

Geopolitical Tensions and AI-Driven Demand in the Global DRAM Supply Chain: An Investigation into the Nature of Semiconductor Shortages and Price Volatility

Author: Divyanshu Singh
Affiliation: Christ Deemed to Be University
Email: divyanshu.singh@law.christuniversity.in
Submitted to: Journal of International Research & Multidisciplinary Innovation (JIRMI)
Date: 21 June 2026

Abstract

The dynamic random-access memory (DRAM) market is going through a complete reversal of recent trends, shifting from a market of plenty and negative margins to one of the most rapid increases in contract pricing in the history of semiconductor memory, as reported contract pricing for commodity DRAM rose by about 700 percent over four years and by over 170 percent year-over-year in one quarter of 2025. This paper addresses two questions that are posed. First, it asks how much of a disruption of supply-chain as opposed to strategic supply management by a small group of manufacturers is being caused by the recent DRAM shortages. Secondly, it addresses the impact on DRAM demand from the swift adoption of artificial intelligence (AI) and whether that acceleration is a direct driver of the price increase. Using facts from industry sources, corporate disclosures, market research reports, and a June 2026 antitrust class action filed in the Northern District of California, the paper suggests that the lack of supply is as much a true structural dislocation caused by demand as it is an "amplified" supply disruption. It also explores as a competing hypothesis the notion that state-backed Chinese capacity growth has been a distorting force in the market and concludes that existing evidence is more supportive to recognizing China as a market opportunist than a builder of scarcity.

I. Introduction

Memory is the connective tissue of modern computation. Dynamic random-access memory is found in most digital devices, including smart phones, cars, and the accelerators that train and run large language models, sandwiched between the processor and long-term storage. The market has been acting much like a commodity market over the years and following a predictable cycle of oversupply, price collapse, capacity discipline, and recovery, “a bit is a bit.” Since 2020, the years have been examining if that historical trend is present. Deep production cuts pave the way for a recovery that then picked up much faster than was typical of a cyclical turnaround, after the pandemic era swings and a brutal downcycle in 2022 and early 2023, when chips were being sold at below cost prices.

By the end of 2025, the recovery had become an escalation, with industry players using superlatives in their descriptions. One senior Micron executive described it as the biggest mismatch between supply and demand in terms of both magnitude and time frame that the industry had witnessed in 25 years. Retail data backed up the wholesale figures: the 32-gigabyte version of DDR5, which had been priced at around ninety-one dollars in the middle of 2025, was priced near to one hundred eighty-three dollars four months later, and major original-equipment manufacturers (OEMs) had predicted personal-computer price hikes of 15 to 20 percent in 2026. This paper explores two questions, formally stated in Part II: the first is about causation and intent on the supply side, the second about the influence on demand-side of artificial intelligence. The methodology used in the study is discussed in Part III. Part IV sets out the structure and concentration of the industry. Part V looks at the demand side, following the path from the AI buildout and capital race between hyperscale buyers to the memory demand and price. Part VI tackles the supply question, balancing real disruption with strategic management and in view of the June 2026 antitrust litigation. Part VII examines the geopolitical aspect and tests as an alternative theory the idea that China has created an artificial crisis. Part VIII concludes.

II. Research Questions

There are two research questions herein that each account for half of the causal chain suggested in the title pairing of geopolitical tension and AI demand.

Research Question-1. How much of the shortfall in DRAM that caused outages recently is due to actual disruptions in the supply chain and to how the leading global semiconductor firms are managing the supply chain? [OBJ]

Research Question-2. What has changed with AI's fast growth in the industry's demand for DRAM and how has it contributed to the recent surge in DRAM prices and scarcity?

Weakly linked the two questions analytically. R2 quantifies the size of the demand shock and its mechanism; R1 asks if the observed shortage is driven entirely without discretionary action by a focused supply base or if, along with the shock, there is some discretionary action on the part of a focused supply base? It is therefore a prerequisite for R1 to be answered first before the question at R2 can be answered.

III. Research Methodology

The approach adopted in this study is a qualitative and evidence-synthesis method appropriate to a fast-changing market where prices of the core quantitative data contracts and the allocation of wafers and the order-fulfilment rates are confidential and only released on a case-by-case basis. The paper doesn't construct a new econometric model from a small set of data, but it tests each of the independent data sources for each question in the research and tries to make a triangulation between them to see whether the research question gets the "pass" or not for the available data.

