IT Hardware

GB200 rack the spotlight for AI server in 2H24-2025F

Key message

1. We expect total AI GPU shipments for Nvidia will rise to 4.0mn units in 2024F, and 4.7-4.8mn units in 2025F, based on a Hopper GPU allocation of 85% in 2024F and 25% in 2025F, and Blackwell GPU allocations of 10% in 2024F and 75% in 2025F.

2. Blackwell production shipments will start in 2Q24F, ramp up in 3Q24F, and customers should have the GPUs in data centers in 4Q24F. We expect Nvidia to see a 10% sales weighting for Blackwell in 2024F and 70-80% in 2025F. CSPs, enterprise, consumer IoT and governments are clients for AI GPUs & the main sales growth drivers.

3. GB200 rack demand is from top CSPs and server brands, and we expect 40-45k racks to ship in 2025F, with the top-4 US CSP comprising 80-85% of demand, and Hon Hai (2317 TT) & Quanta (2382 TT) are the key assemblers. Thermal solution, power supply, DAC cable & high-end optical fiber connector sales will also benefit from this trend.

Event

Per Nvidia's (US) guidance that the Blackwell platform is in production in 2Q24, with a ramp up in 3Q24F, customers should have Blackwell GPUs installed in data centers in 4Q24F. GPU transition hiccups will be limited due to strong AI demand, so we expect AI server supply chain to continue to benefit from growing sales and profits in 2024-25F.

Impact

Nvidia's Blackwell GPUs in the pipeline to boost 2025F AI server sales. On increased demand for Nvidia's AI GPUs, TSMC (2330 TT) has expanded CoWoS capacity from 13k wafers per month (kwpm) in 4Q23 to 40kwpm in 4Q24F and 55kwpm in 4Q25F expected. Total AI GPU shipments for Nvidia will rise to 4.0mn units in 2024F, versus our previous expectation of 3.55mn units, and up to 4.7-4.8mn units in 2025F, based on a Hopper GPU allocation of 85% in 2024F and 25% in 2025F, and Blackwell GPU allocations of 10% in 2024F and 75% in 2025F. The market expects GB200 AI server models to be in high demand on strong client interest. We think Blackwell GPUs should come on stream in September, with some volume in the market in 4Q24F. B100 and B200 GPUs, and the GB200, should launch in the market around the same time, with significant volume available in 2025F. Combined with AMD's (US) MI300X/350, Intel's (US) Gaudi 2 and 3, and ASIC shipments (mainly from Google's (US) TPU and Amazon Web Services' (AWS; US) Trainium), we expect total training AI GPU shipments will grow to 5.5mn units in 2024F and 7.9mn units in 2025F, translating to AI training server shipments of 715k units in 2024F and 1.25mn in 2025F. Training AI servers will comprise 6% of total server shipments in 2024F and 10% in 2025F. The total AI server shipment weighting will be even higher if inference servers are taken into consideration. On AI servers' high ASP, we forecast an AI server revenue weighting of 60-70% of global server revenue in 2024-25F.

AI servers carry higher content value for thermal, power supply & rack assembly sectors. During a May 22 earnings call, Nvidia cited Blackwell production shipments will start in 2Q24F, ramp in 3Q24F, and customers should have the GPUs in data centers in 4Q24F. We expect Nvidia to see a 10% GPU shipments weighting for Blackwell in 2024F and 70-80% in 2025F. CSP, enterprise, consumer IoT and sovereign are clients for the firm's AI GPUs, and the main sales growth drivers. We expect GB200 rack demand among top CSPs and server brands to result in 40-45k racks shipping in 2025F. Microsoft (US) plans to receive 35% of GB200 racks in 2025F, while AWS, Meta (US), Google, Supermicro (US), Dell (US), Oracle (US) are also customers. We expect these CSPs will keep revising up capex in 2025F from upward-revised high level in 2024F. Hon Hai (2317 TT) group's FII (CN) is the major GB200 rack supplier to Microsoft and Oracle, and Quanta Computer (2382 TT) is for AWS, Meta and Google. We expect Hon Hai's rack assembly market share for GB200 servers to be approximately 40%, with Quanta holding around 30%. The remaining 30% will go to Wiwynn (6669 TT), Wistron (3231 TT), ZT Systems (US) and Supermicro. The high ASP of GB200 server racks, of between US$1.8mn-3.5mn for NVL36/ 72 racks, should contribute ODMs' AI server sales weighting to over 50% of their total server revenue in 2025F. Among component sectors, thermal sector will see the greatest content value rise on liquid cooling adoption. With cold plate modules, coolant distribution units (CDU), manifolds, racks, RDHx, fan door and chassis required, we expect the thermal content value will rise to US$40-80k for each GB200 rack, versus an air cooling content value of US$2-3k per rack. On the power supply side, with power consumption per GB200 chip reaching 2.7kW, a NVL72 rack could require power of 120kW, or even higher. We expect the content value for each NVL72 rack will be at least US$18-20k, versus US$6-8K for H100 designs. We are also optimistic about the increase in the number and specifications of NVLinks when GPUs migrate from Hopper-series to Blackwell-series, benefiting DAC cable and high-end optical fiber connector suppliers.

Stocks for Action

We expect key benefacaries of the AI server trend will include Hon Hai, Quanta Computer, Wiwynn, Wistron, Asia Vital Components (AVC; 3017 TT), Auras Technology (3324 TT), Kaori Heat Treatment (8996 TT), Delta Electronics (2308 TT), Chenbro Micom (8210 TT), King Slide Works (2059 TT), Browave (3163 TT), and Jess-Link Products Co. (6197 TT).

Risks

Weak demand; over-ordering of AI servers by CSPs.

Key assumption chart

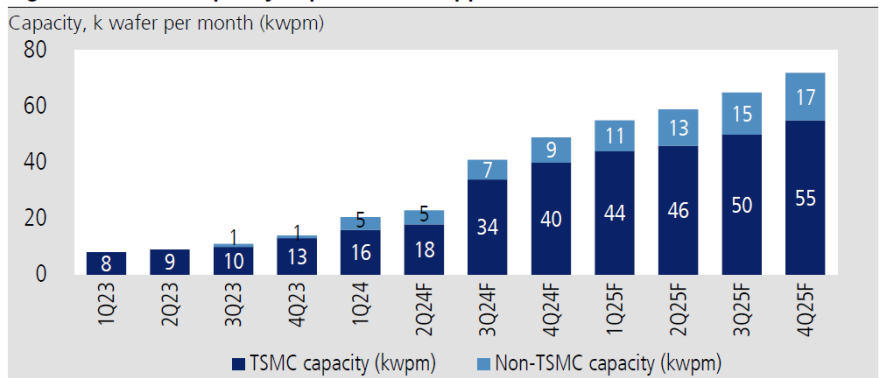**Figure 1: AI GPU & AI server shipments forecast in 2024-25F**

Nvidia's training GPU units (k units)	2024F	2025F
Total Nvidia GPU	4,005	4,760
H100/ A100 shipments	3,605	1,190
Blackwell shipments (B100/B200/GB200)	401	3,570
B100/B200	200	1,474
GB200	200	2,096
Training GPU weighting (%)	**2024F**	**2025F**
H100/ A100 shipments	90	25
Blackwell shipments (B100/B200/GB200)	10	75
B100/B200	50	41
GB200	50	59
Nvidia's AI training server shipments (k units)	**2024F**	**2025F**
H100 /A100 (8 GPU per server)	451	149
B100/B200 (8 GPU per server)	25	184
GB200 (4 GPU per server)	50	524
Nvidia's total AI training server	**526**	**857**
AI training GPU shipments (k units)	**2024F**	**2025F**
Nvidia (H/B-series)	4,005	4,760
AMD (MI300X/350)	420	768
Intel (Gaudi 2/3)	232	450
Subtotal	4,657	5,978
ASIC (Google TPU/AWS Trainium)	864	1,888
Total training GPU	5,521	7,866
AI training server shipments (k units)	**2024F**	**2025F**
Nvidia (H/B-series)	526	857
AMD (MI300X/350)	53	96
Intel (Gaudi 2/3)	29	56
Subtotal	607	1,009
ASIC (Google TPU/AWS Trainium)	108	236
Total AI training server shipments	**715**	**1,245**
Server shipments (k units)	**2024F**	**2025F**
AI training server	715	1,245
Others (General server + AI inferencing)	11,106	11,759
Total server	11,822	13,004
YoY (%)	**2024F**	**2025F**
AI training server	188	74
Others (General server + AI inferencing)	1	6
Total server	5	10
Weighting (%)	**2024F**	**2025F**
AI training server	6	10
Others (General server + AI inferencing)	94	90
Total server	100	100

Source: KGI Research estimates

CoWoS capacity expansion to support AI server GPU demand surge

➢ On increased demand for Nvidia's (US) AI GPUs, TSMC (2330 TT, NT$875, NR) continues to expand CoWoS capacity, along with other suppliers like Amkor (US) and ASE Technology (3711 TT, NT$161.5, NR). We expect TSMC's CoWoS capacity will grow from 13k wafers per month (kwpm) in 4Q23 to 40kwpm in 4Q24F and 55kwpm in 4Q25F, and from 1kwpm, and 9kwpm to 17kwpm from others over the same period (Figure 2).

➢ We therefore expect total AI GPU shipments for Nvidia will grow to 4.0mn units in 2024F, versus our previous expectation of 3.55mn units, and 4.7-4.8mn units in 2025F, based on a Hopper GPU allocation of 85% in 2024F and 25% in 2025F, and Blackwell GPU allocations of 10% in 2024F and 75% in 2025F (Figure 3).

➢ Following Nvidia's showcase of Blackwell GPUs and GB200 superchips during GTC 2024, the market expects GB200 AI server models, under Nvidia's DGX architecture, to achieve high sales volumes on strong client interest. We think Blackwell GPUs may come on stream in September, with some volume in the market in 4Q24F. B100 and B200 GPUs (upgrades from the H100 and H200 under the x86 CPU platform) and the GB200 (Grace CPU and B200 GPU under the ARM platform) should launch in the market around the same time, with significant volume available in 2025F.

➢ In 2025F, we assume 75% of Nvidia's AI GPU sales will be of the Blackwell-series, with 25% of the Hopper-series. Considering CSPs' AI server demand of between 42-43k racks, we expect meeting GB200 demand will require around 2.1mn units of GPUs, comprising 50-60% of the total Blackwell GPU supply, while B100 & B200 will comprise 40-50% of supply in 2025F (Figure 3).

➢ The allocation between Hopper and Blackwell, and the B100, B200 and GB200 are still under discussion, dependant upon Nvidia's market strategy. We will have further review of the volume of Blackwell GPUs and GB200 systems, as they are subject to CoWoS production and yield rate, memory supply (HBM3e) and NVLink switch production.

Figure 2: CoWoS capacity expansion to support AI server demand increase

Capacity, k wafer per month (kwpm)

	TSMC capacity (kwpm)	Non-TSMC capacity (kwpm)
1Q23	8	
2Q23	9	
3Q23	10	1
4Q23	13	1
1Q24	16	5
2Q24F	18	5
3Q24F	34	7
4Q24F	40	9
1Q25F	44	11
2Q25F	46	13
3Q25F	50	15
4Q25F	55	17

Source: KGI Research estimates

Figure 3: Nvidia's AI training server shipments will grow to 526k units in 2024F and 857k units in 2025F

Nvidia's training GPU units (k units)	2024F	2025F
Total Nvidia GPU	4,005	4,760
H100/ A100 shipments	3,605	1,190
Blackwell shipments (B100/B200/GB200)	401	3,570
B100/B200	200	1,474
GB200	200	2,096
Training GPU weighting (%)	**2024F**	**2025F**
H100/ A100 shipments	90	25
Blackwell shipments (B100/B200/GB200)	10	75
B100/B200	50	41
GB200	50	59
Nvidia's AI training server shipments (k units)	**2024F**	**2025F**
H100 /A100 (8 GPU per server)	451	149
B100/B200 (8 GPU per server)	25	184
GB200 (4 GPU per server)	50	524
Nvidia's total AI training server	**526**	**857**

Source: KGI Research estimates

AI server demand keeps growing in 2024-25F; system integration becomes more important with GB200 designs in 2025F

➤ Based on industry CoWoS capacity expansions, we see AI server GPU shipment growth in 2024-25F just meeting AI server demand from CSPs and enterprises. Based on 4mn units of AI GPUs from Nvidia in 2024F, with 90% being A100, H100, and H200 designs and 10% of B100, B200, and GB200 designs in 2024F, we expect AI training server shipments of Nvidia GPUs will reach 526k units, as most AI servers are configured in eight-GPU designs (Figure 3).

