IDC Semiconductor Market Outlook

For SEMI Korea

Soo-Kyoum Kim

Program Associate VP / Memory

September 25, 2025

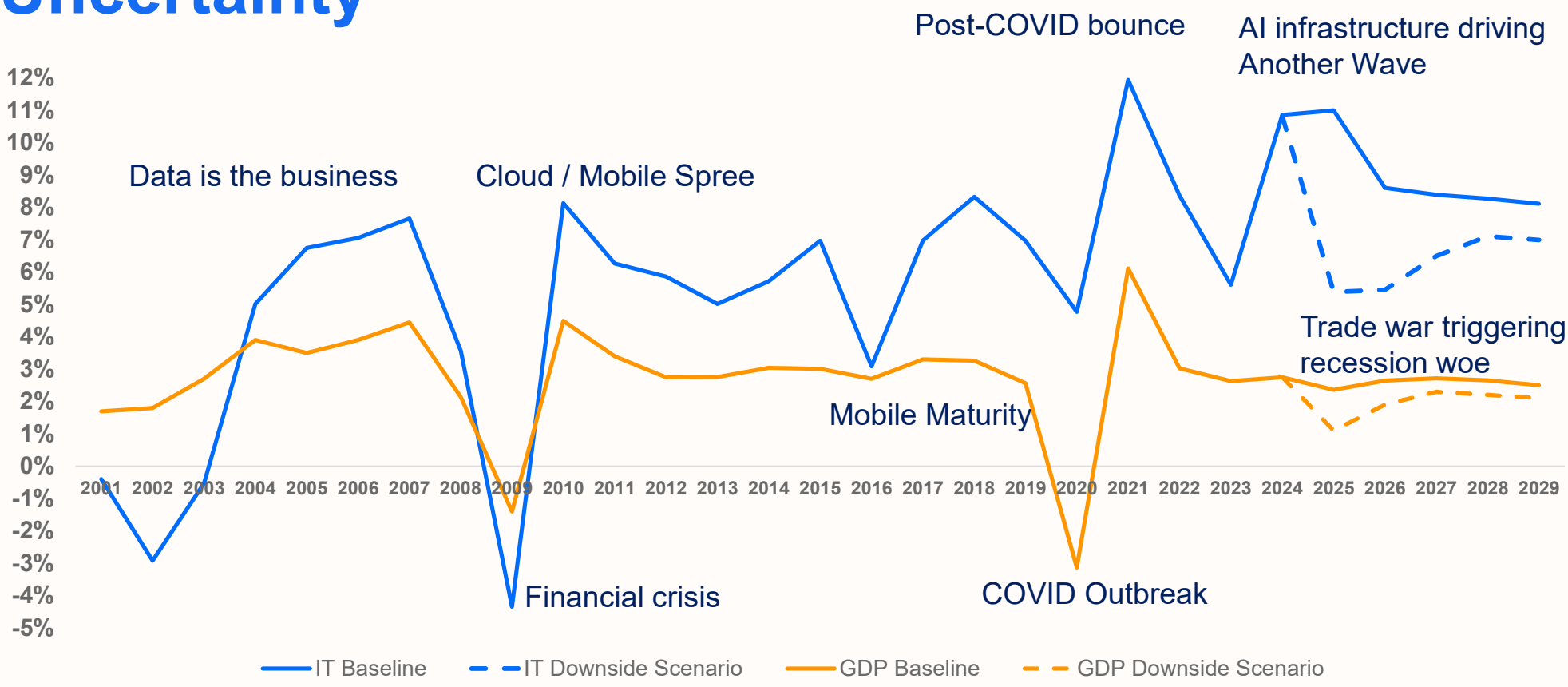
Agenda

- Key Takeaway
- IT Spending
- Enterprise Spending on AI
- Semiconductor Market Outlook
- Memory Market Outlook

Key Takeaways

- ***Tariff negotiations and policy uncertainties*** **affecting supply chains, investment, and cost over the rest of this year. markets have returned from shock, but industry still in search of leadership and clear business outcomes.**
- ***IDC expects the semiconductor industry will grow at strong high single digit, reaching $1T earlier than consensus***
- ***AI inferencing plays a larger role in overall spending as it drives more compute and semiconductor consumption. Edge AI, advanced optimized SoCs, and intelligent devices enable AI at the edge***
- ***Enterprise AI IT spending for platform will increase considerably from 2025-2029, even as non-AI IT spending softens***
- ***Enterprise memory (HBM, eSSD) will lead the memory market growth with less volatility.***

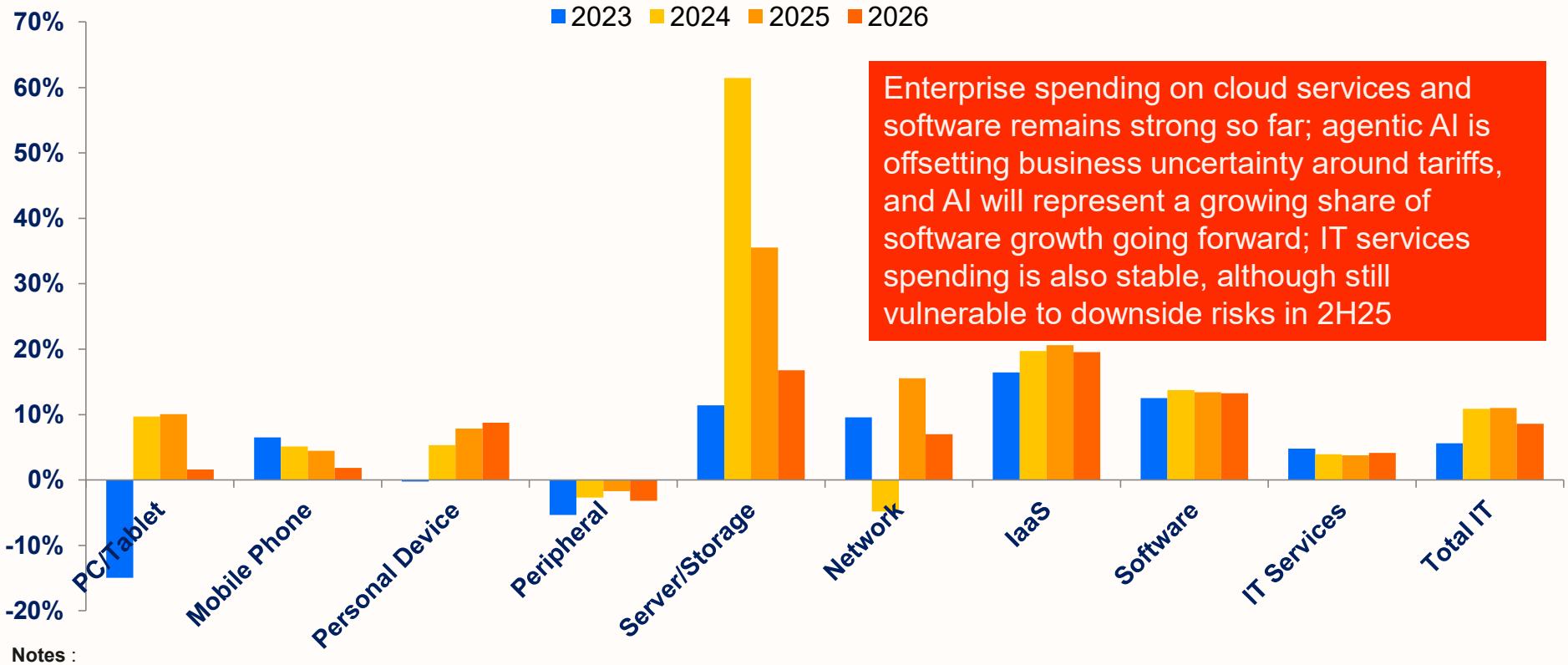
WW IT spending : AI Boom Offsetting Tariff Uncertainty



Notes: Growth in constant currency . IT Spending excluding telecom, Business/OT services

Source: IDC Worldwide Black Book, Jun. 2025

Worldwide IT Spending Growth Forecast



Notes :
- Personal Device includes wearables and AR/VR viewers; Peripheral includes printers, MFPs and PC monitors; Software includes SaaS, PaaS and on-premise software
- Growth in constant currency

Source: Worldwide Black Book, June 2025

Enterprise Spending Drives The Next Wave of Growth for AI Investment and Inferencing

$4.0T
Enterprises (25-29)

+$1.7T
for AI Platforms + applications

Customer Clusters

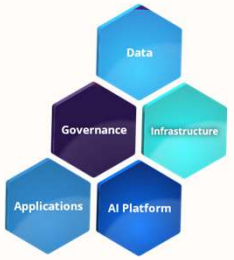
Risk-averse

Balanced

Operationally focused

IDC prediction 2024

By 2026, 75% of G1000 adopt value-based AI economics models covering tech acquisition cost, productivity gains, decision making, and innovation



Source: Future Enterprise Resiliency & Spending Survey Wave 7, IDC, July 2024, N=891 ***Worldwide Artificial Intelligence IT Spending Forecast, 2025–2029, August 2025***

Inference Spending is Shifting Spending Pattern

- **NVIDIA dominates inference as well as training**
  - Newer chips target inference with better INT8/FP8 throughput, increased memory bandwidth
- **AMD MI300X and MI350X targeting inference performance**
- **Customers may seek cheaper, lower-precision, inference optimized chips**
  - Older GPUs, accelerators
  - Adopt startup alternatives
  - Custom ASICs
  - Arm, RISC-V based solutions

- **Edge is logical place for inference**
  - Latency-sensitive
  - Privacy dependent
  - Bandwidth constrained
- **Edge AI hardware beneficiaries**
  - Low-power, AI-capable SoCs (Qualcomm, MediaTek, NXP, Hailo)
  - Inference HW (NVIDIA Jetson, AMD Kria)
  - Specialized silicon (automotive-grade, industrial, medical)
- **FP8/INT4 edge inference**

- **Implications**
  - Distributed processing (pre-processing at sensors)
  - Hybrid inference (maybe first pass at edge – classification, identification) with cloud for heavy workloads
  - Hybrid edge-cloud orchestration
- **Custom SoCs, chiplets, AI accelerator integration**
- **Advanced packaging becomes strategic differentiator to enable flexibility and accommodate workload diversification**

AI is Driving the Bulk of new Data Center Investment

Gen AI is in the spotlight, but more workloads are coming.

