

August 10, 2026 09:00 PM GMT

Asia-Pacific Technology | Asia Pacific

AI Supply Chain: GPU Architecture Optimization Given Memory Shortage and Supernode Challenge

Our checks indicate Rubin Ultra may adopt tiered HBM spec to balance supply, cost & workload requirements; shorter testing times could ease near-term capacity constraints.

GPU: Hardware Configuration Optimization

1. HBM de-specification – investors cross-checking whether this is true:

- We think Nvidia's Rubin Ultra will feature several SKUs with high-end version adopting HBM4e 8Hi lower-end versions using HBM4 12Hi or 8Hi. Based on discussions with memory coverage analysts, we see this reflecting: 1) HBM capacity shortages, 2) cost pressures, and 3) optimization for different workloads based on memory capacity requirements.
- Nvidia should make final decision before end-3Q26. We continue to gather feedback on the performance impact of changes to HBM specifications. Our checks suggest HBM de-specification will have a more meaningful impact on decoding than pre-fill workloads, particularly for long-context-window LLMs.
- However, reducing memory content per chip may increase overall chip shipments, boosting volume-driven suppliers. We do not rule out major ASIC projects adopting different HBM configurations for similar reasons.

2. Kyber server rack delays with no concrete timeline, while inter-rack CPO connectivity remains Nvidia's preferred direction:

- We believe Nvidia's (analyst: Joe Moore) long-term vision for AI server racks is maximizing compute capacity/GPU density and rack-level performance, with GPUs interconnected through scale-up architectures for LLM training & large-scale inference. Since 2023, we have seen the transition from Hopper HGX (8-GPU, or NVL8) systems to Blackwell NVL72.
- To further increase GPU density per rack, Nvidia first disclosed its Kyber blade server design at GTC 2025. However, given ongoing PCB and thermal challenges, first-generation Rubin Ultra server racks will likely continue to use the Oberon solution for NVL72 while adopting CPO/NPO technologies to enable NVL576-scale deployments.
- We remain optimistic about Nvidia's adoption of CPO as a key differentiator in scale-up performance versus China's supernode architectures (see below).

(Continued on page 2)

MORGAN STANLEY TAIWAN LIMITED*

Charlie Chan

Equity Analyst

Charlie.Chan@morganstanley.com

+886 2 0730 1725

Daniel Yen, CFA

Equity Analyst

Daniel.Yen@morganstanley.com

+886 2 0730 1813

MORGAN STANLEY ASIA LIMITED*

Daisy Dai, CFA

Equity Analyst

Daisy.Dai@morganstanley.com

+852 2848 1310

MORGAN STANLEY TAIWAN LIMITED*

Tiffany Yeh

Equity Analyst

Tiffany.Yeh@morganstanley.com

+886 2 771 23073

Lucas Wang

Research Associate

Lucas.Wang@morganstanley.com

+886 2 0730 2875

MORGAN STANLEY ASIA LIMITED*

Ethan Jia

Research Associate

Ethan.Jia@morganstanley.com

+86 13962611817

Henry Zhao

Research Associate

Henry.Zhao@morganstanley.com

+86 13962611731


 Asia Summer School 2026

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

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

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

3. China AI supernodes – Chinese AI GPU vendors are adopting NPO rather than CPO for inter-rack connectivity

- To support larger LLMs (>2tn parameters), Chinese GPU vendors have also introduced supernode architectures. These racks connect through NPO (near-package optics), as local GPU SerDes speeds may remain at 100Gb/s per lane due to foundry process constraints. See: [AI Networking: Scale-up Technology Enables China's AI Supernodes \(27 Jul 2026\)](#).

ASIC – TPU Update

- MediaTek guided that it could achieve a 15–20% share of the ASIC market in 2027, broadly in line with our preview and Joe's assumption that MediaTek could secure a 20% revenue share in TPU programs. That said, MediaTek believes its revenue share could increase further in 2028 alongside the ramp-up of 2nm TPU products.
- Separately, GUC indicated at its recent analyst meeting that it is interested in providing TPU compute die back-end design services for future TPU generations.

2027 GPU/ASIC Total Power Consumption

Strong Google TPU demand: In our previous [MediaTek](#) report, we highlighted upside from Google TPU demand, which is driving strong revenue growth for KYEC in both 2026 and 2027. We now forecast total TPU shipments of 3.7mn units in 2026 and more than 7mn units in 2027. We estimate TPU-related business could contribute 7–8% of KYEC's total revenue in 2026 and slightly more than 10% in 2027, including final testing and a portion of chip probing revenue.

In addition, MediaTek's TPU is expected to adopt burn-in testing for its 3nm AI ASIC. We believe this reflects a broader industry trend, with AI chips increasingly requiring burn-in and longer testing cycles to ensure performance and reliability.