➤ During a May 22 earnings call, Nvidia cited Blackwell production shipments will start in 2Q24F, ramp in 3Q24F, and customers should have the GPUs in data centers in 4Q24F. We expect Nvidia to see a 10% sales weighting for Blackwell in 2024F and 70-80% in 2025F. CSP, enterprise, consumer IoT and sovereign are clients for the firm's AI GPUs, and the main sales growth drivers.

➤ In 2025F, as most (70-80%) AI training servers will be upgraded to B100, B200 and GB200 designs, we currently assume 2.1mn GPU units for GB200 designs, around 1.5mn units for B100 and B200 designs, and 1-1.2mn for H100 and H200 designs (Figure 4). This is based on CSPs' GB200 orders, and the allocation will be clearer after 3Q24F.

➤ As for GB200 rack demand among top CSPs and server brands, we expect 40-45k racks to ship in 2025F, according to the supply chain. Microsoft (US) plans to receive 15k GB200 racks in 2025F, AWS (US) 10k racks, Meta (US) 5k racks, Google (US) 6k racks, Supermicro (US) 5k racks, and Dell (US), Oracle (US) and others combined at 1-2k racks.

➤ Based on current GB200 designs, only Microsoft will adopt the NVL72 design (1U servers), while AWS will design 2 racks of NVL36 (2U) as a set, and Meta and Google will use the NVL36 design. The NVL36 is a 2U server design, and will adopted by CSPs on heat dissipation considerations in 2025F.

➤ The GB200 NVL72 system, which includes 36 Grace CPUs and 72 Blackwell GPUs, and the GB200 NVL36, which includes 18 Grace CPUs and 36 GPUs, are both very likely to use liquid cooled designs.

➤ Microsoft's GB200 AI server assemblers will be Hon Hai (2317 TT, NT$173, OP) and FII (CN), and Quanta (2382 TT, NT$284, OP) for AWS, Meta and Google. We therefore expect Hon Hai's rack assembly market share for GB200 servers will be approximately

40%, while Quanta holding around 30%. The remaining 30% will go to Wiwynn (6669 TT, NT$2,740, OP), Wistron (3231 TT, NT$114, OP), ZT (US) and Supermicro.

➤ With expected Nvidia's total AI GPU shipments of 4.0mn units in 2024F and 4.76mn units in 2025F, Nvidia's AI training server GPU shipments will be 526k units in 2024F and 857k units in 2025F.

➤ Besides Nvidia's high-end AI training GPUs, AMD (US) should produce 420k units of GPUs in 2024F and 768k units in 2025F. Intel's (US) Gaudi 2 and Gaudi 3 will see 232k units produced in 2024F and 450k units in 2025F. Thus, global AI GPU shipments should grow to 4.66mn units in 2024F and 5.98mn units in 2025F (Figure 5).

➤ Combined with ASIC designs, including Google's TPU and AWS' Trainium, we expect total training GPU shipments to be 5.52mn units in 2024F, and 7.87mn units in 2025F. Global training AI server shipments will be 715k units in 2024F and 1.25mn units in 2025F. This forecast is revised up from our previous one, to reflect a higher ASIC volume than we had expected (Figure 6).

➤ With the expected AI training server shipments, we anticipate training servers to comprise 6% of total server shipments in 2024F, and up to 10% in 2025F. Strong growth in AI training server demand should be the major driver of global server demand during this period.

Figure 4: Top-4 CSPs will account for 85% of GB200 demand in 2025F

Units	Racks	Weighting (%)	Major ODM
MSFT	15,000	35	Hon Hai
AWS	10,000	24	Quanta
Google	6,000	14	Celestica / Quanta
Meta	5,000	12	Quanta
Others	6,500	15	
Total	42,500	100	

Source: KGI Research estimates

Figure 5: AI server shipments rising in 2024-25F; we revise up 2024-25F demand

AI training GPU shipments (k units)	2024F	2025F
Nvidia (H/B-series)	4,005	4,760
AMD (MI300X/350)	420	768
Intel (Gaudi 2/3)	232	450
Subtotal	4,657	5,978
ASIC (Google TPU/AWS Trainium)	864	1,888
Total training GPU	5,521	7,866
AI training server shipments (k units)	**2024F**	**2025F**
Nvidia (H/B-series)	526	857
AMD (MI300X/350)	53	96
Intel (Gaudi 2/3)	29	56
Subtotal	607	1,009
ASIC (Google TPU/AWS Trainium)	108	236
Total AI training server shipments	**715**	**1,245**

Source: KGI Research estimates

Figure 6: AI training server weighting of total server shipments up from 2% in 2023 to 10% in 2025F

| Server shipments (k units)	2022	2023	2024F	2025F
AI training server	124	249	715	1,245
Others (General server + AI inferencing)	13,703	11,010	11,106	11,759
Total server	13,827	11,259	11,822	13,004
YoY (%)	**2022**	**2023F**	**2024F**	**2025F**
AI training server		101	188	74
Others (General server + AI inferencing)		(20)	1	6
Total server	7	(19)	5	10
Weighting (%)	**2022**	**2023F**	**2024F**	**2025F**
AI training server	1	2	6	10
Others (General server + AI inferencing)	99	98	94	90
Total server	100	100	100	100

Source: Gartner; KGI Research estimates

Nvidia showcased GB200 NVL72 / NVL36 platforms, a DGX design

➢ Nvidia's GB200 NVL72 is a rack-level solution with liquid-cooling design and is capable of more than 1.4 exaflops. Nvidia GB200 NVL72 is structured as 10 compute trays (1U height) atop nine switch trays (2U each) over eight additional compute trays (1U each), a structure referred to as "10+9+8". There will be a TOR Switch on the top of the rack, and two power trays (3U height) with one at the top, and the other at the bottom of the rack (Figure 7-9).

➢ Nvidia has also introduced the DGX SuperPOD, which consists of eight GB200 NVL72 racks with liquid cooling for optimal efficiency (Figure 10).

➢ The GB200 NVL36 is a customized solution and most CSP will adopt it. The GB200 NVL36 is structured as five compute trays (2U height) atop nine switch trays (2U each), over four additional compute trays (2U each), referred to as "5+9+4".

➢ The major difference between the GB200 NVL72 and GB200 NVL36 is that the latter only has 9 compute trays per rack, with each compute tray at a 2U height to allow for customized designs and better thermal efficiency.

➢ Each compute tray is composed of two GB200 Grace Blackwell Superchips (each with 1 Grace GPU and 2 Blackwell GPU per Superchip), an NVLink and four Smart NIC cards. That is, there will be 36 CPUs and 72 Blackwell GPUs connected together via NVLinks in a GB200 NVL72 rack, and 18 CPUs and 36 GPUs in a GB200 NVL36 rack (Figure 7).

➢ Each Switch tray will include two NVLink Switch chips and provide 14.4TB/s of aggregate bandwidth (Figure 8).

➢ Per our calculations, we estimate the ASP of a GB200 NVL72 to be US$3-3.5mn, and a GB200 NVL36 at US$1.8-2mn per rack, with the GPUs and CPUs accounting for the bulk of BOM cost, at between 75-80% (Figure 11).

➢ In addition, several key components will enjoy a higher content cost in GB200 NVL72/36 servers compared to H100 racks (4xH100 servers per rack), including thermal solutions, power supplies, and rack assembly. We will discuss spec upgrades and content value differences of key components in the following section.

Figure 7: GB200 compute trays are composed of GB200 Superchips, smart NICs and DPU
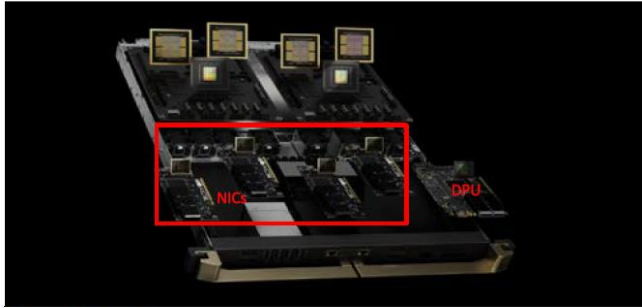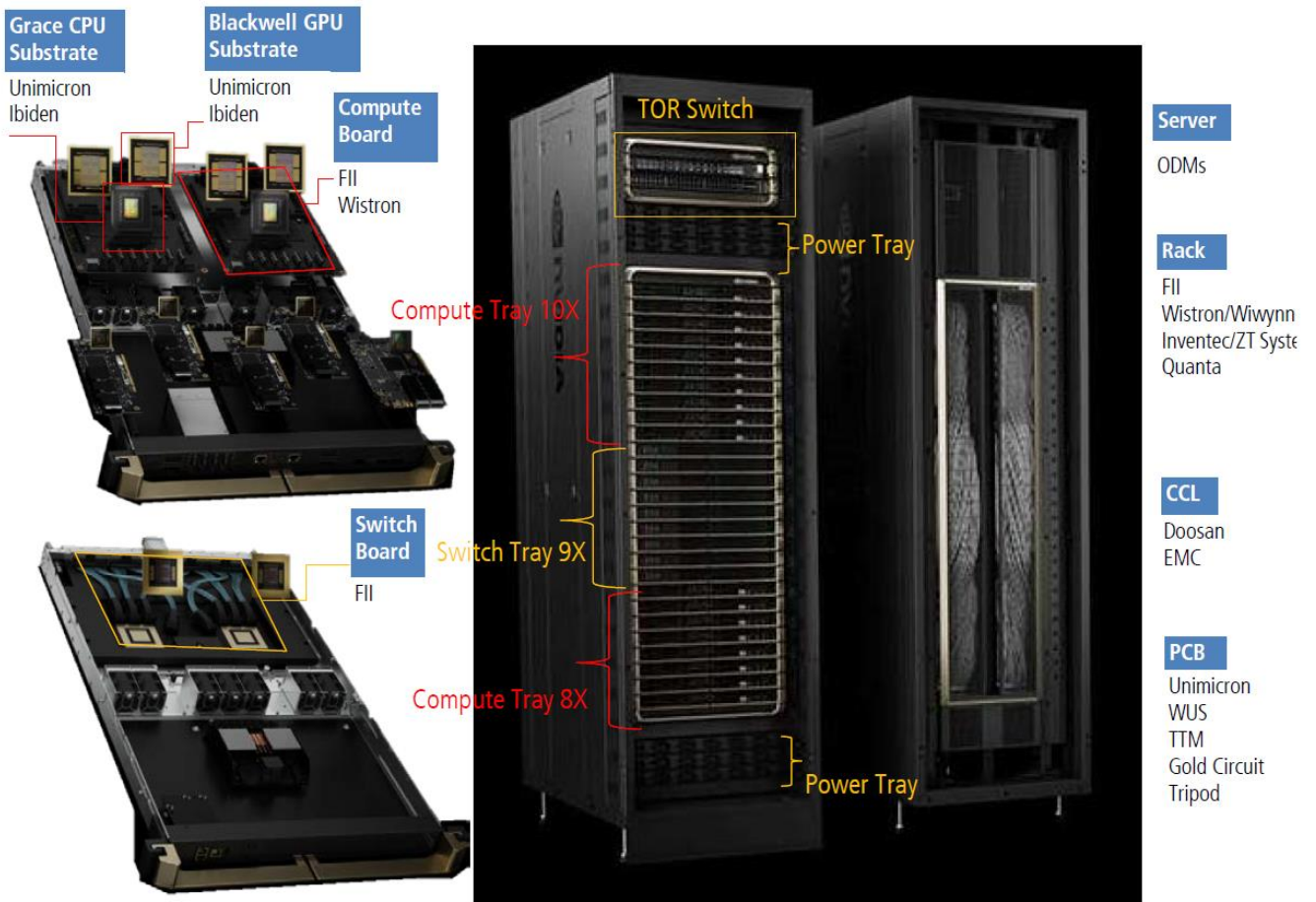