Starting with the problem

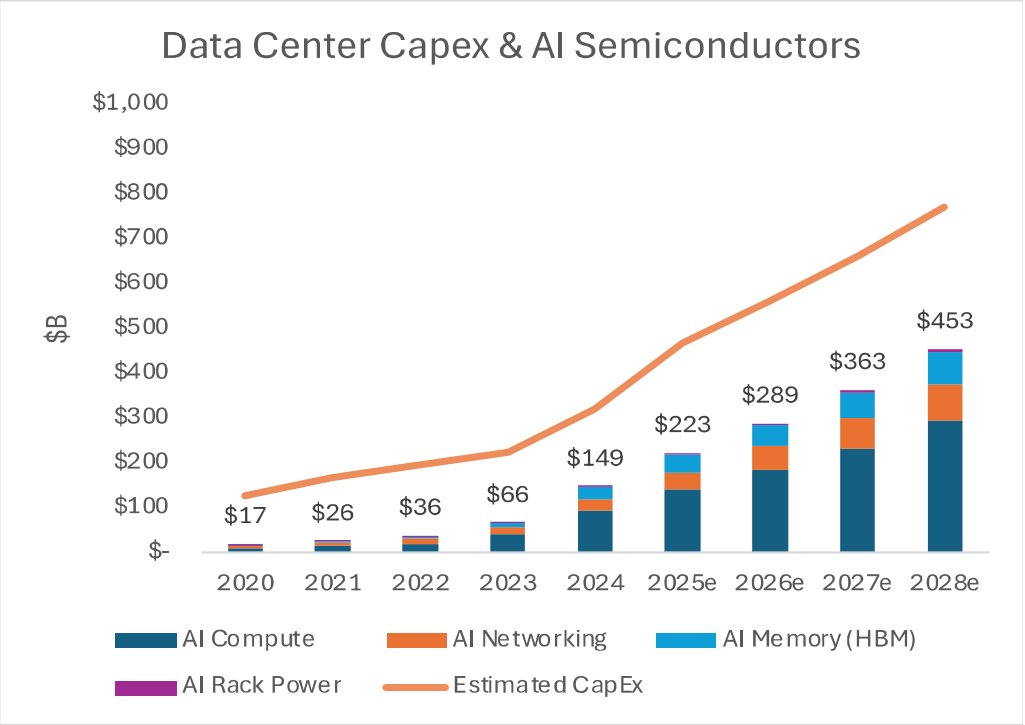
Rewiring organizations for continuous innovation and transformation

Growth

Strong growth due to new workloads in the datacenter

New Investments

Significant increase in demand for compute, networking, and HBM.

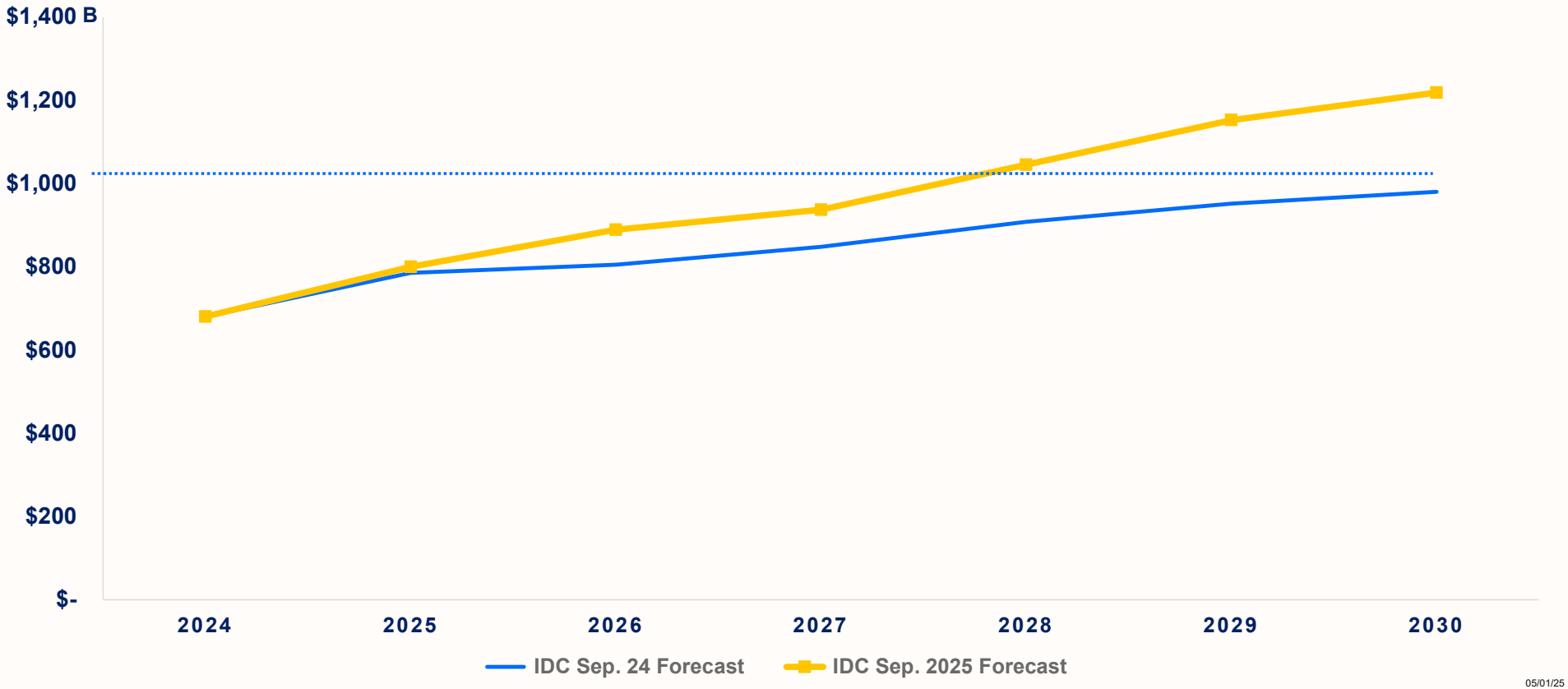


Notes: Capex data from the hyperscalers is a good indicator for the spend on data center Infrastructure,.

Source: IDC, May 2025

Semiconductor Market Outlook

AI Supercycle Accelerates $1T Semiconductor Revenues Earlier Than Expected (~2 years)