There are three kinds of sources that are used. The first is the pricing and wafer-allocation data from a group of market-research firms focused on memory pricing and capacity monitoring: TrendForce, Counterpoint Research, IDC, Omdia and SemiAnalysis; the second is the pricing and wafer-allocation data on which R2 turns. The second is primary corporate disclosure, in this case, that of the DRAM companies (Samsung, SK Hynix, Micron), the hyperscale buyers (Microsoft, Alphabet, Amazon and Meta), and the downstream OEMs (Dell, HP, Lenovo and Apple), respectively. The third is legal and regulatory documentation, including the class action complaint filed on 25 June 2026 in the Northern District of California and the history of prior antitrust action that is documented ¹(the evidentiary context for R1).

¹ TrendForce, Memory Price Rally May Run Past 2028 (Dec. 2, 2025), <https://www.trendforce.com/news/2025/12/02/news-memory-price-rally-may-run-past-2028-as-samsung-sk-hynix->

There are two methodological safeguards twice in it. First, all price information is considered direction only, not a true price as the contract price is only an estimate and differences in price quoted from sources are reported but not quietly reconciled. Second, the paper uses the legal definition of 'unlawful coordination' and 'lawful 'conscious parallelism' as an analytical sieve: any observation of strategic supply management is considered evidence of the withholding theory as opposed to the independent profit-maximizing behaviour which an oligopoly might adopt in the absence of an agreement. The China hypothesis of Part VII is tested as a testable alternative hypothesis and is maintained only so far as the evidence which supports its base (the conduct of the Chinese) is relevant.²

V. Industry Structure and Market Concentration

About 90 to 95 percent of the world production of DRAMs is controlled by three firms: Samsung Electronics, SK Hynix and Micron Technology. This focus is an important characteristic of the analysis to follow: a three-firm market allows for parallel behaviour that is hard to tell apart from coordination. Barriers to entry are extraordinary: the cost of one extreme-ultraviolet lithography line is on the order of 15 billion dollars; new fabrication facilities take three-to-five years to build and ramp.³

There are many points worth making about the fact that a bit of DRAM is fairly interchangeable from supplier to supplier, but the key ones here for this analysis are process technology and cost structure, and the allocation of limited wafer capacity between commodity DRAM and high-margin high-bandwidth memory. The one allocation decision that underlies both research questions is also the through-which AI demand is channelled into commodity scarcity, and the lever by which strategic supply management, if it is possible, would be activated.

reportedly-cautious-on-expansion/ (Samsung's stated preference for profitability over expansion and roughly 70 percent order fulfillment).

² See DRAM Prices Expected to Nearly Double in Q1, The Register (Feb. 2, 2026), https://www.theregister.com/2026/02/02/dram_prices_expected_to_double/ (explaining key-value cache storage across HBM, system memory, and storage tiers during inference).

³ SemiAnalysis, *supra* note 27 (roughly 99 percent DDR/LPDDR revenue and minimal HBM wafer allocation through 2025).

V. Artificial Intelligence and the Demand Shock

A. The mechanism: from accelerators to high-bandwidth memory

The AI demand shock didn't increase DRAM consumption across the board; it focused demand on a niche product known as high-bandwidth memory (HBM) which uses more capacity. Their compute logic is surrounded by stacks of HBM, each stack made up of a dozen or more vertically bonded DRAM dies using silicon vias. The B300, for instance, is Nvidia's accelerator, which is built with eight stacks of 12 dies, while other parts like AMD's MI350 are built in the same way. The sheer size of the compute build-out was another first: Nvidia's data-centre revenue jumped from just one billion dollars in the fourth quarter of 2019 to fifty-one billion dollars in the quarter that ended in October 2025.

The two aspects of HBM that are relevant to the supply story are: First, HBM uses much more wafer capacity than conventional DRAM, since stacking and advanced packaging increase the number of silicon and processing steps that are needed. Second, the profits by HBM are much higher. Advanced packaging is the solution that the DRAM industry has been seeking with the increasing physical difficulty of scaling DRAM. In essence, advanced packaging is using more DRAM to gain the same performance benefit. The result is a diversion effect all those wafers that go into the lucrative HBM contracts are wafers that are too scarce to go into commodity DDR4 and DDR5 modules that fill PCs and smartphones. The attractions of HBM are clearly evident in Micron's own sales, as the company's HBM and related cloud memory business jumped from seventeen per cent of DRAM sales in 2023 to nearly fifty per cent in 2025.