Source: Nvidia; KGI Research

Figure 8: Each GB200 switch tray includes two Nvlink chips

Source: Nvidia; KGI Research

Figure 9: GB200 NVL72 breakdown & supply chain

Source: Nvidia; KGI Research

Figure 10: GB200 Superpod – Contains 8 NVL72 racks

Source: Nvidia; KGI Research

Figure 11: BOM cost analysis - NVL 72 & NVL 36 racks

|	NVL36	NVL72
Rack shipment (units)	26,775	15,725
Weighting of total GB series (%)	63	37
Assumption		
Compute tray	9	18
GPU per tray	4	4
CPU per tray	2	2
Smart NIC per tray	4	4
Swtich tray	9	9
Total shipments (k units)		
GPU	964	1,132
CPU	482	566
ASP		
GPU	35,000	35,000
CPU	2,500	2,500
Smart NIC	2,000	2,000
Content value per rack (US$mn)	**1.8-2**	**3-3.5**
Content value mix per rack (%)		
GPU	67.2	74.9
Grace CPU	2.4	2.7
NIC	3.8	4.3
Compute tray	73.5	81.9
Switch tray	16.0	8.9
Thermal	3.6	2.9
Power, chassis and others	2.1	1.5
Assembly	4.8	4.8
Total	**100.0**	**100.0**

Source: KGI Research estimates

Taiwan ODMs to play important role within AI server supply chain

➤ Taiwan ODMs have dominated the global server supply chain, with assembly market share over 90%. We expect Taiwan ODMs to benefit and see server sales growth in 2024-25 due to the AI server trend.

➤ AI server H100/H200 designs are equipped with eight GPU per server. Due to primarily using air cooling, AI server height is 7U. Thus a rack should contain around four servers on average. This is the mainstream design for AI servers in 2024.

➤ Nvidia has launched its next-generation Blackwell GPU for AI servers, and introduced the GB200 NVL36/ 72 AI GPUs with DGX architecture. This AI GPU is a 1U/2U design, with the lower height enabling it to have 9 or 18 compute trays and 9 switch trays per server rack. We expect the GB200 to be the mainstream AI GPU in 2025F.

➤ We expect shipment of 4.0mn AI GPUs and 526k AI servers by Nvidia in 2024F (Figure 3). Around 85% of AI server shipments will be H100/H200, 10% will be from the Blackwell series, and 5% will be from the Ampere series. Among these AI server shipments, Microsoft (US) will account for the largest share at around 17%, followed by Amazon Web Services (AWS), Google, Meta, and China CSPs. These top CSPs have a combined share of 50-60% (Figure 12).

➤ Based on CoWoS capacity expansions to meet Nvidia's AI GPU growth, we estimate AI server shipments equipped with Nvidia's GPUs will grow to around 857k units in 2025F (Figure 3), with 25% being H100 & H200 and 75% utilizing B100, B200, & GB200 GPUs. Microsoft will continue to be the biggest consumer of Blackwell-series GPU, in particular the GB200. As the GB200 is used in the NVL72 (rack with 72 GPU and 36 CPU in 18 compute trays and nine switch trays), we estimate 15k NVL72 racks will comprise 35% of a total of 42.5k GB200 racks in 2025. AWS, Google, and Meta, have ordered a respective 10k, 6k, and 5k racks with the GB200, but they are primarily using the NVL36 design (rack with 36 GPU and 18 CPU in nine compute trays and nine switch trays).

➤ The top four US CSPs will likely be the major GB200 buyers in 2025 (85% of GB200 racks), with Supermicro, Oracle, and Dell, combined comprising the remaining 15% of GB200 demand.

➤ Based on GB200 AI server rack shipments, Hon Hai, and Quanta will be the major ODMs for rack assemblers. Hon Hai is the main supplier for Microsoft, while Quanta Computer supplies AWS, Meta, and Google. We expect Hon Hai to account for 40% of GB200 racks (for Microsoft, Oracle, and Dell), while Quanta will account for around 30%.

➤ Hon Hai will have higher value in the GB200 supply chain than its peers, with manufacturing of compute board (around 50%), switch board (90%), DPU board, and server rack integration for GB200 DGX/MGX systems (40%).

➤ Wistron will handle compute boards and some peripheral boards for the GB200, with board content value per rack in GB200 above that of H100.

➤ Quanta Computer mainly performs rack assembly for CSPs using the GB200, with rack ASP of around US$2.0-3.0mn, which is much higher than that of the H100 rack (assuming 4 servers in a rack on average, at a cost of US$0.8-1.2mn).

➤ Wiwynn will penetrate the market for GB200 products in 2025F for MGX systems, and will penetrate some sovereign projects using the B100/B200, while it currently supplies ASIC models for AI servers in 2023-1H24F.

➤ Inventec (2356 TT, NT$54, OP) will supply compute boards to assembler ZT Systems, a strategic partner with Inventec for AI servers. We expect most of its GB200 products to be for MGX systems.

➢ Gigabyte Technology (2376 TT, NT$317, OP) is undertaking product design-in and sending samples to clients for B100, B200, and GB200 designs. GB200 projects are rack-level designs with more complicated product development, including system and thermal solution integration. Most CSPs have orders placed for 2025. We think it will take a long time for channel clients to develop and adopt B100 and GB200 designs.

➢ As GB200 AI server racks are expensive (US$2.0-3.0mn for NVL36/ 72), we only expect top CSPs or large enterprises to be able to afford them. Meanwhile, ODMs will require significant capital to run this business. Therefore, ODMs with better financial structure and strong balance sheets will be more capable to win orders.

➢ We have recently seen several ODMs and supply chain members conduct fundraising (CB and ECB) to meet working capital demand for 2025, including Hon Hai, Gigabyte, and Auras (3324 TT, NT$834, OP).

Figure 12: Top-4 CSPs will account for 58% of total AI server demand in 2024F
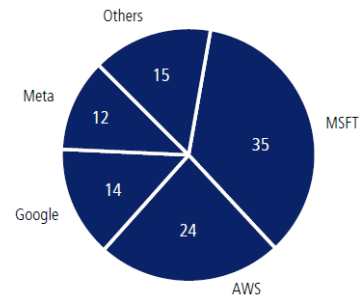
Market share, percent
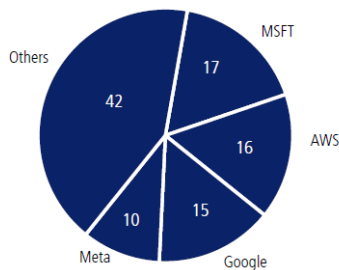
Source: KGI Research estimates

Figure 13: Top-4 CSPs will account for 85% of total GB200 demand in 2025F

Market share, percent

Source: KGI Research estimates

Figure 14: AI server matrix between major CSP and enterprise & ODMs in 2024F; all Taiwan's ODMs are the key assemblers (H100 GPU)

| | AI server client | GPU solution | 2024F AI server shipments allocation (%) - based on L10 | | | | | | | |
			Quanta	Hon Hai	ZT / Inventec	Wiwynn	Wistron	Gigabyte	Supermicro	Dell / Lenovo / Inspur / Others
CSP	Microsoft	Nvidia / AMD	35	30	30	5				
	Google	Nvidia / TPU	50	15	35					
	AWS	Nvidia / Trainium	30	15	35	20				
	Meta	Nvidia / AMD	70	12		18				
	BBAT	Nvidia / Habana		5	10					85
	Oracle	Nvidia / AMD		90						10
Enterprise	Major enterprise	Nvidia		30	10		30		20	10
	Nvidia DGX	Nvidia					100			
	Tesla	Nvidia / Dojo					40		60	
Channels	Coreweave	Nvidia						10	75	15
	Other channels	Nvidia						35	50	15
Total server sales in 2024F (NT$bn)			728	1,222	260	345	448	170	670	
Server sales weighting (%)			50	19	45	100	45	60	100	
AI server sales weighting (%)										
- of server sales			50	40	18	35	41	80		
- of total sales			25	7	8	35	18	48		

Source: KGI Research estimates

Figure 15: AI server sales will account for over 50% of total server sales for most ODMs in 2025F

Products	Ticker	Company	Server sales (NT$bn)			Weighting of total sales (%)			YoY (%)		
			2023	2024F	2025F	2023	2024F	2025F	2023	2024F	2025F
ODM	2317 TT	Hon Hai	1,018	1,222	1,589	17	19	20	(35)	20	30
	2382 TT	Quanta	384	728	1,311	35	50	63	11	89	80
	3231 TT	Wistron	307	448	595	35	45	51	(13)	46	33
	2356 TT	Inventec	208	260	315	40	45	47	(7)	25	21
	6669 TT	Wiwynn	242	345	462	100	100	100	(17)	43	34
Brand	2357 TT	Asustek	16	45	70	3	8	11	60	181	56
	2376 TT	Gigabyte	53	170	187	39	60	60	162	222	10

	Ticker	Company	AI server sales (NT$bn)			Weighting of server sales (%)			YoY (%)		
			2023	2024F	2025F	2023	2024F	2025F	2023	2024F	2025F
ODM	2317 TT	Hon Hai	305.5	488.9	873.8	30	40	55		60	79
	2382 TT	Quanta	76.9	364.2	917.7	20	50	70		374	152
	3231 TT	Wistron	71.9	181.8	327.2	23	41	55		153	80
	2356 TT	Inventec	14.2	46.9	94.5	7	18	30		231	102
	6669 TT	Wiwynn	48.4	120.8	277.3	20	35	60		150	129
Brand	2357 TT	Asustek	5.6	36.0	56.0	35	80	80		543	56
	2376 TT	Gigabyte	32.1	135.8	149.3	61	80	80		323	10

Source: company data; KGI Research estimates

Figure 16: Supply chain flow chart of GB200
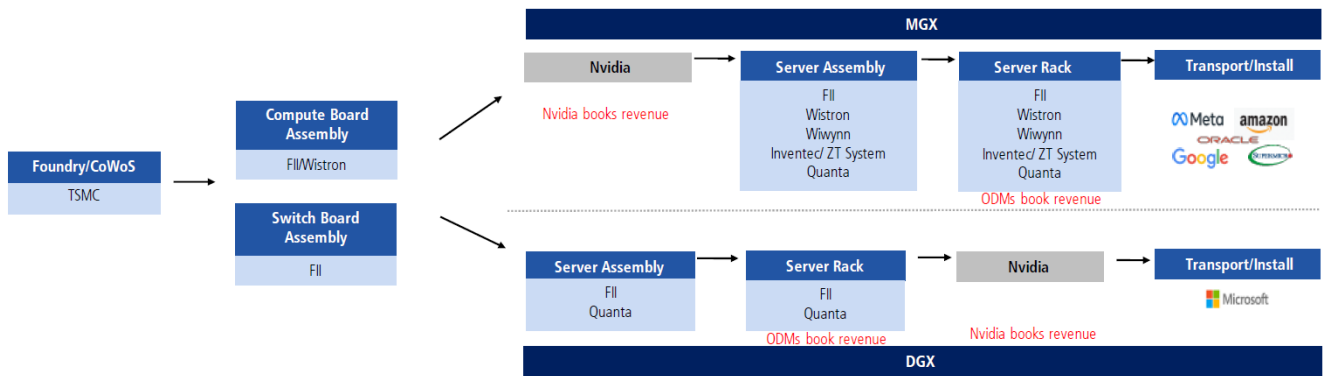
Source: KGI Research

Thermal – liquid cooling becomes more popular in low-U server designs

➢ As the DGX GB200 superchip system design is more compact, it is well suited to liquid cooling thermal solutions. Liquid cooling enhances performance while minimizing thermal constraints across a system. Such designs allow for more sustainable and energy-efficient operations, even under heavy computational loads.

