14 Sep 2026 04:15:22 ET | 29 pages

Global Semiconductors

Continual Learning to Fuel HBM/eSSD Demand (AI Training/Inference)

Deep Dive

Peter Lee AC
Asiya Merchant, CFA AC
Atif Malik AC
Takero Fujiwara AC
Kevin Chen AC
Heath Terry AC
Elizabeth Sun, CFA AC
Laura (Chia Yi) Chen AC
Kelsey Chia, CFA AC
Kyna Wong AC
Masahiro Shibano AC
Papa Sylla AC
Pavan Daswani, CFA AC
Jayden Oh AC
James Bowlin
Michael Cadiz

CITI'S TAKE

Despite memory de-spec and recent AI safety concerns, we believe that continual learning in AI will reshape long-term memory demand. We see it driving memory demand across diverse products including HBM (AI training), server DDR5 and eSSDs (AI inference), leading to memory undersupply worsening throughout 2028E and persisting into 2031E. Amid the structural growth trajectory of the memory market, we highlight major potential beneficiaries.

Entering Era of Continual Learning — We expect continual learning to be a key theme in AI over the next five years, as it would maximize resource efficiency and make AI increasingly intelligent. Continual learning is an AI learning approach that reinforces a model with training on new tasks and knowledge. As a result, continual learning could create persistent need for both model updates and access to previously acquired data, entailing significant usage of memory.

HBM and eSSDs Demand to Increase Exponentially — We expect both training and inference demand to grow simultaneously from 2027E onward, as the continual learning approach will increasingly become sophisticated. This should drive DRAM demand for server DDR5 used in AI inference and HBM, which is needed to acquire new information continuously. In particular, we forecast HBM bit demand to surge +62% YoY to 75.2bn Gb in 2027E and +69% YoY to 127.0bn Gb despite concerns over HBM de-spec and recent AI safety issues. On NAND, we project eSSDs demand to expand as retaining previously learned data requires substantial storage capacity.

27E/28E DRAM S/D Ratio of -8.7%/-9.7% — We forecast global DRAM demand to grow +30%/+35% YoY in 27E/28E driven by continual learning, which should drive server DDR5, and HBM demand simultaneously. With personal/physical AI emerging from 2H28E onwards, we project overall DRAM market undersupply to intensify throughout 2028E, with DRAM S/D ratio projected at -8.7%/-9.7% in 27E/28E amid limited supply growth of +19%/+22% in 27E/28E.

27E/28E NAND S/D Ratio of -6.1%/-5.5% — We project 27E/28E NAND demand to grow +29%/+33% YoY, exceeding supply growth of +21%/+25%, as memory suppliers are expanding DRAM/HBM green-field capacity at the expense of NAND. We anticipate continual learning as well as Nvidia's KV cache offloading trend to drive high-density eSSDs demand where we model +53%/+41% YoY demand growth in 27E/28E. As a result, we see NAND supply as much tighter and expect NAND S/D ratio at -6.1%/-5.5% in 27E/28E. (vs. -0.8% in 2026E)

Memory Top Picks: Samsung Electronics, SK Hynix, Micron, Sandisk, Kioxia; SPE/Materials Top Picks: Montage, AMAT, Lam Research, TES, Eugene Tech, and TechWing — We expect global memory suppliers to strongly benefit from the persistent memory supply shortage driven by continual learning and the potential emergence of personal AI in the mid-to-long-term. As demand is structurally growing, we think memory SPEs and material companies will incrementally grow on the back of massive capex plans and elevated memory production run rates.

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliations



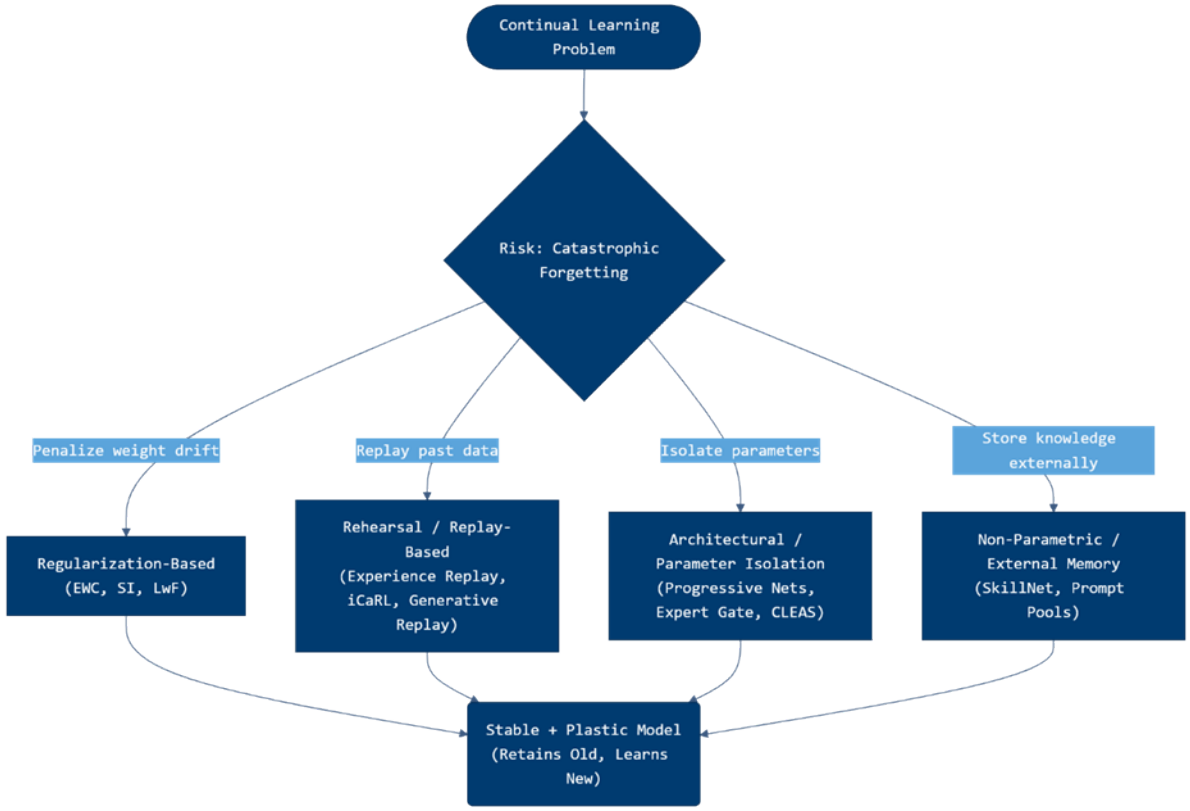
Continual Learning AI Era: HBM, Server DRAM and eSSDs Demand Looks Set to Exponentially Rise

Like the protagonist in the movie *Memento*, today's AI models retain what they have learned but do not automatically incorporate new interactions and data into their model parameters. We expect continual learning to be a key theme in AI over the next five years, as it would maximize resource efficiency and make AI increasingly intelligent.

Continual learning is an AI learning approach that reinforces a model by incrementally training on new tasks and knowledge. It involves training on new data while simultaneously preserving previously learned knowledge.

The key challenge is to balance learning new information with mitigating catastrophic forgetting. In particular, models are trained on a continuous stream of nonstationary data, with the total number of tasks to be learned not fixed in advance. As a result, continued learning could create persistent needs for both ongoing model updates and access to previously acquired data.

Figure 1. Continual Learning



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research

In the memory market, HBM demand grew exponentially in 2022~2024, driven by increasingly complex and advanced training processes, while demand for conventional DRAM and NAND remained relatively soft. Since 2H25, however, demand for server DDR5 and eSSDs has surged as AI inference workloads have grown more complex, while market interest in HBM has cooled.

Figure 2. Evolution of AI Workloads and Memory Demand Implications

	22-1H25	2H25-26E	27E-1H28E	2H28E~
AI Era	AI Training Era	AI Inferencing Era (KV Cache)	Continual Learning Phase 1: AI Server	Continual Learning Phase 2: Physical AI/Personal AI
HBM Demand	Robust HBM Demand	HBM Demand relatively soft	Demand for HBM/Server DRAM/SoCAMM2/eSSDs to increase simultaneously	On-device/Edge Memory Demand to Emerge
DRAM/NAND Demand	Conventional DRAM/NAND relatively muted	Strong Server DRAM/eSSDs		

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

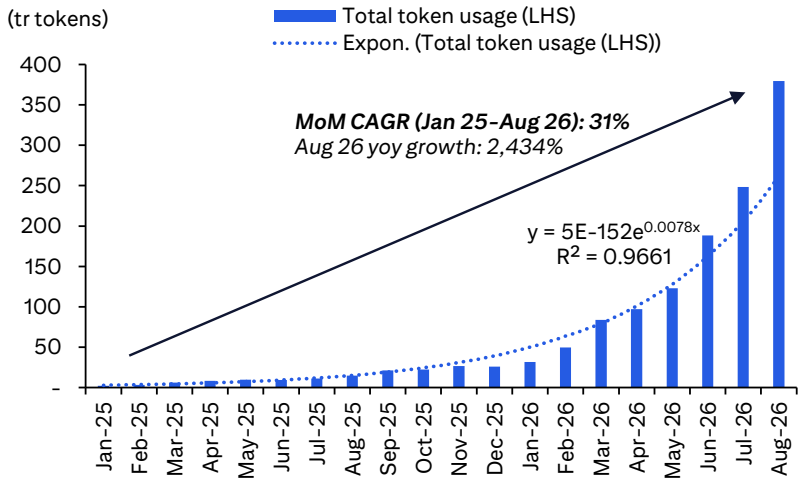
Source: Citi Research

On the other hand, we expect both training and inference demand to grow simultaneously after 2027E, as the continual learning approach itself increasingly becomes sophisticated. This should drive demand for server DDR5 and SoCAMM2 used in AI inference, as well as demand for HBM needed to continuously acquire new information. In addition, eSSDs demand will expand, on our analysis, as retaining previously learned data requires substantial storage capacity.