Strong Google CPU demand too: In our previous [GUC](#) report, we noted that the upgraded demand outlook extends beyond Google TPU to include Google CPUs. We now forecast Google CPU shipments of 1.5mn units in 2026 and 3mn units in 2027. Our checks also indicate that Google's CPU programs require burn-in and system-level testing, which we estimate could contribute 3–4% of KYEC's 2027 revenue.

In summary, we now forecast Google-related demand will account for 10–15% of KYEC's total revenue in 2027, including TPU, CPU, final testing, and a portion of chip probing revenue, versus 8–10% of total revenue in 2026.

Exhibit 1: Strong TPU shipments could contribute >10% of revenue in 2027

x Units	2023	2024	2025e	2026e	2027e	2028e
v5	500	2,400	250			
v6 (Trillium)			1,200			
v7 (Ironwood, by Broadcom)			300	2,300	200	
v8i (Sunfish; 3nm, by Broadcom)				900	4,000	2,500
v8t (Zebrafish; 3nm, by MediaTek)				500	3,000	1,000
v9 (Humufish; 2nm, by MediaTek)					150	3,000
v9a (Merope; 2nm, by US design service)						unknown
v10 (Icefish; 1.4nm, by MediaTek)						unknown
Total	500	2,400	1,750	3,700	7,350	>6500

Source: Morgan Stanley Research estimates

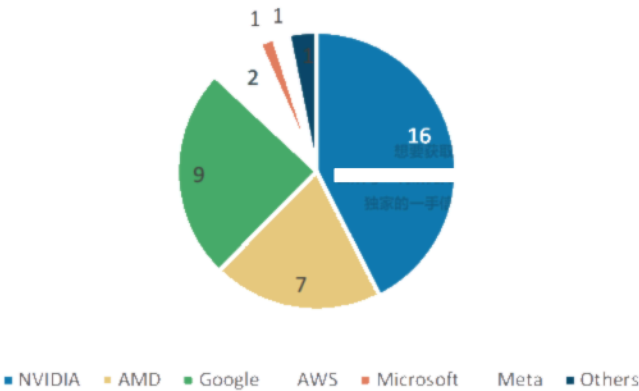
Implied chip volume vs. electricity consumption (GW) frequently discussed, suggesting power supply remains a global bottleneck

We published our latest report, [Asia-Pacific Technology: Connecting Dots: HK investors' feedback on Asian semis \(9 Aug 2026\)](#), yesterday. During our recent Hong Kong marketing trip, investors questioned whether power infrastructure can support our projected 2027 CoWoS-implied chip volume of approximately 19mn units. Assuming an average TDP of 2kW per GPU/ASIC, **the installed base would require roughly 38GW of power capacity.**

Conversely, when SpaceX indicated plans to expand its computing capacity from 2GW by the end of this year to nearly 10GW under an ambitious scenario (see [SpaceX: Strong Beat/Outlook vs. Higher Capex](#)), investors also asked what this would imply for TSMC's CoWoS capacity consumption. Notably, Adam Jonas models only 3GW of incremental computing demand, reaching 5GW by the end of 2027. As a rough illustration, the additional 8GW implied by SpaceX's target would equate to approximately 4mn Rubin GPUs, or roughly 20% of the AI accelerators that TSMC could produce in 2027.

Exhibit 2: 2027 GPU/ASIC chip-implied power consumption - 38GW in total

2027 GPU/ASIC total power consumption (GW)



Source: Morgan Stanley Research estimates

Global CoWoS Capacity Allocation

Exhibit 3: AI HBM consumption: Up to 50bn Gb in 2027

AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	HBM chip density (GB)	HBM chip units	Total HBM size (GB)	HBM generation	HBM vendor	Total HBM demand (k GB)
AI GPU (2027e)										
NVIDIA	B300	40	14	560	36	8	288	HBM3e 12hi	Hynix/Micron	161,280
	Vera CPU	250	23	5,750						
	Rubin R200	740	8	5,920	36	8	288	HBM4 12hi	Hynix/Micron/Samsung	1,704,960
	Rubin Ultra	130	8	1,040	48	8	384	HBM4e 12hi	Hynix/Micron/Samsung	399,360
	MI350 series	24	12	288	36	8	288	HBM3e 12hi	Samsung/Micron	82,944
AMD	MI455	157	7	1,099	36	12	432	HBM4 12hi	Samsung/Micron	474,768
	MI450 (for Meta)	35	14	490	36	6	216	HBM4 12hi	Samsung/Micron	105,840
	MI500 (Arcadia)	24	4	96	48	16	768	HBM4e 12hi	Samsung/Micron	73,728
	Venice CPU	270	21	5,670						
AI ASIC (2027e)										
Google	TPU v7p (Ironwood; AVGO)	13	16	208	24	8	192	HBM3e 8hi	Hynix/Samsung	39,936
	TPU v8i (Sunfish; AVGO)	330	12	3,960	36	8	288	HBM3e 12hi	Samsung/Hynix/Micron	1,140,480
	TPU v8t (Zebrafish; MediaTek)	180	20	3,600	36	6	216	HBM3e 12hi	Samsung/Hynix/Micron	777,600
	TPU v9 (Humufish; MediaTek)			400	48	12	576	HBM4e 12hi	Samsung/Hynix/Micron	230,400
AWS	Trainium 3	140	17	2,380	36	4	144	HBM3e 12hi	Hynix/Samsung/Micron	342,720
GUC's customers		60	20	1,200	24	6	144	HBM3e 8hi	Samsung?	172,800
Microsoft	Maia 300	50	11	550	36	8	288	HBM4 12hi	Samsung	158,400
Meta	MTIA 3 (Iris)	55	10	550	36	8	288	HBM3e 12hi	Samsung/Hynix/Micron	158,400
Others		18								
Total		2,664								6,077,256
Total HBM demand (mn Gb)										48,618