➢ We expect the penetration rate of liquid cooling will grow significantly in 2025F, with benefits for thermal module companies offering cold plate modules, coolant distribution units (CDU), manifolds, and fan doors, such as AVC (3017 TT, NT$708, OP), Auras, and Delta (2308 TT, NT$329.5, OP).

➢ Given this liquid cooling trend, we note that several companies are expanding their cold plate and CDU capacity, including AVC and Auras for cold plates, Nidec CCI (6230 TT, NT$362.5, NR), Kenmec (6125 TT, NT$114, NR), Hon Hai, Vertiv (US) and nVent (US) for CDU, and Kroai (8996 TT, NT$478, OP) for manifolds. All of these makers would like to capture liquid cooling business opportunities, as total liquid cooling rack content value is high, at US$40-80k per rack depending on design, including US$200-400 for each cold plate modules, US$20-30k for each CDU, US$10-15k for manifolds with quick disconnectors (QD), US$2-4k for fan doors and RDHx, and US$5k for each rack itself.

➢ With higher-than-average gross margin on liquid cooling products, rising penetration of liquid cooling solutions will fuel gross margin and EPS growth for thermal companies in 2024-25F. Liquid cooling sales weightings for thermal companies will be low in 2024F, but will surge in 2025-26F.

➢ Most of the liquid cooling products for GB200 GPUs will have two suppliers, and AVC is one of two suppliers for most projects for Microsoft, AWS and Meta. Cooler Master (TW) and AVC are the major suppliers of cold plate modules and fan doors for hyperscale CSPs, while Auras also offers cold plate modules to some brands, like Supermicro.

➢ AVC plans to gain inner manifold and chassis market share. Coolant distribution units (CDU, liquid to air / liquid to liquid) or side car (liquid to air) modules will be supplied to Microsoft by Vertiv initially.

➢ The cold plate module ASP for one GB200 compute tray (1U design) is US$1,500-2,500, based on 2 CPUs and 4 GPUs. Total cold plate module ASP of the NVL36 racks, a 2U design, is estimated to be US$20-25k, based on 36 GPUs, 18 CPUs, and 9 switches with 2 chips each, and around US$35-45k for NVL72 servers, a 1U compute tray design.

Figure 17: Several thermal plays plan to expand liquid cooling production capacity

Company	Liquid cooling CDU capacity expansion plans
Vertiv	Plans to expand CDU capacity by 45x in 12 months to end-4Q24F
nVent	Plans to double liquid cooling capacity by mid-2024F
Auras	Monthly capacity of 1-2k units currently in Thailand, and will expand to 2-3k units depending on clients' demand
AVC	Monthly capacity of 2k units currently in Vietnam, and will keep expanding
Nidec	Expects to increase monthly CDU production capacity 10x to 2k units by mid-2024 in Thailand

Source: company data; KGI Research

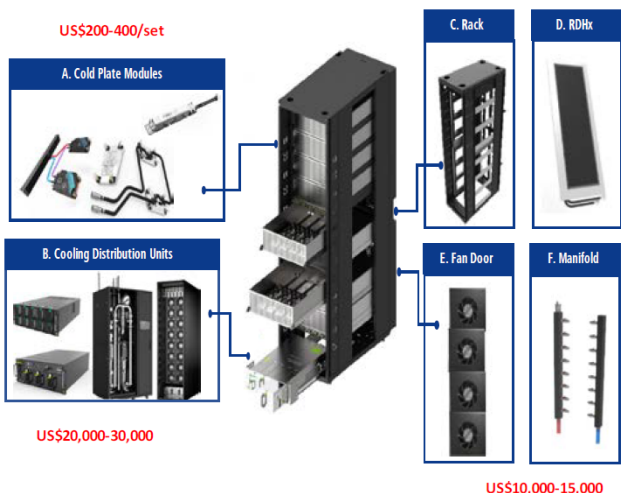**Figure 18: Capacity plans by vendor – CDU**

Capacity	CDU				
(k/month)	Current			Expansion	
	Vietnam	Thailand	Overseas		Schedule
AVC	2		To be announced	Vietnam	Under planning
Auras		1-2	2-3	Thailand	By 3Q24 (dependent on end-client demand)
Nidec		0.2	2	Thailand	By June, 2024F
Kenmec			0.8	Taiwan	
Hon Hai			Begin CDU manufacturing		
Vertiv			Expanding by 45x in 3 plants		Within 12 months to end-2024F

Source: company data; KGI Research

Figure 19: Capacity plans by vendor – Cold plates

Capacity (k/month)	Cold plate					
	Current		Expansion			
	China	Vietnam	China	Vietnam	Thailand	Schedule
AVC	50	65	100	265		Dependent upon end-client demand
Auras	30		30		300	By 3Q24 (dependent on end-client demand)

Source: company data; KGI Research

Figure 20: Thermal solution transition to liquid cooling to create much higher content value for thermal plays

Parts of liquid cooling solution	Major supplier	Cost (US$)
Cold plate module	Cooler Master	1,500-2,500
	CoolIT Systems	(per server)
	AVC	
	Auras	
CDU	CoolIT Systems	20,000-30,000
	Vertiv	(per rack; in-rack)
	nVent	
	AVC	
	Auras	
	Delta	
	Kaori	
	Hon Hai	
	MGCooling	
Manifold & quick connector	CoolIT Systems	10,000-15,000
	AVC	(quick connector 4,000-6,000)
	Colder Product Co. (CPC)	
	CEJN	
	Parker	
	Staubli	
	Kaori	
RDHx & fan door		2,000-4,000
Rack		5000
Total		**40,000-80,000**

Source: Auras; KGI Research

Power Supplies – spec upgrades for Blackwell GPU

➢ With power consumption per GB200 chip reaching 2.7kW, a NVL72 rack could require power of 120kW, or even higher, or 60kW+ for a NVL36 rack.

➢ AC-DC power supply units (PSU) will likely see spec upgrades from 3-3.3kW per unit for H-series servers to 5-5.5kW for B-series servers, with power efficiency reaching 97.5%+.

➢ We believe AC-DC PSU dollar content could be at the higher end of US$0.1-0.2/W, with further upside for liquid-cooled designs.

➢ The adoption of 48V power architecture in GB200s will fuel rising demand for DC-DC power delivery modules.

➢ Per our understanding, Delta Electronics controls 50%+ market share in AC-DC server PSU, and will likely represent an even higher share in Blackwell platform GPU servers given its industry-leading technology. It is also one of the two leading DC-DC power delivery module suppliers worldwide, directly competing with Monolithic Power Systems (US).

➢ Lite-On (2301 TT, NT$112, R) will serve as another key AC-DC server PSU supplier, and will likely gain market share by leveraging its relationship with major US CSPs.

Figure 21: AC-DC power shelf design

Source: Lite-On Technology; KGI Research

Figure 22: The adoption of 48V power architecture will fuel demand for DC-DC power delivery modules
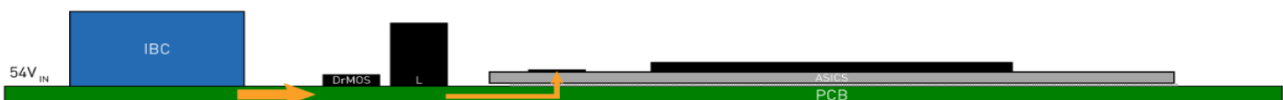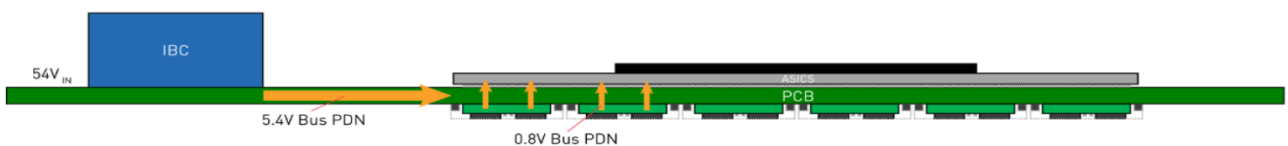
Source: Monolithic Power Systems; KGI Research

Chassis & rail kit with complicated designs in AI server racks.

➢ There will be 18 1U chassis for compute trays and 9 2U chassis for switch trays in a GB200 NVL72 rack, and 18 2U chassis, including compute trays and switch trays, in a GB200 NVL36 rack.

➢ Per our calculation, the ASP of 2U GB200 chassis will be between NT$500-600 (Level 3), 50-100% higher than for general server chassis at the same height. Overall chassis content value may be US$10,000-11,000 for a NVL72/ 36 rack.

➢ Chenbro (8210 TT, NT$300, OP) is the first reference design partner for Nvidia's MGX 2U, including GB200, chassis, while 1U chassis for compute trays will be supplied by a Singaporean company.

➢ We believe Chenbro will have first-mover advantages. The firm will start to ship GB200 chassis in 2H24F, earlier than industry peers, and will have the opportunity to penetrate more clients. In addition, GB200 2U chassis will carry higher gross margin for Chenbro, compared to H100 projects, as most shipments are Level 3 (only chassis), versus Level 5 (including chassis and components such as thermal and power solutions) for H100 projects.

➢ For server rail kits, a GB200 NVL72 rack will have 27 sets of server rail kit, including 18 1U server rail kits for compute trays and nine 2U rail kits for switch trays. Each NVL36 rack will have 18 sets of server rail kits, all 2U.

➢ Based on our supply chain check, the ASP of each 2U and 1U server rail kit for GB200 racks will be between US$50-100. Server rail kit content value per rack will be US$1,500-2,000 for NVL72 / 36 racks, versus US$1,200-1,600 per rack for H100 designs.

➢ Suppliers of rail kit for switch trays is designated by Nvidia, while CSPs have more flexibility in choosing rail kit suppliers for compute trays. We believe King Slide (2059 TT, NT$1,285, OP) will enjoy dominant market share in GB200, H100, and B100 series rail kit. The firm holds more than 50% of market share in the AI server rail kit market, versus around 30% market share in the general server rail kit market. It will benefit from rising AI server demand, which will drive higher ASP and margins on AI server rail kits.

➢ Nan Juen (6584 TT, NT$211.5, NR) is also sampling GB200 2U rail kit currently, and may supply Supermicro and other CSPs after certification. The firm is also testing and certifying AI server rail kits (4U/7U) for CSP clients, which contribute to sales in 2H24F.

➢ Fositek's (6805 TT, NT$727, OP) AI server rail kits (2U, 4U & 7U) are under certification and may begin to ship in 2H24F, at the earliest, if certified in late-2Q24F or 3Q24F.

Figure 23: Chenbro's MGX 2U chassis


Source: Chenbro; KGI Research

Figure 24: Server rail kit in an NVL72 rack



Source: Hon Hai; KGI Research

Networking & PCB

➢ Given growing demand for AI and HPC applications, which has prompted the prevalence of multi-node computing units, and given the fact that large-scale computing can only be performed by systems consisting of multiple GPUs, we believe efficient communication between each GPU is essential.

➢ In the case of the GB200 series, each GPU can support up to 18x Gen 5 NVLink, and every NVLink series spec will be upgraded from 112GB to 224GB, with total bandwidth upgraded from 900GB/s to 1.8TB/s, which is twice of the bandwidth of a Gen 4 NVLink.

➢ In terms of the number of NVLinks, compared to the GH200 system which adopts a fat-tree architecture, the GB200 system uses a single-layer architecture to connect all GPUs, which will increase the number of connections.

➢ The PCB design of a GB200 NVL72 system consists of a computing tray PCB and switch tray PCB, both adopting the 5-to-6-step 22-24L OAM board, instead of 22-24L UBB board. For reference, a 5-step OAM board is commonly used to carry a GPU, but the layer count is lower, at 18.

➢ CCLs are produced with mixed material lamination. The central layers are made with low-end M2-M5 materials, while the outer layers are made with high-end M7-M8 materials.