B. Price inelasticity and the inversion of pricing power

One of the unique features of AI-powered demand is that it doesn't backslide with an increase in price. Even for the hyper-scale buyers who are constructing the training/inference clusters, it will be a non-tranchesome value over hyper-scale as the unit price of a cluster will not stand out. One semiconductor analyst aptly said the AI buyers dictate not just the bottom but as well the top for

the marketplace. ⁴This inelasticity leads to the opposite of the usual A-to-Z cycle. The determination of marginal prices has been replaced by the orders customers accept while reducing that for memory costs when it gets too high. Hyperscale's soar absorption of costs reached up to half in 2025 as of yet, yet the startup has been obtaining segmental fulfilment.

This shift of AI workloads from training to inference has spread the pressure to nearby memory tiers. Inference is stored in a key-value based cache of working memory that remains in fast HBM when in use and switches to slower DRAM or storage when dormant, expanding usage of inference throughout the entire memory hierarchy.

C. The AI arms race and its direct contribution to RAM prices

While AI demand in itself doesn't necessarily drive-up RAM prices, the pressure between buyers and sellers stems from the race to increasingly grow the capacity of computer hardware for AI use among the largest tech companies with a kind of transactional arms race as they all drive compute capacity up at the same time in ways that are not always cost-effective. The four biggest hyperscale's in the United States Amazon, Microsoft, Alphabet and Meta led the charge in capital spending, totalling approximately 725 billion dollars for the 2026 period, an increase of approximately seventy-seven percent over the previous year of 410 billion dollars, according to the forecasts of analysts, and this amount should exceed one trillion dollars in 2027. The structural logic of this race is that, for all four parties, it is more dangerous to be short of compute than it is to be overcommitting. Hence why all four expand simultaneously, turning a coordination problem into a bidding war for the very limited available amount of packaging and memory.

The customers' narratives about price increases in their budget should be the most obvious indicator of how these arms race directly effects memory prices: Chief financial officer Richard Steer hinted at a sizable increase on the near side in his breakdown of the company's most recent capital-expenditure forecast of 190 billion. USD, significantly above consensus, where he said about 25 billion. USD will be allocated to improving memory-chip and component costs, and

⁴ Fusion Worldwide, AI Sets the Price: Why DRAM Shortages Are Rewriting Memory Market Economics (Nov. 6, 2025), <https://www.fusionww.com/insights/blog/ai-sets-the-price-why-dram-shortages-are-rewriting-memory-market-economics> (quoting analyst Claus Aasholm).

warned competition to watch out for capacity constraints at least through 2026. Meta boosted its full-year outlook to \$125 billion to \$145 billion, explicitly highlighting increased component pricing as one of the factors behind the growth. Now, memory has grown to a significant line item, enough to shift a trillion-dollar company's guidance at one estimate, 30 percent of hyperscale spending for data centres is going towards memory⁵. So, the causality is clarified and is two-way – the arms race drives up memory prices, and the obtained memory inflation drives up the cost of others memory that sustains the arms race.⁶

D. Quantifying the AI contribution to price

The price response, however, is stark enough that there's no mistaking that it is being used because of rising demand for AI. According to a data research report released by TrendForce on Wednesday, DRAM prices for a battery of chipsets exceeded the 170 percent year-over-year marked just for the third quarter of 2025. According to Counterpoint Research (CP) the lift was eightieth to ninety per cent while Trend Force predicted a DRAM contract price hike of between ninety and ninety-five per cent quarter-over-quarter (QoQ) in 2026,⁷ which is unprecedented, according to a Trend Force analyst.⁸

That is a need that cannot be overlooked because it was for AI. TrendForce pointed out that as the Q4 2025 shipment volume of PCs rose from looking forward, current stockpiling by OEMs and channels compounded its effect, in turn leading to a serious shortage. These are second order amplifiers on top of a 1st order shock, however. The theory of the case has been compelling, with

⁵ Anton Shilov, Google, Microsoft, Meta, and Amazon Capex Spending to Hit \$725 Billion in 2026, Up 77% From Last Year, Tom's Hardware (Apr. 30, 2026), <https://www.tomshardware.com/tech-industry/big-tech/big-techs-ai-spending-plans-reach-725-billion> (combined 2026 capex and year-over-year increase, per Financial Times compilation of first-quarter earnings); see also Big Tech AI Spending 2026, Value Add VC, <https://valueaddvc.com/blog/big-tech-ai-capex-in-2025-microsoft-google-meta-amazon-and-the-spending-race> (last visited July 4, 2026) (projecting capex above one trillion dollars in 2027).

