

Wealth Flash The Big Picture

Monday, 21 September 2026

The Memory Supercycle: Numbers Don't Lie

Analyst

Qi Wang

Chief Investment Officer

Private Wealth Management

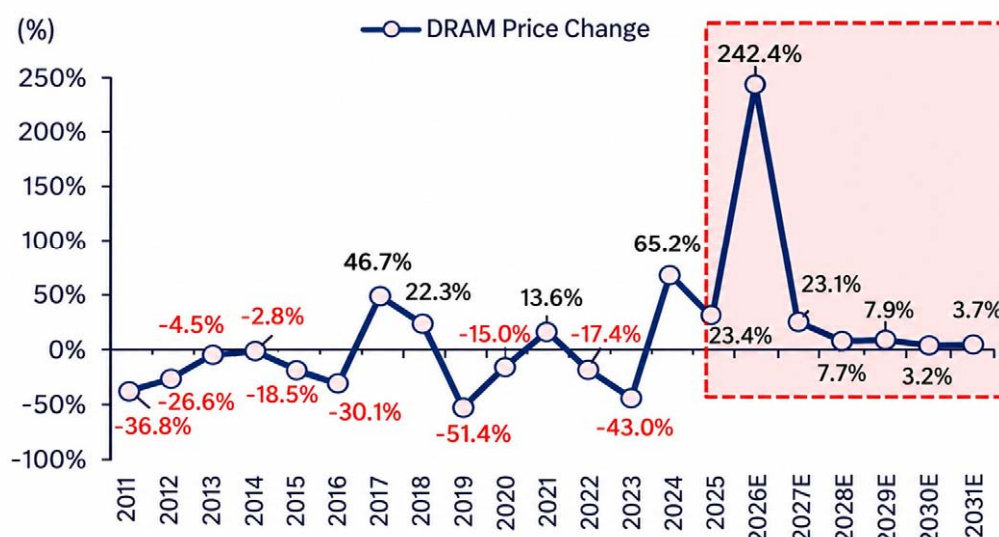
qi.wang@uobkh.com

Based on volatile share prices, investors seem to have a precarious love-hate relationship with memory stocks this year. Yet the industry's fundamentals are nowhere near this volatile, with demand set to outpace supply consistently over the next 3-5 years. This report is the first part of our series designed to help you gain conviction in the memory supercycle and invest in related stocks with clarity.

CIO Summary

- ❖ **Multi-year demand surge driven by AI.** Global DRAM demand is set to grow 14–22% CAGR, powered primarily by AI data centers – the largest and fastest-growing segment. In addition, both enterprise (on-premise) and end-device demand could provide further upside.
- ❖ **Supply growth fundamentally constrained.** Global DRAM supply is expected to grow ~16.6% CAGR, with the top four vendors controlling ~93% of output and China as the key wildcard. The memory industry has a deep moat: USD 15–25bn fab costs, process know-how, IP protection, and multi-year ramp timelines. Even established players cannot accelerate capacity fast enough today, making supply effectively fixed over the next three years.
- ❖ **In all, a sustained memory shortage.** With demand clearly outpacing supply, the industry continues to face a 40–50% structural shortage, supporting a sustained increase in memory prices. This imbalance is unlikely to be resolved over the next 3–5 years, even as China ramps up capacity, reinforcing our view of a durable memory supercycle.

DRAM Price Change Outlook



Global DRAM Demand Forecast (Next 3-5 Years)

Segment	Nominal growth (ex. bit density factors)	Bit density growth (memory content / system)	Aggregate growth (CAGR)	Percent of global bit demand	Contribution to total bit demand growth (CAGR)
Data centers (cloud)	8-12%	15-20%	24-34%	45%	11-15%
Enterprise servers (on-premise)	1-3%	8-12%	9-15%	12%	1-2%
Client devices and others	1-3%	6-10%	7-13%	43%	3-6%
Total	-	-	-	100%	14-22%

Source: Gemini, Gartner, TrendForce, IDC, Micron. Base unit is memory bits.