➢ Comparing to H100 and B100 series racks, total PCB consumption (by area) in GB200 racks is 10-13% greater, but the content value is roughly the same.
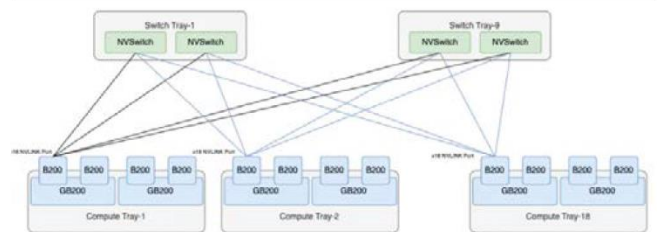
Figure 25: NVLink performance
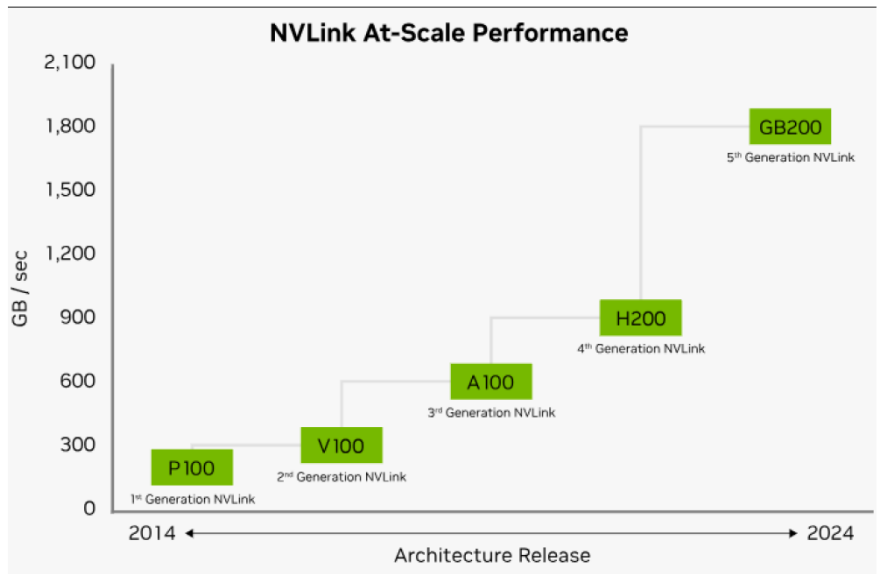
Source:Nvidia; KGI Research

Figure 26: : GB200 is adopting a fat-tree architecture
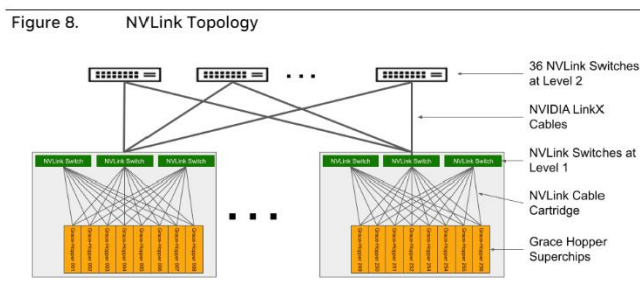
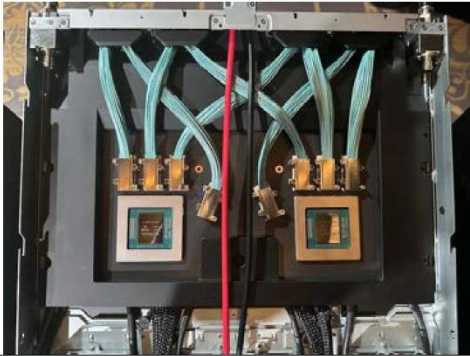

Source: Nvidia; KGI Research

Figure 27: The GB200 uses a single-layer architecture
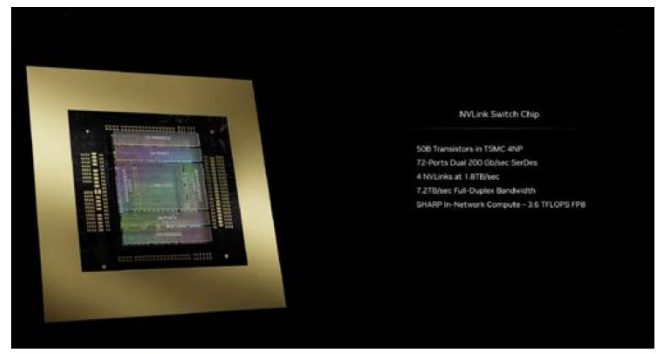

Source: naddod; KGI Research

Figure 28: Each NVSwitch tray consists of two NVSwitch chips


Source: NVidia; KGI Research

**Figure 29: Specifications of an NVLink Switch chip**

Source: Nvidia; KGI Research

**AI server continues to boost overall server demand growth in 2024-25F**

➢ Global server shipments fell by 19% YoY in 2023 to 11.3mn units in 2023. This weakness was related to budget cannibalization form CSPs, as high AI infrastructure cost took a big portion of their budget that was not put AI into consideration in early of 2023.

➢ AI server was just in a initial stage in 2023 due to CoWoS capacity constraint. After expansion from TSMC and other interposer, packaging companies in the past few quarters, we expect AI server supply will become more in 2024F to meet market demand. In addition, top US hyperscale CSPs all revise up their capex plan in 2024F for rising spendings of AI server. This implies they all have strong intention to expand data centers for AI infrastructure setup.

➢ Based on our research on the supply side, we expect global training AI server shipments in 2024F to grow to 715k units, including 526k units of Nvidia's GPUs, 53k units of AMD's, 29k of Intel's and 108k units of ASIC (Figure 5).

➢ In 2025F, we expect AI server in Nvidia's GPU design would see different form factors and platform in H100/B100/B200 (x86 CPU) and in GB200 (ARM CPU), while we assume AMD and ASIC should be still in a 8 GPU or accelerators in a server. Therefore, we forecast global training AI server shipments of 1.25mn units, including 857k units in Nvidia's GPUs, 96k units of AMD's, 56k of Intel's and 236k units of ASIC (Figure 5).

➢ Under solid sales growth for training AI servers in 2024-25F, we expect inference AI servers will be also see a rising trend. However, general server demand has grown relatively moderately in the last two years, because AI servers are CSPs' and enterprises' first priority for expansion. We thus expect total global server shipments to grow 5% YoY in 2024F to 11.8mn units, with non-AI server sales growth flat, while global server shipments will grow 10% in 2025F to 13mn units with non-training AI server growth of 6% YoY.

➢ We expect training AI server will comprise 6% of total server demand in 2024F, and rise to 10% in 2025F (Figure 33). Inference server demand should grow more robustly after 2025F. In terms of server revenue conirtbuion, AI servers will comprise 60-70% of the total server market in 2024-25F, on much higher ASP.

Figure 30: Global server shipments to grow 5% YoY in 2024F and 10% YoY in 2025F

Global server shipments, k units (LHS); YoY, percent (RHS)
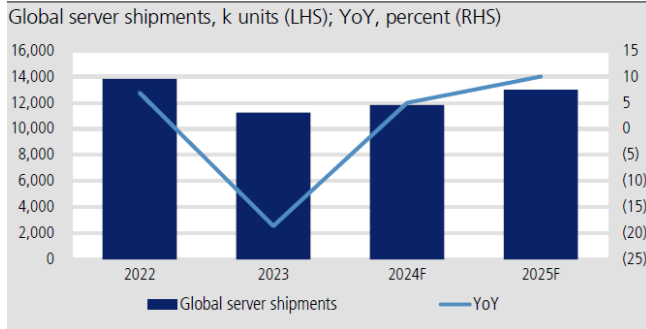
Source: Gartner; KGI Research estimates

Figure 31: AI servers sales will outgrow general server sales in 2023-25F

YoY growth of server shipments, percent
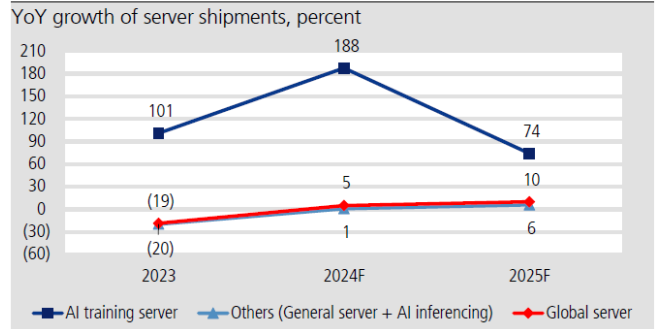
Source: Gartner; KGI Research estimates

Figure 32: Training server shipments to grow from 715k units in 2024F to 1245k in 2025F

Global training server shipments, k units (LHS); YoY, percent (RHS)
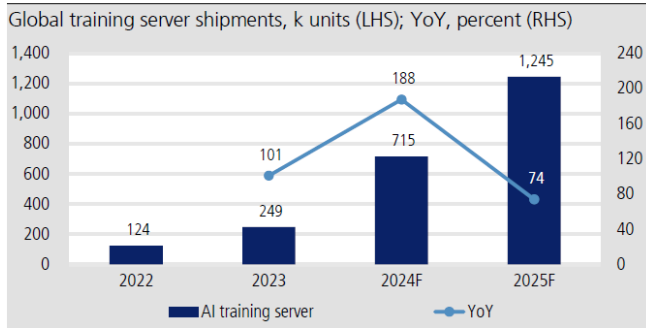
Source: Gartner; KGI Research estimates

Figure 33: AI servers will account for 63% of the total server market value in 2024F and 70% in 2025F

Shipments and revenue weighting of AI server, percent
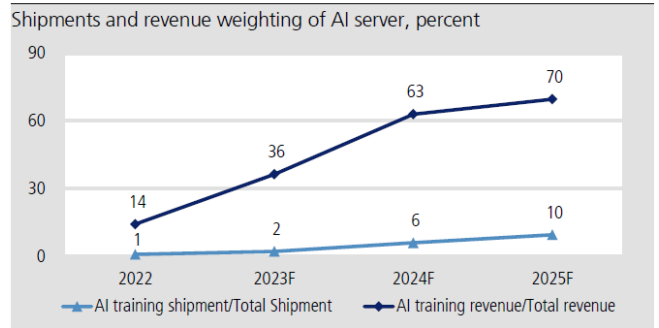
Source: Gartner; KGI Research estimates

Figure 34: AI server and new CPU platform launch will drive global server shipments growth in 2024-25F

Global server shipment, mn units (LHS); YoY growth, percent (RHS)
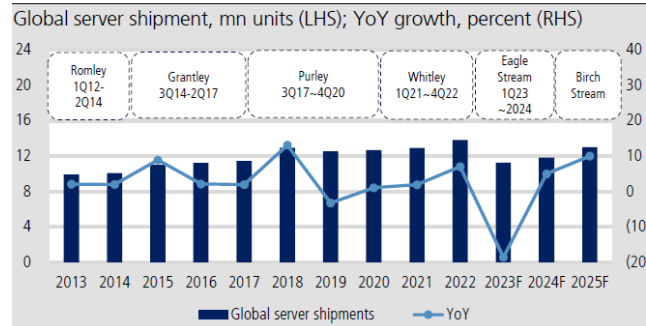
Source: Gartner; KGI Research estimates

Figure 35: Server ASP in uptrend on computing performance upgrades & AI server demand

Global server shipment, mn units (LHS); server ASP, US$ (RHS)
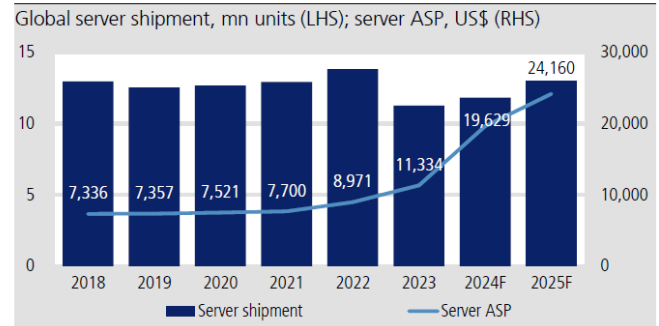
Source: Gartner, KGI Research estimates

Figure 36: Intel to see Eagle Stream penetration increase in 1H24 and new Birch Stream will be launched in 2H24