Beyond continual learning, we believe the next major shift in AI will lie in the era of personal AI/physical AI (See our previous note, Global Semiconductors - The AI Hardware Shift in IT Devices: The Era of Personal AI Server Begins). Specifically, we expect technology enabling edge AI execution and local data management to gain traction from 2H28E, allowing AI devices and robots to maintain persistent memory and adapt to their environments over time.

Meanwhile, we see AI chip architecture increasingly shifting from scale-up to scale-out. Additionally, continual learning should further raise memory intensity, as AI workloads increasingly incorporate continued training and model updates, which drives renewed focus on memory optimization. As shown in Figure 3, however, AI token demand has continued to grow exponentially.

Figure 3. Monthly AI Token Usage Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research, OpenRouter

We believe the recent de-specification issues indicate resource efficiency optimization amid memory shortage rather than demand weakness (See our previous note, Global Semiconductors - AI Chipmakers to Pivot from Scale-Up to Scale-Out Amid HBM Shortage). In particular, we expect HBM bit demand to surge 62% YoY to 75.2bn Gb in 2027E and 69% YoY to 127.0bn Gb in 2028E, double our previous 27E/28E estimates. This should be driven by robust demand growth not only from Nvidia but also from ASICs manufacturers such as Broadcom and Google.

Figure 4. HBM Demand Outlook

		2021	2022	2023	2024	2025	2026E	2027E	2028E
HBM Total(1Gb eq)		736	1,664	3,350	10,096	20,965	46,409	75,157	126,990
	Nvidia	512	1,024	1,728	4,160	8,400	22,003	32,503	40,939
	Broadcom	96	154	365	1,440	3,840	9,002	13,001	41,003
	Google	51	256	576	1,728	2,765	4,702	10,300	14,003
	AMD	64	128	384	832	2,160	4,200	7,000	8,200
	AWS	6	45	144	640	1,512	2,501	2,802	6,391
	Others	6	58	154	1,296	2,288	4,001	9,550	16,455
HBM DRAM (Mpcs, 1Gb eq)		736	1,664	3,350	10,096	20,965	46,409	75,157	126,990
Breakdown	HBM1(1Gb eq.)	368	832	-	-	-	-	-	-
	HBM2(1Gb eq.)	368	832	1,104	-	-	-	-	-
	HBM2E(1Gb eq.)	-	-	2,246	1,338	-	-	-	-
	HBM3(1Gb eq.)	-	-	-	6,560	6,598	1,601	1,364	-
	HBM3E 8hi(1Gb eq.)	-	-	-	1,680	3,840	6,104	1,005	-
	HBM3E 12hi(1Gb eq.)	-	-	-	518	10,526	25,313	14,153	12,628
	HBM4 8hi(1Gb eq.)	-	-	-	-	-	2,316	6,056	4,330
	HBM4 12hi(1Gb eq.)	-	-	-	-	-	11,075	32,107	47,379
	HBM4 16hi(1Gb eq.)	-	-	-	-	-	-	-	-
	HBM4e 8hi(1Gb eq.)	-	-	-	-	-	-	11,305	21,170
	HBM4e 12hi(1Gb eq.)	-	-	-	-	-	-	7,011	33,756
	HBM4e 16hi(1Gb eq.)	-	-	-	-	-	-	2,154	4,557
	HBM5 16hi(1Gb eq.)	-	-	-	-	-	-	-	3,169
Portion	HBM1(%)	50%	50%	0%	0%	0%	0%	0%	0%
	HBM2(%)	50%	50%	33%	0%	0%	0%	0%	0%
	HBM2E(%)	0%	0%	67%	13%	0%	0%	0%	0%
	HBM3(%)	0%	0%	0%	65%	31%	3%	2%	0%
	HBM3E 8hi(%)	0%	0%	0%	17%	18%	13%	1%	0%
	HBM3E 12hi(%)	0%	0%	0%	5%	50%	55%	19%	10%
	HBM4 8hi(%)	0%	0%	0%	0%	0%	5%	8%	3%
	HBM4 12hi(%)	0%	0%	0%	0%	0%	24%	43%	37%
	HBM4 16hi(%)	0%	0%	0%	0%	0%	0%	0%	0%
	HBM4e 8hi(%)	0%	0%	0%	0%	0%	0%	15%	17%
	HBM4e 12hi(%)	0%	0%	0%	0%	0%	0%	9%	27%
	HBM4e 16hi(%)	0%	0%	0%	0%	0%	0%	3%	4%
	HBM5 16hi(1Gb eq.)	0%	0%	0%	0%	0%	0%	0%	2%
Total DRAM Demand(Mpcs)		191,673	193,221	190,282	246,759	305,484	390,057	510,101	630,035
	HBM portion(%)	0.4%	0.9%	1.8%	4.1%	6.9%	11.9%	14.7%	20.2%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates, TrendForce

Figure 5. Customers' HBM Roadmap by GPU and ASIC product

Company	AI Chips	External ASIC/ Design Partner	2024				2025				2026E				2027E				2028E			
			1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26E	4Q26E	1Q27E	2Q27E	3Q27E	4Q27E	1Q28E	2Q28E	3Q28E	4Q28E
Nvidia	H100	In-house	HBM3 8hi 80GB (16GB*5)																			
	GH200 (CPU+GPU)	In-house	HBM3e 8hi 144GB (24GB*6)																			
	H20	In-house	HBM3 8hi 96GB (16GB*6)																			
	H200	In-house	HBM3e 8hi 141GB (24GB*6)																			
	B100/ B200	In-house			HBM3e 8hi 192GB (24GB*8)																	
	GB200 (CPU+GPU)	In-house			HBM3e 8hi 192GB/ 384GB (24GB*8/ 192GB*2)																	
	B200A Ultra (B300A)	In-house					HBM3e 12hi 144GB (36GB*4)															
	Blackwell Ultra (B300)	In-house							HBM3e 12hi 288GB (36GB*8)													
	GB300 (CPU+GPU)	In-house							HBM3e 12hi 576GB (288GB*2)													
	Rubin	In-house									HBM4 12hi 288BG (36GB*8)											
	Rubin Ultra	In-house															HBM4 8hi 192GB (24GB*8)						
	Feynman	In-house																	custom HBM (possibly NVHBM)			

Company	AI Chips	External ASIC/ Design Partner	2024				2025				2026				2027				2028			
			1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26	1Q27	2Q27	3Q27	4Q27	1Q28	2Q28	3Q28	4Q28
AMD	MI200	In-house	HBM2e 8hi 128GB (16GB*8)																			
	MI300X	In-house	HBM3 12hi 192GB (24GB*8)																			
	MI300A (CPU+GPU)	In-house	HBM3 8hi 128GB (16GB*8)																			
	MI325X	In-house				HBM3e 8hi 256GB (32GB*8)																
	MI350	In-house						HBM3e 12hi 288GB (36GB*8)														
	MI400	In-house									HBM4 12hi 432GB (36GB*12)											
	MI500	In-house															HBM4e (TBD) (within 2027)						
	MI600	In-house																	HBM5 (TBD; within 2028)			

