Computex 2024

AI to accelerate the future; liquid cooling to bring energy green

Key message

1. Computex 2024 focused on "Connecting AI". The keynote and forum gathered a record number of CEOs, amidst a global AI craze.
2. Most CPU and GPU makers announced their AI chip roadmap and emphasized performance and energy saving product attributes.
3. AI servers dominated exhibitions with every ODM showing off GB200 rack designs with liquid cooling solutions. Shipment commencement in 4Q24F will fuel sales growth for the supply chain.
4. AI PCs were another spotlight at Computex with several NB model introductions with Copilot+ capable systems; to benefit brands, memory and power supply plays.

Event

Computex 2024 was held during June 4-7 in Taipei, with "Connecting AI" as the main theme. The keynotes and forums gathered a record number of CEOs, amidst a global AI craze.

Impact

AI front & center. AI has become a core global technology across multiple industries. While Microsoft (US) invested a substantial US$10bn in OpenAI, it also comprehensively integrated AI into its products, workforce and data management strategies to establish an early lead in this domain. Amazon (US) and Google (US) also followed suit, launching a series of AI application services. From their keynote speeches and in forums during Computex, Nvidia (US), AMD (US) and Intel (US) all elaborated upon their cloud AI platform and generation launch schedules. Nvidia will launch the Blackwell GPU in 2024F, Blackwell Ultra in 2025F and Rubin in 2026F, and AMD will launch the MI325X, MI350X and MI400X over the same time period. AI PC CPU offerings were also highlighted, such as AMD's Ryzen AI 300 CPU (Strix Point with 50 TOPS) and Qualcomm's Snapdragon X Elite CPU (45 TOPS). After Intel launched Meteor Lake CPU last (NPU/GPU/CPU 34 TOPS) year, it will launch Lunar Lake CPU this year with 120 TOPS. Nvidia also showcased RTX AI PC, which uses its consumer GPUs to execute AI functions. Mediatek (2454 TT, NT$1,290, NR) demonstrated the capacities of its AI technology.

Focus on GB200 rack designs with liquid cooling solutions. Several server brands and ODMs showcased GB200 server racks at their booths, including Supermicro Computer (US), Ingrasys (a subsidiary of Hon Hai Group (2317 TT, NT$177.5, OP)), Wiwynn (6669 TT, NT$2,620, OP), Inventec (2356 TT, NT$53.7, OP) Asustek (2357 TT, NT$493, OP), and Gigabyte (2376 TT, NT$317.5, OP), Asrock (3515 TT, NT$232, N). GB200 server racks showcased included 72 GPUs and 36 CPUs designs, and most companies also showcased liquid cooling solutions integrated with the racks. Liquid cooling solution exhibitions at Ingrasys' booth included cold plate modules, rack manifolds, in-rack and in-row coolant distribution units (CDU), quick dis-connectors (QD), and sidecars. Ingrasys will integrate all these parts in GB200 NVL72 rack designs, which will start shipping from 4Q24F. We also saw Koari's (8996 TT, NT$479.5, OP) manifolds and CDU designs at the firm's booth, which extolled the virtues of vacuum brazing technology, a more advanced method for manifold manufacturing than traditional gas tungsten arc welding (GTAW), due to better control of solder and welding beads. Cooler Master (unlisted) demonstrated its 75kw liquid-to-air sidecar co-design with Asustek. Kenmec Mechanical Engineering (6125 TT, NT$114.5, NR) showcased a sidecar and two-phase immersion cooling solution, Forcecon Technology (3483 TT, NT$196, NR) demonstrated a single-phase immersion cooling tank in 4U 7kw and 25U 60kw designs, and Chenming Electronic Technology Corp (3013 TT, NT$91.1, NR) showcased its liquid cooling rack designs featuring manifold and QD collaborations with Koari and Lotes (3533 TT, NT$1,580, OP).