Platform	Intel Whitley	Intel Eagle Stream	Intel Eagle Stream	Intel Birch Stream	Intel Birch Stream	Intel Birch Stream	Intel Oak Stream
Time of launch	2Q21	1Q23	4Q23	2H24F	2H24F	1H25F	2H25F
CPU	Ice Lake	Sapphire Rapids (Intel 7)	Emerald Rapids (Intel 7)	Sierra Forest (Intel 3, E-core)	Granite Rapids (Intel 3, P-core)	Clearwater Forest (Intel 18A, E-core)	Diamond Rapids (Intel 20A, P-core)
Process	10nm	10nm	10nm++	3nm (equivalent to TSMC's 5nm node)	3nm (equivalent to TSMC's 5nm node)	1.8nm (equivalent to TSMC's 2nm node)	2nm (equivalent to TSMC's 3nm node)
CPU sockets	LGA 4189	LGA 4677	LGA 4677	LGA 7529	LGA 7529	LGA 7529	TBA
CPU cores	26	60	64	288	136	288	TBA
DRAM	8-channel DDR4	8-channel DDR5	DDR5	DDR5	DDR5	TBA	TBA
PCIe	PCIe 4.0	PCIe 5.0	PCIe 5.0	PCIe 5.0	PCIe 5.0	TBA	TBA
CPU TDP	up to 270W	up to 350W	350-400W	up to 500W	up to 500W	TBA	TBA

Source: Company data, KGI Research

Figure 37: AMD to see Genoa and Bergamo CPU penetration increase in 1H24 and will have new Turin CPU introduction in 2H24

Platform	AMD Zen 3	AMD Zen 4	AMD Zen 4c	AMD Zen 4	AMD Zen 5	AMD Zen 6
Time of launch	1Q21	4Q22	1H23	3Q23	2H24F	2H25F
CPU	Milan	Genoa	Bergamo	Siena	Turin	Venice
Process	7nm+	5nm	5nm	5nm	3nm	3nm
CPU sockets	FC LGA 4094	FC LGA 6096	FC LGA 6096	FC LGA 4844	FC LGA 6096	TBA
CPU cores	64	96	128	64	256	TBA
DRAM	8-channel DDR4	12-channel DDR5	DDR5	DDR5	TBA	TBA
PCIe	PCIe 4.0	PCIe 5.0	PCIe 5.0	PCIe 5.0	TBA	TBA
CPU TDP	225-280W	320-400W	320-400W	70-225W	480-600W	TBA

Source: Company data, KGI Research

CSP capex becoming more aggressive in 2024F

➤ The top four US CSPs have intensified their capex to build technical infrastructure and AI offerings, evidenced by rising 2024F capex guidance during recent CY1Q24 earnings calls.

➤ Meta has increased 2024F capex guidance from US$30-37bn to US$35-40bn, representing a 12% YoY increase at the midpoint, due to greater investment in AI.

➤ Google's capex guidance for 2Q24-3Q24F suggests full-year capex growth of over 50% YoY in 2024F.

➤ Microsoft guides capex to grow YoY in FY2025F (year-end June), driven by growing AI and cloud demand.

➤ Amazon guides a significant YoY capex increase in 2024F from US$48.1bn in 2023, fueled by infrastructure investment to support AWS and high demand for generative AI.

➤ Consensus now forecasts the top four US CSPs' capex to increase by 38% YoY in 2024F, up from the 26% forecast before the 1Q24 earnings season, and expects a further increase of 9% in 2025F.

➤ Based on estimated GB200 server shipments and costs (US$3-3.5mn for NVL72; US$1.8-2mn for NVL36), combined with 2025F GB200 rack demand from CSPs, we think Microsoft may see a further upward revision possibility for its 2025F capex from the current consensus estimation of US$57,358mn (or up 14% YoY). As for other CSPs, expected spending on GB200 racks should be less than 50% of their capex in 2025F, quite a reasonable level.

Figure 38: Capex of top-four US CSPs fell 2% YoY in 2023, but will grow 38% YoY in 2024F

Capex (US$mn)	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	1Q24	2019	2020	2021	2022	2023	2024F	2025F
Meta (Facebook)	5,441	7,572	9,375	9,043	6,842	6,216	6,543	7,665	6,400	15,102	15,115	18,567	31,431	27,266	37,383	41,418
Amazon	14,951	15,724	16,378	11,268	14,207	11,455	12,479	14,588	14,925	16,861	35,044	55,396	58,321	48,133	62,228	67,213
Microsoft	5,340	6,871	6,283	6,274	6,607	8,943	9,917	9,735	10,952	13,546	17,592	23,216	24,768	35,202	50,363	57,358
Google	9,786	6,828	7,276	7,595	6,289	6,888	8,055	11,019	12,012	23,548	22,281	24,640	31,485	32,251	46,817	48,241
US hyperscale subtotal	35,518	36,995	39,312	34,180	33,945	33,502	36,994	43,007	44,289	69,057	90,032	121,819	146,005	142,852	196,791	214,229
YoY (%)	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	1Q24	2019	2020	2021	2022	2023	2024F	2025F
Meta (Facebook)	27.4	64.2	117.4	68.4	25.7	(17.9)	(30.2)	(15.2)	(6.5)	8.0	0.1	22.8	69.3	(13.3)	37.1	10.8
Amazon	23.7	10.1	4.0	(40.5)	(5.0)	(27.1)	(23.8)	29.5	5.1	25.6	107.8	58.1	5.3	(17.5)	29.3	8.0
Microsoft	4.9	6.5	8.1	7.0	23.7	30.2	57.8	55.2	65.8	(4.8)	29.9	32.0	6.7	42.1	43.1	13.9
Google	64.7	24.2	6.7	19.0	(35.7)	0.9	10.7	45.1	91.0	(6.3)	(5.4)	10.6	27.8	2.4	45.2	3.0
US Hyperscale subtotal	29.7	19.9	20.3	(6.5)	(4.4)	(9.4)	(5.9)	25.8	30.5	3.4	30.4	35.3	19.9	(2.2)	37.8	8.9
QoQ (%)	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	1Q24	2019	2020	2021	2022	2023	2024F	2025F
Meta (Facebook)	1.3	39.2	23.8	(3.5)	(24.3)	(9.1)	5.3	17.1	(16.5)							
Amazon	(21.0)	5.2	4.2	(31.2)	26.1	(19.4)	8.9	16.9	2.3							
Microsoft	(9.0)	28.7	(8.6)	(0.1)	5.3	35.4	10.9	(1.8)	12.5							
Google	53.3	(30.2)	6.6	4.4	(17.2)	9.5	16.9	36.8	9.0							
US Hyperscale subtotal	(2.8)	4.2	6.3	(13.1)	(0.7)	(1.3)	10.4	16.3	3.0							

Source: Company data; Bloomberg; KGI Research

Figure 39: Top four US CSP capex outlook – Positive growth, with better-than-expected guidance

Company	Time	Actual & Guidance
Microsoft	1Q24	• Capex grew 80% YoY and 22% QoQ to $14bn to support cloud demand, inclusive of the need to scale AI infrastructure
	2Q24F	• Expect capex to grow materially QoQ, driven by cloud and AI infrastructure investments and seasonality • BBG consensus: US$13.1bn (+ 20% QoQ)
	2024F	• Guide capex to grow YoY in FY25F (year end June), to meet the growing demand signal for cloud and AI products. • BBG consensus: US$50.32bn (+43% YoY)
Google	1Q24	• 1Q24 capex rose 91% YoY to US$12bn, beating consensus by 17%, driven overwhelmingly by investment in technical infrastructure with the largest component for servers followed by data centers, especially for AI
	2024F	• Guide capex in 2Q24-4Q24 to be roughly at or above the Q1 level. That's it, 2024F capex will grow over 50% YoY to above US$48bn, beating consensus forecast by 6% (US$45bn, up 40% YoY) • Capex will be largely used for technical infrastructure in 2024, while investment in offices will be flat YoY, accounting for less than 10% of total capex in 2024
Meta	1Q24	• 1Q24 capex was US$6.4bn, down 7% YoY and 17% QoQ, below consensus by 9%, driven by investments in servers, data centers, and network infrastructure.
	2024F	• Revise up 2024 capex guidance by 12%, up from US$30-37bn to US$35-40 (up 33% YoY at midpoint), beating consensus forecast by 9% as Meta continues to accelerate infra investments to support AI roadmap
	2025F	• Expect capex to grow YoY in 2025F as the company invests aggressively to support AI research and product development efforts • BBG consensus: $40.9bn (+11% YoY)
amazon	1Q24	• Capex was US$14.9bn, up 5% YoY and 2% QoQ, in line with consensus
	2024F	• Expect capex to meaningfully increase YoY in 2024, driven by infrastructure to support AWS's reaccelerating growth including high demand for gen AI • Expect 1Q24 capex to be the lowest quarter of 2024F, implying full-year capex will increase YoY to over $59.6bn (over 23% YoY), ahead of consensus • Consensus forecast 2024F capex of $62.05bn (+29% YoY)

Source: Company data, Bloomberg, KGI Research

Figure 40: Top four US CSP cloud business outlook

Company	Time	Actual & Guidance
Microsoft	1Q24	• Intelligent Cloud revenue was $26.7bn, up 21% YoY and 3% QoQ, beating both the company guidance ($26-26.3bn) and the consensus ($26.2bn) • Azure and other cloud service revenue growth of 31% YoY in cc, beating both the consensus and the company's guidance of 28%
	2Q24F	• Guides Azure sales to grow 30-31% YoY, ahead of consensus of 28% YoY
	FY24F	• Consensus expects Intelligent Cloud sales to grow 19% YoY in FY2024 • Consensus expects Azure sales to grow 28% YoY in cc in FY2024
Google	1Q24	• Google Cloud sales rose 28% YoY and 4% QoQ to $9.5bn, ahead consensus by 2%, driven by demand for GCP infrastrucutre and solutions
	2024F	• Consensus expects Google Cloud sales to grow 25% YoY in 2024F
amazon	1Q24	• AWS revenue grew 17% YoY and 3% QoQ to US$25bn in 1Q24, beating consensus by 4% • Witnessed growth in both Gen AI and non-Gen AI workloads across a diverse group of customers and across different industries, since companies are migrating more workloads to the cloud, while signing up for longer deals, making bigger commitments.
	2024F	• The company continues to see the impact of cost optimization diminishing, causing customers turning their attention to newer initiatives and re-accelerating existing • Consensu AWS sales growth of 15% YoY to $104bn

Source: Company data, Bloomberg, KGI Research

Figure 41: CPU vendor server business outlook

Company	Time	Actual & Guidance
intel	1Q24	• Data center & AI (DCAI) sales was US$3.03bn, down 24% QoQ and 18% YoY, in line with market consensus, driven by higher Xeon ASPs and improved enterprise demand. • More than 5 million AI PCs have shipped since the launch of Intel Core Ultra processors.
	2Q24F	• Guides 2Q24F DCAI sales will be flat QoQ at the low-end of seasonality, shy of consensus by 7%. • Expect to release the 1.0 PDK for Intel 18A in 2Q24F.
	2024F	• Guides sales growth across all segments in 2H24, led by improving demand for general servers from both cloud and enterprise customers • Expects sales to grow by QoQ throughout 2024F and into 2025F, driven by 1) the beginnings of an enterprise refresh cycle and growing momentum for AI PCs, 2) a datacenter recovery with a return to more normal CPU buying patterns and ramping of accelerator offering. • Consensus DCAI sales growth of 8% YoY to $13.7bn in 2024 • Granite Rapids (Birch Stream) will launch in 3Q24, and Clearwater Forest (Birch Stream) will launch in 2025F.
AMD	1Q24	• Data center sales grew 2% QoQ and 81% YoY to $2.3bn, beating consensus by 1% and prior guidance of flat QoQ, driven by the ramp up of MI300 and double-digit YoY server CPU sales growth, partially offset by a seasonal decline in server CPU sales. • MI300 total sales passed $1bn in less than two quarters.
	2Q24F	• Guides double-digit QoQ data center sales growth double and significant YoY growth, driven by the continuing ramp up of MI300 and market share gains for EPYC processor. • AMD saw early signs of recovery in the enterprise, and some large customers started to refresh programs.
	2024F	• Raise data center GPU sales guidance from $3.5bn to exceed $4bn in 2024F, implying over 30% sales weighting of data center sales in 2024F. • Guides double-digit YoY data center sales growth in 2024, versus consensus of $12.2bn (up 87% YoY) • Positive on traditional server outlook in 2024 and beyond thanks to server CPU replacement cycle and the launch of Turin CPU in 2H24
NVIDIA	1Q24	• Data center sales grew 427% YoY and 23% QoQ to US$22.6bn, beat consensus by 7%, driven by strong demand for the Hopper GPU computing platform. • Strong data center sales growth was led by enterprise and consumer Internet companies. CSPs continue to drive strong growth as they deploy and ramp AI infrastructure at scale and represented around 45% of data center revenue.
	2Q24F	• Guides 2Q24F Data center sales to see a sequential growth, production shipments will start in 2Q24, ramp in 3Q24F, and customers should have data centers stood up in 4Q24F.
	2024F	• Expects H200 and Blackwell demand will exceed supply and continue into 2025. • Accelerating demand for Gen-AI training and inference on the Hopper platform is boosting data center growth.