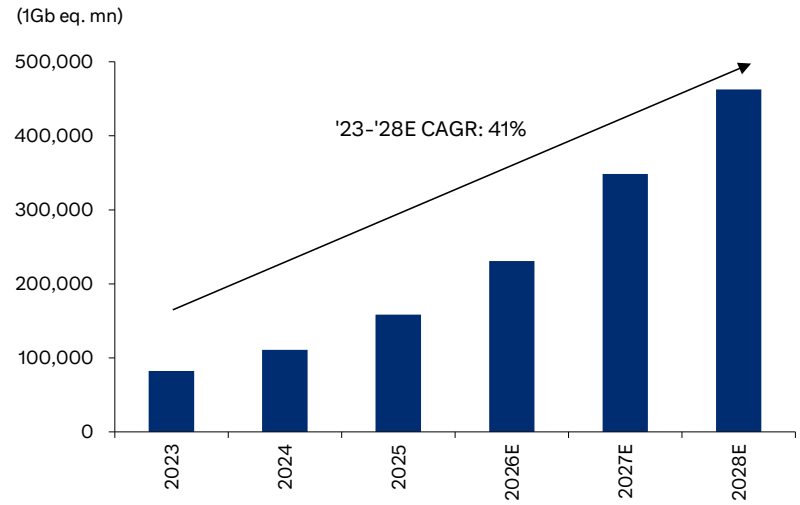
Company	AI Chips	External ASIC/ Design Partner	2024				2025				2026				2027				2028			
			1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26	1Q27	2Q27	3Q27	4Q27	1Q28	2Q28	3Q28	4Q28
AWS	Trainium 1	Alchip	HBM2e 32GB (16GB*2)																			
	Inferentia 2	Alchip	HBM2e 32GB (16GB*2)																			
	Trainium 2	Marvell				HBM3 96GB (24GB*4)																
	Trainium 2.5	Marvell							HBM3e 144GB (36GB*4)													
	Trainium 3	Alchip										HBM3e 12hi 144GB (36GB*4)										
	Trainium 4	Alchip, possibly Marvell																HBM4 12hi 288GB (36GB*8)				
Google	TPU v5i	Broadcom	HBM2e 16GB (8GB*2)																			
	TPU v5t	Broadcom	HBM2e 96GB (16GB*6)																			
	TPU v6i	Broadcom				HBM3 32GB (16GB*2)																
	TPU v7t	Broadcom							HBM3e 192GB (24GB*8)													
	TPU v8i	Broadcom											HBM3e 12hi 288GB (36GB*8)									
	TPU v8t	MediaTek											HBM3e 12hi 216GB (36GB*6)									
	TPU v9 Base (Humufish)	Marvell																	HBM4 (TBD)			
	TPU v9i (Triggerfish)	MediaTek																	HBM4e 8hi 384GB (32GB*12)			
	TPU v9t	MediaTek																	HBM4e 8hi 384GB (32GB*12)			
	TPU v10 (Icefish)	MediaTek / AMD																				TBD
	Merope (variant of TPU v10 or LPU)	Marvell / AMD																				
Microsoft	Maia 100	GUC	HBM2e 64GB (16GB*4)																			
	Maia 200	GUC									HBM3e 12hi 216GB (36GB*6)											
	Maia 300	Marvell															HBM4 12hi 288GB (36GB*8)						
Meta	MTIA 200	Broadcom			LPDDR5 128GB																	
	MTIA 300	Broadcom							HBM3e 216GB (36GB*6)													
	MTIA 400	Broadcom										HBM3e 12hi 288GB (36GB*8)										
	MTIA 450	Broadcom													HBM4 12hi 288GB (36GB*8)							
	MTIA 500	Broadcom																HBM4e 12hi 384GB (48GB*8)				
	MTIA 600	Broadcom; possibly Alchip / GUC / MediaTek																	TBD (2028E+)			
OpenAI	Jalapeño	Broadcom													HBM4 12hi 216 GB (36GB*6)							
	TBD	Broadcom																	TBD (2028E+)			
Apple	Baltra	Broadcom																TBD				
Anthropic	TBD	MatX																	TBD (2028E+)			

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: TrendForce, Companies, Citi Research Estimates

We forecast AI inference demand to rise continuously due to the impact of continual learning. Specifically, we expect server DRAM demand to grow more than 50% YoY in 2027E, with eSSDs demand rising ~50% YoY. Server DRAM demand should rise alongside growing CPU demand. We believe this improving multi-year demand visibility is also consistent with memory customers increasingly extending LTAs from 3 to 5 years.

Figure 6. Server DRAM Demand Outlook

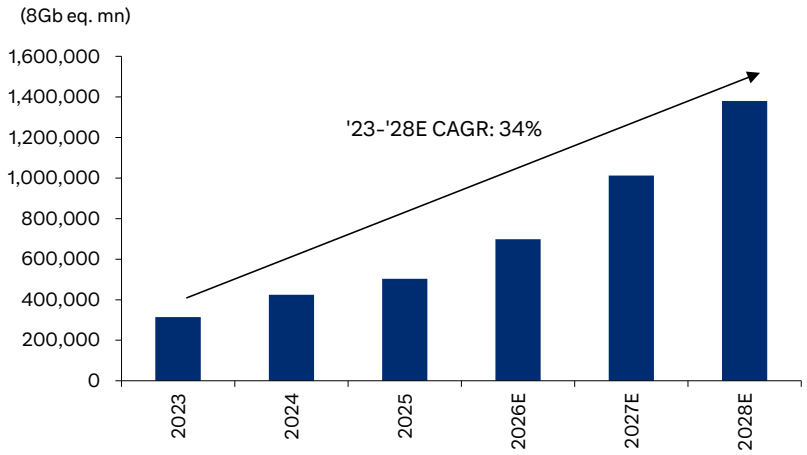


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research Estimates; TrendForce

In terms of NAND, we expect growing demand for advanced solutions such as XL Flash and HBF with storage-complementary products such as CXL. We also expect near GPU storage solutions using QLC SSDs to gain traction as AI systems require larger pools of storage closer to accelerators to enhance memory retention.

By contrast, consumer NAND demand could soften amid weaker mobile and PC market conditions, though we expect this to be offset by rising AI inference-related NAND demand. In addition, AI data centers, particularly in China, are increasingly considering SSDs over HDDs, potentially creating further upside to eSSDs demand from 2027E.

Figure 7. SSD Demand Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates; TrendForce

In conclusion, we believe continual learning will drive simultaneous growth in both training and inference related memory demand from 2027E. HBM should benefit from ongoing training and AI model updates, server DRAM/SoCAMM2 from rising inference demand and CPU workloads, and eSSDs from persistent storage requirements. As a result, we expect AI memory demand to strengthen beyond current market expectations, with demand for server DRAM/SoCAMM2/eSSDs particularly robust in 2027E.

Memory Supply/Demand Outlook

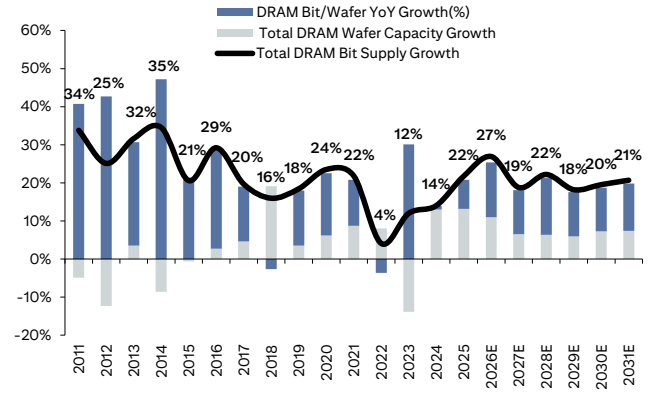
Sustained demand and constrained supply to drive an undersupplied memory market in 27E

We expect memory market tightness to intensify in 2027E, driven by limited supply growth given higher HBM production volume allocation and increasing difficulty in tech migration.

In terms of DRAM, we expect DRAM supply growth of ~19% in 2027E, constrained by higher HBM capacity allocation and slowing node migration. On the demand side, despite HBM de-specification issues, we forecast ~30% growth, driven by continued HBM demand from AI chipmakers' shift toward scale-out from scale-up and robust server DRAM demand.

Regarding NAND, we expect NAND demand to grow ~29% in 2027E amid constrained industry capacity. This is supported by rising AI inference demand and incremental eSSDs demand from offloading as lower HBM content shifts more memory workloads to external storage.

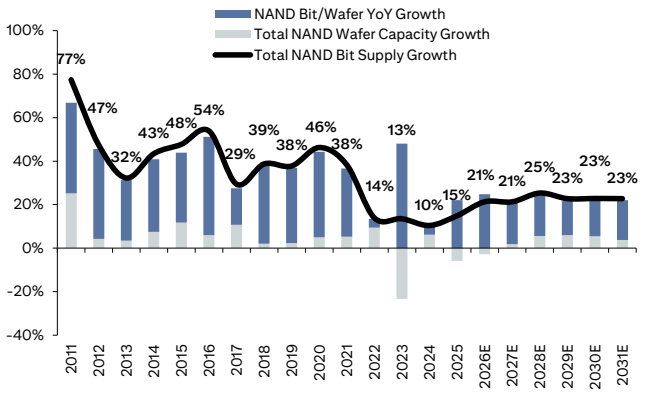
Figure 8. DRAM – Bit Supply Growth Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