⁶ See Shilov, *supra* note 15 (reporting memory's projected share of hyperscaler data-center spending).

⁷ The Register, *supra* note 13 (PC shipments compounding shortage); Techwire Asia, Are Chinese Memory Chips About to Reshape Global NAND and DRAM Supply? (Apr. 9, 2026), <https://techwireasia.com/2026/04/chinese-memory-chips-ymtc-cxmt-nand-dram-expansion/> (Counterpoint's Hwang noting stockpiling had not produced inventory excess).

⁸ Moore, *supra* note 7 (Counterpoint estimate of 80–90 percent); The Register, *supra* note 13 (TrendForce projection of 90–95 percent); AI Memory Is Sold Out, Causing an Unprecedented Surge in Prices, CNBC (Jan. 10, 2026), <https://www.cnbc.com/2026/01/10/micron-ai-memory-shortage-hbm-nvidia-samsung.html> (TrendForce analyst Tom Hsu describing the increase as “unprecedented”).

the surge in DRAM prices driving the adoption of AI and transformation of wafers, and price measurements into a buyer's constraint that leads hyperscale customers to purchase HBM⁹.

VI. Genuine Disruption or Strategic Supply Management?

The real shock of the demand shock studied in Part V is real, but is this shock alone sufficient to answer R1, what is the nature of the shortage (real or intentionally managed)? The two explanations are not mutually exclusive, the interesting question being how much of the relative scare is related to constraints on the constellatory not under the control of the manufacturers, and how much is related to the choices they made, and could more or less properly have made.

A. The case for genuine, structural disruption

There are a few facts that suggest disruption reading. There simply isn't enough cleanroom capacity for the DRAM industry, TrendForce said that would be Samsung and SK Hynix who can increase their cleanroom capacity a little in 2026, while also expecting a new DRAM fab in the U.S. before sometime in 2027. Even a significant escalation in capital expenditure (capex) on DRAM behemoths was predicted to lead to an increase in bit supply of only a couple of billion in 2026, as new capacity does not appear quickly¹⁰. On this basis, the shortage is fully expected as it is driven by a relatively hard-to-satisfy demand growth and a supply base hardwired in time for the multi-year forecasts that executives Chairman of SK Hynix estimated the shortage may last through 2030.¹¹

B. The case for strategic supply management

⁹ The Register, *supra* note 13 (PC shipments compounding shortage); Techwire Asia, Are Chinese Memory Chips About to Reshape Global NAND and DRAM Supply? (Apr. 9, 2026), <https://techwireasia.com/2026/04/chinese-memory-chips-ymtc-cxmt-nand-dram-expansion/> (Counterpoint's Hwang noting stockpiling had not produced inventory excess).

¹⁰ *Id.* (capex figures and minimal near-term bit-supply impact).

¹¹ See 2025–present Global Memory Supply Shortage, Wikipedia, https://en.wikipedia.org/wiki/2024%E2%80%93present_global_memory_supply_shortage (last visited July 4, 2026) (collecting executive forecasts, including SK Hynix chairman Chey Tae-won's estimate that the shortage would persist until 2030).

Competing reading focuses on discretion. For example, manufacturers have decided to use approximately a quarter of the DRAM wafer production capacity for HBM starting from 2022, and they have opted to phase out commodity lines, as opposed to continuing them. In corporate statements, this isn't an accident, but rather it is a purposive capital discipline.¹² Samsung told investors it would focus on profitability rather than a rush to cut its DRAM capacity and would try to avoid any oversupply threats, but said it would be able to meet approximately seventy percent of its orders for DRAM. But Micron ended its consumer Crucial brand, and indicated it will not "significantly increase" new commodity DRAM fabs for 2026, effectively rejecting expansion of its business into the shortage kingdom it was profiting from.¹³

C. The antitrust litigation and the limits of the strategic reading

On 25 June 2026, Now, seventeen plaintiffs individuals and small business owners have sued the three manufacturers in U.S. District Court for the Northern District of California, claiming that they employed the artificial intelligence technology of the HBM transition in order to artificially reduce the supply of commodity DDR3 and DDR4 and create artificial scarcity. The claim is that commodity DRAM prices have increased by an estimated 700 percent over four years, saying that none of the three firms pulled back to gain share in the market without collusive involvement "makes economic sense" without collusion.

