# IT Hardware

## Spotlight on ASIC server market growth in 2025-26F

### Key message

1. AI server demand will continue growing in 2H25-2026F, with ASIC demand picking up more significantly.
2. We expect a 2024-26F CAGR of 55%, for total AI training chip shipments (GPU & ASIC) with GPU shipments (60-70% market share) at a CAGR of 47% and ASIC shipments (30-40%) at 70%.
3. We expect the 2024-26F CAGR of the total AI training server TAM (GPU and ASIC) to be 37%, with GPU AI server TAM (90% market share) at 34% and ASIC TAM (10%) at 92%.
4. We believe the major content value upgrades for the ASIC AI market will benefit thermal components and PCB/ CCL firms. Our picks for growing ASIC AI market are Wiwynn (6669 TT, NT$2,455, OP), Quanta (2382 TT, NT$277, OP), Asia Vital Components (AVC; 3017 TT, NT$762, OP), Auras Technology (3324 TT, NT$635, OP), Fositek (6805 TT, NT$787, OP), Elite Material (EMC; 2383 TT, NT$901, OP) and Gold Circuit Electronics (2368 TT, NT$298, OP).

**Event**

The AI training market has grown rapidly since 2023, using mainly Nvidia's (US) training GPUs. We note that US CSPs are actively developing their own ASIC designs for AI servers with different applications, and expect the ASIC market to be a highlight in 2025-26F.

**Impact**

**Rapid AI server market growth.** The AI training market will enjoy rapid growth in 2025-26F, and we forecast AI training GPU and ASIC shipments to grow from 8.57mn units in 2025F (up 69% YoY) to 12.15mn units in 2026F (up 42% YoY), driven by CSP demand. GPU shipments will account for 60-70% of the total AI training market in 2025-26F, and Nvidia still enjoys the bulk of market share (80-90%), with GB200/ 300 NVL72 shipments of approximately 20,000 racks this year and 30,000 racks in 2026F, not only to hyperscaler CSPs, but to neo-cloud and sovereign AI developers, as well as enterprises. Nvidia's HGX B200/300 and AMD's (US) MI355/400 will also see demand grow, and ASIC server demand will be robust in 2025-26F (largely from the top-4 US CSPs), growing at a CAGR of 70% in 2024-26F chip shipments. These will drive the global AI training TAM value, including GPU and ASIC designs, to grow at a 37% CAGR in 2024-26F, with training GPU sales at a 34% CAGR and ASIC sales at a 92% CAGR over the same period.

**More ASIC rack designs to be launched in 2025-26F.** Amazon Web Services (AWS; US) developed the "Teton" 1 rack (Trainium 2 chip) in 2024, and will deploy the Teton 2 rack (Trainium 2 or 2.5 chips) in 2H25F, with a similar structure to the previous version, including an air-cooled design (Figure 16). In 2026F, Teton 2 Max and Teton 3 racks will migrate to a liquid-cooled design with new chips, and each rack will contain 18 compute trays and 10 switch trays, implying 72 ASICs per rack, a higher rack density than the air-cooled designs, at 32 ASIC chips per rack. We expect AWS Trainium chip production to grow from 1.6mn units in 2025F to 2.28mn units in 2026F, up 43% YoY, benefiting rack assembler Wiwynn (6669 TT, NT$2,455, OP) and the related supply chain. Meta (US) will start to ramp up Minerva rack production in 2H25F, and expand it further in 2026F. Each Minerva rack has 16 compute blades, with each having one MTIA chip and one CPU in a liquid-cooled design, as well as 6 network blades and 1 chassis management module blade. Key assemblers for Meta include Celestica (US) and Quanta Computer (2382 TT, NT$277, OP). Google (US) will also launch TPU v7 and Microsoft (US) will debut Maia 200 in 2026F. Hyperscaler CSPs are the major driver for growing training ASIC chip demand, and we forecast ASIC shipments to them to grow by 74% to 2.77mn units in 2025F, and by 67% to 4.61mn units in 2026F, at a 2024-26F CAGR of 70%.

**Liquid cooling & PCB/CCL firms to benefit from ASIC trend.** We believe the major content value upgrades will be to thermal components and PCB/ CCL for the ASIC AI market. AWS' ASIC racks will migrate from air to liquid-cooling in 2026F, while Google and Microsoft may utilize both air- and liquid- cooling solutions. Meta's Minerva rack will ramp up after 2H25F, also with a liquid-cooled design. The thermal supply chain will benefit from growing liquid cooling adoption in GPU racks (GB200/ 300s and B300s) and ASIC racks, with increasing usage of cold plates, rack manifolds, quick-disconnectors (QD) and sidecar/ CDUs, benefitting Asia Vital Components (AVC; 3017 TT, NT$762, OP), Auras Technology (3324 TT, NT$635, OP), Fositek (6805 TT, NT$787, OP), and Delta Electronics (2308 TT, NT$450, OP). Furthermore, more ASIC AI servers will adopt advanced materials and multi-layer PCB, and AWS' Trainium 2 and 2.5 chips will adopt a 26L PCB design and M8 materials for CCL. Google's TPU and Meta's ASIC servers will adopt 22L and 30-40L PCB designs, respectively, and M7-M8 materials for CCL, pointing to a definite trend toward high-speed multi-layer PCB, benefiting Elite Material (EMC; 2383 TT, NT$901, OP) and Gold Circuit Electronics (GCE; 2368 TT, NT$298, OP).

**Stocks for Action**

With AI server market growth continuing, we expect ASIC server demand will accelerate and key beneficiaries include ODM, liquid cooling and PCB/ CCL plays. Key ASIC AI beneficiaries include Wiwynn, Quanta, AVC, Auras, Fositek, Elite Material and GCE.

**Risks**

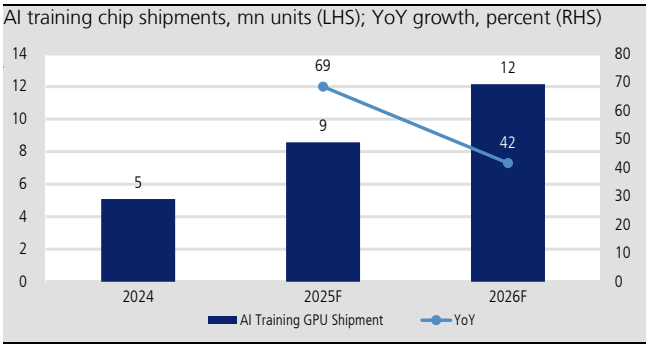
NT dollar appreciation; AI server production yield; rising production or financing costs.

### Total training AI market

The AI training market will enjoy rapid growth in 2024-26F, and we forecast AI training GPU and ASIC shipments to grow from 8.57mn units in 2025F (up 69% YoY) to 12.15mn units in 2026F (up 42% YoY) (Figure 1). GPU shipments will account for 60-70% of the total AI training GPU market (including Nvidia and AMD's GPUs) (Figure 2), with a 2024-26F shipments CAGR of 47%. The remainder of the AI training market is for ASIC chips, accounting for a smaller share of 30-40%, but we expect solid growth from ASIC shipments, mostly to the top-4 US CSPs, at a 2024-26F CAGR of 70%.