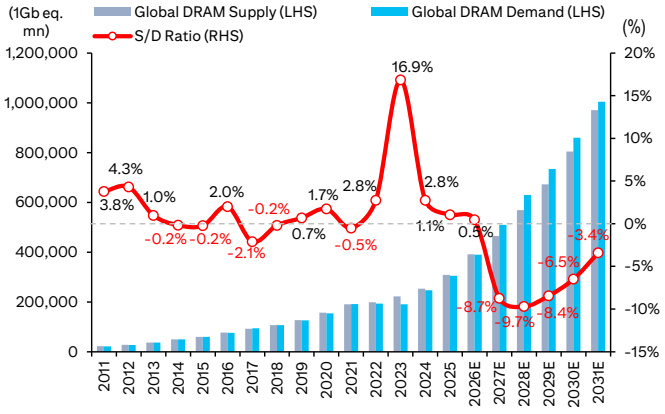
Figure 9. NAND – Bit Supply Growth Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

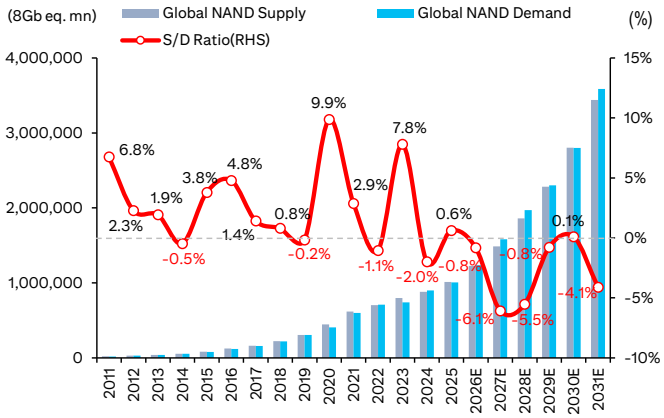
Figure 10. DRAM - Supply/Demand Outlook (2027E S/D Ratio: -8.7%)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates, DrameXchange

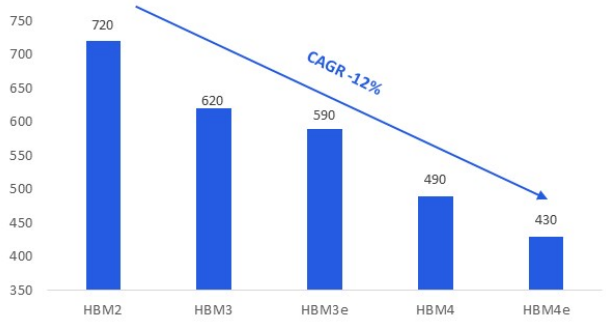
Figure 11. NAND - Supply/Demand Outlook (2027E S/D Ratio: -6.1%)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

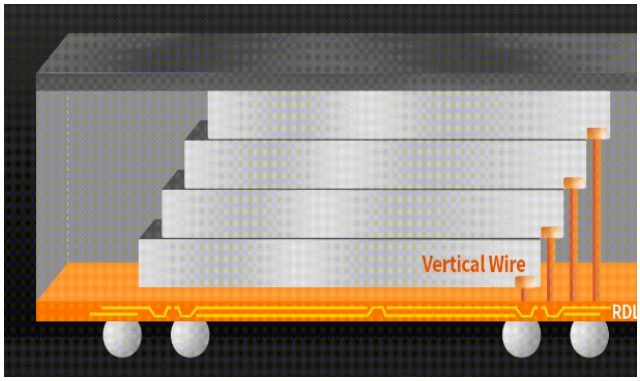
Source: Citi Research Estimates, DrameXchange

Figure 12. # of Dies per HBM Wafer



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research estimates; TrendForce

Figure 13. Wide I/O (Mini HBM) for Mobile with VFO



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: SK Hynix, Citi Research

DRAM Supply/Demand Outlook

DRAM to be undersupplied in 27E

In 2027E, we forecast overall DRAM S/D to be undersupplied driven by [1] continued HBM demand from AI chipmakers' shift toward scale-out, [2] incremental demand for server DRAM, [3] constrained supply growth due to higher HBM capacity allocation, and [4] slowing node migration.

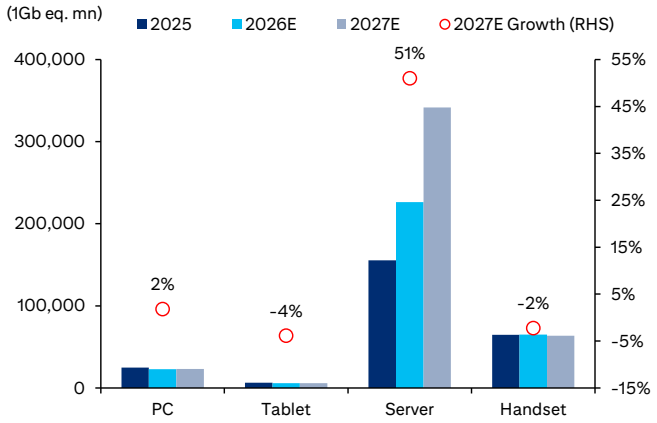
DRAM demand outlook: +30.2% YoY in 27E

We expect 27E DRAM bit demand to grow +30.2% YoY, significantly outpacing bit supply growth of +18.8% YoY, driving the industry S/D ratio to -8.7% from +0.5% in 26E. On the demand side, we expect robust growth to be driven by continued HBM demand as AI chipmakers shift toward scale-out, together with strong server DRAM demand, while PC demand remains relatively soft and mobile demand sees a modest recovery.

We view recent HBM de-specification issues as a response to constrained HBM availability rather than weaker underlying demand. Given AI chipmakers are pursuing a larger number of accelerator units amid limited HBM supply, we believe it will support continued HBM demand at a system level.

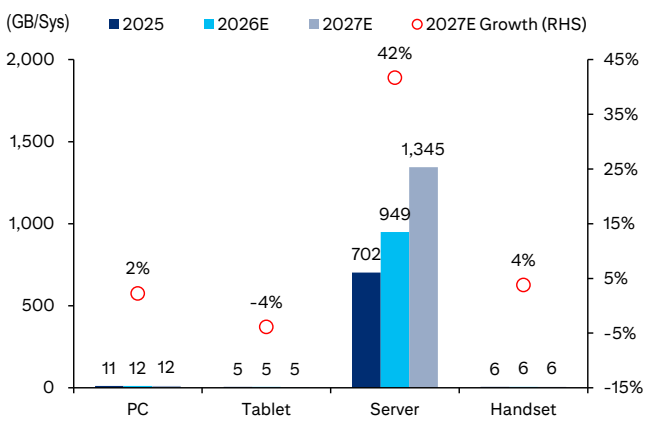
At the same time, server DRAM will benefit from memory workload offloading as lower HBM content should increase KV-cache and provide incremental demand outside HBM. Specifically, we forecast server DRAM demand to grow from 226.3bn 1Gb equivalent units in 2026E to 341.7bn units in 2027E (+51% YoY). In 2027E, we expect server will continue to account for the largest DRAM demand at 67%.

Figure 14. DRAM Set Application Forecasts



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research Estimates, DRAMeXchange

Figure 15. DRAM Avg. Content Adoption Forecasts

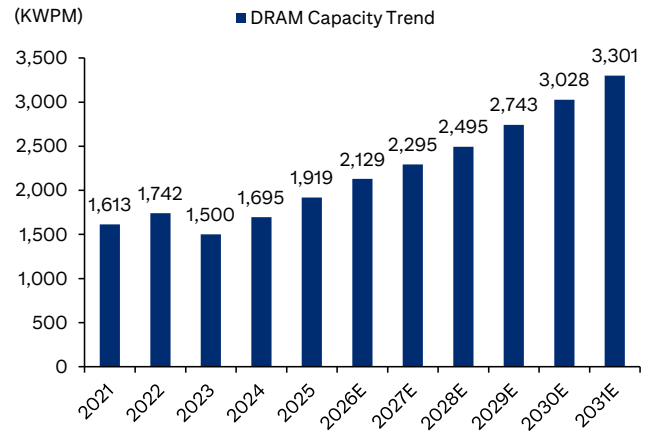


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research Estimates, DRAMeXchange

DRAM Capex outlook: +8% YoY in 27E

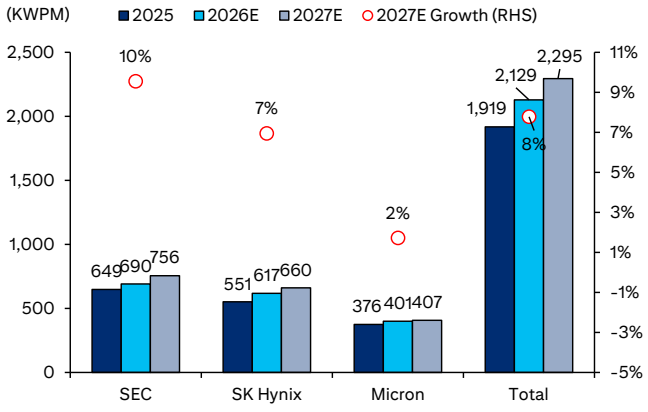
On the supply side, we expect average DRAM industry wafer capacity to increase only ~8% YoY to 2,295KWPM in 2027E from 2,129KWPM in 2026E. By supplier, we expect Samsung Electronics' DRAM greenfield capacity to increase by ~70KWPM in end 4Q27E vs. end-4Q26E. For SK Hynix, we project its greenfield capacity to increase by ~70KWPM in end 4Q27E vs. end-4Q26E