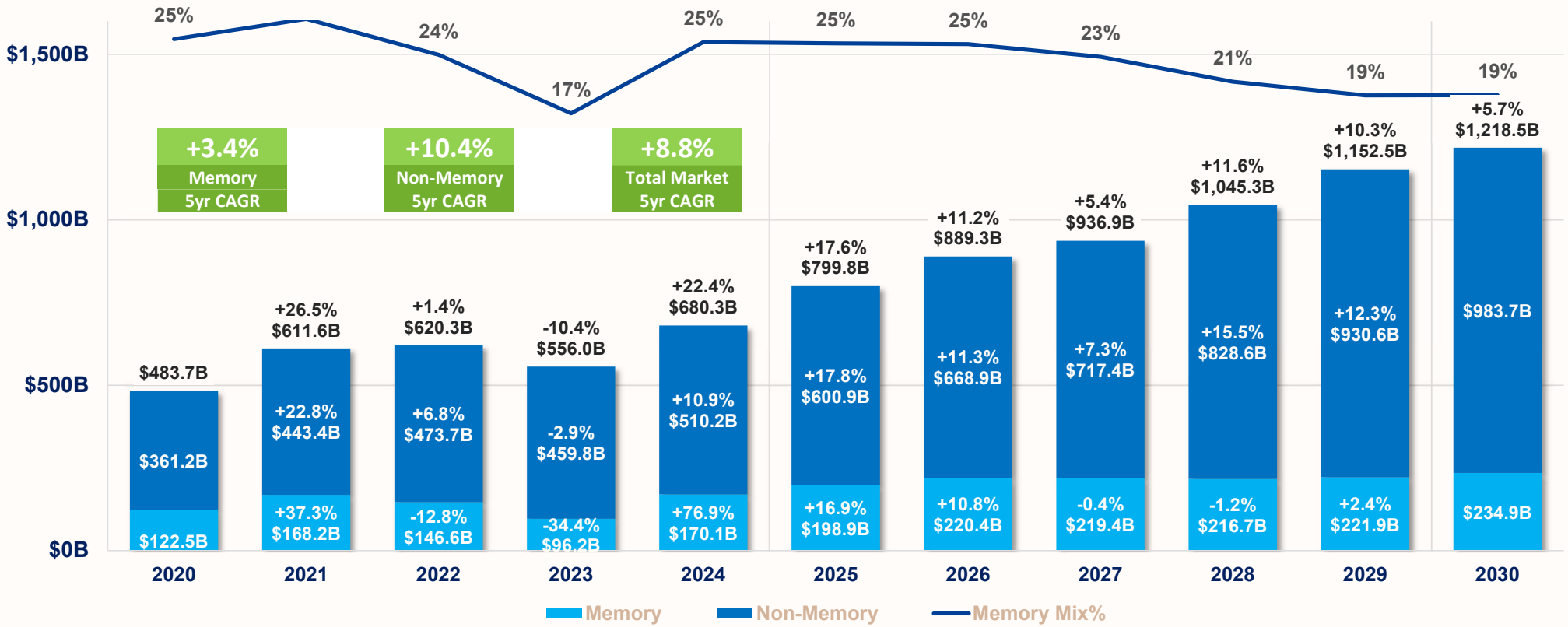


05/01/25

Source: Company Filings, IDC Estimates

Source: IDC CY2Q25 Semiconductor Application Forecast, Sep. 2025

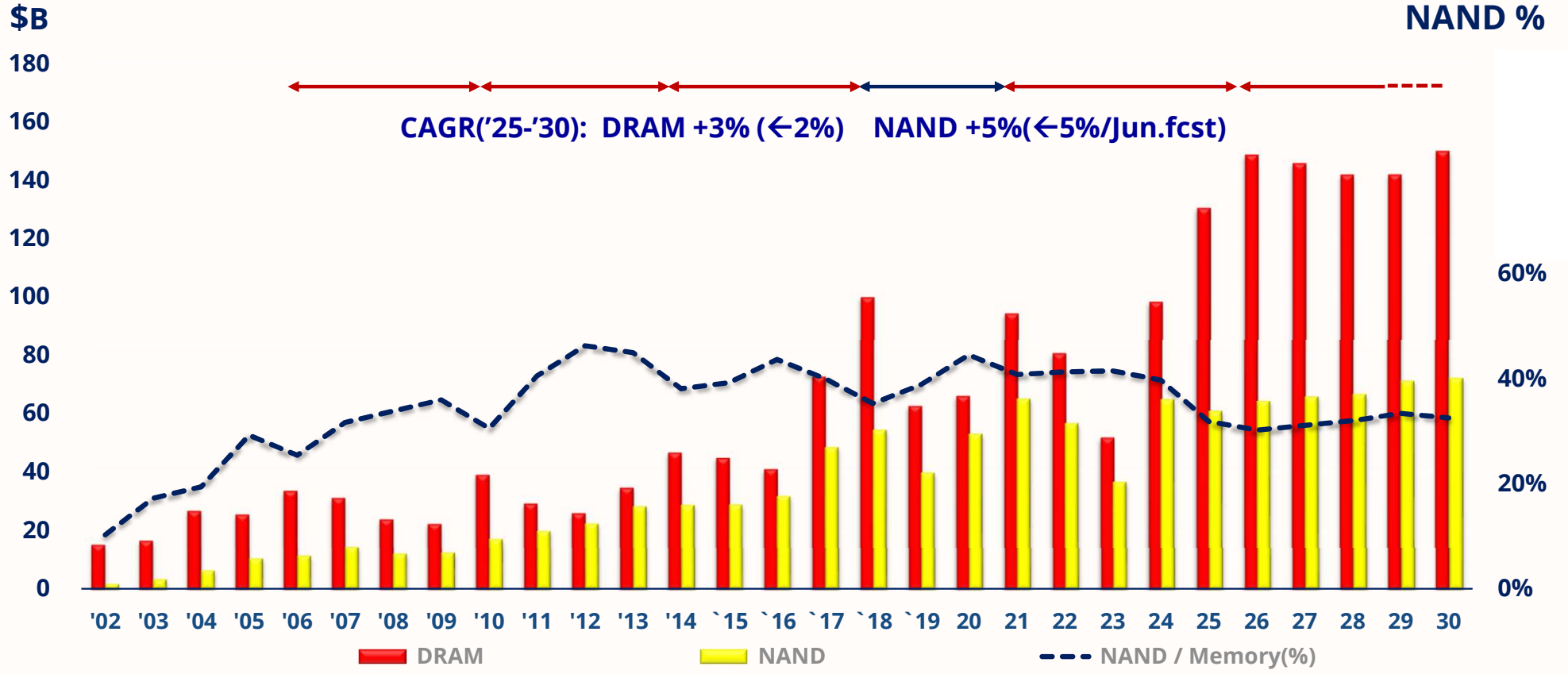
Worldwide Semiconductor Market 2030 Outlook



Notes : CAGR is for 2025-2030

Source: IDC CY2Q25 Semiconductor Application Forecast, Sep. 2025

Memory Revenue Forecast

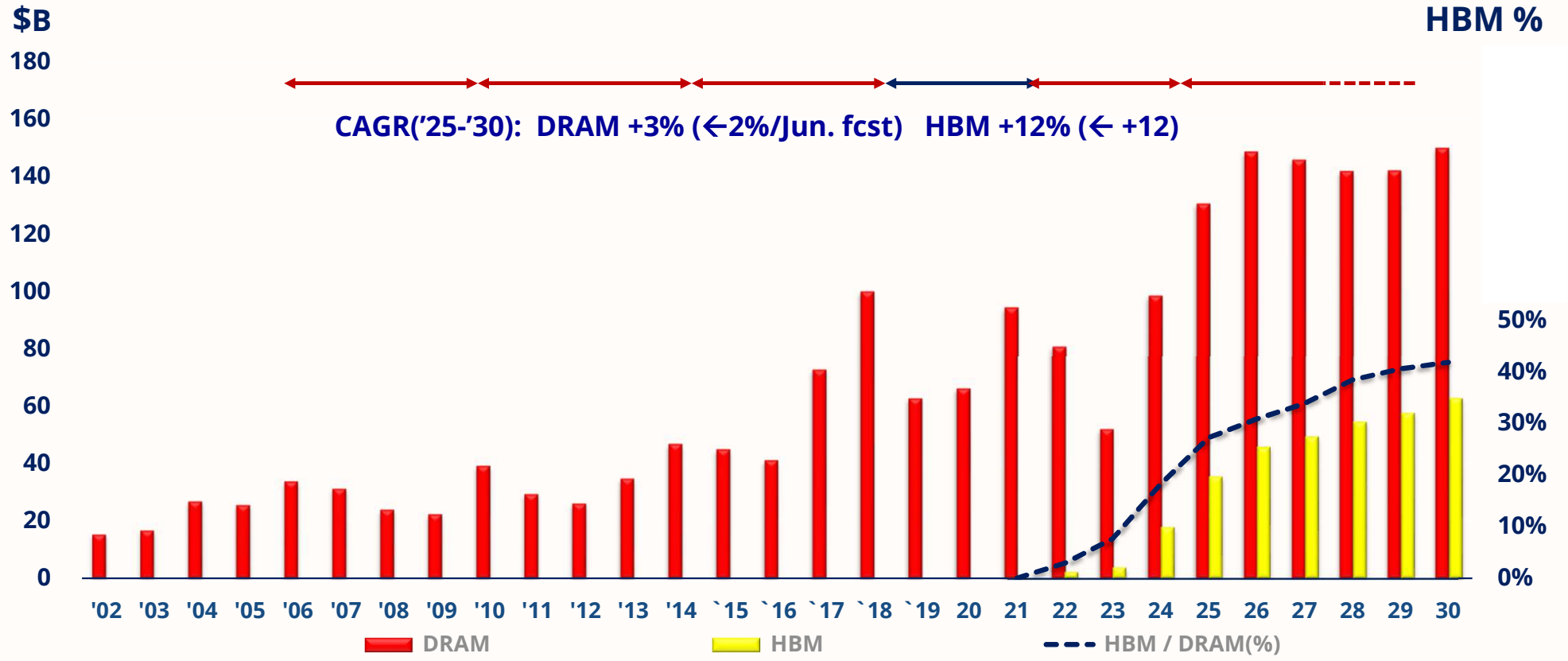


Source: IDC CY2Q25 Update, Sep. 2025

Long Term Semiconductor Market Drivers

- ***CSP, Enterprises, and Sovereign AI** drive large investment in compute*
- ***Enterprise refresh cycle** factored in 2026 and 2027 for clients and GP/AI servers*
- ***CE and IoT refresh cycle** in late 2026 and 2027 as AI changes UI and there a new introduction of intelligent devices*
- *IDC expects a **longer than expected recovery for industrial and automotive**; EV sector pivoting in favor of Chinese ecosystem putting more price pressure on market leaders*
- *IDC raises 2028 Market Outlook to $1.0T. It will **open $1T era**, 2 yrs earlier than consensus.*
- ***Edge AI** is coming as new SoCs, controllers, and tuned models enable AI inferencing where data is being generated - Inference at scale*
- *The next major markets will take off on AI inferencing being adopted in **Healthcare, Transportation, Industrial, and Robotics** (2029 and beyond)*
- ***Physical AI** becomes an enabler for robotics inflection point beyond 2030*

