Not for redistribution without written consent of Morgan Stanley

Morgan Stanley | RESEARCH

IDEA

July 8, 2026 06:45 AM GMT

Asia-Pacific Technology | Asia Pacific

AI Supply Chain: Further Updates on 2027 CoWoS Allocation and ASIC Dynamics

AMD's 2027 CoWoS allocation remains at 240k for now, with combined MI455/MI450 shipments at 1.5mn. TPU MP timing skews toward 4Q. Blackwell chip "inventory" turned out to be a supply chain buffer and will be fully consumed by 2026. We expect ~7mn Rubin chips and 90k NVL72 racks in 2027.

Further updates on our AMD CoWoS allocation: We published our preliminary 2027 CoWoS allocation last week and received many investor queries on AMD's (covered by Joe Moore) CoWoS number. Some MI300 series and upcoming MI500 production will still run in 2027. However, the MI400 series will come in two versions: (1) MI455: the standard version with 2 compute dies and 12 HBM4 12hi, paired with the Helios rack (18 CPUs and 72 GPUs), with Microsoft, AWS, and Oracle as key customers; (2) MI450: a Meta-customized, half-size chip paired with 1 compute die and 6 HBM4 12hi (9 CPUs and 36 GPUs). We still expect AMD's 2027 CoWoS at 240k, but do not rule out execution risk, given its prior record of trimming CoWoS bookings in 2026. 2027 chip shipments: MI455 1mn and MI450 500k. For CPU, Venice is AMD's first CPU adopting CoWoS, and we see CoW production concentrated in OSATs—ASE/SPIL, Amkor, and Powertech. Total CPU chip shipments could reach 5.7–6mn units in 2027 vs. 1mn in 2026.

Google TPU progress update: Our industry checks suggest KYEC's 3Q26 revenue could grow close to 10% Q/Q, below our initial 15% Q/Q expectation, mainly due to slight Rubin and Sunfish delays and MediaTek's smartphone SoC order trim. We do see Sunfish shipments largely in 4Q, with full-year volume at 960k units. With Sunfish and Rubin pushed out, we believe more revenue will concentrate in 4Q26 or 1Q27 without order cuts. Many investors asked about Zebrafish delays; we still see its 4Q26 volume ramp unchanged.

Comparing the Blackwell chip/rack shipment and Rubin chip/rack shipment: We provided a chip vs. rack shipment comparison for Blackwell in last June's AI tracker. We now include HGX (8-GPU per server) in our chip consumption model, and treat 9 HGX servers as equivalent to one NVL72 rack. We expect 5.4mn Blackwell units in 2026, and chip volume could meet Grace Blackwell NVL72 demand by 2H26. Our latest checks also suggest close to 7mn Rubin and Rubin Ultra units in 2027, while total Rubin NVL72 server racks could reach 90k in 2027 (Exhibit 12). All Blackwell chip "inventory" turned out to be a supply-chain buffer and will be fully consumed by 2026, so we see no need to worry about inventory issues. We think Rubin will follow a similar pattern.

(continued on the next page)

MORGAN STANLEY TAIWAN LIMITED+
Charlie Chan
Equity Analyst
Charlie.Chan@morganstanley.com +886-2-7730-7125

Daniel Yen, CFA
Equity Analyst
Daniel.Yen@morganstanley.com +886-2-7730-7368

MORGAN STANLEY ASIA LIMITED+
Daisy Dai, CFA
Equity Analyst
Daisy.Dai2@morganstanley.com +852 2848-7310

MORGAN STANLEY TAIWAN LIMITED+
Tiffany Yeh
Equity Analyst
Tiffany.Yeh2@morganstanley.com +886-2-7730-3032

Lucas Wang
Research Associate
Lucas.Wang.Zhe@morganstanley.com +886-2-2730-2875

MORGAN STANLEY ASIA LIMITED+
Ethan Jia
Research Associate
Ethan.Jia@morganstanley.com +852 3963 2187

Henry Zhao
Research Associate
Henry.Zhao@morganstanley.com +852 2239 7231



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.