Figure 16. Average DRAM Industry Capacity Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research Estimates, TrendForce

Figure 17. Average DRAM Capacity Change by Company

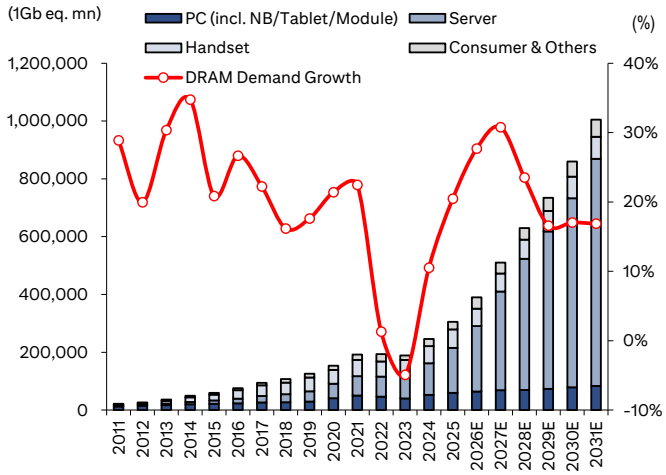


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research Estimates, TrendForce

DRAM supply growth: +18.8% in 27E

We forecast DRAM industry bit supply growth of +18.8% YoY in 2027E. By company, we forecast Samsung to realize YoY DRAM supply B/G of +17.6% in 2027E, Hynix to increase its DRAM B/G by +20.6% and expect Micron to increase its supply B/G by +21.1% in 2027E. While industry DRAM Capex rises, a large portion of incremental investment could continue to be directed toward HBM. Additionally, slower technology migration limits bit growth from existing wafer capacity. Moreover, the long lead time for greenfield additions should prevent the sharp increase in bit supply. As a result, we expect DRAM supply to remain constrained throughout 2027E.

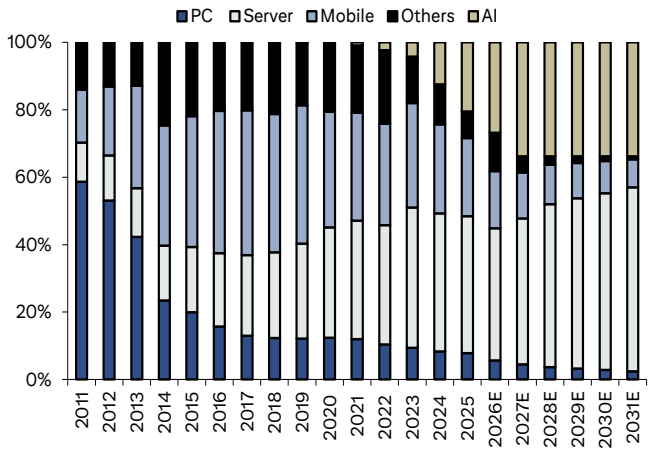
Figure 18. DRAM Demand Forecast by Application



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates, TrendForce

Figure 19. DRAM Demand Portion by Application



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates, TrendForce

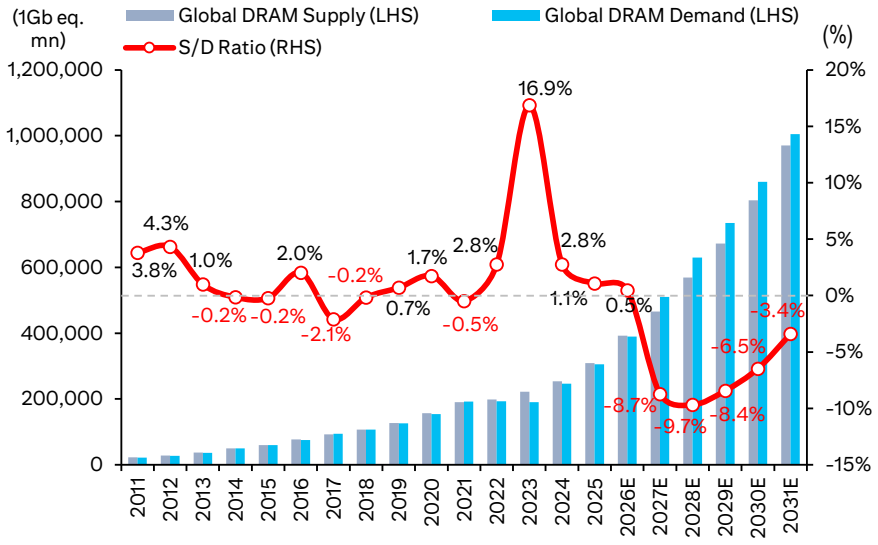
Figure 20. Quarterly/Annual DRAM Supply / Demand Forecasts

(1Gb eqiv. mn units)	1Q26	2Q26	3Q26E	4Q26E	1Q27E	2Q27E	3Q27E	4Q27E	1Q28E	2Q28E	3Q28E	4Q28E	2026E	2027E	2028E
DRAM Supply	88,212	96,103	102,411	105,166	105,274	111,354	121,625	127,307	126,993	134,665	150,071	157,240	391,891	465,560	568,969
DRAM Demand	88,049	95,749	102,294	103,966	108,824	118,010	135,373	147,894	142,892	149,641	165,370	172,132	390,057	510,101	630,035
DRAM S/D Ratio (%)	0.2%	0.4%	0.1%	1.2%	-3.3%	-5.6%	-10.2%	-13.9%	-11.1%	-10.0%	-9.3%	-8.7%	0.5%	-8.7%	-9.7%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

Figure 21. DRAM Supply / Demand Forecasts



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research Estimates

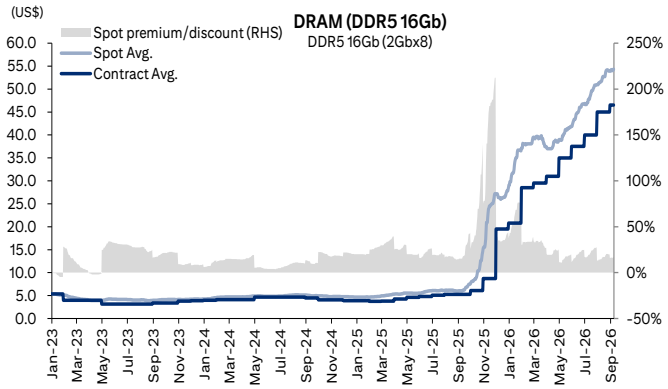
DRAM price outlook

DRAM ASP outlook: +23.1% YoY in 27E

Given the sharp deterioration in the DRAM S/D balance, we expect favorable pricing conditions to persist through 2027E. We forecast blended DRAM ASP to increase +23.1% YoY in 2027E, following an outsized +242.4% YoY increase in 2026E. While the magnitude of price increases should moderate from the exceptional 2026E level, we expect persistent supply tightness in both HBM and conventional DRAM to support further annual pricing upside.

By quarter, we expect blended DRAM ASP to move +20% QoQ in 3Q26E, and +13% QoQ in 4Q26E. On an annual basis, we expect annual average DRAM ASP to rise +242.4%/+23.1% in 2026E/2027E.

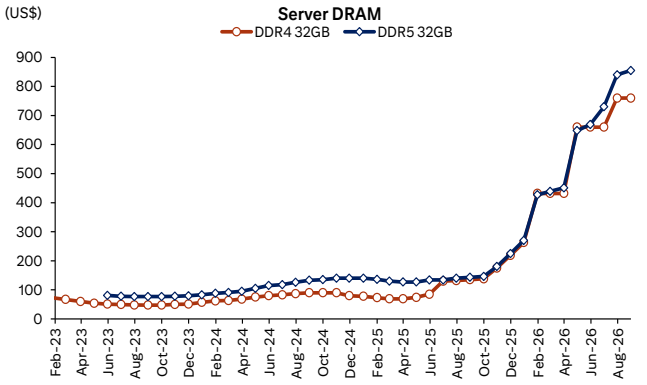
Figure 22. DDR5 PC DRAM Price Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

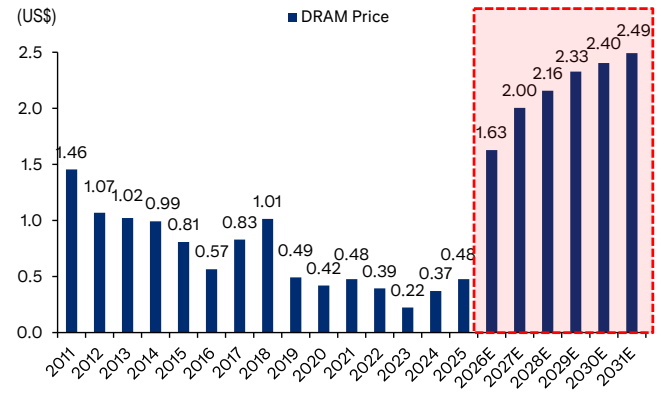
Figure 23. Server DRAM Price Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