Source: Company data, Bloomberg, KGI Research

Figure 42: Decelerating CSP capex growth in 2023, but the market expects CSP capex to resume YoY growth in 2024F

Capex, US$mn	2019	2020	2021	2022	2023	2024F	2025F
Meta	15,102	15,115	18,567	31,431	27,266	37,383	41,418
Amazon	16,861	35,044	55,396	58,321	48,133	62,228	67,213
Microsoft	13,546	17,592	23,216	24,768	35,202	50,363	57,358
Google	23,548	22,281	24,640	31,485	32,251	46,817	48,241
Baidu	931	738	1,689	1,586	1,687	1,669	1,745
Alibaba	6,517	6,379	8,311	5,014	5,286	6,011	6,232
Tencent	3,927	5,719	4,808	4,611	4,371	7,100	6,659
Hyperscale subtotal	**80,432**	**102,867**	**136,627**	**157,216**	**154,196**	**211,571**	**228,865**
Apple	9,247	8,702	10,388	11,692	9,564	10,918	11,921
IBM	2,286	2,618	2,062	1,346	1,488	1,720	1,934
Oracle	1,591	1,833	3,118	6,678	6,935	9,636	9,965
Paypal	704	866	908	706	759	800	946
eBay	508	463	444	420	455	500	504
Salesforce	643	710	717	798	813	739	821
Netflix	253	498	525	408	349	428	465
Uber	588	616	298	252	238	304	338
Enterprise subtotal	**15,820**	**16,306**	**18,460**	**22,300**	**20,601**	**25,046**	**26,894**
Total	**96,793**	**119,173**	**155,086**	**179,516**	**174,797**	**236,617**	**255,759**

YoY growth, percent	2019	2020	2021	2022	2023	2024F	2025F
Meta	8.5	0.1	22.8	69.3	(13.3)	37.1	10.8
Amazon	25.6	107.8	58.1	5.3	(17.5)	29.3	8.0
Microsoft	6.0	29.9	32.0	6.7	42.1	43.1	13.9
Google	(6.3)	(5.4)	10.6	27.8	2.4	45.2	3.0
Baidu	(29.9)	(20.7)	129.1	(6.1)	6.3	(1.1)	4.6
Alibaba	(11.9)	(2.1)	30.3	(39.7)	5.4	13.7	3.7
Tencent	17.0	45.6	(15.9)	(4.1)	(5.2)	62.4	(6.2)
Hyperscale subtotal	**4.0**	**27.9**	**32.8**	**15.1**	**(1.9)**	**37.2**	**8.2**
Apple	(26.7)	(5.9)	19.4	12.6	(18.2)	14.2	9.2
IBM	(32.7)	14.5	(21.2)	(34.7)	10.5	15.6	12.4
Oracle	8.4	15.2	70.1	114.2	3.8	38.9	3.4
Paypal	(14.5)	23.0	4.8	(22.2)	7.6	5.4	18.3
eBay	(22.0)	(8.9)	(4.1)	(5.3)	8.2	9.9	0.8
Salesforce	8.1	10.4	1.0	11.3	1.8	(9.1)	11.1
Netflix	45.5	96.8	5.4	(22.3)	(14.5)	22.9	8.7
Uber	5.4	4.8	(51.6)	(15.4)	(5.6)	27.7	11.2
Enterprise subtotal	**(22.0)**	**3.1**	**13.2**	**20.8**	**(7.6)**	**21.6**	**7.4**
Total	**(1.3)**	**23.1**	**30.1**	**15.8**	**(2.6)**	**35.4**	**8.1**

Source: Company data; Bloomberg; KGI Research

Figure 43: Cloud peer comparison – Valuation

Sector	Company	Ticker	Market cap. (US$mn)	Share price (LCY)	Rating	Target Price (LCY)	EPS (LCY) 2023	EPS (LCY) 2024F	EPS (LCY) 2025F	EPS YoY (%) 2023	EPS YoY (%) 2024F	EPS YoY (%) 2025F	PE (x) 2023	PE (x) 2024F	PE (x) 2025F	PB (x) 2023	PB (x) 2024F	PB (x) 2025F	ROE (%) 2023	ROE (%) 2024F	ROE (%) 2025F	Cash yield (%) 2023	Cash yield (%) 2024F
ODM	Hon Hai	2317 TT	74,391	173.0	Outperform	205.0	10.25	11.38	14.19	0.4	11.0	24.8	16.9	15.2	12.2	1.6	1.5	1.4	9.7	10.3	12.2	3.1	3.5
	Inventec	2356 TT	6,009	54.0	Outperform	62.0	1.71	2.69	3.47	0.0	57.2	29.1	31.6	20.1	15.6	3.2	3.1	3.0	10.2	15.2	19.0	2.8	3.9
	Quanta	2382 TT	34,027	284.0	Outperform	320.0	10.29	13.97	18.02	37.0	35.8	29.0	27.6	20.3	15.8	5.9	5.6	5.2	22.3	28.1	34.1	3.2	3.9
	Wistron	3231 TT	10,243	114.0	Outperform	140.0	4.08	5.72	8.24	1.7	40.1	44.0	27.9	19.9	13.8	3.2	2.8	2.5	11.4	14.7	18.7	2.3	3.2
	Wiwynn	6669 TT	14,860	2,740.0	Outperform	2,900.0	68.88	123.87	139.79	(15.0)	79.8	12.9	39.8	18.9	16.8	11.3	7.3	5.8	29.7	44.1	38.5	1.5	2.9
	Gigabyte Tech	2376 TT	6,251	317.0	Outperform	348.0	7.46	16.11	18.73	(27.5)	115.9	16.3	42.5	19.7	16.9	5.4	5.2	5.1	12.9	27.0	30.4	2.1	4.6
	Asrock	3515 TT	838	222.0	Neutral	242.0	7.54	10.00	12.00	(13.2)	32.7	20.0	29.4	22.2	18.5	3.3	3.2	3.0	11.3	14.6	16.7	3.1	3.0
	Mitac Holdings*	3706 TT	1,796	48.0	Not rated	N.A.	1.48	1.92	N.M.	(80.9)	29.7	N.A.	32.4	25.0	N.A.	0.9	N.A.	N.A.	2.9	N.M.	N.M.	2.7	N.A.
Socket/ Connector/cable	Lotes	3533 TT	5,792	1,675.0	Outperform	1,660.0	50.65	71.33	89.68	(13.7)	40.8	25.7	33.1	23.5	18.7	6.7	5.9	5.1	22.1	26.5	29.0	1.6	2.1
	Fit Hon Teng Ltd*	6088 HK	1,979	2.12	Not rated	N.A.	0.08	0.13	0.20	222.3	70.5	47.4	27.2	15.9	10.8	6.3	5.8	5.3	5.4	7.5	12.8	0.0	0.0
	Argosy*	3217 TT	461	165.0	Not rated	N.A.	8.11	10.65	N.M.	19.3	31.3	N.A.	20.3	15.5	N.A.	3.5	3.3	N.A.	17.9	20.5	N.M.	3.2	5.1
	Alltop	3526 TT	434	236.5	Outperform	290.0	11.68	15.53	18.82	24.2	32.9	21.2	20.2	15.2	12.6	5.4	5.3	5.3	26.9	35.1	42.4	4.9	6.5
Rail kit	King Slide Works	2059 TT	3,798	1,285.0	Outperform	1,590.0	28.38	49.07	56.93	(33.3)	72.9	16.0	45.3	26.2	22.6	7.5	6.4	5.5	16.9	26.3	26.1	1.1	2.0
Thermal module	Sunonwealth	2421 TT	963	113.5	Outperform	160.0	5.16	6.32	8.02	18.9	22.5	26.8	22.0	18.0	14.2	4.3	4.0	3.7	21.5	23.1	27.3	3.1	4.0
	Auras	3324 TT	2,352	834.0	Outperform	920.0	14.28	21.10	32.71	(2.7)	47.8	55.0	58.4	39.5	25.5	15.5	9.7	8.0	20.2	25.5	33.6	0.8	1.1
	AVC	3017 TT	8,418	708.0	Outperform	800.0	14.11	18.91	26.66	19.7	34.0	41.0	50.2	37.4	26.6	12.3	10.6	8.8	28.3	30.4	36.2	1.0	1.3
	Kaori	8996 TT	1,325	478.0	Outperform	570.0	6.45	7.97	16.95	91.5	23.5	112.8	74.1	60.0	28.2	15.8	15.3	10.8	24.2	26.0	45.1	0.8	1.8
Heat spreader	Jentech*	3653 TT	4,237	970.0	Not rated	N.A.	16.59	N.M.	N.M.	(15.1)	N.A.	N.A.	58.5	N.A.	N.A.	11.9	N.A.	N.A.	21.9	N.M.	N.M.	1.0	N.A.
BBU	Simplo Tech	6121 TT	2,341	408.0	Neutral	490.0	30.67	30.53	34.76	(21.6)	(0.5)	13.8	13.3	13.4	11.7	2.2	2.1	2.0	16.9	16.2	17.5	5.3	5.3
	AES-KY	6781 TT	1,677	633.0	Neutral	760.0	23.04	26.03	34.79	(38.8)	13.0	33.6	27.5	24.3	18.2	2.9	2.7	3.5	15.4	16.5	20.1	1.8	2.1
Chassis	Chenbro	8210 TT	1,122	300.0	Outperform	340.0	9.03	14.52	18.90	8.5	60.8	30.1	33.2	20.7	15.9	6.5	5.7	4.9	20.8	29.4	32.9	1.7	2.5
BMC	Aspeed Tech*	5274 TT	3,942	3,360.0	Not rated	N.A.	26.66	47.96	75.15	(52.2)	79.9	56.7	126.0	70.1	44.7	33.1	26.7	21.5	24.2	44.7	54.8	0.6	1.1
Silicon photonics	Land Mark Opto	3081 TT	377	132.5	Neutral	118.0	(2.31)	0.67	7.15	N.M.	N.M.	959.5	N.M.	196.4	18.5	3.2	3.1	3.1	(5.1)	1.6	16.7	0.4	0.4
CCL	Iteq*	6213 TT	1,233	109.5	Not rated	N.A.	1.86	4.67	6.96	(62.3)	150.8	49.1	58.9	23.5	15.7	2.0	1.9	1.9	3.4	8.8	11.7	1.4	2.4
	Elite Material	2383 TT	4,597	430.5	Outperform	590.0	16.35	27.41	31.55	7.3	67.6	15.1	26.3	15.7	13.6	5.5	4.9	4.3	22.5	33.0	33.6	2.3	3.9
ABF	Unimicron Tech	3037 TT	9,035	191.0	Outperform	235.0	7.88	9.45	16.48	(60.7)	19.8	74.4	24.2	20.2	11.6	3.2	2.9	2.5	13.6	15.1	23.4	1.6	2.0
PCB	Gold Circuit	2368 TT	3,074	201.5	Outperform	270.0	6.99	12.79	17.21	(21.1)	83.1	34.5	28.8	15.7	11.7	6.5	5.4	4.2	23.2	37.5	40.6	1.7	3.0
Power	Delta	2308 TT	26,548	329.5	Outperform	352.0	12.86	13.00	15.29	2.2	1.1	17.6	25.6	25.3	21.6	4.3	3.8	3.5	17.3	16.0	17.0	2.0	2.0
	Lite-On Tech	2301 TT	8,156	112.0	Restricted	N.A.	6.75	7.90	9.04	9.0	17.1	14.4	16.6	14.2	12.4	2.9	2.7	2.5	18.2	19.9	21.3	4.5	5.4
	Chicony Power	6412 TT	1,919	155.0	Not rated	N.A.	8.27	9.94	11.88	0.6	20.2	19.4	18.7	15.6	13.0	4.5	3.7	3.7	24.9	29.7	31.5	3.9	4.1
	AcBel Polytech*	6282 TT	993	37.2	Not rated	N.A.	0.07	0.36	1.19	(94.0)	414.3	230.6	530.7	103.2	31.2	1.4	N.A.	N.A.	0.3	1.4	4.4	1.3	N.A.
Foundry	TSMC	2330 TT	703,904	875.0	Not rated	N.A.	32.3	38.15	44.21	(17.5)	18.0	15.9	27.1	22.9	19.8	6.5	5.4	4.5	26.0	25.6	24.9	1.3	1.5
Design service	Alchip Tech	3661 TT	6,660	2,720.0	Outperform	4,600.0	45.5	84.39	99.44	77.0	85.6	17.8	59.8	32.2	27.4	11.2	9.3	7.8	21.4	32.1	30.8	0.8	1.6
	Global Unichip	3443 TT	5,944	1,430.0	Neutral	1,360.0	26.2	27.23	33.55	(5.5)	4.0	23.2	54.6	52.5	42.6	19.8	16.7	13.7	39.5	34.5	35.3	1.0	1.0
IC package/testing	Winway	6515 TT	949	880.0	Outperform	865.0	15.9	34.57	N.A.	(50.7)	117.8	N.A.	55.4	25.5	N.A.	8.6	7.0	N.A.	15.1	30.4	N.A.	1.3	2.9
	Chroma	2360 TT	3,694	280.0	Not rated	N.A.	10.9	12.72	N.A.	(10.3)	16.8	N.A.	25.7	22.0	N.A.	5.2	4.8	N.A.	20.7	22.5	N.A.	2.6	3.0
Networking	Luxnet	4979 TT	627	143.5	Underperform	98.0	3.3	4.23	5.51	63.0	26.7	30.1	43.0	33.9	26.1	6.7	5.8	4.8	19.9	17.9	19.3	1.0	1.4
	Accton Tech	2345 TT	8,864	510.0	Outperform	620.0	16.0	19.28	22.05	9.2	20.6	14.4	31.9	26.4	23.1	11.3	10.0	8.7	39.1	39.9	40.0	2.0	2.5