Meta ASIC progress update: On the other hand, Meta cancelled its original 2nm ASIC Olympus earlier this year; a new 2nm chip, code-named Apollo, succeeds it. Broadcom (covered by Joe Moore) will continue design services, with mass production in 1Q28. For GUC, we think it is likely to win one of the Meta ASIC projects led by the Rivos team, possibly a side project from the usual MTIA product line. The AI ASIC targets tape-out in 1H27, with potential CoWoS out by 2027 year-end or 1H28.

Introducing Our 2027 Global CoWoS Capacity Update And Comparing Chip Shipments vs. Racks

We published our preliminary 2027 CoWoS allocation last week and received lots of investor queries around AMD's (covered by Joe Moore) CoWoS number. There will still be some production of the MI300 series and the upcoming MI500 in 2027.

However, for the MI400 series, there will be two versions coming: 1) MI455: the standard version, including 2 compute dies and 12 HBM4 12hi, pairing with the Helios rack (18 CPUs and 72 GPUs), with key customers being Microsoft, AWS and Oracle; 2) MI450: customized version for Meta, with the chip being half size pairing with 1 compute die and 6 HBM4 12hi (9 CPUs and 36 GPUs) (see Exhibit 1). We still expect AMD's CoWoS to be 240k in 2027, but do not rule out execution risk, given its prior record of trimming CoWoS bookings in 2026. We estimate chip shipments in 2027 will be 1mn for MI455 and 500k for MI450.

For CPU, Venice is AMD's first CPU to adopt CoWoS, and we see CoW production concentrating in OSATs, including ASE/SPIL, Amkor and Powertech. We expect total chip shipments of 5.7-6mn units in 2027 vs. 1mn units in 2026.

Exhibit 1: AI HBM consumption: up to 50bn Gb in 2027

		General assumptions			HBM assumptions					
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
AMD	MI350 series	24	12	288	36	8	288	HBM3e 12hi	Samsung/Micron	82,944
	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: Estimates are compiled using our Asian supply chain checks

Exhibit 2: AI wafer consumption: at least US$47bn 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	58	27,300	1,588
AMD	MI350 series	24	12	288	110	3nm	8	8	27,300	229
	MI455	157	7	1,099	850	2nm	2	13	30,000	400
	MI450 (for Meta)	35	14	490	850	2nm	1	3	30,000	89
	MI500 (Arcadia)	24	4	96		2nm			30,000	
	Venice CPU	270	21	5,670		2nm				
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	2	32	27,300	867
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**						**1,700**		**46,208**