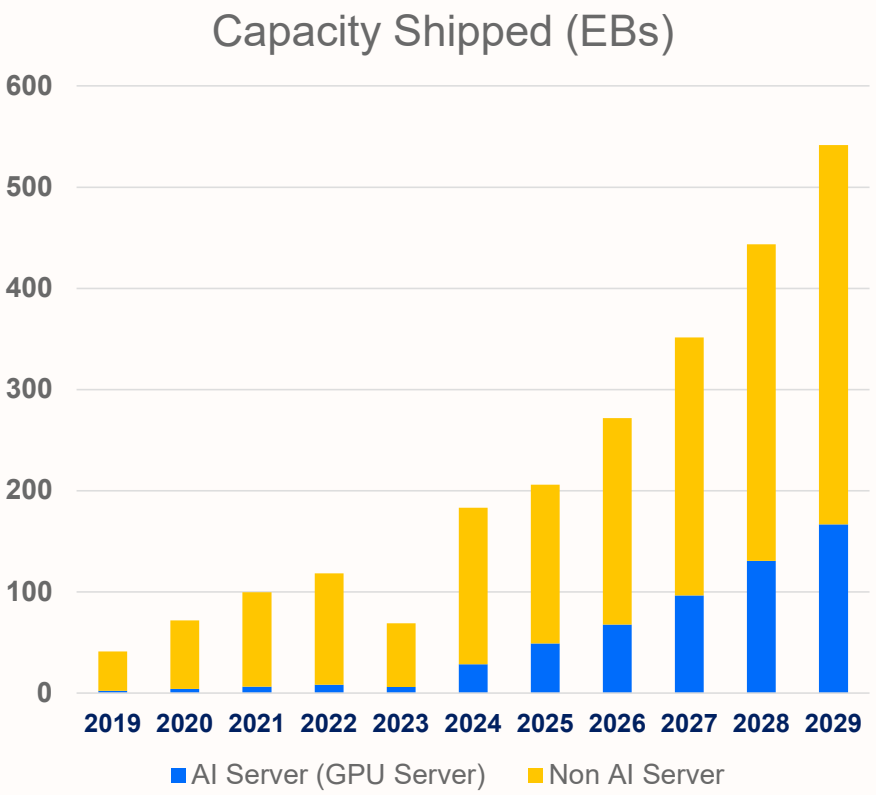
HBM Revenue Forecast

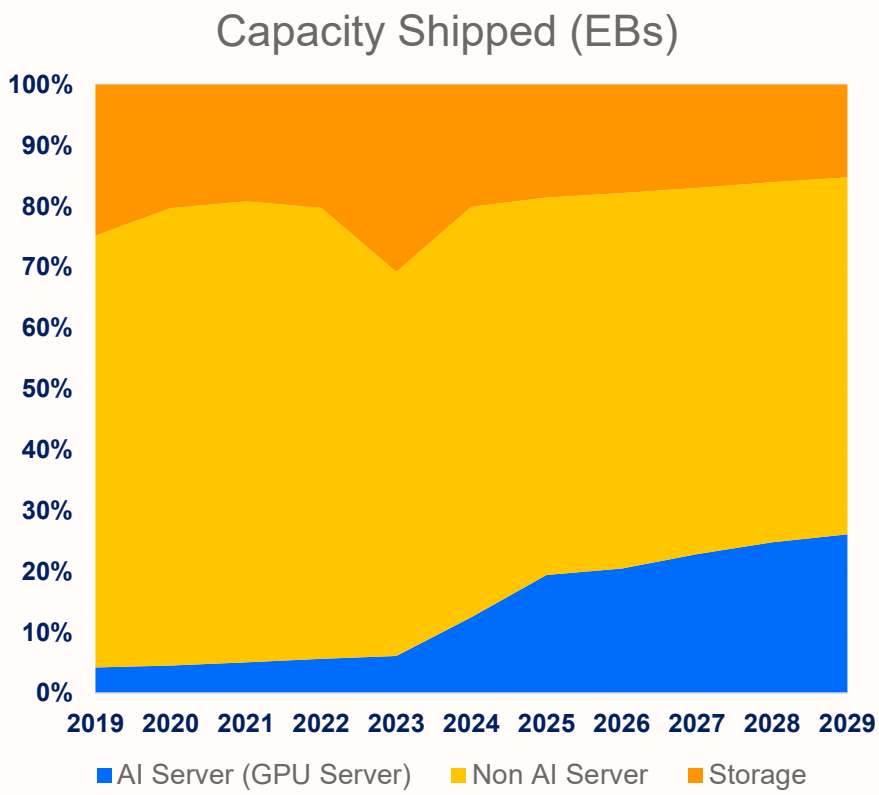


Notes: DRAM includes HBM

Source: IDC CY2Q25 Update, Sep. 2025

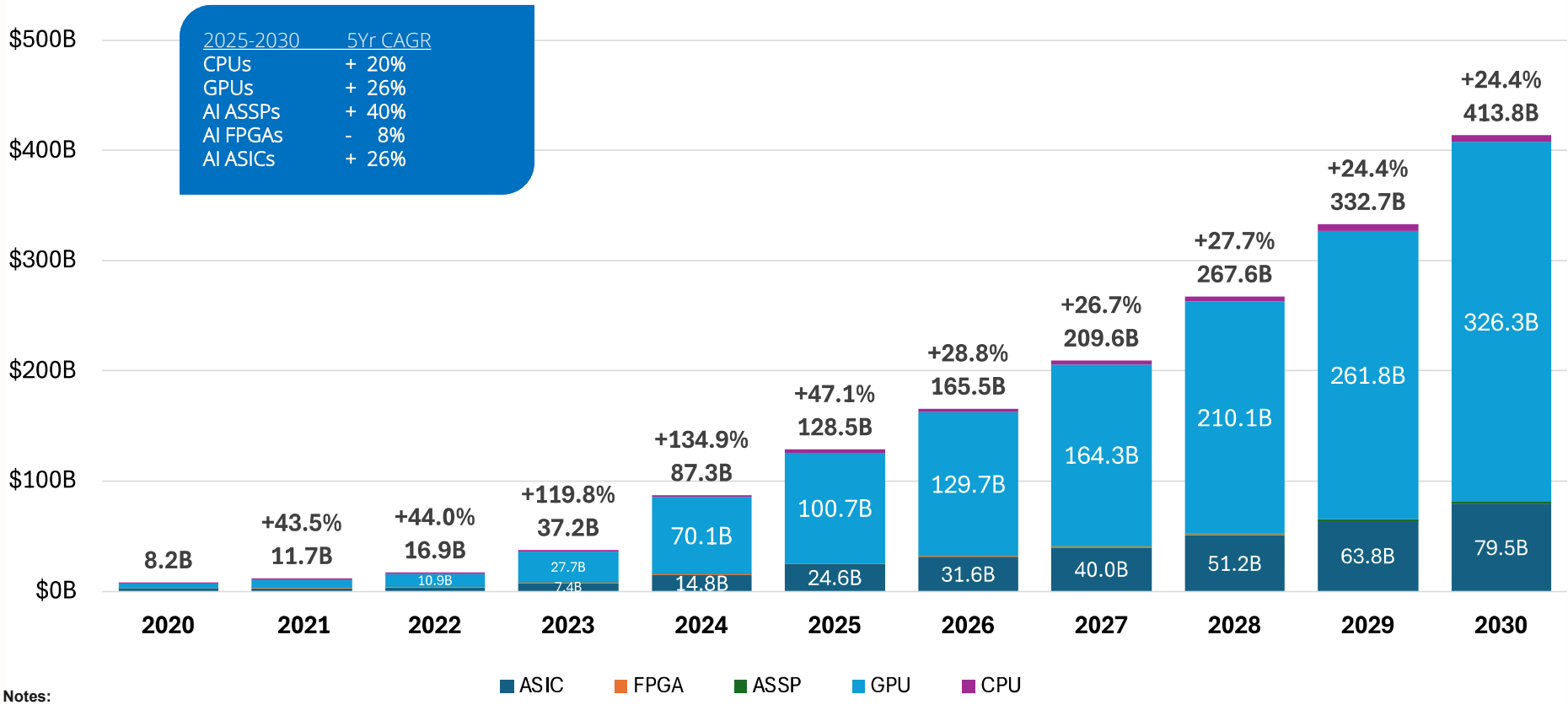
Enterprise SSDs : AI(GPU server) Outlook





Source :Worldwide Solid State Drive 2024-2029 Forecast Update (Kioxia Custom Study)

Processors and Architectures:
All AI Semiconductor Shipments (Rev.)