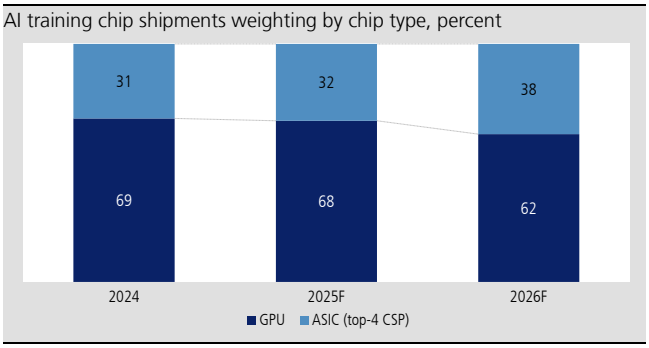
Based on the ASP of Nvidia's GB200/ 300 AI server rack, at around US$3mn, the HGX server system at US$250k, AMD's MI300/ 350 server systems at around US$200k, and ASIC server racks at US$50-600k, depending on the different designs by different CSPs, we estimate the global AI training TAM, including GPU and ASIC designs, will grow to around US$300bn in 2025-26F, with 2024-26F CAGR of 37%. (Figure 3) Training GPU demand (mostly benefiting Nvidia) started to accelerate in 2023 and the TAM will grow at a 34% CAGR over 2024-26F, accounting for around 90% of the global AI training market, despite a GPU shipments share of 60-70%. However, training ASIC demand will grow strongly in 2024-26F, at a stronger CAGR of 92% (Figure 4), despite comprising only around 10% of total AI market TAM. Therefore, we suggest investors keep an eye on ASIC beneficiaries with content value upgrades in store, such as liquid cooling and PCB/ CCL plays, in addition to GPU beneficiaries.

**Figure 1: AI training GPU & ASIC shipments to grow by 69% in 2025F and 42% in 2026F**



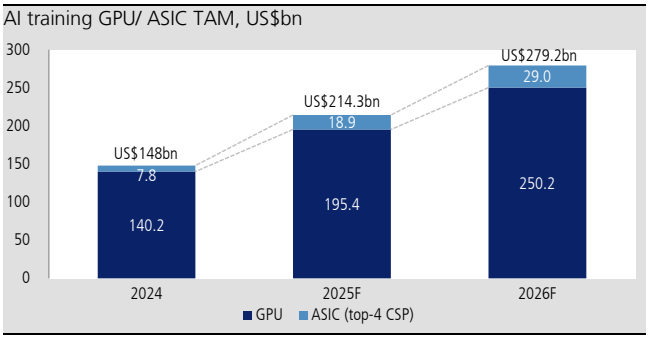

*Source: KGI Research estimates*

**Figure 2: AI training GPU shipments account for the bulk of market share, at 60-70%**



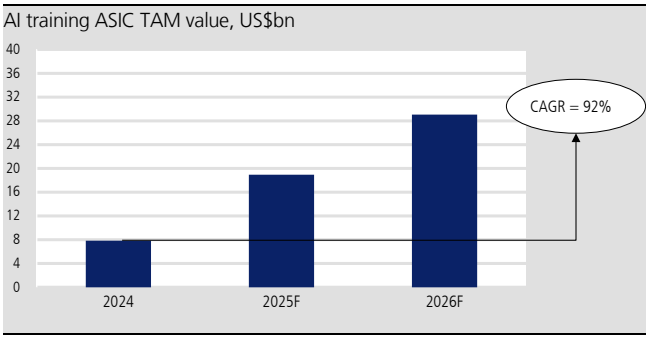

*Source: KGI Research estimates*

**Figure 3: Surging AI GPU & ASIC TAM growth on chip and spec upgrades in 2025-26F**




*Source: KGI Research estimates*

**Figure 4: Steady & strong AI ASIC TAM growth in 2025-26F**




*Source: KGI Research estimates*

**Figure 5: Both Nvidia's GPU & CSPs' ASIC chip (training & inference) shipments will grow in 2025-26F, with Google and AWS the leading ASIC CSPs**

| GPU (mn units) | 2024 | 2025F | 2026F |
|---|---|---|---|
| Nvidia | 2.8 | 5.2 | 6.3 |
| AMD | 0.5 | 0.6 | 1.3 |
| **ASIC (mn units)** | **2024** | **2025F** | **2026F** |
| Google TPU | 2.1-2.2 | 2.4-2.5 | 2.5-2.6 |
| AWS Tranium | 0.3-0.4 | 1.5-1.6 | 2.2-2.3 |
| Meta MTIA | 0.5-0.6 | 0.4-0.5 | 0.6-0.7 |
| MSFT MAIA | 0.0 | 0.0 | 0.4-0.5 |

*Source: KGI Research estimates*

**Figure 6: Both Nvidia's GPU & CSPs' ASIC chip shipments will grow in 2025-26F, and benefit the server supply chain**

| | GPU | | ASIC | | | |
|---|---|---|---|---|---|---|
| | GB200 | HGX | AWS-air | AWS-liquid | Meta | Google |
| **Chip** | Blackwell / Blackwell Ultra | | Traninum 2/2.5 | Trainium 2.5/3 | MTIA 2.5 (2T) | v5p/v5e |
| **System name** | GB200/300 | B200/300 | Teton 1/2 | Teton 3 | Minerva | TPU |
| **GPU per server** | 4 | 8 | 2 | 4 | 1 | 8 |
| **Server per rack** | 18 | 4 | 16 | 18 | 16 | ? |
| **Design service vendor** | Nvidia | Nvidia | Marvell<br>Alchip | Marvell<br>Alchip | Broadcom | Broadcom<br>MediaTek |
| **L10/11 assembler** | Hon Hai<br>Quanta<br>Wistron | Supermicro<br>Gigabyte<br>Quanta<br>Hon Hai<br>Inventec<br>Asustek<br>Asrock | Wiwynn | Wiwynn<br>Jabil | Quanta<br>Celestica | Celestica<br>Quanta<br>Flex |

*Source: KGI Research*

**Figure 7: Most ODMs have GPU & ASIC sales exposure with diverse clients**

| | AI server L10 shipments mix (%) | | | Major client | | |
|---|---|---|---|---|---|---|
| | GB200 | HGX | ASIC | GB200 | HGX | ASIC |
| Quanta | 55 | 40 | <5 | Meta, AWS | Meta, AWS | Meta |
| Hon Hai | 80-90 | 10-20 | | Microsoft, Oracle | Oracle, HPE | |
| Wistron | 100 | | | Dell / xAI | | |
| Wiwynn | <10 | | 90+ | YTL | | AWS |
| Gigabyte | <10 | 90+ | | | Coreweave, Nebius, OVH Cloud | |
| Asrock | <10 | 90+ | | | Nebius | |
| Celestica | V | | V | Google | | Meta |
| Flex | | V | V | | | |
| Jabil | | | V | | | AWS |
| SMCI | <20 | >80 | | | Coreweave, Oracle, xAI | |