The plaintiffs call on the history in print. In the mid-2000s, both Samsung and SK Hynix's forebears agreed to plead to a Department of Justice criminal price-fixing charge, being fined more than 700 million dollars total by the industry, including jail time for several executives of the companies, for practices that had been going on since around 1999 through 2002. Significantly, however, a previous class action with similar allegations was dismissed in the same 2018 case, and

¹² TrendForce, Memory Price Rally May Run Past 2028 (Dec. 2, 2025), <https://www.trendforce.com/news/2025/12/02/news-memory-price-rally-may-run-past-2028-as-samsung-sk-hynix-reportedly-cautious-on-expansion/> (Samsung's stated preference for profitability over expansion and roughly 70 percent order fulfillment).

¹³ IntuitionLabs, RAM Shortage 2025: How AI Demand Is Raising DRAM Prices (Dec. 2, 2025), <https://intuitionlabs.ai/articles/ram-shortage-2025-ai-demand> (Micron's retreat from consumer DRAM and absence of new commodity fabs).

the Ninth Circuit upheld the dismissal in March of 2022 on the grounds there was no plausible evidence of an actual agreement under the Sherman Act.¹⁴

This Litigation History offers the context for the analytical crux of R1. Under United States antitrust law, it is not illegal to make a price cut and/or chasm reduction by themselves, but it is illegal for firms to coordinate such actions. The exact same results are likely to be obtained by rival firms as it were without collusion in a market comprising three suppliers. It is possible for firms to come to similar conclusions independently, which is called conscious parallelism, without any illegal collaboration in a market with three suppliers. It is not just a legal issue! It is analytically centre-stage because it would simultaneously require that the same basic facts lead to the capacity change, the price change, and the disciplined capital expenditure, which would be consistent with both the strategic-collusion hypothesis and the lawful response to a genuine demand shock by an oligopoly maximizing the profits on its existing tools. More importantly, the market structure provides each firm with a rational, unilateral incentive to give greater weight to high margin HBM, and not put a fire under a commodity segment; without an accord, as it were, no explanation would suffice.¹⁵

Given this current state of evidence, the more defensible conclusion is that the shortage is a material dislocation, a demand-driven shortage, and that this shortage is severe, but is exacerbated by the strategic, but lawful allocation decisions made by a concentrated industry. There is no claim that the manufacturers have hyped the need for AI to create the displacement of capacity the need for AI is documented by itself. This is precisely what they didn't do, namely reduce or lighten up the ensuing compressor of prices, for a price that is tactical, economic in nature, but not there yet one piece of illegal coordination. But the allegations of collusion yet to be proven and may be put to the test in the court battle in California that also is pending in the European Commission may be at the centre of the issue.¹⁶

¹⁴ BigGo Finance, Samsung, SK Hynix, and Micron Face Class-Action Lawsuit Over Alleged DRAM Price Fixing—Prices Up 700% (June 30, 2026), https://finance.biggo.com/news/202606301021_DRAM_Price_Fixing_Lawsuit_Samsung_SK_Hynix_Micron (quoting the complaint's allegation that the parallel retreat makes no economic sense absent collusion).

¹⁵ TechTimes, *supra* note 21 (explaining that conscious parallelism among a few suppliers is lawful and that plaintiffs must prove an actual agreement).

¹⁶ See Mordor Intelligence, *supra* note 3 (noting European Commission scrutiny of parallel price hikes); BigGo Finance, *supra* note 22 (allegations remain unproven; case assigned to Judge Noel Wise).

VII. Geopolitics and the “Artificial Crisis” Hypothesis

The shortage has been happening amid export restrictions, industrial policy and the swift expansion of Chinese memory manufacturers, including Chang-Xin Memory Technologies (CXMT). In some commentary there is a provocative notion that there's an AI memory crisis of sorts being created or exploited by China as part of its plans for “self-sufficiency.” This Part tests such framing as a hypothesis against, and finds it only partly convincing in this regard.