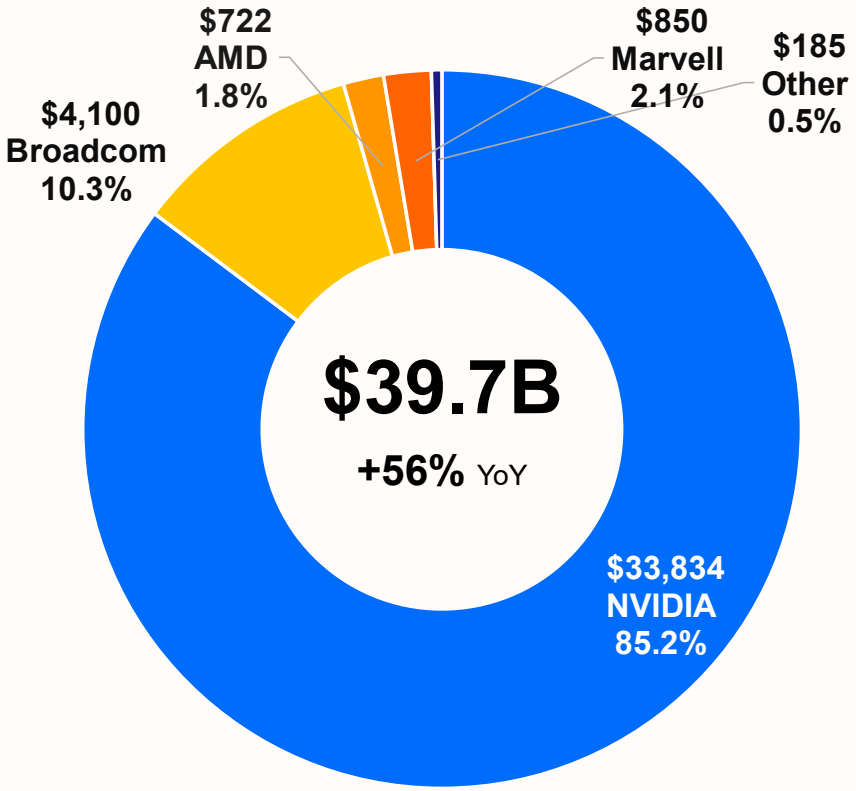


Notes:

- Data shown represent GPU silicon only. Revenues do not include memory, other semiconductors, PCBs, mechanicals, software or other components of the server GPU-based end solution.
- Server discrete GPUs include general-purpose GPUs used for offloading tasks from the server's host microprocessor. Hyperscaler ASICs estimates the value if Google sold the TPU to others, not what they pay Broadcom.
- NVIDIA's server discrete GPU products include B200, B100, H100, H800, A100, A800, T4, A30 and others. AMD's server discrete GPU products include MI300X. Broadcom and Marvell produce XPUs (ASICs).

Source: IDC CY1Q25 Semiconductor Application Forecast, June. 2025

AI Accelerator Vendor Share(CY2Q25)

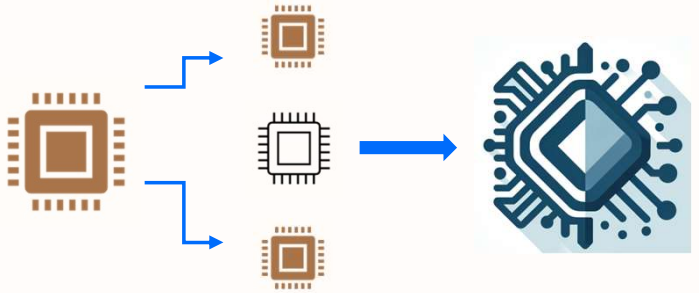


- **Nvidia** is #1 with 85.2% market share with GPUs
  - 50% y/y growth
  - $28B in Blackwell (GB200) Revenue – 71% of datacenter accelerated compute
- **Broadcom** is #2 and gains market share with 10.3% market share with AI ASICs
  - Includes long term customers Google, Meta, and five additional customers ramping or starting to ramp
  - Projected a market SAM of $60B-$90B in 2027 from three customers
- **Marvell** is #3 with 2.1% market share with AI ASICs

Source: IDC CY1Q25 Semiconductor Application Forecast, June. 2025

Processor Diversity Grows, Fitting for Purpose

- Processors able to run more advanced AI like Agentic AI will become increasingly fit for purpose to include NPU blocks, GPUs
- Accelerators will be tailored for targeted workloads for energy efficiency
- Memory, memory bandwidth, and interconnect will also be tailored for AI workloads
- Heterogenous processors, chiplet designs, new advanced packaging are enabling the customization of processors using common chip blocks



CPU

- **Strengths:** CPUs excel at logic-based decision making, orchestration, symbolic/rules-based reasoning for Agentic AI. Lower power consumption.
- **Limitations:** limited parallelism and ability to handle larger AI models. Potential memory bandwidth limitations.

GPU

- **Strengths:** ability to handle large models (compute / energy), able to handle batch inferencing for simultaneous decision making.
- **Limitations:** Power consumption, heat generation. Limited flexibility for AI workloads involving symbolic or logic processing

NPU

- **Strengths:** designed for high-efficiency neural network processing, lower power consumption for AI tasks. Likely to be integrated into purpose-built SoCs
- **Limitations:** less flexible and have a narrower range of applications, often limited to specific AI tasks.

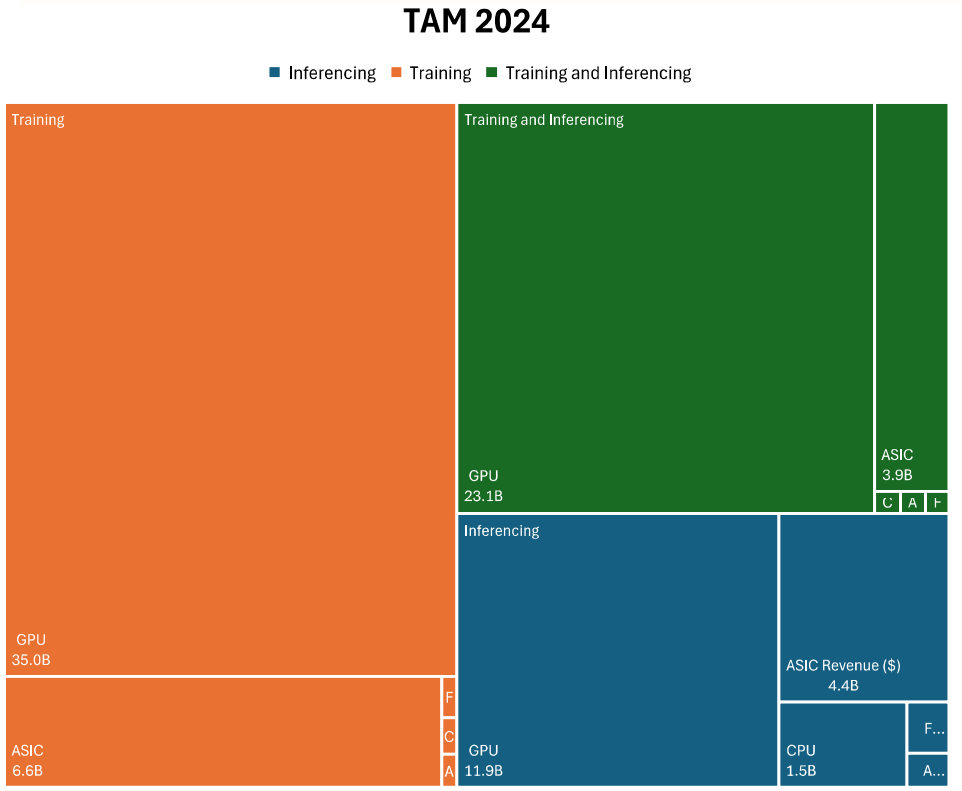
APU

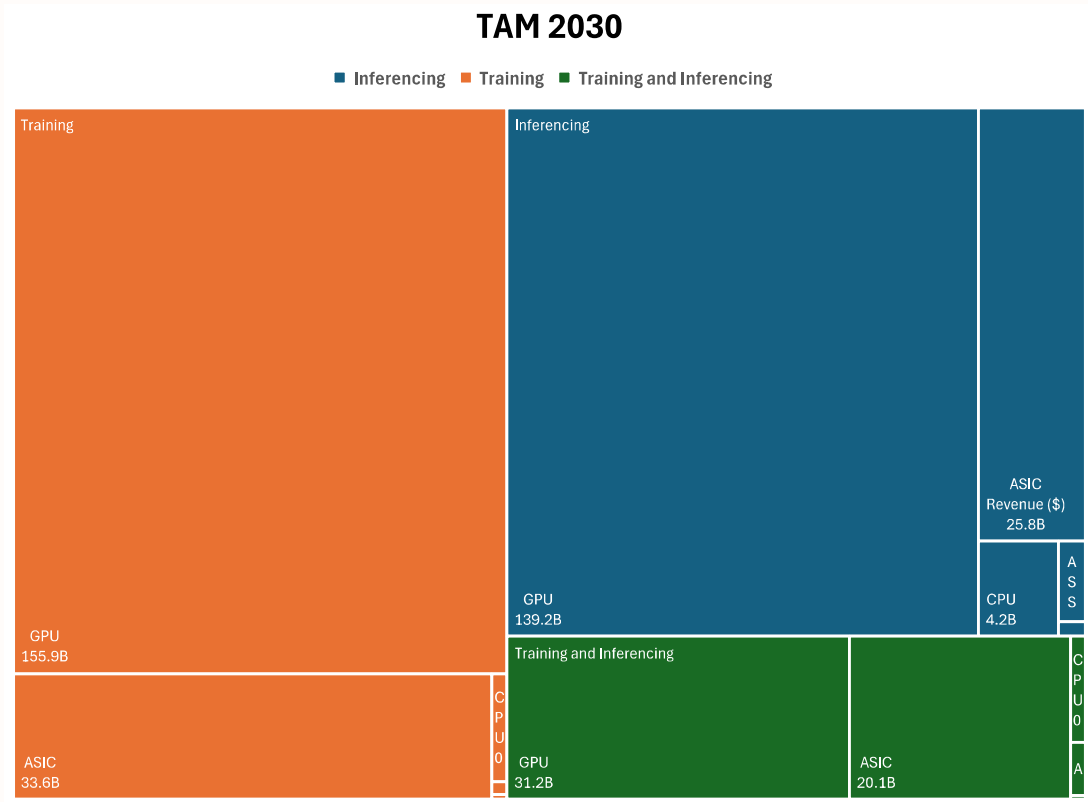
- **Strengths:** balancing performance with CPU and GPU cores allow for customization for workload. Reduced latency between cores. Energy efficient
- **Limitations:** less raw performance, limited scalability flexibility

Source: IDC CY1Q25 Semiconductor Application Forecast, June. 2025

Update!