*Source: KGI Research estimates*

### GPU AI market

- We expect AI training GPU shipments to grow by 66% YoY to 5.81mn units in 2025F, and 30% YoY to 7.54mn units in 2026F, and account for 68% and 62% of total AI training GPU and ASIC shipments in 2025F and 2026F, respectively.
- Nvidia will account for 80-90% of AI training GPU shipments in 2025-26F. We estimate shipment of 20k GB200/ 300 NVL72 racks in 2025F and 30k racks in 2026F.
- In system level design, Nvidia offers high-density GB200/ 300 designs, with GB200s entering into mass production in 2Q-3Q25F, while GB300s will likely ramp up in 4Q25-1H26F. NVL72 Oberon racks with 72 GPUs per AI server rack will be the mainstream design for GB200/ 300s, and the next generation Rubin GPU. We expect the rack design will change to Kyber rack architecture with the blade servers when adopting Rubin Ultra GPUs (Figure 11)
- In addition, HGX AI server shipments will keep growing thanks to increasing demand from enterprise as well as from neo-cloud and sovereign AI developers. Thus, we expect Nvidia's training GPU shipments (including the GB and HGX series)

will grow by 85% YoY to 5.19mn units in 2025F, and by 21% YoY to 6.28mn units in 2026F. (Figure 5)

- B300 racks will increasingly adopt liquid cooling designs, and thus the HGX AI server design will reach a height of 4-6U, down from the B200's 7-8U.
- AMD will launch the MI400 AI GPU with HBM4 memory in 2026F, which will have 10 times the AI compute performance of the MI355X. The firm will also launch the "Helio" AI rack, equipped with MI400 GPUs, Venice CPUs and Vulcano NICs, in a bid to compete with Nvidia's Vera Rubin racks. AMD is trying to catch up in AI development, and we forecast AMD's training GPU shipments to double YoY to 1.26mn units in 2026F, capturing 15-20% of the AI training GPU market, after expected 20-25% YoY growth in 2025F.
- The supply chain for AI servers equipped with GB200/300 and HGX GPUs is quite mature. This high-density AI server rack design, compared to H200, benefits the content value for some components, including liquid cooling, power supply, PCB, casing, and rail kits. The ASP for assembly manufacturers has a significantly increase accordingly, with Hon Hai, Quanta, and Wistron being the main US hyperscale CSP assemblers for GB200/300 AI server racks (L10/L11)

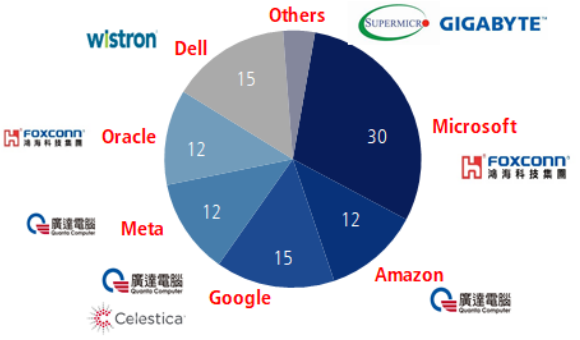
**Figure 8: Nvidia and AMD's data center GPU roadmaps**

| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025F | 2026F | 2027F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NVIDIA | 2Q16 P100 | 2Q17 V100 | 3Q18 T4 | | 2Q20 A100 | | 3Q22 H100 | 3Q23 H200 | 2Q-3Q24 GB200 | 2Q25 GB300 | | |
| AMD | 4Q16 MI6, MI8, MI25 | | 4Q18 MI50, MI60 | | 4Q20 MI100 | 4Q21 MI250X | | 4Q23 MI300X | 4Q24 MI325X | 2Q25 MI350X | 2026F MI400X | |

*Source: KGI Research*

**Figure 9: GB200 AI server racks are mainly assembled by Hon Hai, Quanta & Wistron**

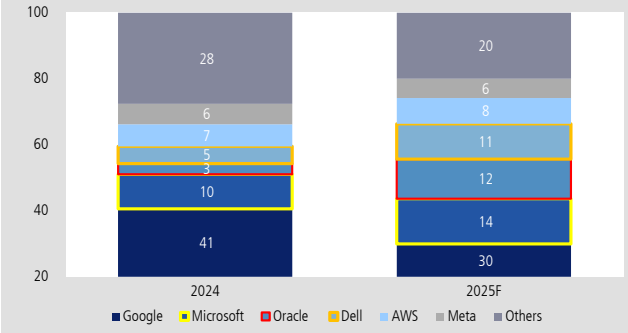
GB200 AI server rack shipments weighting, percent




*Source: KGI Research estimates*

**Figure 10: AI server demand has grown from Oracle, Dell & Microsoft in 2025F, mainly due to the GB200 ramp-up**

2025F AI server shipments weighting, percent



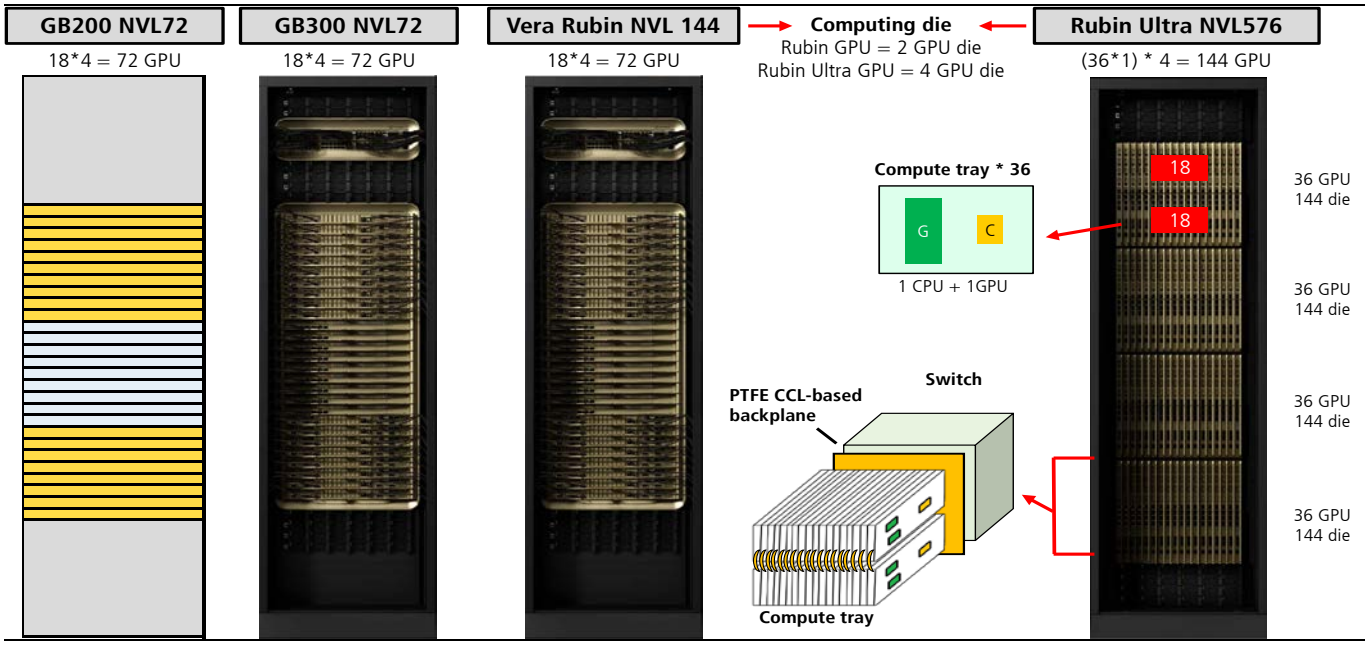

*Source: KGI Research estimates*

**Figure 11: L11 rack shipment allocation by end-customer**

| Allocation (%) | Hon Hai | Quanta | Wiwynn | Wistron | Celestica | Gigabyte | Asustek | Supermicro |
|---|---|---|---|---|---|---|---|---|
| Microsoft | 95 | 5 | | | | | | |
| Amazon | | 100 | | | | | | |
| Google | | 10 | | | 90 | | | |
| Meta | | 100 | | | | | | |
| Oracle | 100 | - | | | | | | |
| Dell | | | | 100 | | | | |
| Others | 10 | 10 | 25 | | | 20 | 20 | 15 |