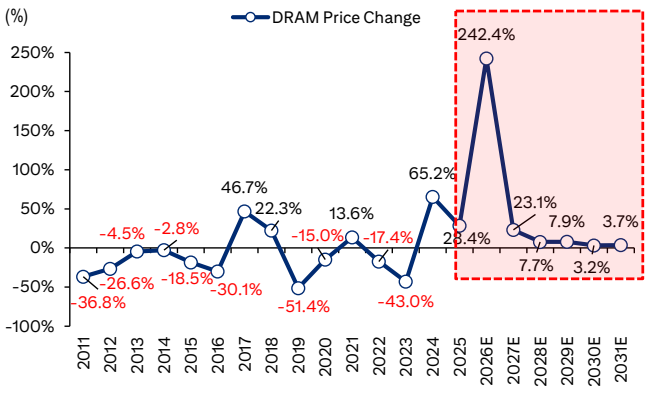
Figure 24. DRAM Price Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

Figure 25. DRAM Price Change Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

HBM supply/demand outlook

HBM S/D: -21% in 27E

We expect the HBM market to remain significantly undersupplied in 2027E despite the continued capacity expansion by memory suppliers. Our HBM demand forecast rises +62% YoY from 46.4bn 1Gb-equivalent units in 2026E to 75.2bn in 2027E, while supply increases from 36.1bn to 59.3bn, +64% YoY. As a result, we forecast the HBM S/D ratio to stay negative at -21% in 2027E, broadly unchanged vs -22% in 2026E. As the HBM demand from accelerated ASIC shipments is expected to grow further in 2028E, we project the 28E HBM S/D to worsen at -36%.

At the system level, total HBM unit demand will increase from 178.5mn units in 2026E to 269.1mn units in 2027E, we estimate, supported by continued growth from Nvidia as well as expanding ASIC demand. Therefore, we continue to view recent HBM de-specification issues as a practical way of maximizing accelerator shipments amid constrained supply rather than a sign of weaker HBM demand.

Figure 26. HBM Supply and Demand Outlook

		2021	2022	2023	2024	2025	2026E	2027E	2028E
HBM	Demand(Mpcs, 1Gb eq)	736	1,664	3,350	10,096	20,965	46,409	75,157	126,990
	Supply(Mpcs, 1Gb eq)	752	1,699	3,104	9,200	17,068	36,137	59,347	80,936
	S/D Ratio	2%	2%	-7%	-9%	-19%	-22%	-21%	-36%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates; TrendForce

Figure 27. HBM Supply Outlook

		2021	2022	2023	2024	2025	2026E	2027E	2028E
Year-End TSV Capacity(Kwafer/month)		32	53	106	180	280	420	700	870
	Samsung	20	28	49	70	110	155	240	300
	SK Hynix	8	20	50	75	110	150	270	290
	Micron	4	6	7	35	55	95	140	200
	CXMT	-	-	-	-	5	20	50	80
HBM Supply(Mpcs, 1Gb eq)		752	1,699	3,104	9,200	17,068	36,137	59,347	80,936
	Samsung	483	896	1,424	3,027	5,788	11,000	25,000	33,006
	SK Hynix	206	661	1,481	4,887	8,543	16,999	21,002	27,025
	Micron	63	142	200	1,285	2,719	8,034	13,009	20,001
	CXMT	-	-	-	-	19	104	336	904
M/S		100%	100%	100%	100%	100%	100%	100%	100%
	Samsung	64%	53%	46%	33%	34%	30%	42%	41%
	SK Hynix	27%	39%	48%	53%	50%	47%	35%	33%
	Micron	8%	8%	6%	14%	16%	22%	22%	25%
	CXMT	0%	0%	0%	0%	0%	0%	1%	1%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates; TrendForce

Figure 28. HBM Demand Outlook

		2021	2022	2023	2024	2025	2026E	2027E	2028E
HBM Total(1Gb eq)		736	1,664	3,350	10,096	20,965	46,409	75,157	126,990
	Nvidia	512	1,024	1,728	4,160	8,400	22,003	32,503	40,939
	Broadcom	96	154	365	1,440	3,840	9,002	13,001	41,003
	Google	51	256	576	1,728	2,765	4,702	10,300	14,003
	AMD	64	128	384	832	2,160	4,200	7,000	8,200
	AWS	6	45	144	640	1,512	2,501	2,802	6,391
	Others	6	58	154	1,296	2,288	4,001	9,550	16,455
HBM DRAM (Mpcs, 1Gb eq)		736	1,664	3,350	10,096	20,965	46,409	75,157	126,990
Breakdown	HBM1(1Gb eq.)	368	832	-	-	-	-	-	-
	HBM2(1Gb eq.)	368	832	1,104	-	-	-	-	-
	HBM2E(1Gb eq.)	-	-	2,246	1,338	-	-	-	-
	HBM3(1Gb eq.)	-	-	-	6,560	6,598	1,601	1,364	-
	HBM3E 8hi(1Gb eq.)	-	-	-	1,680	3,840	6,104	1,005	-
	HBM3E 12hi(1Gb eq.)	-	-	-	518	10,526	25,313	14,153	12,628
	HBM4 8hi(1Gb eq.)	-	-	-	-	-	2,316	6,056	4,330
	HBM4 12hi(1Gb eq.)	-	-	-	-	-	11,075	32,107	47,379
	HBM4 16hi(1Gb eq.)	-	-	-	-	-	-	-	-
	HBM4e 8hi(1Gb eq.)	-	-	-	-	-	-	11,305	21,170
	HBM4e 12hi(1Gb eq.)	-	-	-	-	-	-	7,011	33,756
	HBM4e 16hi(1Gb eq.)	-	-	-	-	-	-	2,154	4,557
	HBM5 16hi(1Gb eq.)	-	-	-	-	-	-	-	3,169
Portion	HBM1(%)	50%	50%	0%	0%	0%	0%	0%	0%
	HBM2(%)	50%	50%	33%	0%	0%	0%	0%	0%
	HBM2E(%)	0%	0%	67%	13%	0%	0%	0%	0%
	HBM3(%)	0%	0%	0%	65%	31%	3%	2%	0%
	HBM3E 8hi(%)	0%	0%	0%	17%	18%	13%	1%	0%
	HBM3E 12hi(%)	0%	0%	0%	5%	50%	55%	19%	10%
	HBM4 8hi(%)	0%	0%	0%	0%	0%	5%	8%	3%
	HBM4 12hi(%)	0%	0%	0%	0%	0%	24%	43%	37%
	HBM4 16hi(%)	0%	0%	0%	0%	0%	0%	0%	0%
	HBM4e 8hi(%)	0%	0%	0%	0%	0%	0%	15%	17%
	HBM4e 12hi(%)	0%	0%	0%	0%	0%	0%	9%	27%
	HBM4e 16hi(%)	0%	0%	0%	0%	0%	0%	3%	4%
	HBM5 16hi(1Gb eq.)	0%	0%	0%	0%	0%	0%	0%	2%
Total DRAM Demand(Mpcs)		191,673	193,221	190,282	246,759	305,484	390,057	510,101	630,035
	HBM portion(%)	0.4%	0.9%	1.8%	4.1%	6.9%	11.9%	14.7%	20.2%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research Estimates; TrendForce

NAND Supply/Demand Outlook

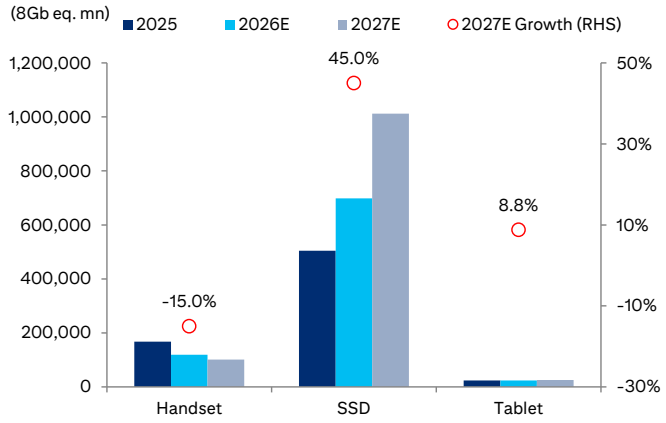
NAND S/D ratio: -6.1% in 27E

We forecast NAND demand to grow +29.1% YoY in 2027E, ahead of supply growth of +21.2% YoY, tightening the NAND S/D ratio to -6.1% from -0.8% in 2026E. We expect demand growth to be increasingly driven by AI inference and eSSDs applications, offsetting relatively softer consumer NAND demand.

NAND demand mainly driven by eSSDs

In particular, we expect eSSDs demand to grow at +52.9% YoY (total SSD demand growth: +45.0% YoY) in 2027E, given strong demand from HBM offloading, as lower HBM content per accelerator shifts a greater portion of KV-cache and other memory workloads toward external storage. Stronger AI inference demand should also continue to support higher-capacity QLC eSSDs adoption at AI servers. This is consistent with recent memory suppliers plans that NAND production is increasingly being allocated toward higher-value eSSDs products despite softer consumer demand.