Processors and Architectures:
All AI Semiconductor Shipments (Rev.)





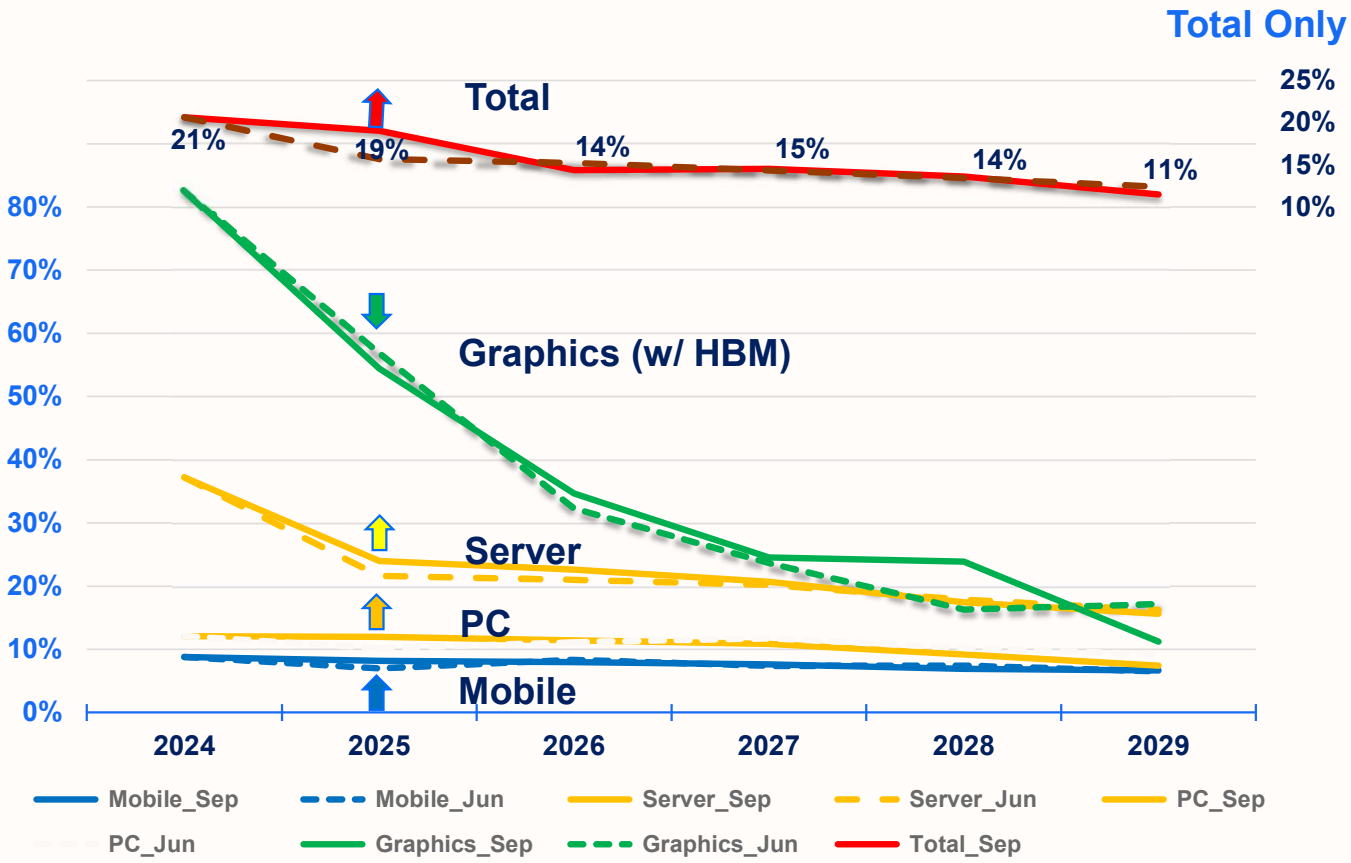
Notes:

- Data shown represent GPU silicon only. Revenues do not include memory, other semiconductors, PCBs, mechanicals, software or other components of the server GPU-based end solution.
- Server discrete GPUs include general-purpose GPUs used for offloading tasks from the server's host microprocessor. Hyperscaler ASICs estimates the value if Google sold the TPU to others, not what they pay Broadcom.
- NVIDIA's server discrete GPU products include B200, B100, H100, H800, A100, A800, T4, A30 and others. AMD's server discrete GPU products include MI300X. Broadcom and Marvell produce XPUs (ASICs).

Source: IDC CY1Q25 Semiconductor Application Forecast, June. 2025

Memory Market Outlook

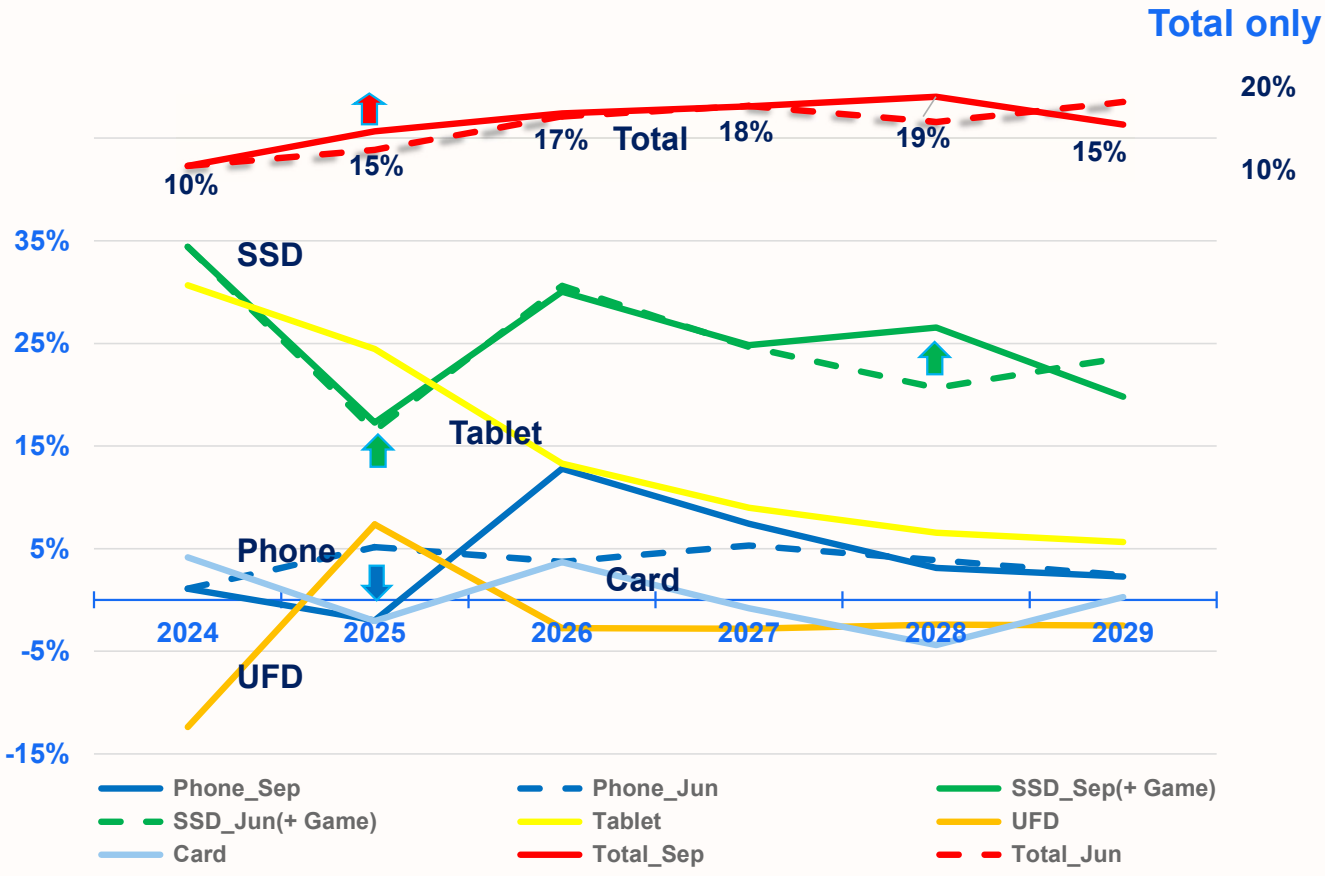
DRAM Demand Growth(Bit %)



Source: IDC CY2Q25 Update, Sep. 2025

- **Common**
  - AI (Server + HBM) will lead growth
  - Saturated consumer(PC, Mobile) demand
- **Mobile (6% ← 7% of CAGR (24-29) /Jun. fcst)**
  - China in gradual recovery
  - Apple 12GB & On-device AI from flagship
- **Server (20% ←19%)**
  - Transition to inferencing service
  - GPU server at CAGR of 14% (←16%), ASIC server at 31%(←37%)
  - 32Gb rapidly replacing TSV based DIMM
  - SOCAMM vs MRDIMM
- **Graphics (29% ←28%)**
  - HBM at CAGR of 34% (← 34%)
  - HBM % of total demand ; 10%/'29← 7%/'25
  - Next gen (PS6, X-box3) in 2H27~
  - H4/E'26 ~, H5/E'28~
- **PC (+10% ← 10%)**
  - Win10 EOL & Replacement/'25
  - AI PC in cost pressure/'25, ramping in '26~
- **Total Demand (15% ←14%)**
  - 14%/'26 & 19%/'25 ← 15 & 16/Jun. fcst
  - Long-term growth to low teens