*Source: KGI Research estimates*

**Figure 12: Rack design comparison – GB200, GB300 (Vera Rubin NVL144) & Rubin Ultra NVL576**




*Source: Nvidia; Semi Vision; KGI Research*

**ASIC AI market**

- The training ASIC market will grow at an accelerated pace in 2025-26F, due to demand from the top-4 US CSPs (Figure 13-15). We forecast training ASIC shipments to grow by 74% YoY to 2.77mn units in 2025F, and 67% YoY to 4.61mn units in 2026F.
- AWS developed the Teton 1 rack (Trainium 2 chip, air-cooled) in 2024, and the firm will deploy the Teton 2 rack (air-cooled) in 2H25F, the Teton 2 Max rack (air- or liquid-cooled) in 4Q25-1Q26F, and the Teton 2 Max and Teton 3 rack (new chips, liquid-cooling) in 2026F (Figure 16). Teton 1 and 2 have a similar structure, with 16 compute trays (2 GPUs in each tray) and 2 CPU trays in a rack. Teton 3 racks will use a liquid-cooled design, and each rack may contain 18 compute trays (4 GPUs in each tray) and 10 switch trays, implying 72 ASICs per rack, with higher rack density than the air-cooled designs, at 32 ASICs per rack. Increasing usage of cold plates, manifolds and quick-disconnectors will benefit thermal plays AVC (3017 TT, NT$762, OP), and Auras (3324 TT, NT$635, OP), more sidecar and CDU demand may benefit Delta (2308 TT, NT$450, OP), and quick-disconnector demand will benefit Fositek (6805 TT, NT$787, OP). Key assembler beneficiaries will be Wiwynn (6669 TT, NT$2,455, OP), and Jabil (US).
- We estimate AWS Trainium chip shipments will rise from 1.6mn units in 2025F to 2.28mn units in 2026F, up 43% YoY. Currently, Wiwynn is the primary assembler for AWS Trainium racks, and Jabil may also enter this market in the future.
- Meta (US) launched MTIA 1 in 4Q23 and MTIA 2/ 2.5 in 4Q24-1Q25, and plans to launch MTIA 3 in 2026F. The firm's Minerva rack, equipped with MTIA 2.5 (2T)

chips, will enter mass production in 2H25F. Each Minerva rack has 16 compute blades, with each having one MTIA 2.5 (2T) chip and one CPU in a liquid-cooled design, as well as 6 network blades and 1 chassis management module blade. Key assemblers for Meta include Celestica (US) and Quanta Computer (2382 TT, NT$277, OP). (Figure 17)

- Google's v5p, v6P and v7P TPU are for training purposes, with each server containing 8 TPUs. Key assemblers for Google include Celestica, Flex (US) and Quanta Computer. Microsoft introduced Maia 100 chips in 1Q23. The Maia 100 rack uses a liquid-cooled design, and there are Sidekick cooling units near the main server racks, to cool the Maia chips and CPUs.
- We believe the major content value upgrades in ASIC AI servers will be to thermal components, PCB/ CCL and assemblers like Wiwynn and Quanta.

**Figure 13: US CSPs' ASIC roadmaps**

| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026F | 2027F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| amazon | | | | | 4Q20 Trianium | | 1Q22 Inferntia2 | 4Q23 Trainium2 | | 3Q25-4Q25F Trainium 2.5 | 2Q26F Trainium 3 | 2Q27F Trainium4 |
| Meta | | | | | | | | 4Q23 MTIA | 2Q24 MTIA 2 | 1Q25 MTIA 2.5 | 1Q26F MTIA 3 | 2Q27F MTIA 4 |
| Google | 1Q16 TPU v1 | | 2Q18 TPU v2 | 2Q19 TPU v3 | | 2Q21 TPU v4 | | 1Q23 TPU v5p / 3Q23 TPU v5e / 4Q23 TPU v6e | 3Q24 TPU v6p | | 2Q26F TPU v7p / 3Q26F TPU v7e | 2Q27F TPU v8 |
| Microsoft | | | | | | | | 4Q23 Maia 100 | | | 2Q26F Maia 200 | |

*Source: KGI Research*

**Figure 14: US CSPs' ASIC spec comparison**

| | Amazon | | | | | | | META | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Project** | **Trainium** | **Inferntia 2** | **Trainium 2** | **Inferentia 2.5** | **Trainium 2.5/R** | **Trainium 3** | **Trainium 4** | **MTIA** | **MTIA 2** | **MTIA 2.5** | **MTIA 3** | **MTIA 4** |
| **Launch Date** | 4Q20 | 1Q22 | 4Q23 | NA | 3Q-4Q25F | 2Q26F | 2Q27F | 2Q23 | 2Q24 | 1Q25 | 1Q26F | 2Q27F |
| **Nodes** | 7nm | 7nm | 5nm | 5nm | 3nm | 3nm | 2nm | 7nm | 5nm | 5nm | 3nm | 2nm |
| **Memory Type** | HBM2/2E | HBM2E | HBM3 | HBM3 | HBM3 | HBM3 | HBM4 | LPDDR5 | LPDDR5 | LPDDR5 | HBM3E | HBM4 |
| **Memory Capacity** | 32GB | 32GB | 96GB | 48GB | 96GB | 96/128 | N.A. | 64GB | 128GB | N.A. | 216GB | N.A. |
| **Memory Stack** | 2*8hi*16GB | 2*8hi*12GB | 4*8hi*24GB | 2*8hi*24GB | 4*8hi*24GB | 4*8/12hi*24GB | N.A. | N.A. | N.A. | N.A. | 6*12hi*24GB | 8*12hi*24GB |
| **Architecture** | CoWoS-R | CoWoS-R | CoWoS-R | CoWoS-R | CoWoS-R | CoWoS-R | CoWoS-R | Flip Chip | Flip Chip | Flip Chip | CoWoS-S | CoWoS-L |
| **Hybrid Bonding** | N | N | N | N | N | N | N | N | N | N | N | N |
| **TDP** | N.A. | N.A. | 500W | N.A. | N.A. | N.A. | N.A. | 35W | 90W | N.A. | N.A. | N.A. |
| **Cooling Technology** | N.A. | N.A. | Air Cooled | Air Cooled | Air Cooled | Liquid Cooled | Liquid Cooled | N.A. | N.A. | N.A. | Air Cooled & Liquid Cooled | Liquid Cooled |