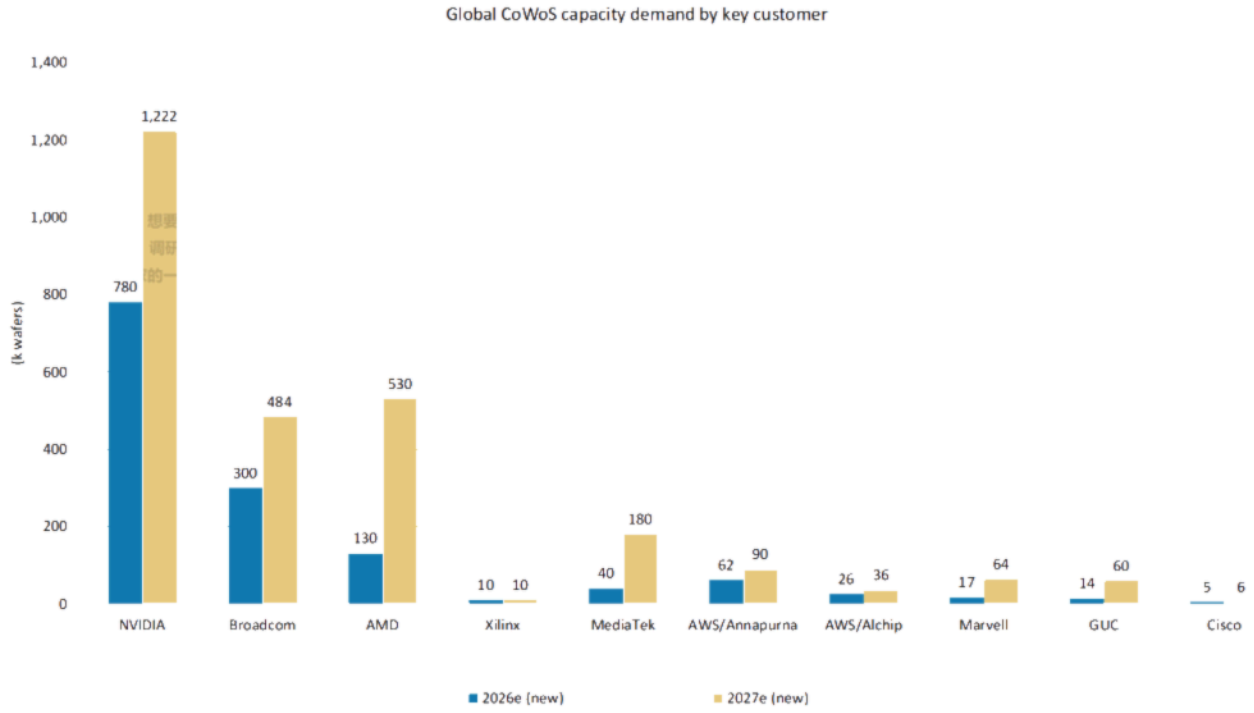
Source: Company data, Morgan Stanley Research (e) estimates. Note: Estimates are compiled using our Asian supply chain checks

Exhibit 3: 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			390	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					
tf			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	17	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 4: Global CoWoS demand breakdown: 2026e vs. 2027e



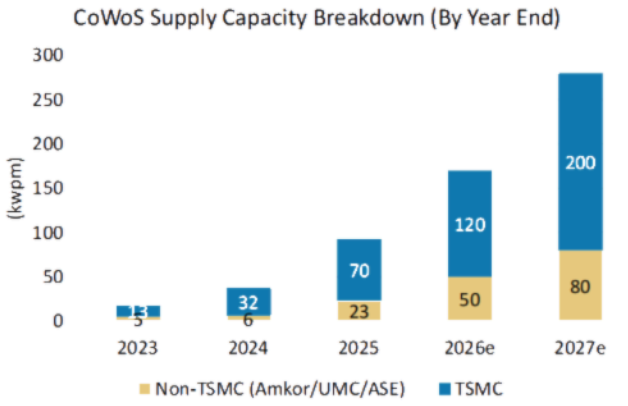
Source: Company data, Morgan Stanley Research (e) estimates; note: estimates are compiled using our supply chain checks

Exhibit 5: 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%**

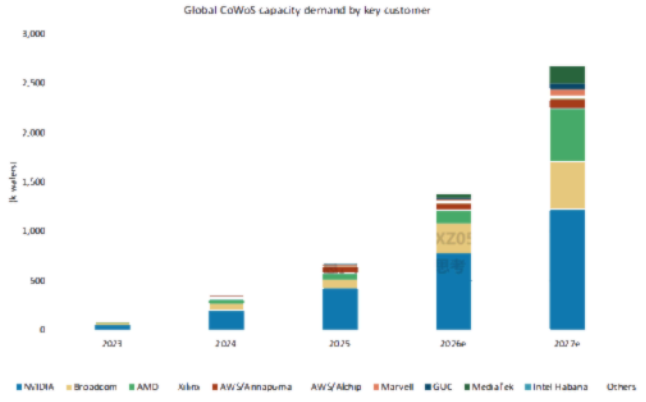
Source: Company data, Morgan Stanley Research (e) estimates; note: estimates are compiled using our supply chain checks