- **Background:** Our DRAM demand forecast of 14-22% CAGR is based on a select set of credible third-party estimates, cross-checked against vendor guidance and channel checks. Our goal is not to offer absolute accuracy (no one can at this stage), but to establish a consistent reference point with clear underlying drivers. This framework can help us gain better understanding of demand-supply imbalances, run sensitivity analysis and deepen company-specific research.
- **All about AI data centers.** Key end markets for DRAM include data centers (cloud), enterprise servers (on-premise) and client devices (PCs, smartphones, tablets etc). Today, data centers represent both the largest and fastest-growing segment, driven by surging demand for AI cloud computing. This is expected to grow at a CAGR of 24-34% over the next 3-5 years.
- **AI memory content: the real driver.** This structural growth reflects both an increasing number of servers (+8-12% p.a.) and higher memory content per system (+15-20% p.a.). This is because AI workloads require substantially larger models and datasets to be stored and processed, resulting in significantly higher memory requirements. For example, a typical AI training server requires 2-4 TB DDR5 memory today - doubled from 1-2 TB in 2022. That's just DD5 - it does not include the material increase in HBM, SRAM and VRAM caused by AI workloads.
- **Safety concerns add upside to on-premise demand.** We recently wrote about the [“AI pacing debate”](#) amid rising safety concerns for frontier models. AI assurance requirements may actually expedite [on-premise server deployments](#), as enterprises seek tighter control over model safety and data governance. This will likely lead to upside in enterprise memory demand, making our 9-15% CAGR forecast look conservative.
- **Device as the key swing factor.** Because AI adoption is still in a nascent stage, there remains significant uncertainty around both the memory content per device and brand-new categories of devices coming online. For example, the top-of-the-line iPhone 18 Pro (Sep 2026) comes with a maximum 12 GB RAM – sufficient to run a small language model (SLM) with 3-7 billion parameters, but still unable to support typical large language models (LLMs) locally. Beyond smartphones, entirely new categories of devices are emerging – from smart glasses to humanoid robots to AI-enabled

home appliances – all of which require large memory footprints and expand the addressable market for memory. Taken together, our overall 7-13% CAGR forecast could be conservative.

- **Key risks.** Needless to say, any potential slowdown in the AI industry could negatively impact memory demand, whether stemming from safety concerns, regulatory risks, financial constraints or monetarization setbacks. More importantly, we are closely watching AI innovations that could improve computing efficiency and lower memory requirements. For example, Nvidia's next generation architecture, Vera Rubin Ultra, uses co-packaged optics to connect GPUs across racks at massive scale, thus reducing the HBM content per accelerator. As a result, Nvidia uses HBM 8-Hi for Vera Rubin Ultra vs. HBM 12-Hi for Vera Rubin.

Global DRAM Supply Forecasts (Next 3-5 Years)

Vendor	Est. production bit growth (CAGR)	Percent of global bit supply	Contribution to global bit supply growth (CAGR)
Samsung	12.2%	33%	4.2%
Hynix	17.4%	31%	5.4%
Micron	18.8%	20%	3.8%
CXMT	25.2%	9%	2.3%
Other	12.9%	7%	0.9%
Total		100%	16.6%

Source: companies and RHCC Consulting

- **Background:** Our DRAM supply forecast is constructed in a similar manner. The table above lists the four largest DRAM producers globally, which together account for roughly 93% of industry output and aggregate supply growth of approximately 16.6% p.a.
- **Fundamental barriers prevent new entrants.** Despite surging memory demand (see above), no new large-scale players entering the market today – not even in China. This is mainly because memory manufacturing is extremely capital intensive, with a single leading-edge fab costing USD 15-25bn to build and an additional 8-12% of that amount in annual maintenance. This industry is further protected by deep process know-how, extensive IP protection, a matured ecosystem and high cyclicalities – all of which discourage new entrants and reinforce the moat.
- **Capacity expansion can't be rushed.** Even for established players, supply takes time to build out. New capacity requires multi-year corporate commitment, fab construction, tool installation, yield ramp-up, and ecosystem qualification before meaningful output can be added. For example, SK Hynix (000660.KS / SKHY) broke ground on its M15X fab in April 2024, with tool move-in in 2025-2026, yield ramp-up in 2026-2027 and customer qualification and volume ramp only expected in 2027-2028 – more than three years later. This means that, from the vantage point of 2026, industry supply is largely fixed until 2029 (the next three years).