Chinese opportunism has a strong element of proof. The incumbents shifted capacity towards HBM, but CXMT built up its DRAM capacity, got it high and ready to be sold when the market was short. It was reported to have logged in sales of approximately eight billion dollars in 2025, an increase of almost 130 percent on the previous year, while in first quarter 2026 it reported sales of 7.3 billion dollars, with operating margins around seventy percent, something that shows, according to an analyst, the passage of the cycle rather than the latest technological innovation.

But it is wrong for this opportunistic lie to be called authorship¹⁷. Yet there are two facts pointing to the contrary, that China caused the crisis. Let's start by saying that, according to the filings, CXMT did not have a significant contribution of HBM revenue in its 2025 operating history a point that holds especially true considering that typically only 10 percent of DRAM prices are for HBM devices. First,¹⁸ CXMT is largely still a DRAM company: its filings show that approximately 99 percent of its revenue for 2025 came from DDR and LPDDR products, with essentially no HBM capacity, making it an aside to the HBM market that is actually at the epicenter of scarcity. Secondly it expands as a reliever and does not contract as a restriction: Western OEMs like HP and Dell have opened qualification of CXMT parts specifically for the purposes of diversifying supply and helping to put some space between the U.S. and the markets now "under attack".

The actual distortionary political reality is contrary to the hypothesis. China is blocked by U.S. export controls from getting access to high-end manufacturing equipment, which means that

¹⁷ SemiAnalysis, supra note 27 (roughly 99 percent DDR/LPDDR revenue and minimal HBM wafer allocation through 2025).

¹⁸ Semicon Electronics, HP, Dell and Other PC Giants Consider Chinese Memory Chips for the First Time (Feb. 10, 2026), <https://www.semicon.com/article-392.html> (HP and Dell qualification of CXMT DRAM to secure supply).

CXMT sits in the shadowy rear of the queue and far behind U.S. plans to produce and bundle HBMs en masse in a few years, and a proposed tariff for Korean and Taiwanese chips could drive up the cost for everyone. It's through this policy that global supply becomes tighter. Rather, China is an opportunistic beneficiary of something that it did not create and cannot satisfy with current technology. The “artificial crisis” thesis is the exact opposite in failing to find the evidence that ties the shortage to China's actions; it lives only in the weaker sense that once capabilities are developed fully, they might eventually have a dampening effect on the pricing power of the incumbents¹⁹.

VIII. Conclusion

To understand the DRAM market of 2020-2026, one needs to answer two research questions sequentially. The DRAM market of 2020-2026 has a clear storyline that can be defied by answering these two research questions in order. The evidence is overwhelming and consistent: The surge in demand is driven mainly by the rapid adoption of AI, and the price increase is driven by both the competition for scarce water supply by high-bandwidth memory users, as well as by the price inelasticity of hyperscale buyers, which set a floor and a ceiling on the market.

On R1, the answer is qualified. The lack is not artificial, but real, stemming from a real and structurally based demand and a real and physically limited supply base due to fab timelines and capital intensity. However, the extent of the commodity memory scarcity is an indication of strategic management of the supply, the will to phase-out legacy lines, and the capacity to be built slowly by a three-firm oligopoly. That management was unlawful, but parallel, not coordinated, is the exact issue now before a federal court, and the history of prior convictions is not at odds with a more recent dismissed lawsuit. The truth, of course, is that there are no competitive answers to the question of genuine disruption and strategic management; both are layered on one another: a real demand shock, compounded by rational and possibly but not yet demonstrated coordinated restraint.

¹⁹ See Wccfttech reporting, collected in 2025–present Global Memory Supply Shortage, supra note 18 (former Samsung chip executive suggesting China’s DRAM expansion could ease the price spike within a year).

Journal of International Research & Multidisciplinary Innovation

The geopolitical thesis that China created a "crisis" is unsustainable, since there is no evidence that it was the author of the crisis but merely the beneficiary of it, and that the really distortive policy levers export controls and threatened tariffs actually restrict supply, not create it. That's not just an eye-opener for consumers, enterprises and policymakers it's a warning: The days of cheap and plentiful memory are behind us, at least in the near term, and the configuration of a three-firm market will continue to drive the cost of the AI build-out.