Source: Company data; Morgan Stanley Research (e) estimates. Note: Est. mates are compiled using our Asian supply chain checks.

Exhibit 4: AI wafer consumption: At least US\$59bn in 2027

AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	Compute die size	Geometry	Compute die units	Wafer consumption (k wafers)	Wafer price (US\$)	Wafer revenue TAM (US\$ mn)
AI GPU (2027e)										
NVIDIA	B300	40	14	560	850	4nm	2	44	23,042	1,024
	Vera CPU	250	23	5,750	850	3nm	1	228	27,300	6,229
	Rubin R200	740	8	5,920	850	3nm	2	470	27,300	12,827
	Rubin Ultra	130	8	1,040	850	3nm	2	83	27,300	2,253
	MI350 series	24	12	288	110	3nm	8	8	27,300	229
AMD	MI455	157	7	1,099	850	2nm	2	116	30,000	3,489
	MI450 (for Meta)	35	14	490	850	2nm	1	26	30,000	778
	MI500 (Arcadia)	24	4	96	850	2nm	2	10	30,000	305
	Venice CPU	270	21	5,670	155	2nm	8	227	30,000	6,804
AI ASIC (2027e)										
Google	TPU v7p (Ironwood; AVGO)	13	16	208	700	3nm	2	14	27,300	371
	TPU v8i (Sunfish; AVGO)	330	12	3,960	800	3nm	2	296	27,300	8,075
	TPU v8t (Zebrafish; MediaTek)	180	20	3,600	800	3nm	2	269	27,300	7,341
	TPU v9 (Humufish; MediaTek)			400	850	2nm	4	63	30,000	1,905
AWS	Trainium 3	140	17	2,380	700	3nm	2	127	27,300	3,465
GUC's customers		60	20	1,200	645	4nm	1	29	23,042	674
Microsoft	Maia 300	50	11	550	850	2nm	1	29.1	30,000	873
Meta	MTIA 3 (Iris)	55	10	550	850	3nm	2	44	27,300	1,192
Others		18								
Total		2,664						2,119		58,798

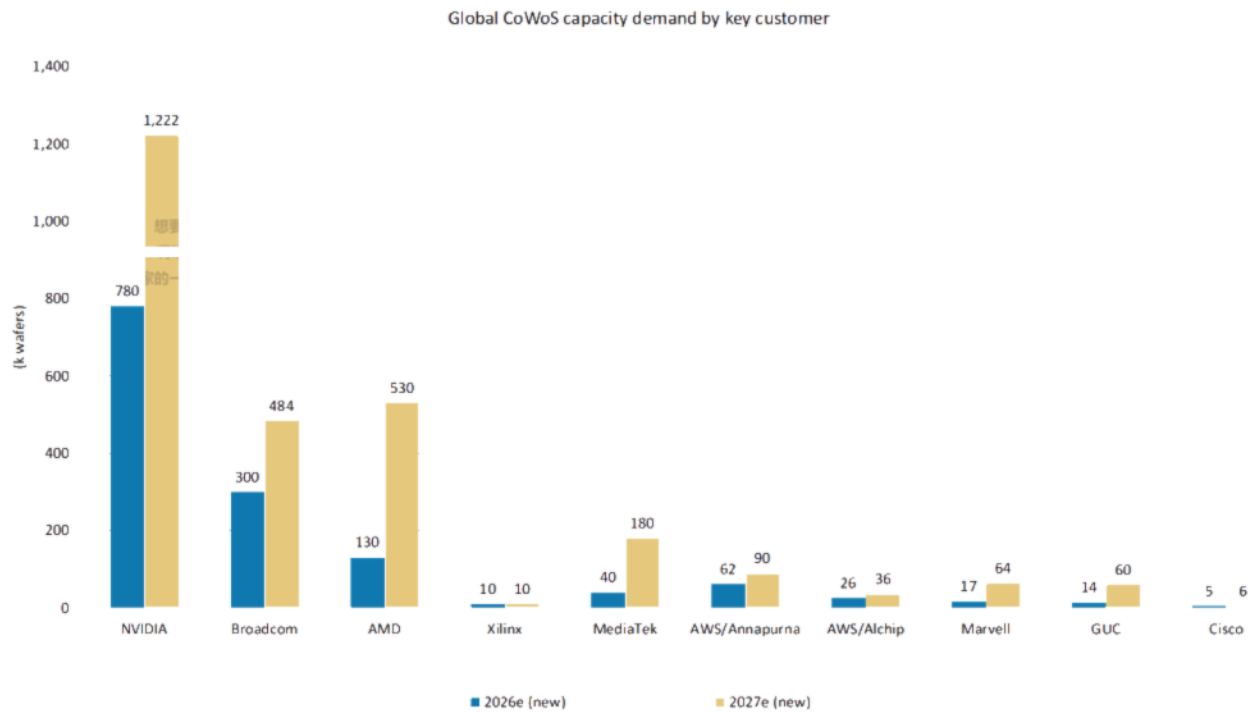
Source: Company data; Morgan Stanley Research (e) estimates. Note: Est. mates are compiled using our Asian supply chain checks.