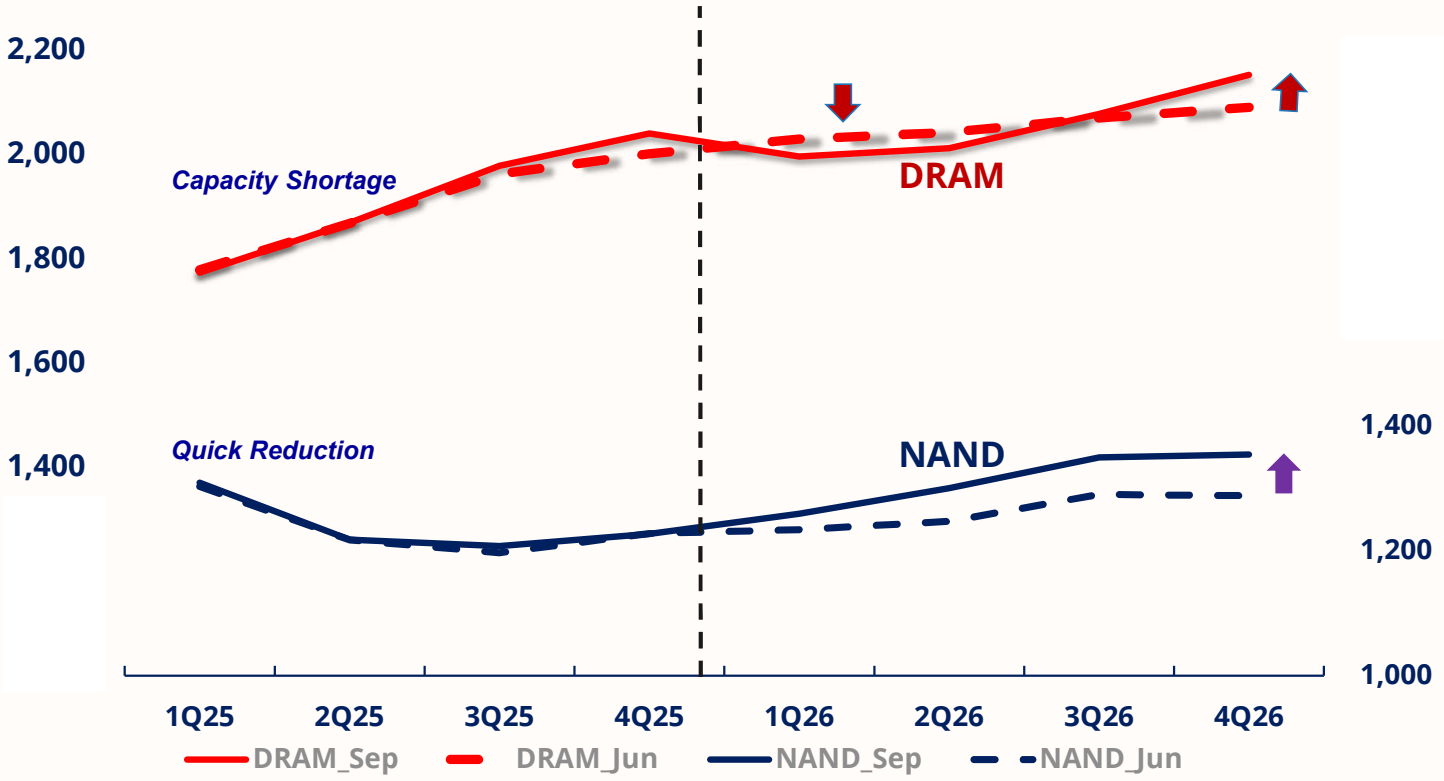
NAND Demand Growth(Bit %)



Source: IDC CY2Q25 Update, Sep. 2025

- **Common**
  - CSP will have strong growth in '25
  - OEM server in slow recovery
  - Dmd strength; CSP>> OEM> PC>>> Mobile
- **SSD (24%** ← 23% of CAGR (24-29) /Jun. fcst)
  - eSSD growth > cSSD
  - High capacity QLC eSSD growth in 2H26
  - 1TB PC & 2TB console pervasion in '26
  - Next gen. consoles (P6, X-box3) in 2H27~.
- **Phone (5%** ← 4%)
  - China/ Apple in still weakness
  - 2TB system in 2H25 ; Apple
  - QLC storage (1/2TB) will start in 2026~
- **Tablet (12%** ← 9%)
  - Robust content growth
- **UFD(-1%), Card (- 1%)**
  - Low growth in long-term
- **Total Demand (17%** ←16%)
  - 17%/'26 & 15%/'25← 16% & 12%/Jun. fcst
  - Inventory consumed much in 2Q25

Memory Capacity Forecast



Notes : Wafer Output base

Source: IDC CY2Q25 Update, Sep. 2025

NAND ▪ WW

- D : +8%/'26 ← +11%/'25 (← +8 & +11/Jun. fcst)
  - w/o China +5 ← +7
- N : +6%/'26 ← -7%/'25 (← +2 & -8/ Jun. fcst)
  - w/o YMTC, +3 ← -11

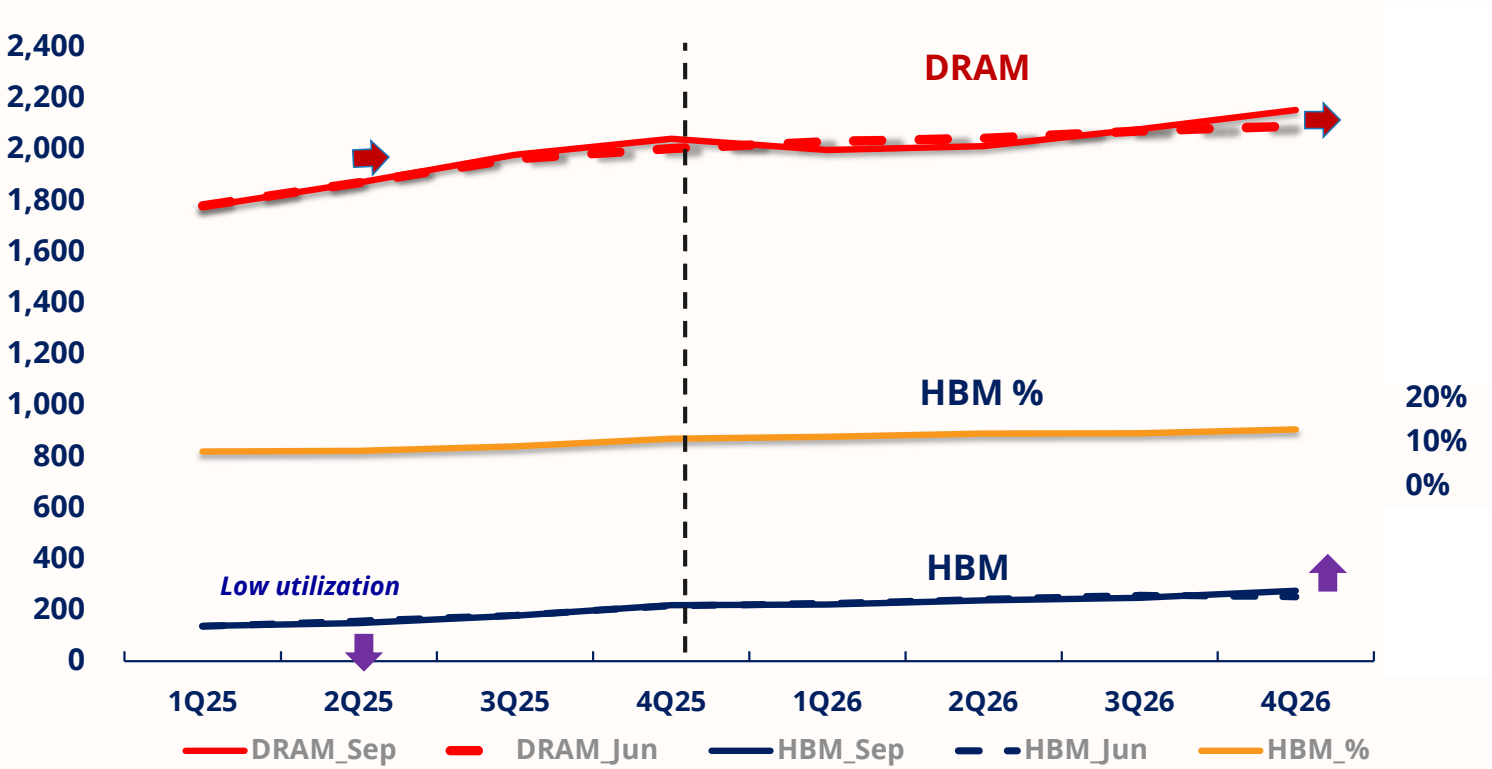
▪ **New Fabs in '25 ~'26**

- DRAM : SS P4, HY M15x, MU Hiroshima
- NAND : SS P3/P4, JV K-2

▪ **New Fabs in '27-'29**

- Yong-in(N, 4 fabs, '27~),
- MU Boise (D, F2, '28~), NY(D, 4 fabs, '29~),
- NTC(Fab3B/'27~)
- CXMT (A3/'27, C2/'29~)