| | Google | | | | | | | | | | Microsoft | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Project** | **TPU v3** | **TPU v4i** | **TPU v4** | **TPU v5p** | **TPU v5e** | **TPU v6e (Trillium)** | **TPU v6p (Ironwood)** | **TPU v7p** | **TPU v7e** | **TPU v8** | **Maia 100** | **Maia 200** |
| **Launch Date** | 2Q19 | NA | 2Q21 | 1Q23 | 3Q23 | 4Q23 | 3Q24 | 2Q26F | 3Q26F | 2Q27F | 4Q23 | 2Q26F |
| **Nodes** | 16nm | 7nm | 7nm | 5nm | 5nm | 5nm | 5nm | 3nm | 3nm | 2nm | 5nm | 3nm |
| **Memory Type** | HBM2 | HBM2E | HBM2E | HBM3 | HBM2E | HBM3E | HBM3E | HBM3E | HBM3E | HBM4 | HBM2E | HBM4 |
| **Memory Capacity** | 8GB | 16GB | 16GB | 96GB | 32GB | 128GB | 192GB | 288GB | 216GB | NA | 64GB | 288GB |
| **Memory Stack** | 4*4hi*2GB | 4*8hi*4GB | 4*8hi*4GB | 6*8hi*16GB | 2*4hi*16GB | 8*8hi*16GB | 8*8hi*24GB | 8/12hi*24GB | 6*12hi*24GB | NA | 4*8hi*16GB | 8*12hi*24GB |
| **Architecture** | N.A. | CoWoS-S | CoWoS-S | CoWoS-S | CoWoS-S | CoWoS-S | CoWoS-S | CoWoS-S | CoWoS-S | NA | CoWoS-S | CoWoS-R |
| **Hybrid Bonding** | N | N | N | N | N | N | N | N | N | Y | N | N |
| **TDP** | 220W | N.A. | 170W | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. | 700W | N.A. |
| **Cooling Technology** | Liquid Cooled | N.A. | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled |

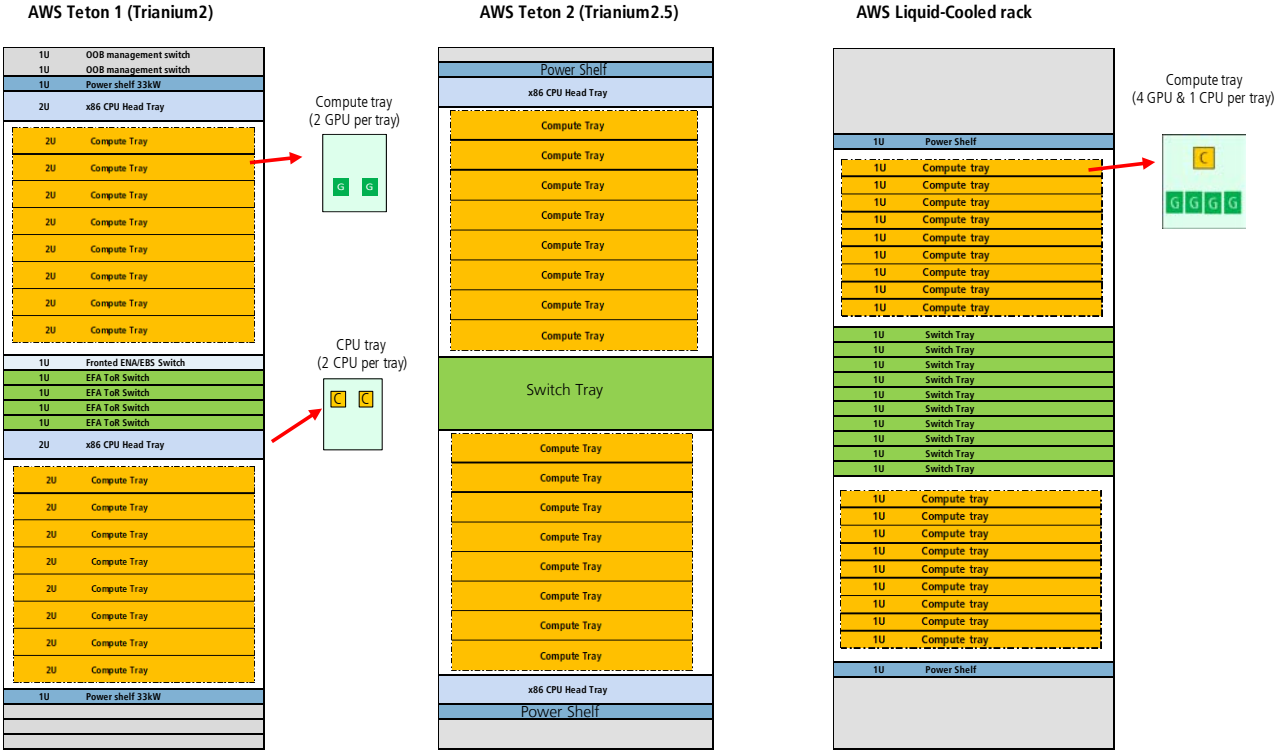
*Source: company data; KGI Research estimates*

## Figure 15: US CSPs' ASIC rack design comparison

| Company | Amazon | | | Meta | Microsoft | Google | | | |
|---|---|---|---|---|---|---|---|---|---|
| Rack Name | Teton 1 | Teton 2 | Teton 3 | Minerva | Athena | TPU v5p | TPU v5e | TPU v6e | TPU v6p |
| Chip | Trainium 2 | Trainium 2/2.5 | Trainium 3 | MTIA 2.5 (2T) | MAIA 100 | TPU v5p | TPU v5e | TPU v6e | TPU v6p |
| Launch time | 4Q23 | 3Q-4Q25F | 2Q26F | 1Q25 | 4Q23 | 1Q23 | 3Q23 | 4Q23 | 3Q24 |
| Process node | 5nm | 5/3nm | 3nm | 5nm | 5nm | 5nm | 5nm | 5nm | 5nm |
| Compute tray | 16 | 16 | 18 | 16 | 8 | 16 | 16 | TBA | TBA |
| Switch tray | 4 | 4 | 10 | 6 | 3 | TBA | TBA | TBA | TBA |
| GPU per rack | 32 | 32 | 72 | 16 | 32 | 64(8960 per pod) | (256 per pod) | (256 per pod) | TBA |
| CPU per rack | 4 | 4 | 18 | 16 | TBA | 16 | TBA | TBA | TBA |
| Power shelf per rack (33kW*6) | 2 | 2 (?) | TBA | 2 | TBA | TBA | TBA | TBA | TBA |
| CCL Level | Compute Tray M8 OAM M7 | Compute Tray M8 OAM M7 | TBA | Compute Tray M8 | TBA | TBA | TBA | TBA | TBA |
| PCB Layer Count | Compute Tray 26 OAM 18 | Compute Tray 26 OAM 18 | TBA | Compute Tray 30-40 | TBA | TBA | TBA | TBA | TBA |
| Networking | 800G | TBA | TBA | 800G | TBA | TBA | TBA | TBA | TBA |
| TDP – Chip | 500W | TBA | TBA | 800W | 700W | TBA | TBA | TBA | TBA |
| TDP – Rack | 26.8kW | TBA | TBA | 27.5kW | 20-40kW | TBA | TBA | TBA | TBA |
| Thermal solution | Air Cooled | Air Cooled | Liquid Cooled | Liquid Coolded & Air Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled | Liquid Cooled |