Exhibit 5: CoWoS allocation breakdown by CoW vendor

(k wafer)	2023	2024	2025	2026e	2027e	2023	2024	2025e	2026e	2027e
NVIDIA	53	200	425	780	1,222	45%	54%	62%	56%	45%
TSMC			370	680	1,090					
CoWoS-L			390	650	910					
CoWoS-S			0	20	50					
CoWoS-R			0	10	130					
Non-TSMC			35	100	132					
Amkor			35	100	132					
CoWoS-S			20	20	12					
CoWoS-R			15	80	120					
ASE/SPIL			0	0	0					
CoWoS-S			0	0	0					
CoWoS-R			0	0	0					
Broadcom	23	68	85	300	484	20%	18%	12%	22%	18%
TSMC			83	260	420					
CoWoS-L			0	15	55					
CoWoS-S			83	245	365					
ASE/SPIL			2	30	40					
CoWoS-L			0	0	0					
CoWoS-S			2	30	40					
Amkor				10	24					
CoWoS-S				10	24					
AMD	7	40	60	130	530	6%	11%	9%	9%	20%
TSMC			60	80	240					
CoWoS-L			0	70	230					
CoWoS-S			60	10	10					
ASE/SPIL			0	50	170					
CoWoS-L			0	50	170					
Amkor				0	120					
CoWoS-L				0	120					
Xilinx	3	10	10	10	10	3%	3%	1%	1%	0%
MediaTek				40	180				3%	7%
TSMC				40	180					
CoWoS-S				40	180					
AWS/Annapurna			60	62	90				4%	3%
TSMC			60	32	54					
CoWoS-R			60	32	54					
ASE/SPIL				30	36					
CoWoS-R				30	36					
AWS/Alchip	9	16	5	26	36	8%	4%	1%	2%	1%
Intel Habana	0	7	9	0	0	0%	2%	1%	0%	0%
Marvell	1	18	15	17	64	1%	5%	2%	1%	2%
TSMC				5	50					
CoWoS-L				5	50					
CoWoS-R			15	0	0					
ASE/SPIL				12	14					
CoWoS-R				12	14					
GUC	1	1	2	14	60	1%	0%	0%	1%	2%
Cisco		2	3	5	6		1%	0%	0%	0%
Others	20	10	15	10	12	17%	3%	2%	1%	0%
Total demand	117	372	689	1,394	2,694	100%	100%	100%	100%	100%

Source: Company data, Morgan Stanley Research (e) estimates. Note: Estimates are compiled using our Asian supply chain checks.

Exhibit 6: Global CoWoS demand breakdown: 2026e vs. 2027e

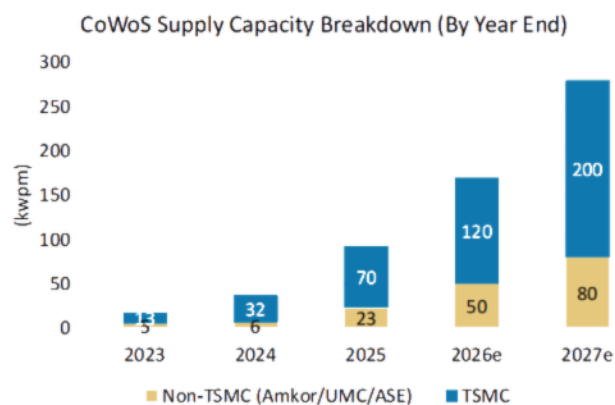


Source: Company data; Morgan Stanley Research (e) estimates; note: estimates are compiled using our supply chain checks