Exhibit 6: Global CoWoS capacity expansion by year end and by vendor



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

Exhibit 7: Global CoWoS consumption, by customer



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

Exhibit 8:

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)										
	B300	390	14	5,460	36	8	288	HBM3e 12hi	Hynix/Micron/Samsung	1,572,480
	Vera CPU	90	23	2,070						
NVIDIA	Spectrum/CPX	60		-						
	Rubin R200	260	8	2,080	36	8	288	HBM4 12hi	Hynix/Micron/Samsung?	599,040
	H200	20	27	540	24	6	141	HBM3e 8hi	Hynix	76,140
	MI300	3	12	36	24	8	192	HBM3e 12hi	Samsung	6,912
AMD	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)										
	TPU v7p (Ironwood; AVGO)	145	16	2,320	24	8	192	HBM3e 8hi	Hynix/Samsung	445,440
Google	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) estimates. Note: Estimates are compiled using our Asian supply chain checks

Exhibit 9: AI wafer consumption: at least US$26bn in 2026

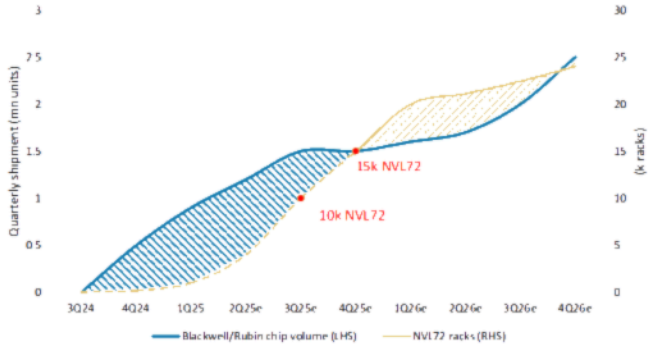
AI chip vendor	Product name	CoWoS capacity allocation (k wafers)	Chips per CoWoS wafer	Implied shipments (k)	Compute die size	Geometry	Compute die units	Wafer consumption (k wafers)	Wafer price (US$)	Wafer revenue TAM (US$ mn)
AI GPU (2026e)										
	B300	390	14	5,460	850	4nm	2	433	21,945	9,510
	Vera CPU	90	23	2,070		3nm				
NVIDIA	Spectrum/CPX	60		-						
	Rubin R200	260	8	2,080	850	3nm	2	165	26,000	4,292
	H200	20	27	540	814	4nm	1	15	21,945	331
	MI300	3	12	36	110	5nm	8	1	18,000	19
AMD	MI350/375	7	12	84	110	3nm	8	2	26,000	64
	MI455	65	7	455	110	2nm	8	22	28,125	620
	Venice	50	21	1,050						
AI ASIC (2026e)										
	TPU v7p (Ironwood; AVGO)	145	16	2,320	700	3nm	2	152	26,000	3,942
Google	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**						**1,066**		**25,842**