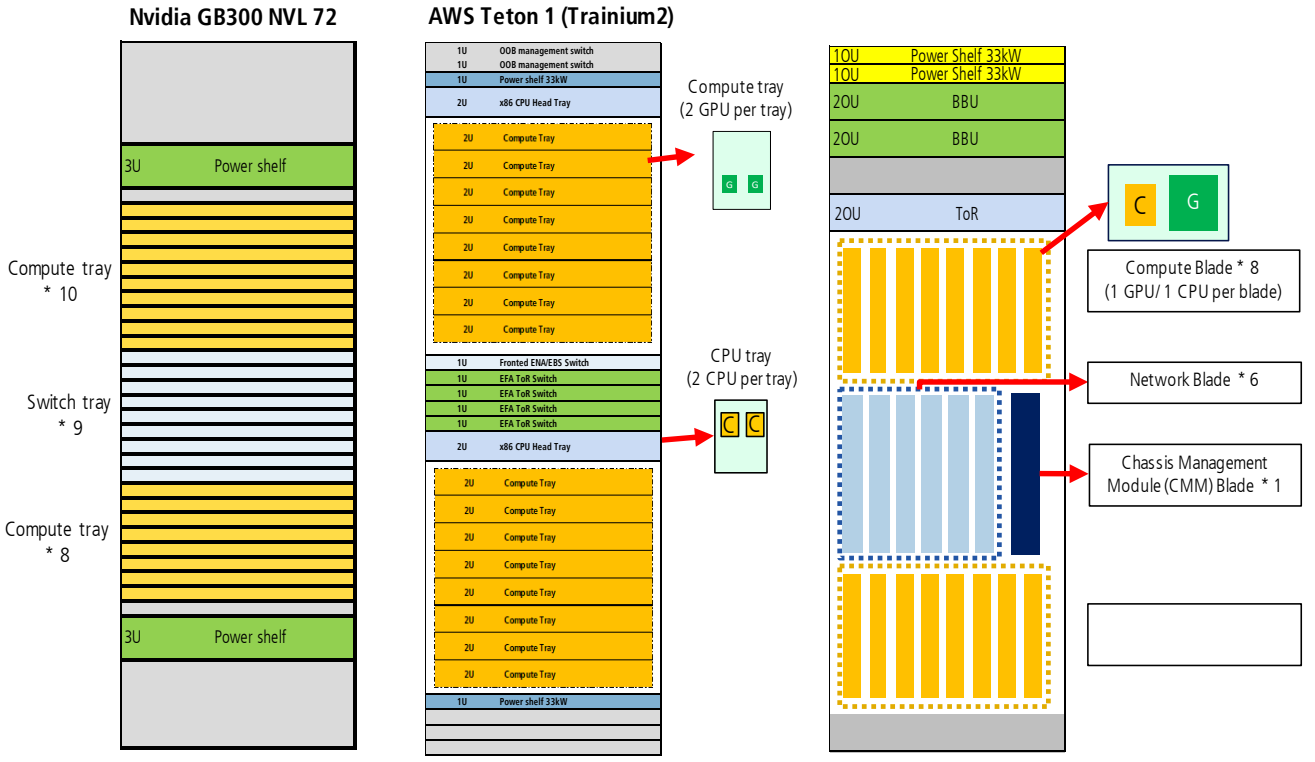
*Source: KGI Research*

## Figure 16: AWS ASIC rack designs




*Source: KGI Research*

**Figure 17: Rack design comparison between GPU servers & ASIC servers**




*Source: KGI Research*

**Key beneficiaries of growing GPU & ASIC demand**

- We believe the major content value upgrades in ASIC AI servers will be to thermal components and PCB/ CCL. AWS' ASIC racks will migrate from air-cooling to liquid-cooling in 2026F, while Google and Microsoft may have both air- and liquid-cooling solutions. Meta's Minerva rack will ramp up after 2H25F, also with a liquid-cooled design. The thermal supply chain will benefit from growing liquid cooling adoption in GPU racks (GB200/ 300s and B300s) and ASIC racks. Increasing usage of cold plates and rack manifolds will benefit AVC, and Auras. Quick-disconnector (QD) demand growth will benefit Fositek, and increased sidecar and CDU demand will benefit Delta Electronics (2308 TT, NT$450, OP). Major ASIC assemblers, Wiwynn and Quanta, will also benefit.
- In addition, more ASIC AI servers are adopting advanced materials and multi-layer PCB. According to our industry survey, AWS' Trainium 2 and 2.5 will each adopt a 26L PCB design and M8 materials for CCL. Google's TPU and Meta's ASIC servers will adopt 22L and 30-40L PCB designs, respectively, and M7-M8 materials for CCL, pointing to a definite trend toward high-speed multi-layer PCB (Figure 19). All of these factors will boost PCB and CCL plays' earnings growth, with key beneficiaries being EMC (2383 TT, NT$901, OP) and GCE (2368 TT, NT$298, OP).

**Figure 18: Taiwan & the US – Main ODM players server sales comparison**

| | | | Server sales (US$bn) | | Weighting of total sales (%) | | YoY (%) | |
|---|---|---|---|---|---|---|---|---|
| Sector | Company | Ticker | 2024 | 2025F | 2024 | 2025F | 2024 | 2025F |
| Taiwan ODM | Hon Hai | 2317 TT | 70 | 109 | 29.5 | 40.0 | 49.0 | 56.2 |
| | Quanta | 2382 TT | 29 | 60 | 60.0 | 74.2 | 120.4 | 105.0 |
| | Wiwynn | 6669 TT | 12 | 24 | 100.0 | 100.0 | 49.0 | 91.6 |
| | Wistron | 3231 TT | 16 | 33 | 45.4 | 57.0 | 56.9 | 101.9 |
| US ODM & Brand | Jabil* | JBL US | 5 | 7 | 15.9 | 25.6 | (4.2) | 60.9 |
| | Flex* | FLEX US | 2 | 4 | 8.8 | 13.6 | N.M. | 50.0 |
| | Celestica** | CLS US | 6 | 8 | 67.3 | 70.8 | 39.8 | 19.0 |
| | Dell | DELL US | 44 | 54 | 45.6 | 49.8 | 28.6 | 22.7 |
| | HPE | HPE US | 17 | 19 | 54.9 | 51.2 | 28.1 | 9.4 |
| | SMCI | SMCI US | 20 | 21 | 96.1 | 96.3 | 132.8 | 4.9 |
| | | | AI Server sales (US$bn) | | Weighting of server sales (%) | | YoY (%) | |
| Sector | Company | Ticker | 2024 | 2025F | 2024 | 2025F | 2024 | 2025F |
| Taiwan ODM | Hon Hai | 2317 TT | 25 | 55 | 36.5 | 50.6 | 150.7 | 116.6 |
| | Quanta | 2382 TT | 13 | 43 | 43.3 | 72.0 | 490.0 | 240.7 |
| | Wiwynn | 6669 TT | 3 | 11 | 26.6 | 48.0 | 99.1 | 246.0 |
| | Wistron | 3231 TT | 6 | 15 | 35.5 | 44.6 | 141.4 | 153.6 |
| US ODM & Brand | Jabil* | JBL US | 2 | 5 | 50.0 | 68.9 | N.M. | 121.7 |
| | Flex* | FLEX US | 1 | 3 | 51.4 | 71.4 | N.M. | 108.3 |
| | Celestica** | CLS US | 3 | 4 | 42.7 | 54.9 | 62.8 | 53.0 |
| | Dell | DELL US | 10 | 17 | 22.5 | 32.2 | 512.5 | 75.5 |
| | HPE | HPE US | 5 | 6 | 26.8 | 32.1 | 228.6 | 31.0 |
| | SMCI | SMCI US | 15 | 15 | 73.5 | 73.6 | 200.7 | 5.0 |