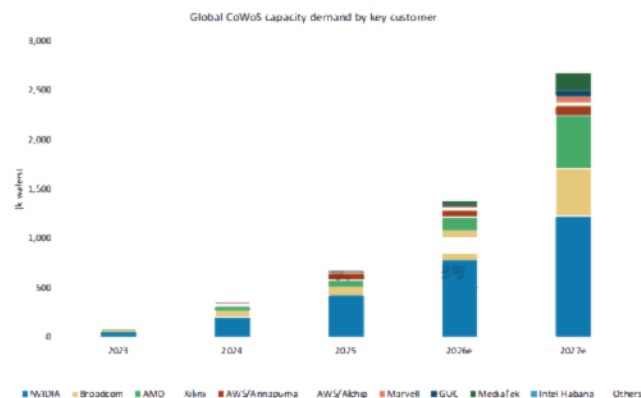
Exhibit 7: Global CoWoS demand Y/Y growth profile

Y/Y	2023	2024	2025	2026e	2027e
NVIDIA	119%	280%	113%	84%	57%
Broadcom	56%	191%	25%	253%	61%
AMD	485%	470%	50%	117%	308%
Xilinx	63%	242%	0%	0%	0%
AWS/Annapurna				3%	45%
AWS/Alchip		71%	(69%)	420%	38%
Marvell	(22%)	1438%	(17%)	13%	276%
GUC		(15%)	300%	600%	329%
Cisco			36%	67%	10%
Others	23%	(49%)	50%	(33%)	20%
Total demand	95%	218%	85%	102%	93%

Source: Company data; Morgan Stanley Research (e) estimates; note: estimates are compiled using our supply chain checks

Exhibit 8: Global CoWoS capacity expansion by year-end and by vendor

Source: Company data; Morgan Stanley Research (e) est. mates

Exhibit 9: Global CoWoS consumption, by customer

Source: Company data; Morgan Stanley Research (e) est. mates; note: estimates are compiled using our supply chain checks

Exhibit 10:

AI HBM consumption: up to 30bn Gb in 2026

AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	HBM chip density (GB)	HBM chip units	Total HBM size (GB)	HBM generation	HBM vendor	Total HBM demand (k GB)
AI GPU (2026e)										
NVIDIA	B300	390	14	5,460	36	8	288	HBM3e 12hi	Hynix/Micron/Samsung	1,572,480
	Vera CPU	90	23	2,070						
	Spectrum/CPX	60		-						
	Rubin R200	260	8	2,080	36	8	288	HBM4 12hi	Hynix/Micron/Samsung?	599,040
AMD	H200	20	27	540	24	6	141	HBM3e 8hi	Hynix	76,140
	MI300	3	12	36	24	8	192	HBM3e 12hi	Samsung	6,912
	MI350/375	7	12	84	36	8	288	HBM3e 12hi	Samsung/Micron	24,192
	MI455	65	7	455	36	12	432	HBM4 12hi	Samsung/Micron	196,560
	Venice	50	21	1,050						
AI ASIC (2026e)										
Google	TPU v7p (Ironwood; AVGO)	145	16	2,320	24	8	192	HBM3e 8hi	Hynix/Samsung	445,440
	TPU v8i (Sunfish; AVGO)	80	12	960	36	8	288	HBM3e 12hi	Hynix/Samsung/Micron	276,480
	TPU v8t (Zebrafish; MediaTek)	40	20	800	36	6	216	HBM3e 12hi	Hynix/Micron	172,800
AWS	Trainium 3	100	17	1,700	36	4	144	HBM3e 12hi	Hynix/Samsung/Micron	244,800
GUC's customers		10	20	200	24	6	144	HBM3e 8hi	Samsung?	1,152
Microsoft	Maia 200	4	29	116	16	4	64	HBM3	Samsung	7,424
	Maia 300	5	11	55	36	8	288	HBM4 12hi	Samsung	15,840
Meta	MTIA 3 (Iris)	15	10	150	36	8	288	HBM3e 12hi	Hynix/Samsung	43,200
Total		1,424								3,741,260
Total HBM demand (mn Gb)										29,930

Source: Company data; Morgan Stanley Research (e) est. mates; Note: Est. mates are compiled using our Asian supply chain checks