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

Aligning Blackwell and Rubin chip to HGX/Rack Shipments

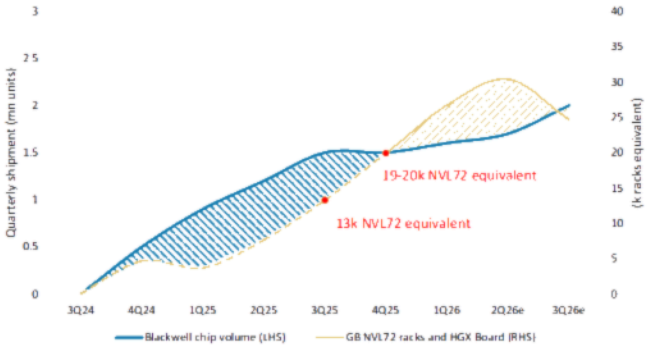
We did an analysis of Blackwell chip shipments vs. racks in our last June AI tracker. We think TSMC's Blackwell chip output (calculated by CoWoS-L capacity) aligns more closely with strong demand (as measured by AI capex). We now include HGX (8-GPU per server) in our chip consumption model and treat 9 HGX servers as equivalent to one NVL72 rack. We expect 5.4mn Blackwell units in 2026, and chip volume could meet Grace Blackwell NVL72 demand by 2H26. Our latest checks also suggest close to 7mn Rubin and Rubin Ultra units in 2027, while total Rubin NVL72 server racks could reach 90k in 2027 (Exhibit 12). All Blackwell chip "inventory" turned out to be a supply-chain buffer and will be fully consumed by 2026, so we see no need to worry about inventory issues. We think Rubin will follow a similar pattern. Rubin will start ramping in 3Q26, with rack shipments starting in 4Q26.

Exhibit 10: (old) Blackwell chip vs. GB NVL/2 rack shipment



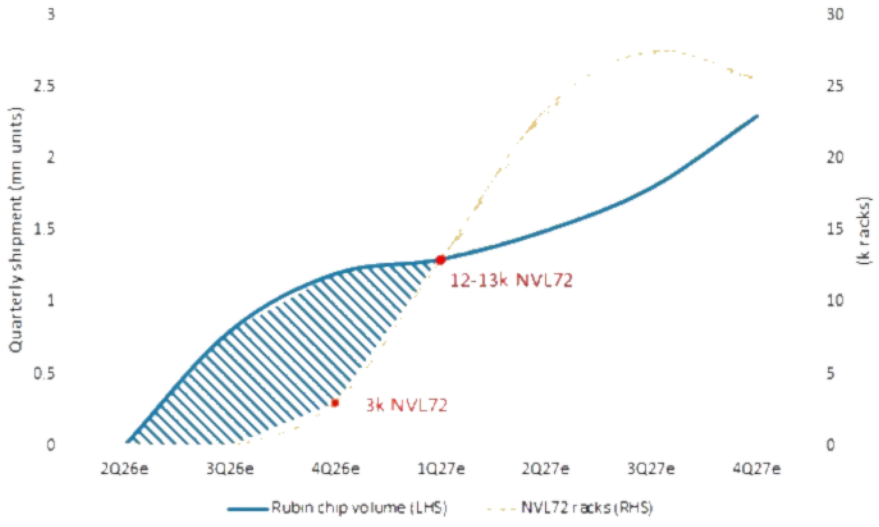
Source: Morgan Stanley Research estimates

Exhibit 11: (new) Blackwell chip vs. GB NVL/2 and HGX equivalent rack shipment



Source: Morgan Stanley Research estimates

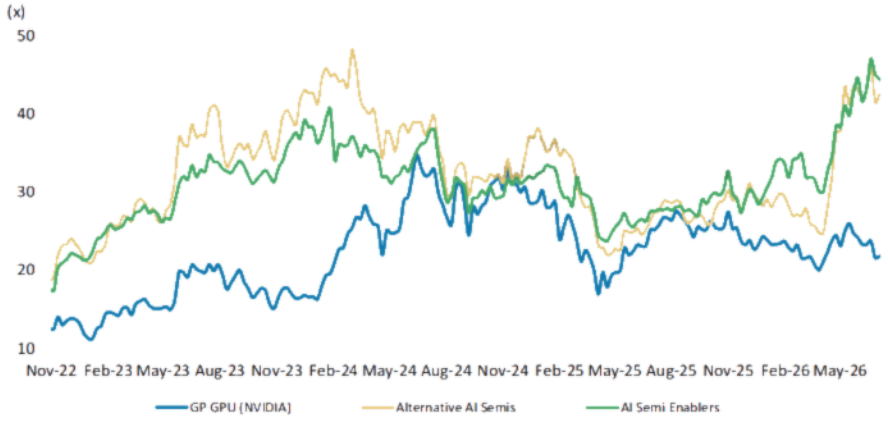
Exhibit 12:
Rubin chip shipments vs. VR NVL72 rack shipments



