by far the most wafer-intensive component of the supply chain.

For the GB200 NVL72, total DRAM wafer demand runs at approximately 153kwpm per 1GW, comprising roughly 110kwpm for HBM and 43kwpm for LPDDR. For the Vera Rubin NVL72, aggregate DRAM intensity is broadly similar at approximately 144kwpm, with HBM at ~89kwpm and LPDDR rising to ~55kwpm – reflecting the higher LPDDR content of the Vera Rubin architecture. In WFE dollar terms, Logic WFE remains the single largest component at the GB200 platform level (~\$14.2bn vs ~\$13.8bn for DRAM), though the split is platform-dependent – Vera Rubin's higher LPDDR content shifts the balance modestly toward DRAM (~\$14.4bn VS ~\$11.6bn). The two components are broadly comparable in aggregate WFE intensity.

End-market fungibility and allocation dynamics

An important nuance for modelling consumer hardware exposure is that DRAM bits are highly fungible across end markets: PC and server DRAM are essentially the same product, and LPDDR now sees substantial use in GPU servers. As a result, the Big 3 suppliers do not formally ring-fence capacity by end market – all three intend to continue serving non-data-center customers, but incremental capacity will flow toward whichever end market is growing fastest. In practice, this means consumer and PC allocations face a structurally tightening ceiling as AI server demand crowds out non-server bits, even without any formal reallocation decision by suppliers.

Agentic AI: a new incremental demand vector

Beyond traditional GPU server deployments, the shift from inference to agentic AI introduces a structurally new demand category for CPU-attached DRAM. Unlike GPU training or batch inference, agentic workloads run continuous, multi-step reasoning chains that require large, persistent memory states at the orchestration layer. Our framework for sizing this opportunity is:

Agentic AI DRAM demand = (Incremental orchestration CPU shipments) × (average DRAM per CPU)

We have previously identified standalone agentic orchestration servers as a new hardware category, with DRAM content per box potentially reaching 1.5-3TB – levels that were unimaginable for a CPU-class server two years ago. KV cache offload dynamics compound this further: whereas prior GenAI models offloaded KV cache to NAND on a session basis. Agentic workloads require weeks-to-months of persistent state, materially increasing both DRAM and NAND content per server. This equation is still evolving and is not yet captured in consensus DRAM TAM models, representing a source of upside to our already constructive demand outlook.

Illustrative WFE scale: TeraFab as a stress test

To illustrate the scale of memory wafer demand that large-scale AI compute build-outs imply, our SPE team's TeraFab analysis is instructive. A facility ramping to 12GW of annual compute output would require approximately \$33bn of WFE in year one – of which DRAM alone accounts for roughly \$14bn. Even at a more conservative 3GW ramp cadence, annual DRAM WFE demand would reach approximately \$3.5bn from a single facility. This underscores that AI DC deployment is not merely a chip demand story but a full-stack memory capacity story, with HBM remaining the most constrained and strategically differentiated product in the stack.

Exhibit 70: SPE to GW

Monthly Build 12	HBM3e	HBM4
	GB200 NVL72 Blackwell	Vera Rubin NVL72 Vera Rubin
kW per Rack	120	200
Racks per 1GW	8,333	4,167
# of GPUs	600,000	300,000

HBM (DRAM)

HBM	3e	4
GB per GPU	192	288
HBM Module GB	24	36
Node	1Beta	1Beta/1C
HBM per Rack		
Packages per GPU	8	8
DRAM dies per stack	8	12
DRAM dies per GPU	64	96
DRAM dies per Rack	4,608	6,912
Die per 1GW	38,400,000	28,802,304
Bits per 1GW (2Gb mn)	461	346

Monthly Build 12	HBM3e	HBM4
	GB200 NVL72 Blackwell	Vera Rubin NVL72 Vera Rubin
kW per Rack	120	200
Racks per 1GW	8,333	4,167
# of GPUs	600,000	300,000

LPDDR

Node	1Beta	1Gamma
LPDDR per Rack (TB)	17	54
LPDDR per GW (TB)	141,667	225,018
LPDDR per die (Gb)	16	32
LPDDR die per 1GW	70,833,333	56,254,500
DRAM Wafer		
Consumption		
Die Size (mm2)	38	58
Gross Die per Wafer	1,752	1,131
Yield	95%	90%
Good Die per Wafer	1,664	1,018
Wafers per 1GW	42,557	55,255
LPDDR WFE (\$mn)	319	460

Source: Morgan Stanley Research

Share prices for stocks referenced: Advanced Micro Devices 510.13; AP Memory Technology Corp 1,075.00; Apple, Inc. 306.31; Applied Materials Inc. 458.17; Arm Holdings plc 408.85; ASM International NV 867.40; ASML Holding NV 1,394.00; Aspeed Technology 18,095.00; BE Semiconductor Industries NV 278.20; Egis Technology Inc 122.50; Everpure, Inc. 84.45; FIT Hon Teng Ltd 9.05; Global Unichip Corp 4,535.00; Hon Hai Precision 301.50; Ibiden 21,135.00; Intel Corporation 109.33; KIOXIA Holdings 77,540.00; KLA Corp 1,940.04; Lotes Co. Ltd. 2,615.00; MEC 10,930.00; Micron Technology Inc. 1,035.50; Montage Technology Co Ltd 372.40; Murata Manufacturing 10,230.00; Nan Ya PCB 862.00; Nitto Boseki 21,900.00; NVIDIA Corp. 224.36; Renesas Electronics 4,455.00; Samsung Electro-Mechanics 1,813,000.00; Samsung Electronics 360,500.00; SanDisk Corporation. 1,761.43; Seagate Technology 921.26; SK Hynix 2,360,000.00; TDK 3,849.00; Tokyo Electron 53,710.00; TSMC 2,380.00; Ulvac 8,913.00; Unimicron 975.00; Western Digital 546.20; Wiwynn Corp 5,475.00; Wonik IPS Co Ltd 97,900.00; WPG Holdings 119.00; Yageo Corp. 846.00

This report references U.S. Executive Order 14032 and/or entities or securities that are designated thereunder. U.S. persons may be prohibited from buying certain securities of entities named in this report. Readers are solely responsible for ensuring that their investment activities are carried out in compliance with applicable laws.

This report references export controls and/or entities that may be subject to export control restrictions. Readers are solely responsible for ensuring that their investment or trade activities are carried out in compliance with applicable laws.

This report references U.S. Executive Order 14105 and/or entities that may be in scope of such order. U.S. persons may be prohibited from engaging in certain transactions or otherwise require certain other transactions be notified to the U.S. Department of Treasury. Readers are solely responsible for ensuring that their investment or trade activities are carried out in compliance with applicable laws.