Exhibit 11: AI wafer consumption: At least US\$26bn in 2026

AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	Compute die size	Geometry	Compute die units	Wafer consumption (k wafers)	Wafer price (US\$)	Wafer revenue TAM (US\$ mn)
AI GPU (2026e)										
NVIDIA	B300	390	14	5,460	850	4nm	2	433	21,945	9,510
	Vera CPU	90	23	2,070		3nm				
	Spectrum/CPX	60		-						
	Rubin R200	260	8	2,080	850	3nm	2	165	26,000	4,292
AMD	H200	20	27	540	814	4nm	1	15	21,945	331
	MI300	3	12	36	110	5nm	8	1	18,000	19
	MI350/375	7	12	84	850	3nm	2	5	26,000	139
	MI455	65	7	455	850	2nm	2	48	28,125	1,354
	Venice	50	21	1,050	155	2nm	8	70	28,125	1,969
AI ASIC (2026e)										
Google	TPU v7p (Ironwood; AVGO)	145	16	2,320	700	3nm	2	152	26,000	3,942
	TPU v8i (Sunfish; AVGO)	80	12	960	800	3nm	2	72	26,000	1,864
	TPU v8t (Zebrafish; MediaTek)	40	20	800	800	3nm	2	60	26,000	1,554
AWS	Trainium 3	100	17	1,700	700	3nm	2	91	26,000	2,357
GUC's customers		10	20	200	645	4nm	1	5	21,945	107
Microsoft	Maia 200	4	29	116	700	3nm	1	3.0	26,000	79
	Maia 300	5	11	55	850	2nm	1	2.9	28,125	82
Meta	MTIA 3 (Iris)	15	10	150	850	3nm	2	11.9	26,000	310
Total		1,424						1,165		28,619

Source: Company data; Morgan Stanley Research (e) est. mates; Note: Est. mates are compiled using our Asian supply chain checks

Aligning Blackwell and Rubin Chip Shipments with HGX and Rack Shipments

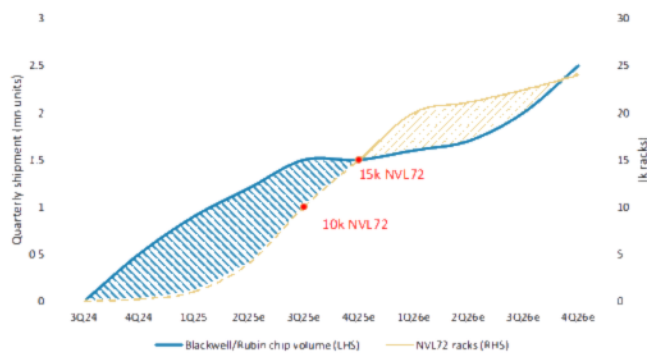
We analyzed Blackwell chip shipments versus rack shipments in [our last June AI tracker](#). We believe TSMC's Blackwell chip output, based on CoWoS-L capacity, aligns more closely with underlying demand as reflected in AI capex trends.

We now incorporate HGX systems (8 GPUs per server) into our chip consumption model and assume that nine HGX servers are equivalent to one NVL72 rack. We forecast 5.4mn Blackwell units in 2026, with chip supply sufficient to meet Grace Blackwell NVL72 demand by 2H26.

Our latest checks also suggest shipments of nearly 7mn Rubin and Rubin Ultra units in 2027, while total Rubin NVL72 rack shipments could reach 90k units in 2027 ([Exhibit 14](#)).

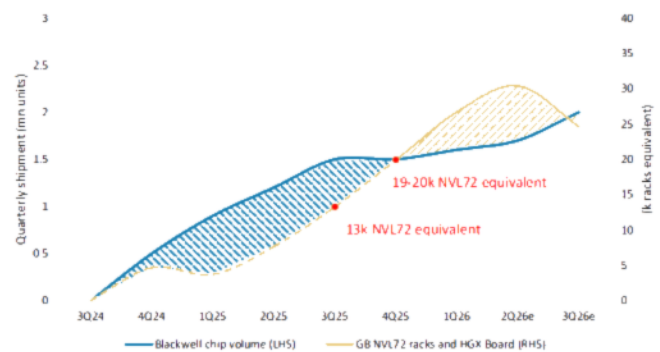
All Blackwell chip "inventory" has proven to be supply-chain buffer stock and should be fully absorbed by 2026. As a result, we see little reason for concern over inventory levels. We expect Rubin to follow a similar pattern. Rubin should begin ramping in 3Q26, with rack shipments starting in 4Q26.

Exhibit 12: (old) Blackwell chip vs. GB NVL72 rack shipment



Source: Morgan Stanley Research estimates

Exhibit 13: (new) Blackwell chip vs. GB NVL72 and HGX equivalent rack shipment



Source: Morgan Stanley Research estimates

Exhibit 14:

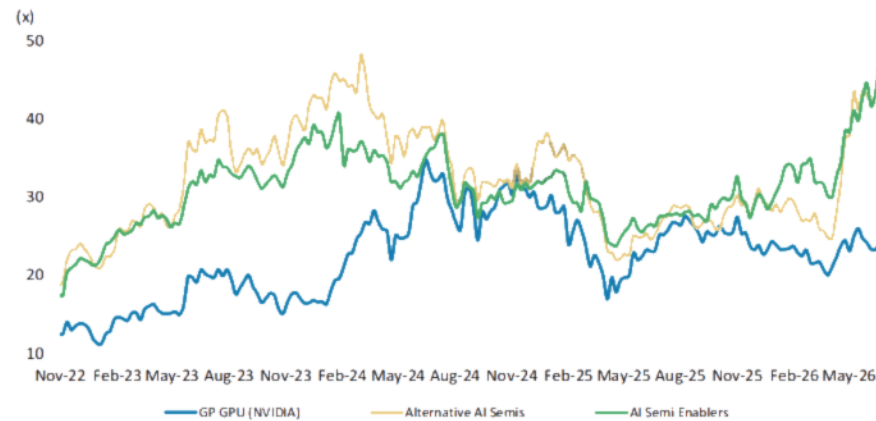
Rubin chip shipments vs. VR NVL72 rack shipments