- **China risk is real but still limited.** The consensus is that China remains the key risk to global memory supply, with CXMT (688825.SH) potentially disrupting the DRAM market and YMTC (soon to be listed) posing a similar threat in NAND through aggressive expansion and competitive pricing. However, there are a few caveats.
 - We estimate that CXMT accounts for only 2% of global supply growth (see table above), far too small to change the severe demand-supply imbalance.
 - CXMT operates at lower geometric density (fewer bits per wafer). While it can reach DDR5-level performance through multi-patterning and larger die sizes, this approach drives lower yields and higher costs, further constraining meaningful supply contribution.
 - CXMT lacks EUV and is therefore limited to 1x/1y (16 nm) nodes, while leaders Samsung and Hynix are well ahead at 1c/1d (10–11 nm). This node gap materially restricts CXMT's ability to produce high-bandwidth memory (HBM), critical to modern AI compute architecture. Therefore, CXMT's impact on HBM is negligible.
 - Having said that, we must monitor China's memory supply ramp-up closely, as Beijing has a proven track record of disrupting global industries through sustained policy support and durable manufacturing advantages.
- **A structural shortage not easily resolved.** The consensus is for a 40-50% shortage in the global memory market, reflecting the severe demand-supply imbalance and supporting a sustained increase in memory prices. Based on our demand forecast of 14-22% CAGR (midpoint: 18%) and supply growth of 16.6%, the overall shortage is unlikely to be resolved over the next 3-5 years, giving us further confidence in the durability of the memory supercycle.

Finally, our Products Team has produced a series of primers on AI data centers and the hardware supply chain, including the critical roles played by memory in the broader compute architecture. Please ping grp_sg_pwm_prod_team@uobkayhian.com (Singapore) or grp_hk_pwm_prod_team-hk@uobkh.com (Hong Kong) if you'd like a copy.

IMPORTANT NOTICE - DISCLOSURES AND DISCLAIMERS

This report is provided subject to various disclosures and disclaimers (the "Disclosures / Disclaimers") which form an integral part of this report and are available at the following link:

<https://research-api.uobkayhian.com/assets/disclaimer/b2112181-0bf2-4c07-af37-3d7129735e61>

or by scanning the QR code below:



The Disclosures / Disclaimers contain important information, including without limitation, (a) exclusions of liability, (b) confidentiality obligations, (c) restrictions on publication, circulation, reproduction, distribution and use of the report, (d) potential conflicts of interest, and (e) disclosures and requirements specific to recipients in the United States and other applicable jurisdictions.

Specifically, this report is intended for general circulation and informational purposes only and does not take into account the specific investment objectives, financial situation, or particular needs of any individual person. It is not intended to constitute personal investment advice or a recommendation to buy or sell any investment product or security. You should independently evaluate the information and, where necessary, seek advice from a qualified financial adviser regarding the suitability of any investment, taking into account your specific objectives, financial situation and needs, before making any investment decision. Analyst certifications required under applicable regulations, including SEC Regulation AC (where relevant), are included in this report.

Recipients of this report must carefully read, review and understand the full Disclosures / Disclaimers before using or relying on any information in this report. By accessing, receiving or using this report, you acknowledge and confirm that you have read, understood, accepted and agreed to be bound by the Disclosures / Disclaimers (as may be amended or updated from time to time) in full."