Source: Morgan Stanley Research estimates

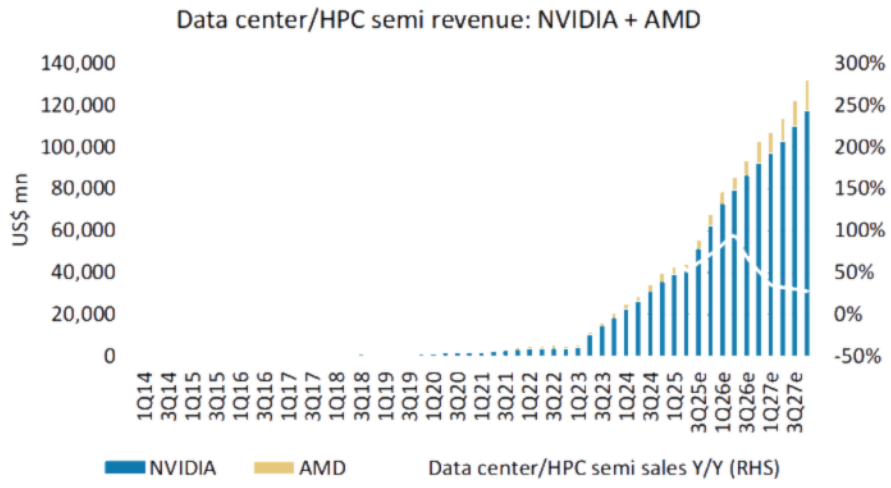
AI semis – stock implications, P/E multiples, revenue exposure

Exhibit 13: P/E multiple trend of AI semis



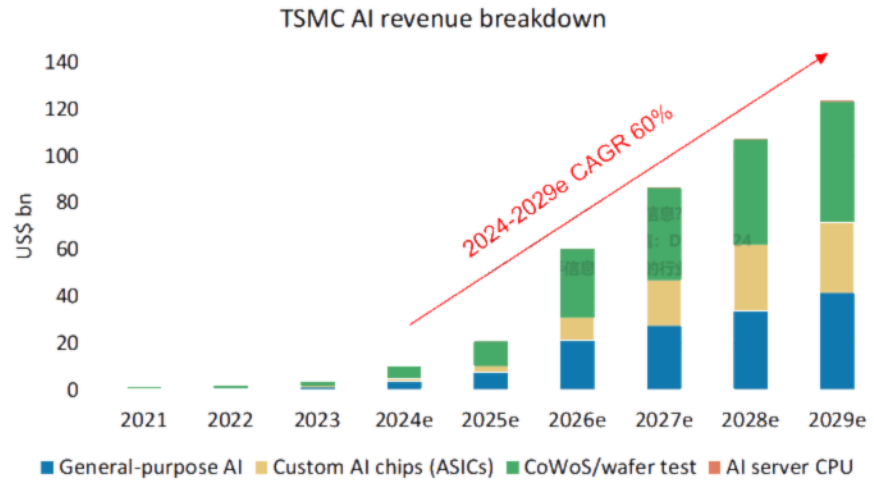
Note: Alternative AI semis group: AMD, Alchip, Andes, Marvell, Broadcom. AI semi enablers group: TSMC, Synopsys, Cadence, ASML, BESI, Ibiden, KYEC, Advantest. Source: Company data, Morgan Stanley Research