Source: Morgan Stanley Research estimates

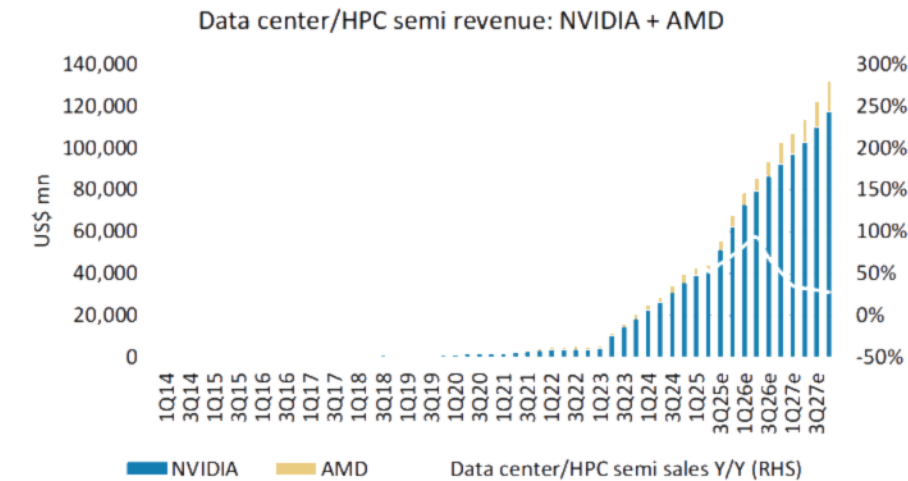
AI semis – stock implications, P/E multiples, revenue exposure

Exhibit 15: P/E multiple trend of AI semis



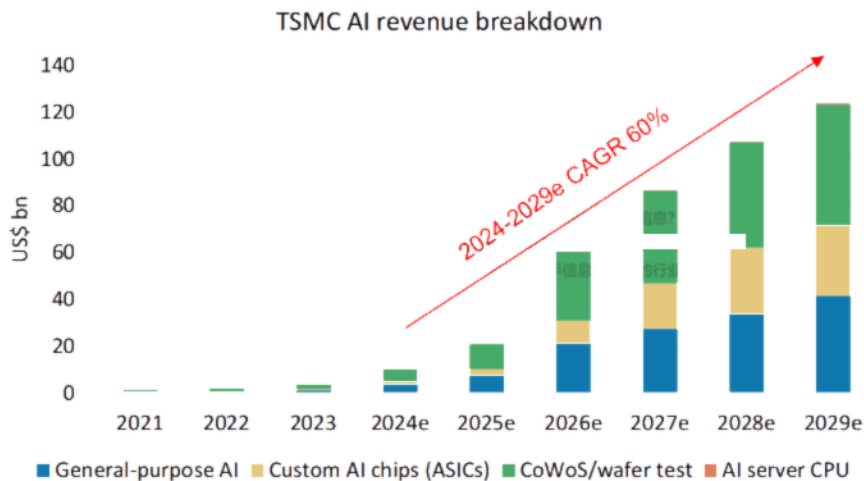
Note: Alternative AI semis group: AMD, Alchip, Andes, Marvell, Broadcom; AI semi enablers group: TSMC, Synopsys, Cadence, ASML, BESS, Ibrsen, KYEC, Advantest. Source: Company data, Morgan Stanley Research

Exhibit 16: We still expect AI chip revenue to rise QoQ



Source: Company data, Refinitiv, Morgan Stanley US Research (e) estimates

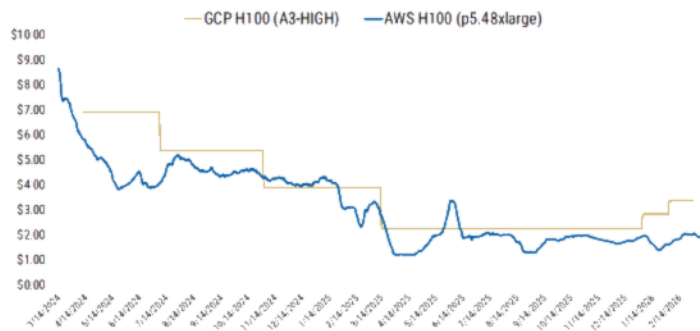
Exhibit 17: TSMC's AI-related revenue 2024-29e CAGR could reach 60%