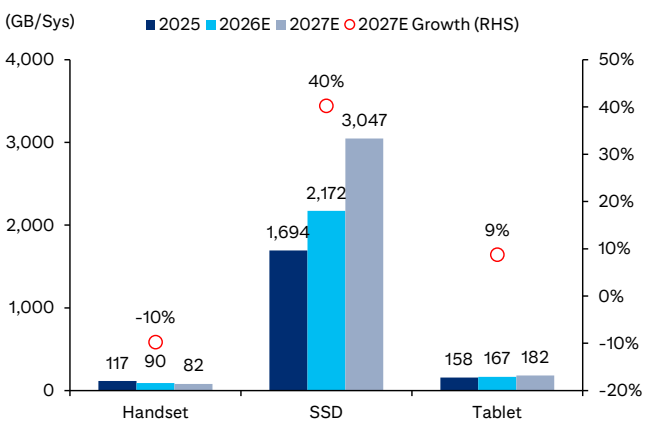
Figure 29. NAND Set Application Forecasts



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

Figure 30. NAND Avg. Content Adoption Forecasts



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

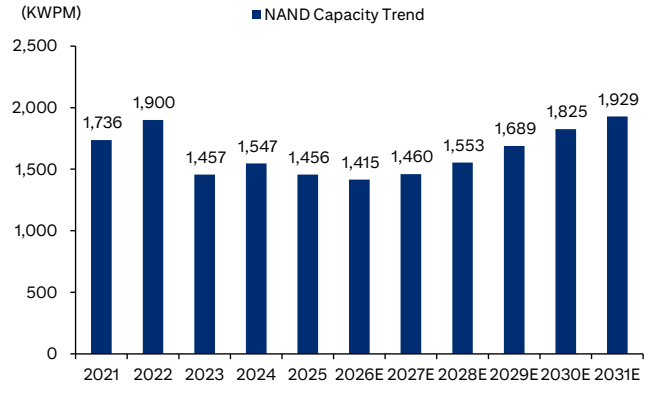
Source: DRAMeXchange, Citi Research estimates

NAND capacity outlook: +3.2% YoY in 27E

Meanwhile, NAND capacity expansion remains relatively disciplined. We forecast average industry average NAND wafer capacity to increase only ~3.2% YoY to 1,460KWPM in 2027E from 1,415KWPM in 2026E. By supplier, we project Samsung Electronics' average NAND capacity to decrease by -4.7% to 331KWPM in 2027E. We forecast average NAND capacity of both Hynix and Micron to remain unchanged in 2027E.

Despite stronger pricing and AI-related demand, we expect suppliers to remain cautious on incremental greenfield NAND capacity, supporting a sustained supply deficit in 2027E.

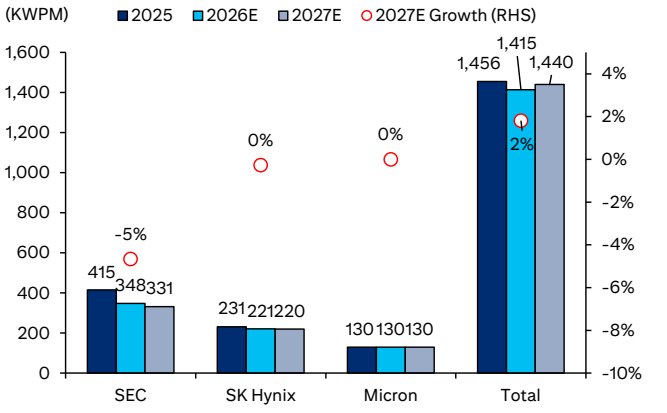
Figure 31. Average NAND Industry Capacity Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: DRAMeXchange, Citi Research estimates

Figure 32. NAND Capacity Change by Company



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

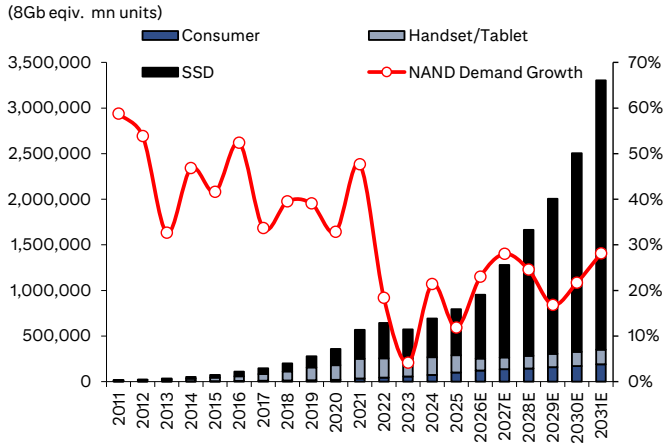
Source: DRAMeXchange, Citi Research estimates

NAND demand growth: +29.1% YoY in 27E

We forecast NAND industry supply/demand growth at +21.2% YoY/+29.1% YoY in 2027E, driven by strong eSSDs demand and limited capex investment. In bit terms, we expect NAND demand to reach 1,486bn units of 8Gb equivalent in 2027E, driven by incremental eSSDs demand on the back of strong AI inference demand and incremental eSSDs demand from KV Cache offloading.

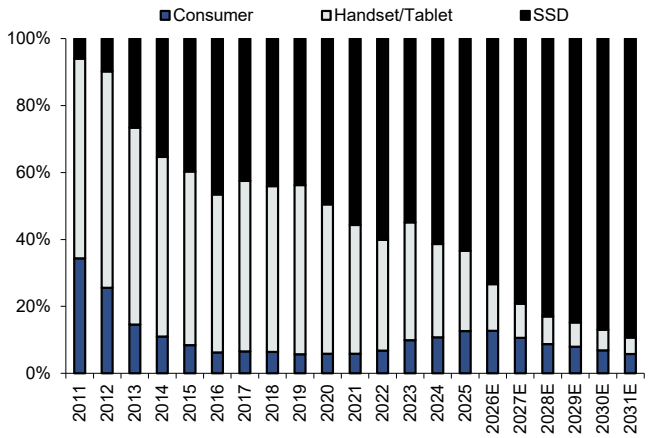
Looking ahead, we expect NAND demand to increasingly favor high-bandwidth, high-density, and AI-optimized solutions. In terms of NAND, we expect growing demand for advanced solutions such as XL Flash and HBF, alongside storage-complementary products such as CXL, as major suppliers position these technologies to utilize for AI workloads. As AI systems require larger pools of storage closer to accelerators to enhance memory retention, we expect near-GPU storage solutions using QLC SSDs to gain traction, reinforcing NAND's expanding role across in AI memory architectures.

Figure 33. NAND Demand Forecast by Application



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: DRAMeXchange, Citi Research estimates

Figure 34. NAND Demand Portion by Application

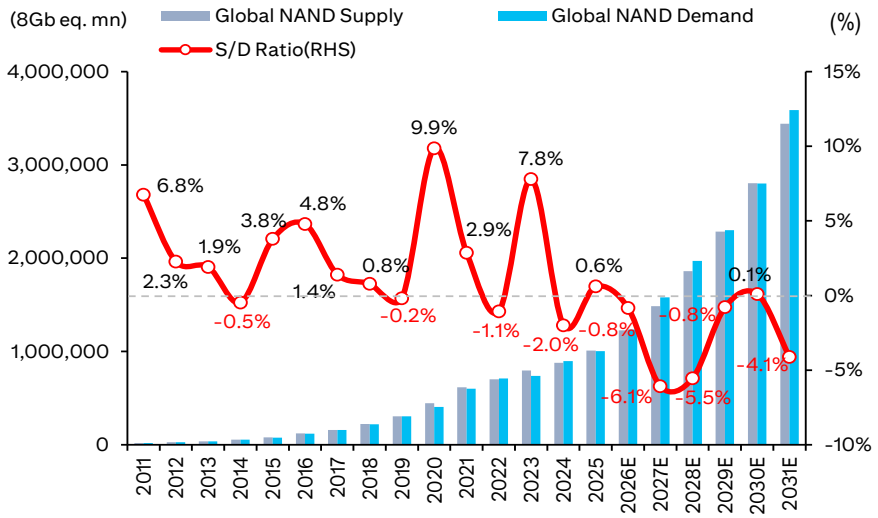


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research; TrendForce

NAND S/D Ratio -6.1% in 2027E

We forecast the NAND S/D ratio to become more undersupplied, widening to -6.1% in 2027E from -0.8% in 2026E, driven by robust eSSDs demand against limited NAND supply growth on constrained capex.