Exhibit 14: We still expect AI chip revenue to rise QoQ



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

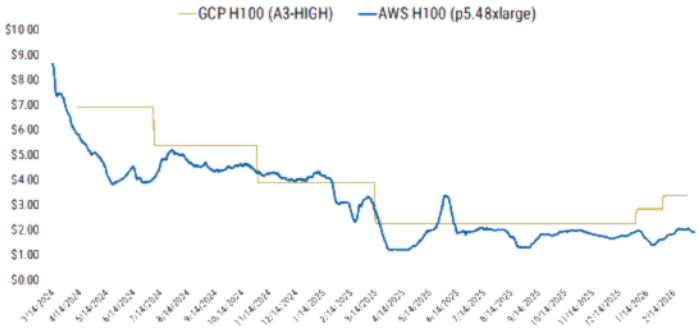
Exhibit 15: TSMC's AI-related revenue 2024-29e CAGR could reach 60%



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

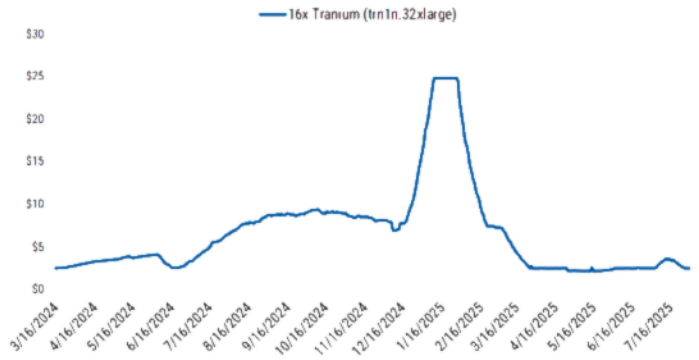
AI GPU and ASIC rental price tracker

Exhibit 16: AI GPU H100 per GPU per hour as of end-March



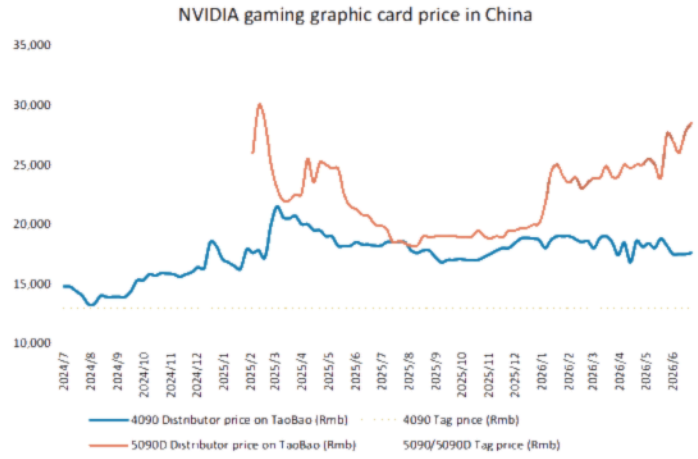
Source: Company data, Morgan Stanley Research

Exhibit 17: AI ASIC equivalent computing power – 16x Inferentia 2 per hour



Source: Company data, Morgan Stanley Research

Exhibit 18: 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., Vice, Jan 3, 2026)

人在草木间

Key Featured Reports on the AI Supply Chain

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)

Asia-Pacific Technology: AI Supply Chain: CES implications, ASIC production, China AI chips (6 Jan 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

WinWay Technology Co Ltd: AI/HPC Continue to Drive Stronger 3Q Revenue; OW (6 Jul 2026)

ASMPT Ltd: 2Q26 Preview: OSAT and PCB – Dual Growth Engines (6 Jul 2026)

Asia-Pacific Technology: Connecting Dots in Tech Supply Chain: CoPoS, T-Glass/TPU, Glassbridge/FAU (6 Jul 2026)

Greater China Semiconductors: CPO Supply Chain Updates; More on GlassBridge (5 Jul 2026)

Hygon Information Technology Co., Ltd.: Integrated CPU+GPU Compute Platform; Initiate at OW (2 Jul 2026)

Greater China Semiconductors: Old memory: Near-term outlook remains positive (1 Jul 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.