*Source: KGI Research*
**Fiscal year (Jabil FY ends in August; Flex FY ends in March)*
***Consensus*

**Figure 19: Comparison – Major CPU & ASIC server CCL & PCB specifications**

| Company | Chip module | Chip number per rack | Main board type | PCB layer count | PCB build up | PCB supplier | CCL level | CCL supplier |
|---|---|---|---|---|---|---|---|---|
| Nvidia | GB200 NVL72 | 72 | Compute Tray | 22 | 5+12+5 HDI | VGT、UMT | M8 | Doosan |
| | | | Switch Tray | 22 | HLC | WUS | M8 | EMC |
| | B200/300 | | UBB | 18/22 | HLC | WUS,TTM, ISU | M8 | Doosan |
| | | | OAM | 20/20 | 5+10+5 HDI | VGT、UMT | M7 | Doosan |
| AWS | Trainium2/2.5 | 64 | Compute Tray / OAM | 26 | HLC | GCE, Shenyi, WUS, First high-tech, TTM | M8 | EMC |
| | | | OAM | 18 | 5+8+5 HDI | WUS | M7 | Panasonic |
| Google | TPU v6e/v6p | 64 | TPU main board | 22 | HLC | ISU, WUS, TTM, LCS | M7 | EMC |
| Meta | MTIA v2/v3 | 72 | Compute Tray | 30-40 | HLC | GCE, WUS,TTM, ISU | M8 | EMC |
| Microsoft | Maia 100/200 | 32 | Compute Tray | 20+ | HLC | GCE, WUS,TTM, ISU | M7 | EMC |