* Bloomberg consensus

Source: Bloomberg; KGI Research estimates

Figure 44: Global server shipments by brand

Vendor	Server shipments ('000 units)																	
	1Q21	2Q21	3Q21	4Q21	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	2018	2019	2020	2021	2022	2023
Dell EMC	489	529	499	548	542	496	526	484	366	362	343	366	2,267	2,050	1,950	2,064	2,048	1,437
HPE	322	339	316	342	300	269	345	380	252	238	209	212	1,691	1,546	1,392	1,319	1,294	911
Inspur Electronics	231	348	356	388	343	340	348	337	212	285	333	275	1,061	1,131	1,230	1,323	1,368	1,105
Lenovo	169	202	216	186	126	197	236	216	173	152	160	150	765	755	702	773	775	635
Supermicro	133	157	152	171	171	208	227	217	151	120	116	161	N.A.	N.A.	496	614	823	548
xFusion Digital Technologies	N.A.	N.A.	N.A.	154	120	136	139	144	90	114	105	135	N.A.	N.A.	N.A.	154	539	444
H3C	86	109	129	134	88	109	92	132	73	90	79	96	178	254	345	458	422	339
Cisco	55	48	46	51	40	48	40	31	35	29	36	37	280	279	228	200	160	136
PowerLeader	20	22	27	81	61	60	62	75	47	54	57	65	86	109	119	150	258	222
ZTE	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	44	102	77	77	N.A.	N.A.	N.A.	N.A.	N.A.	300
Sugon	30	34	48	101	28	36	50	46	52	52	50	55	333	347	209	213	160	186
Others	1,023	1,333	1,248	1,585	1,452	1,455	1,268	1,634	1,138	1,163	1,189	1,294	5,173	5,107	5,048	5,189	5,808	4,785
Total	2,739	3,245	3,140	3,794	3,307	3,387	3,379	3,754	2,664	2,809	2,805	2,980	12,956	12,537	12,672	12,918	13,827	11,259
	YoY (%)																	
Dell EMC	3.1	22.3	(0.9)	1.3	11.0	(6.3)	5.6	(11.7)	(32.6)	(26.9)	(34.8)	(24.2)	10.9	(9.6)	(4.9)	5.8	(0.8)	(29.8)
HPE	(0.2)	(7.1)	(4.6)	(8.3)	(7.1)	(20.6)	9.0	11.2	(16.0)	(11.6)	(39.3)	(44.2)	(7.5)	(8.6)	(10.0)	(5.2)	(2.0)	(29.6)
Inspur Electronics	(3.0)	(4.6)	10.3	27.4	48.8	(2.3)	(2.3)	(13.2)	(38.2)	(16.2)	(4.2)	(18.5)	44.0	6.6	8.8	7.5	3.4	(19.2)
Lenovo	9.2	6.0	34.1	(4.9)	(25.4)	(2.4)	9.1	16.2	37.2	(22.7)	(32.3)	(30.6)	22.4	(1.3)	(7.1)	10.1	0.3	(18.1)
Supermicro	16.9	7.2	37.7	37.4	28.8	32.2	49.0	26.6	(12.1)	(42.3)	(48.7)	(25.7)	N.A.	N.A.	N.A.	23.8	34.1	(33.4)
xFusion Digital Technologies	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	(6.3)	(25.0)	(15.8)	(24.9)	(6.4)	N.A.	N.A.	N.A.	N.A.	250.5	(17.7)
H3C	56.6	23.6	36.4	24.3	2.4	0.4	(28.3)	(1.3)	(16.8)	(17.3)	(14.2)	(27.4)	20.2	43.1	35.7	32.6	(7.8)	(19.7)
Cisco	3.5	(12.4)	(23.2)	(15.4)	(27.0)	0.4	(12.4)	(39.5)	(13.4)	(39.4)	(11.7)	18.9	(11.5)	(0.2)	(18.0)	(12.4)	(20.3)	(14.6)
PowerLeader	7.7	(3.4)	(19.9)	84.2	206.8	171.3	126.1	(7.2)	(22.6)	(10.6)	(8.2)	(13.3)	10.7	26.4	9.6	25.9	71.4	(13.6)
ZTE	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Sugon	(52.6)	(46.4)	19.3	137.8	(5.0)	6.0	3.9	(54.7)	4.8	43.2	(0.9)	20.3	16.4	4.3	(39.8)	1.9	(24.8)	16.1
Others	(17.6)	2.7	20.6	7.6	41.9	9.1	1.6	3.1	(21.6)	(20.1)	(6.2)	(20.8)	16.7	(1.3)	(1.2)	2.8	11.9	(17.6)
Total	(6.1)	(0.1)	7.6	5.7	20.7	4.4	7.6	(1.1)	(19.4)	(17.1)	(17.0)	(20.6)	13.1	(3.2)	1.1	1.9	7.0	(18.6)
	QoQ (%)																	
Dell EMC	(9.6)	8.3	(5.8)	9.9	(1.0)	(8.6)	6.2	(8.1)	(24.4)	(1.0)	(5.2)	6.8						
HPE	(13.6)	5.0	(6.7)	8.2	(12.4)	(10.2)	28.1	10.4	(33.9)	(5.4)	(12.1)	1.4						
Inspur Electronics	(24.3)	51.0	2.3	9.0	(11.6)	(0.9)	2.3	(3.2)	(37.1)	34.4	17.0	(17.5)						
Lenovo	(13.8)	19.5	7.3	(14.0)	(32.3)	56.3	20.0	(8.4)	(20.1)	(11.9)	5.0	(6.1)						
Supermicro	6.8	18.4	(3.3)	12.4	0.1	21.5	9.0	(4.6)	(30.5)	(20.3)	(3.0)	38.3						
xFusion Digital Technologies	N.A.	N.A.	N.A.	N.A.	(22.0)	13.1	2.8	3.3	(37.6)	26.9	(8.3)	28.7						
H3C	(20.1)	26.0	18.7	3.9	(34.1)	23.5	(15.2)	43.1	(44.5)	22.9	(12.1)	21.1						
Cisco	(10.1)	(11.7)	(4.9)	12.1	(22.4)	21.3	(17.0)	(22.7)	11.1	(15.0)	21.0	4.2						
PowerLeader	(55.2)	11.8	23.9	196.9	(25.4)	(1.2)	3.2	21.8	(37.8)	14.3	6.0	15.0						
ZTE	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	130.4	(23.8)	(1.2)						
Sugon	(30.3)	15.3	40.9	109.9	(72.1)	28.6	38.1	(8.4)	(35.6)	75.9	(4.4)	11.1						
Others	(30.5)	30.3	(6.4)	27.0	(8.4)	0.2	(12.9)	28.9	(30.3)	2.2	2.2	8.9						
Total	(23.7)	18.5	(3.2)	20.8	(12.8)	2.4	(0.2)	11.1	(29.0)	5.4	(0.1)	6.2						
	Global market share (%)																	
Dell EMC	17.8	16.3	15.9	14.4	16.4	14.6	15.6	12.9	13.7	12.9	12.2	12.3	17.5	16.4	15.4	16.0	14.8	12.8
HPE	11.8	10.4	10.1	9.0	9.1	7.9	10.2	10.1	9.4	8.5	7.5	7.1	13.1	12.3	11.0	10.2	9.4	8.1
Inspur Electronics	8.4	10.7	11.3	10.2	10.4	10.0	10.3	9.0	8.0	10.1	11.9	9.2	8.2	9.0	9.7	10.2	9.9	9.8
Lenovo	6.2	6.2	6.9	4.9	3.8	5.8	7.0	5.8	6.5	5.4	5.7	5.0	5.9	6.0	5.5	6.0	5.6	5.6
Supermicro	4.9	4.9	4.9	4.5	5.2	6.1	6.7	5.8	5.7	4.3	4.2	5.4	N.A.	N.A.	3.9	4.8	6.0	4.9
xFusion Digital Technologies	N.A.	N.A.	N.A.	4.1	3.6	4.0	4.1	3.8	3.4	4.1	3.7	4.5	N.A.	N.A.	N.A.	1.2	3.9	3.9
H3C	3.1	3.3	4.1	3.5	2.7	3.2	2.7	3.5	2.8	3.2	2.8	3.2	1.4	2.0	2.7	3.5	3.1	3.0
Cisco	2.0	1.5	1.5	1.4	1.2	1.4	1.2	0.8	1.3	1.0	1.3	1.2	2.2	2.2	1.8	1.5	1.2	1.2
PowerLeader	0.7	0.7	0.9	2.1	1.8	1.8	1.8	2.0	1.8	1.9	2.0	2.2	0.7	0.9	0.9	1.2	1.9	2.0
ZTE	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	1.7	3.6	2.8	2.6	N.A.	N.A.	N.A.	N.A.	N.A.	2.7
Sugon	1.1	1.1	1.5	2.7	0.9	1.1	1.5	1.2	1.1	1.8	1.8	1.8	2.6	2.8	1.7	1.6	1.2	1.7
Others	37.4	41.1	39.7	41.8	43.9	43.0	37.5	43.5	42.7	41.4	42.4	43.4	39.9	40.7	39.8	40.2	42.0	42.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: Gartner, KGI Research

All the above named KGI analyst(s) is SFC licensed person accredited to KGI Asia Ltd to carry on the relevant regulated activities. Each of them and/or his/her associate(s) does not have any financial interest in the respectively covered stock, issuer and/or new listing applicant.