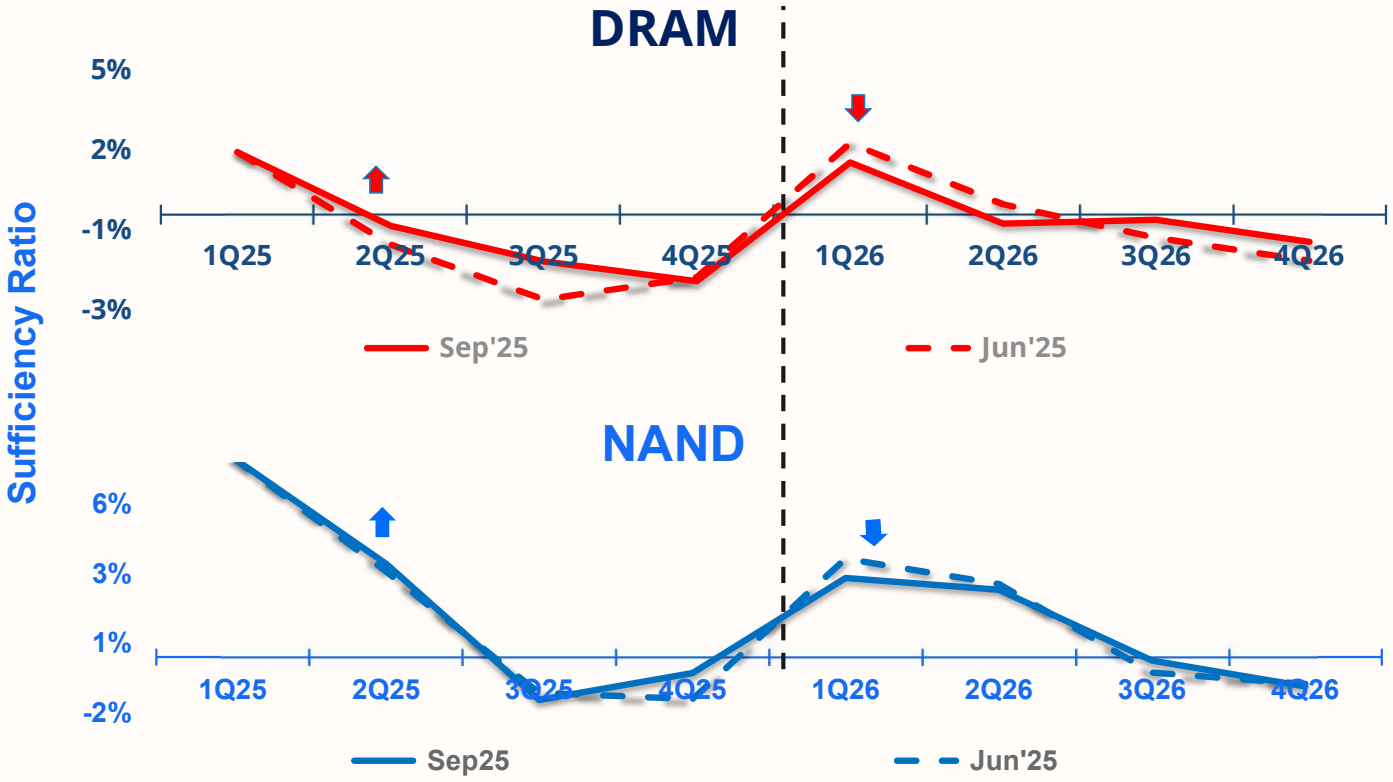
HBM Capacity Forecast



Notes : Wafer Output base, DRAM includes HBM, HBM % = HBM/DRAM ***Source: IDC CY2Q25 Update, Sep. 2025,***

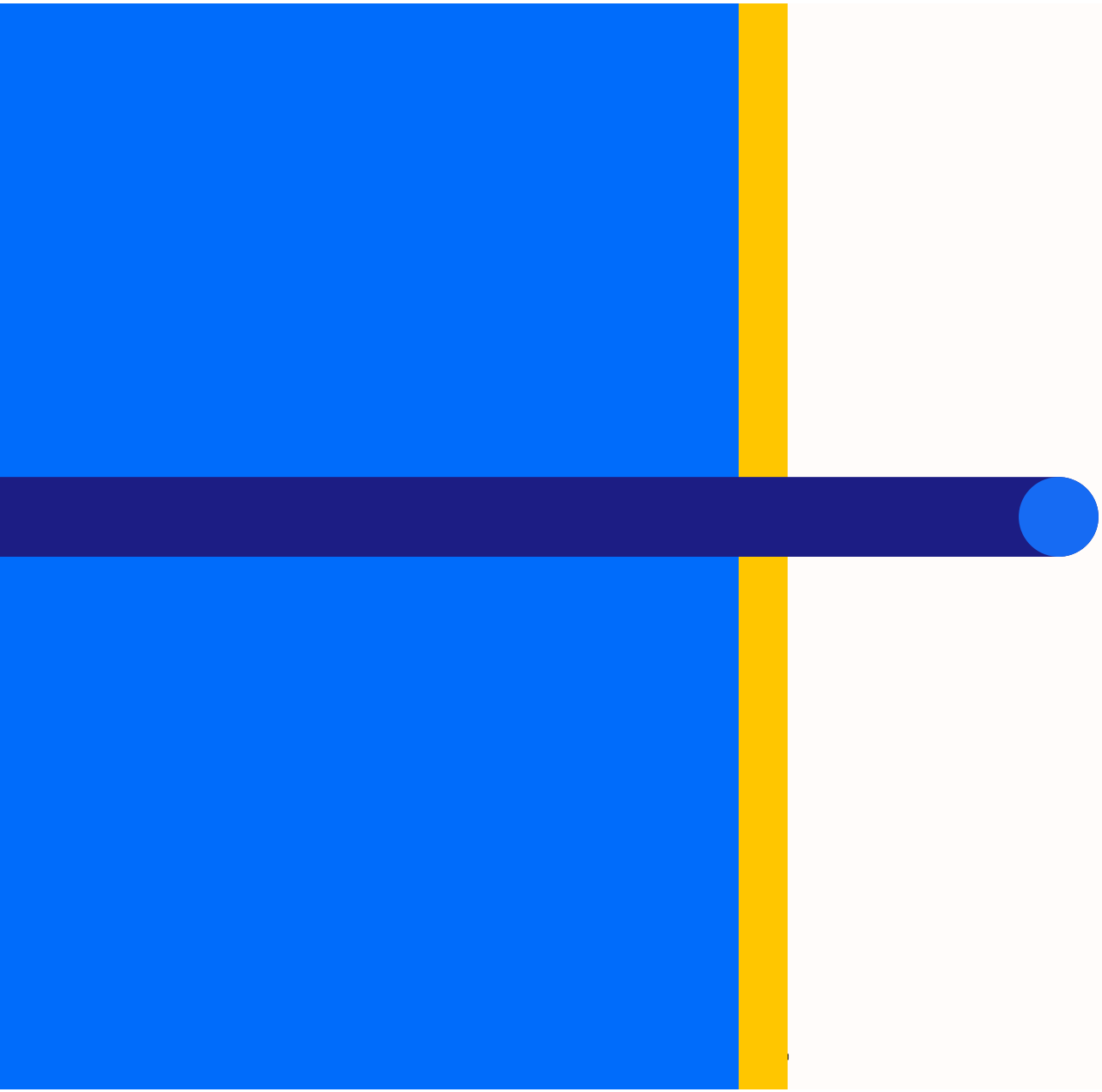
- **WW**
  - D : +8%/'26 ← +11%/'25 (← +8 & +11%/Jun. fcst)
  - HBM : +44%/'26 ← +42%/'25 (← +42 & +44%/ Jun. fcst)
- **HBM Fabs in '25-'26**
  - SS :
    - F17(H2e/3 1z)
    - P1(H3e 1a)
    - P3(H4 1c)
    - P4(H4/4e 1c)
  - HY :
    - M16(H3 1z, H3e/4 1b)
    - M15x (H4e 1c)
  - MU:
    - A2(H3e/4 1b, H4e 1r)
    - Japan (H4e 1r)

Memory Market Scenario, 1Q25 – 4Q26



Source: IDC CY2Q25 Update, Sep. 2025

- **Demand Up**
  - CSP demand will continue strong growth longer than expected ; D (1H26) > N (2H25)
  - PC & Mobile will follow the trend for capacity shortage in DRAM, inventory consumption + capacity control in NAND.
  - Bit growth; Jump in '25, still mid-teens in '26
- **Supply well Managed**
  - D : HBM4 + DDR5 with priority in '26
  - N: Capacity will flatten out in '26
  - D4 shortage will be covered by China/Taiwan.
  - New Layers ; 3xx /2H'25, 4xx /E'26
- **Considerable growth in '25 and better than normal market in '26**
  - AI diffusion to agent service stage from D/C training phase
    - HBM enabling another growth
    - QLC high capacity eSSD up in '26
  - Compelling balance by strict supply control
  - Uncertainties will surge ; US Tech Ban
  - Better than normal market in '26



Soo-Kyoum Kim
Program Associate Vice President
skim@idc.com
+822-550-4347