Source: Company data, Morgan Stanley Research (e) estimates

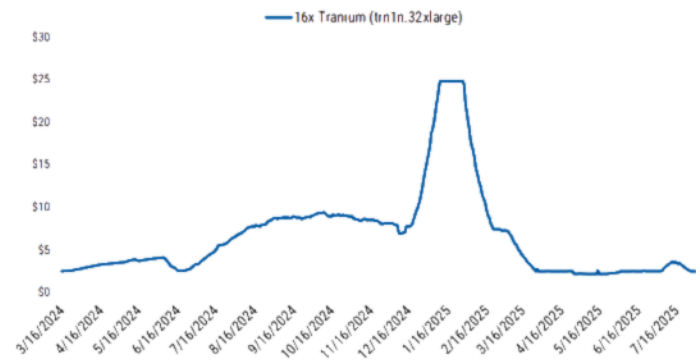
AI GPU and ASIC rental price tracker

Exhibit 18: AI GPU H100 per GPU per hour as of end-March



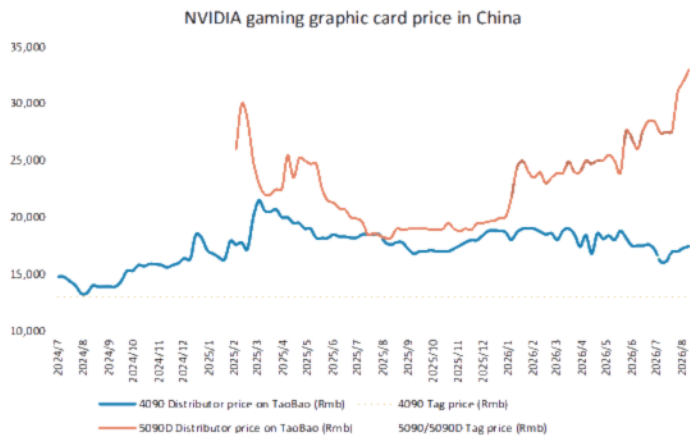
Source: Company data, Morgan Stanley Research

Exhibit 19: AI ASIC equivalent computing power – 16x Inferentia 2 per hour



Source: Company data, Morgan Stanley Research

Exhibit 20: NVIDIA 5090 graphic cards pricing has rebounded recently, mainly in response to market expectations for price hikes and strong AI inference demand from China



Source: TaoBao, Morgan Stanley Research, various media (e.g., Video, Jan. 5, 2026)



Key Featured Reports on the AI Supply Chain

Asia-Pacific Technology: AI Supply Chain: Further Updates on 2027 CoWoS Allocation and ASIC Dynamics (8 Jul 2026)

Asia-Pacific Technology: AI Supply Chain: Preliminary 2027 TSMC CoWoS Allocation (23 Jun 2026)

Asia-Pacific Technology: AI Supply Chain: Cloud Capex Strength; More CoWoS Allocation for TPU (7 May 2026)

Asia-Pacific Technology: AI Supply Chain: Addressing Questions on Nvidia GPU/LPU and Google TPU (9 Apr 2026)

Greater China Semiconductors: AI Supply Chain Tracker: Key Investor Feedback from GTC/OFC (23 Mar 2026)

Asia Technology: AI Supply Chain: CPO and ASIC Dynamic Update (3 Mar 2026)

Foundation

Technology: Rise of the AI Agent – Global Implications (19 Apr 2026)

Global Technology: Supply-chain Reorientation (24 Jul 2025)

Global Technology: China – AI: The Sleeping Giant Awakens (13 May 2025)

Global Technology: AI Cloud Capex in the Spotlight (26 Feb 2025)

Global Semiconductors: AI ASIC 2.0: Potential winners (15 Dec 2024)

Global Technology: Global Technology – Dawn of the AI Smartphone Era: Edge AI – Apple Intelligence Fuels Innovation – More Charts, Fewer Words (17 Jul 2024)

Global Technology: AI PCs To Usher In The Next Leg Of PC Market Growth (21 May, 2024)

Key upstream AI supply chain companies

Asia-Pacific Technology: Connecting Dots: HK investors' feedback on Asian semis (9 Aug 2026)

King Yuan Electronics Co Ltd: AI revenue pushed out to 4Q26 and 2027 (7 Aug 2026)

Cambricon Technology Corporation: 2Q26 Revenue Miss, Profit in Line; Supply visibility Remains Key (7 Aug 2026)

AP Memory Technology Corp: More upside ahead; move to Top Pick (6 Aug 2026)

FOCI Fiber Optic Communications Inc: We Expect 2Q Earnings to Mark the Near-Term Trough – Stay OW (6 Aug 2026)

WinWay Technology Co Ltd: Revising Full-year Revenue, But Tempering GM Expectations (6 Aug 2026)

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

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

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