*Source: KGI Research estimates*

**Figure 20: Valuation table**

| Sector | Company | Ticker | Market cap. (US$mn) | Share price (LCY) | Rating | Target Price (LCY) | EPS (LCY) | | EPS YoY (%) | | PE (x) | | PB (x) | | ROE (%) | | Cash yield (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 2025F | 2026F | 2025F | 2026F | 2025F | 2026F | 2025F | 2026F | 2025F | 2026F | 2024 | 2025F |
| ODM | Hon Hai | 2317 TT | 76,736 | 161.0 | Outperform | 219.0 | 12.85 | 14.46 | 16.8 | 12.5 | 12.5 | 11.1 | 1.3 | 1.2 | 10.6 | 11.3 | 3.6 | 4.2 |
| | Inventec | 2356 TT | 5,373 | 43.7 | Neutral | 45.0 | 2.29 | 2.70 | 13.1 | 17.8 | 19.1 | 16.2 | 2.1 | 2.1 | 11.4 | 13.1 | 3.9 | 4.2 |
| | Quanta | 2382 TT | 36,712 | 277.0 | Outperform | 338.0 | 17.75 | 19.82 | 14.6 | 11.6 | 15.6 | 14.0 | 4.5 | 4.3 | 29.9 | 31.4 | 4.7 | 5.1 |
| | Wistron | 3231 TT | 11,721 | 118.0 | Outperform | 150.0 | 8.47 | 9.95 | 38.6 | 17.5 | 13.9 | 11.9 | 2.3 | 2.0 | 17.3 | 18.1 | 3.2 | 4.3 |
| | Wiwynn | 6669 TT | 15,655 | 2,455.0 | Outperform | 3,070.0 | 186.28 | 207.66 | 47.2 | 11.5 | 13.2 | 11.8 | 4.4 | 3.8 | 36.1 | 34.3 | 3.0 | 4.2 |
| | Gigabyte Tech | 2376 TT | 6,264 | 272.5 | Outperform | 348.0 | 20.29 | 23.16 | 35.0 | 14.1 | 13.4 | 11.8 | 3.2 | 3.1 | 24.7 | 26.8 | 4.3 | 6.0 |
| | Asustek Computer | 2357 TT | 16,133 | 633.0 | Outperform | 808.0 | 53.29 | 54.46 | 26.1 | 2.2 | 11.9 | 11.6 | 1.7 | 1.6 | 14.4 | 14.3 | 5.4 | 6.7 |
| | Asrock | 3515 TT | 1,151 | 271.5 | Outperform | 302.0 | 17.12 | 23.18 | 62.4 | 35.4 | 15.9 | 11.7 | 3.2 | 2.9 | 21.5 | 25.9 | 2.0 | 3.2 |
| Socket/ Connector/cable | Lotes | 3533 TT | 5,125 | 1,340.0 | Outperform | 1,515.0 | 84.21 | 112.31 | 1.7 | 33.4 | 15.9 | 11.9 | 3.7 | 3.2 | 24.7 | 28.8 | 3.1 | 3.1 |
| | Bizlink Holding | 3665 TT | 5,600 | 861.00 | Outperform | 788.00 | 37.97 | 45.74 | 49.4 | 20.5 | 22.7 | 18.8 | 3.8 | 3.4 | 18.4 | 19.2 | 1.4 | 2.1 |
| | Aces | 3605 TT | 296 | 56.30 | Outperform | 75.00 | 4.74 | N.A. | 88.8 | N.A. | 11.9 | N.A. | 1.2 | N.A. | 10.5 | N.A. | 1.3 | 2.5 |
| | Argosy * | 3217 TT | 459 | 148.5 | Not rated | N.A. | 12.44 | N.M. | 10.6 | N.A. | 11.9 | N.A. | N.A. | N.A. | N.M. | N.M. | 5.9 | N.A. |
| | Alltop | 3526 TT | 517 | 231.5 | Outperform | 295.0 | 17.91 | 22.69 | 6.6 | 26.7 | 12.9 | 10.2 | 3.9 | 3.9 | 30.4 | 38.4 | 7.1 | 7.8 |
| Rail kit | King Slide Works | 2059 TT | 6,621 | 2,025.0 | Outperform | 2,530.0 | 92.78 | 109.70 | 43.6 | 18.2 | 21.8 | 18.5 | 7.4 | 6.1 | 37.4 | 36.5 | 1.6 | 2.4 |
| Thermal module | Sunonwealth | 2421 TT | 932 | 99.3 | Outperform | 127.0 | 7.21 | 8.64 | 32.1 | 19.8 | 13.8 | 11.5 | 3.2 | 3.0 | 24.0 | 26.7 | 3.7 | 5.2 |
| | Auras | 3324 TT | 2,000 | 635.0 | Outperform | 791.0 | 33.56 | 49.44 | 58.1 | 47.3 | 18.9 | 12.8 | 5.5 | 4.5 | 31.0 | 38.1 | 1.6 | 2.5 |
| | AVC | 3017 TT | 10,149 | 762.0 | Outperform | 820.0 | 36.26 | 45.78 | 71.0 | 26.2 | 21.0 | 16.6 | 8.2 | 6.5 | 43.2 | 43.4 | 1.3 | 2.3 |
| | Kaori | 8996 TT | 767 | 244.5 | Outperform | 282.0 | 10.10 | 14.58 | 54.1 | 44.3 | 24.2 | 16.8 | 6.2 | 5.1 | 26.2 | 33.2 | 1.6 | 2.5 |
| Heat spreader | Jentech* | 3653 TT | 7,135 | 1,455.0 | Not rated | N.A. | 42.46 | N.M. | 75.8 | N.A. | 34.3 | N.A. | N.A. | N.A. | N.M. | N.M. | 1.0 | N.A. |
| BBU | Simplo Tech | 6121 TT | 2,310 | 364.0 | Outperform | 465.0 | 31.00 | 32.98 | 7.4 | 6.4 | 11.7 | 11.0 | 1.8 | 1.7 | 15.3 | 15.5 | 5.6 | 6.0 |
| | AES-KY | 6781 TT | 3,034 | 1,035.0 | Outperform | 1,085.0 | 40.66 | 49.75 | 60.2 | 22.3 | 25.5 | 20.8 | 5.4 | 4.8 | 22.3 | 24.3 | 1.2 | 2.0 |
| Chassis | Chenbro | 8210 TT | 2,080 | 501.0 | Outperform | 600.0 | 25.65 | 34.31 | 59.8 | 33.8 | 19.5 | 14.6 | 7.0 | 5.7 | 39.0 | 42.6 | 1.5 | 2.6 |
| BMC | Aspeed Tech* | 5274 TT | 6,495 | 5,005.0 | Not rated | N.A. | 89.70 | 116.14 | 31.8 | 29.5 | 55.8 | 43.1 | 29.3 | 23.0 | 55.4 | 57.9 | 1.0 | 1.3 |
| Silcon photonics | Land Mark Opto | 3081 TT | 1,003 | 316.0 | Neutral | 300.0 | 5.21 | 10.02 | N.M. | 92.1 | 60.6 | 31.6 | 7.7 | 7.5 | 12.8 | 24.1 | 0.2 | 1.6 |
| CCL | Iteq* | 6213 TT | 1,044 | 83.8 | Not rated | N.A. | 4.56 | 5.62 | 101.9 | 23.2 | 18.4 | 14.9 | 1.4 | 1.4 | 8.6 | 8.9 | 2.1 | 3.6 |
| | Elite Material | 2383 TT | 10,721 | 901.0 | Outperform | 1,000.0 | 41.52 | 49.44 | 49.3 | 19.1 | 21.7 | 18.2 | 8.2 | 6.6 | 39.3 | 40.2 | 1.9 | 2.8 |
| ABF | Unimicron Tech | 3037 TT | 6,036 | 115.0 | Outperform | 114.0 | 5.64 | 9.52 | 68.7 | 68.9 | 20.4 | 12.1 | 1.8 | 1.7 | 9.0 | 14.3 | 1.3 | 2.0 |
| PCB | Gold Circuit | 2368 TT | 5,029 | 298.0 | Outperform | 330.0 | 16.28 | 20.33 | 41.1 | 24.9 | 18.3 | 14.7 | 6.1 | 4.9 | 35.4 | 37.3 | 2.0 | 2.7 |
| Power | Delta | 2308 TT | 40,108 | 450.0 | Outperform | 520.0 | 17.25 | 20.79 | 27.2 | 20.5 | 26.1 | 21.6 | 4.5 | 4.0 | 18.2 | 19.5 | 1.6 | 2.0 |
| | Lite-On Tech | 2301 TT | 9,158 | 114.0 | Outperform | 134.0 | 6.87 | 7.90 | 31.9 | 15.0 | 16.6 | 14.4 | 2.7 | 2.6 | 16.8 | 18.3 | 4.0 | 4.4 |
| | Chicony Power | 6412 TT | 1,719 | 125.0 | Not rated | N.A. | 8.06 | 9.79 | (3.3) | 21.5 | 15.5 | 12.8 | 3.0 | 3.2 | 19.6 | 25.0 | 4.8 | 4.8 |
| | AcBel Polytech* | 6282 TT | 744 | 25.3 | Not rated | N.A. | N.M. | N.M. | N.M. | N.A. | N.A. | N.A. | N.A. | N.A. | N.M. | N.M. | 0.0 | N.A. |
| Foundry | TSMC | 2330 TT | 960,999 | 1,080.0 | Outperform | 1,105.0 | 57.23 | 66.97 | 26.5 | 17.0 | 18.9 | 16.1 | 5.3 | 4.3 | 31.0 | 29.5 | 1.6 | 1.9 |
| Design service | Alchip Tech | 3661 TT | 8,922 | 3,220.0 | Outperform | 4,020.0 | 75.82 | 129.66 | (6.8) | 71.0 | 42.5 | 24.8 | 6.0 | 5.1 | 14.5 | 21.7 | 1.2 | 1.2 |
| | Global Unichip | 3443 TT | 5,794 | 1,260.0 | Neutral | 1,340.0 | 29.95 | 35.80 | 16.3 | 19.5 | 42.1 | 35.2 | 12.8 | 10.9 | 32.8 | 33.5 | 1.3 | 1.5 |
| Networking | Luxnet | 4979 TT | 894 | 185.0 | Outperform | 225.0 | 8.97 | N.A. | 137.0 | N.A. | 20.6 | N.A. | 5.8 | N.A. | 31.4 | N.A. | 0.8 | 2.2 |
| | Accton Tech | 2345 TT | 14,209 | 738.0 | Outperform | 950.0 | 37.99 | N.A. | 76.8 | N.A. | 19.4 | N.A. | 9.0 | N.A. | 51.8 | N.A. | 1.6 | 2.8 |

*Source: Company data; Bloomberg; 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.**

**Disclaimer**

All the information contained in this report is not intended for use by persons or entities located in or residing in jurisdictions which restrict the distribution of this information by KGI Asia Limited ("KGI") or an affiliate of KGI. Such information shall not constitute investment advice, or an offer to sell, or an invitation, solicitation or recommendation to subscribe for or invest in any securities or investment products or services nor a distribution of information for any such purpose in any jurisdiction. In particular, the information herein is not for distribution and does not constitute an offer to sell or the solicitation of any offer to buy any securities in the United States of America, or to or for the benefit of United States persons (being residents of the United States of America or partnerships or corporations organised under the laws of the United States of America or any state, territory or possession thereof). All the information contained in this report is for general information and reference purpose only without taking into account of any particular investor's objectives, financial situation or needs. Such information is not intended to provide professional advice and should not be relied upon in that regard.

Some of KGI equity research and earnings estimates are available electronically on www.kgi.com.hk. Please contact your KGI representative for information. The information and opinions in this report are those of KGI internal research activity. KGI does not make any representation or warranty, express or implied, as to the fairness, accuracy, completeness or correctness of the information and opinions contained in this report. The information and opinions contained in this report are subject to change without any notice. No person accepts any liability whatsoever for any loss however arising from any use of this report or its contents. This report is not to be construed as an invitation or offer to buy or sell securities and/or to participate in any investment activity. This report is being supplied solely for informational purposes and may not be redistributed, reproduced or published (in whole or in part) by any means for any purpose without the prior written consent of KGI. Members of the KGI group and their affiliates may provide services to any companies and affiliates of such companies mentioned herein. Members of the KGI group, their affiliates and their directors, officers and employees may from time to time have a position in any securities mentioned herein.