Figure 35. NAND Supply/Demand Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: DRAMeXchange, Citi Research Estimates

Figure 36. NAND Supply / Demand Forecasts

(8Gb eqiv. mn units)	1Q26	2Q26	3Q26E	4Q26E	1Q27E	2Q27E	3Q27E	4Q27E	1Q28E	2Q28E	3Q28E	4Q28E	2026E	2027E	2028E
NAND Supply	279,454	297,395	318,953	329,554	325,920	354,904	394,187	410,625	411,362	445,526	493,595	510,598	1,225,356	1,485,636	1,861,081
NAND Demand	285,896	299,454	319,392	330,808	342,803	381,628	416,256	441,104	448,510	485,207	518,871	517,540	1,235,550	1,581,792	1,970,128
NAND S/D Ratio	-2.3%	-0.7%	-0.1%	-0.4%	-4.9%	-7.0%	-5.3%	-6.9%	-8.3%	-8.2%	-4.9%	-1.3%	-0.8%	-6.1%	-5.5%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

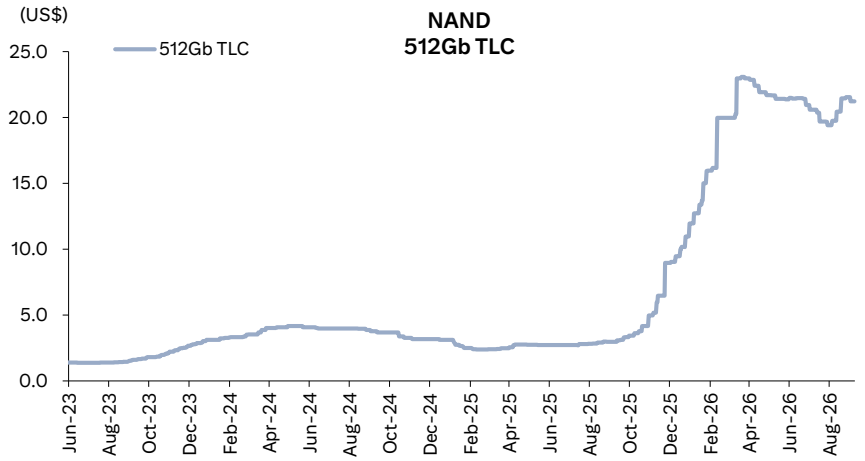
Source: DRAMeXchange, Citi Research Estimates

NAND price outlook

NAND ASP to rise +45.3% YoY in 27E

We also expect NAND pricing to remain constructive as AI-led eSSDs demand strengthens while supply expansion remains limited. By quarter in 2026E, we project blended NAND ASP to grow +34%/ +15% in 3Q/4Q26E. On an annual basis, NAND ASP is expected by us to rise +45.3% YoY in 2027E, following +236.2% YoY in 2026E. Despite weaker consumer NAND demand, sustained eSSDs growth and a tighter industry S/D balance should support continued pricing strength.

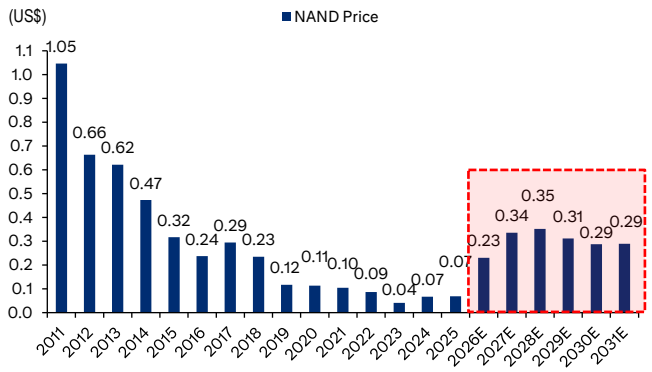
Figure 37. NAND Wafer (512Gb TLC) Spot Price Trend



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Bloomberg, Citi Research; Note: Based on Spectrum's 512Gb spot price index

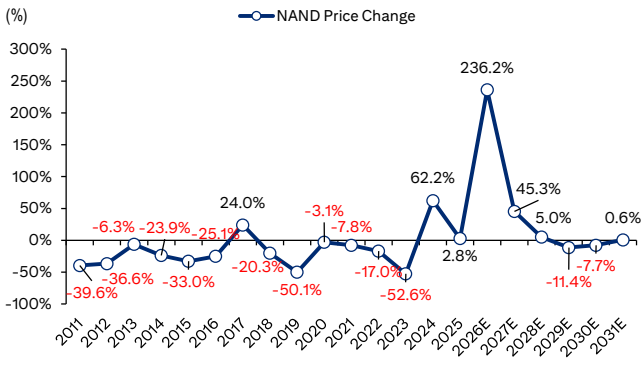
Figure 38. NAND ASP Outlook



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research; TrendForce

Figure 39. NAND ASP Change



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research; TrendForce

Total Memory Capex to Increase +46.5% YoY to US$80.4bn in 2027E

Memory WFE Capex to rise +46.5% YoY in 27E

We forecast global memory Capex to increase +46.5% YoY to US$80.4bn in 2027E, accelerating from +39.5% YoY growth to US$54.9bn in 2026E, as memory suppliers respond to structurally stronger AI-driven demand.

DRAM WFE Capex +51.6% YoY in 27E

NAND WFE Capex +34.2% YoY in 27E

We expect DRAM Capex to rise +51.6% YoY to US$58.6bn in 2027E, led by Samsung Electronics at US$20.6bn, SK Hynix at US$17.5bn, and Micron at US$15.8bn. We expect a meaningful portion of incremental DRAM investment to remain directed toward HBM capacity and advanced DRAM production, limiting the supply response in conventional DRAM.

On the NAND side, we forecast Capex to increase +34.2% YoY to US$21.8bn in 2027E, versus US$16.3bn in 2026E. We forecast Samsung/SK Hynix/ Kioxia/Sandisk/Micron NAND Capex at US$7.9bn/US$3.1bn/US$3.7bn/ US$2.9bn/US$2.1bn, respectively. Despite the sizable headline Capex increase, we believe the long lead time for additional capacity means that higher spending should not be sufficient to close the 2027E supply-demand gap.

Figure 40. Total DRAM+NAND CAPEX

Total DRAM+NAND Capex (US$mn)	2025	2026E	2027E	2028E	2029E	2030E	2031E
Total DRAM+NAND Capex	**39,364**	**54,905**	**80,417**	**118,622**	**168,795**	**234,043**	**322,531**
DRAM+NAND Capex YoY %	7.5%	39.5%	46.5%	47.5%	42.3%	38.7%	37.8%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Companies, Citi Research estimates

Figure 41. DRAM CAPEX by Player

DRAM Capex by Player (US$mn)	2025	2026E	2027E	2028E	2029E	2030E	2031E
Samsung	9,595	13,625	20,574	32,918	48,389	66,293	90,822
SK Hynix	7,211	10,889	17,532	27,174	39,402	54,769	76,129
Micron	7,100	10,630	15,828	22,317	32,360	47,570	66,598
Total DRAM Capex (US$mn)	**26,875**	**38,640**	**58,594**	**89,150**	**129,910**	**182,766**	**254,031**
DRAM Capex YoY %	15.2%	43.8%	51.6%	52.1%	45.7%	40.7%	39.0%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Companies, Citi Research estimates

Figure 42. NAND CAPEX by Player

NAND Capex by Player (US$mn)	2025	2026E	2027E	2028E	2029E	2030E	2031E
Samsung	4,672	5,466	7,926	11,651	17,477	24,468	34,989
SK Hynix	1,987	2,384	3,099	4,432	6,205	8,997	12,596
Kioxia	2,069	3,200	3,700	4,200	4,500	4,800	5,000
Sandisk	1,324	2,202	2,900	2,999	3,485	4,008	4,609
Micron	1,500	1,650	2,100	2,940	3,969	5,755	8,057
Total NAND Capex (US$mn)	**12,488**	**16,264**	**21,823**	**29,471**	**38,885**	**51,277**	**68,500**
NAND Capex YoY %	-5.9%	30.2%	34.2%	35.0%	31.9%	31.9%	33.6%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Companies, Citi Research estimates

Companies mentioned in an investment context:

Applied Materials Inc (AMAT.O; US$456.49; 1; 11 Sep 26; 16:00) | EugeneTechnology (084370.KQ; W131500.0; 1; 11 Sep 26; 15:45) | Kioxia Holdings (285A.T; ¥54030.0; 1H; 11 Sep 26; 15:30) | Lam Research Corp (LRCX.O; US$298.22; 1; 11 Sep 26; 16:00) | Micron Technology Inc (MU.O; US$975.26; 1; 11 Sep 26; 16:00) | Montage Technology (6809.HK; HK$258.0; 1; 11 Sep 26; 16:10) | Samsung Electronics (005930.KS; W259500.0; 1; 11 Sep 26; 15:45) | Sandisk Corp (SNDK.O; US$1633.35; 1; 11 Sep 26; 16:00) | SK Hynix (000660.KS; W1812000.0; 1; 11 Sep 26; 15:45) | TechWing (089030.KQ; W46350.0; 1H; 11 Sep 26; 15:45) | TES (095610.KQ; W143500.0; 1; 11 Sep 26; 15:45)