AI PCs another bright spot. Microsoft defines a Copilot+ PC-capable system as a PC with greater than 40 TOPS processing power, bundled with at least 16GB RAM and a 512GB SSD. AMD (US) showcased Ryzen AI 300 processors, codenamed Strix Point, and expects them to be adopted by branded models for launches in 3Q24F. Intel (US) announced the 48 TOPS processor Lunar Lake will debut in September. In addition, Qualcomm introduced the Snapdragon X Elite chip in May, to be adopted by branded models slated for launch from June 18, with Acer (2353 TT, NT$50.6, NR), Asustek, Dell (US), HP (US), Lenovo (US), Microsoft and Samsung (KR) being the first batch. Various brands demonstrated AI PC functions, including Microsoft's Recall and Co-Creator. Besides hardware, branded vendors also provided software solutions, such as StoryCube & MuseTree by Asustek, Acer Sense and Lenovo AI Now. Consumer and commercial AI PCs will come to market soon, with more Copilot+ PC models in 2H24F. An increase of 10-15% in retail price will boost NB brands' sales and profitability, while related component upgrades to power supplies and memory bode well for related manufacturers. We expect the launch of AI PCs the end of support for Windows 10 and a wave of replacement for commercial models will boost PC sector growth in 2024-25F.

Rating

In the AI server and PC supply chain, we favor Hon Hai, Quanta, Wiwynn, Wistron (3231 TT, NT$112, OP), Asustek, AVC (3017 TT, NT$660, OP), Auras (3324 TT, NT$795, OP), Koari, Delta (2308 TT, NT$345, OP), and Chenbro (8210 TT, NT$308.5, OP).

Risks

GPU supply tightness; lower-than-expected AI demand

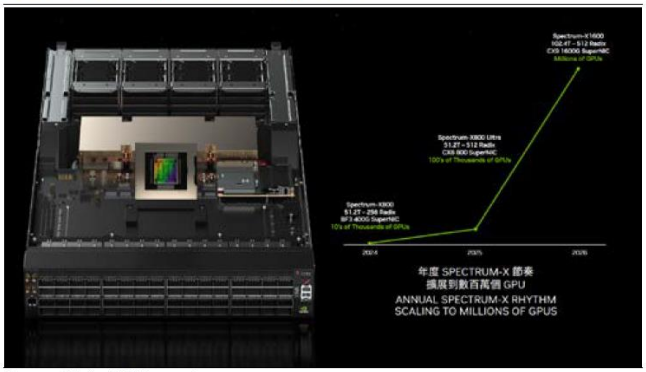
Nvidia (US) Keynote

- Nvidia will maintain its "Data center scale — One-year rhythm — Technology limits — One architecture" strategy. For GPU products, the firm intends to launch one new platform every other year, and make one upgrade to the new platform in the interval year, as with H100 and H200 under the Hopper platform. Regarding the latest Blackwell platform for 2024-25, Nvidia aims to introduce a Blackwell GPU (8S HBM3e) in 2024, followed by an Ultra version (8S HBM3e 12H) in 2025. Blackwell GPUs are designed to be paired with Grace CPUs to power Gen 5 NVLink and CX8 SuperNIC. As for switches, the firm will launch a Spectrum-X800 ethernet switch and Quantum-X800 switch in 2024, followed by a Spectrum Ultra X800 ethernet switch in 2025.
- New products will be launched under the upcoming Rubin platform in 2026-27, including the Rubin GPU (8S HBM4) in 2026 and Rubin Ultra GPU (12S HBM4) in 2027, and Vera CPU to run Gen 6 NVLink, which will offer a data transfer rate of 3,600GB/s, and CX9 SuperNIC (1,600GB/s), along with the launch of the X1600 IB/ethernet switch.
- Spectrum-X is an ethernet platform that Nvidia has created specifically for AI applications. Compared to traditional ethernet architecture, Spectrum-X can boost the performance of generative AI by 60%. Upcoming products to be launched include the Spectrum-X800 in 2024, X800 Ultra in 2025, and X1600 after 2026.
- DGX Blackwell will adopt liquid cooling thermal solutions. Compared to DGX Hopper, the new system provides 9x more NVLink domain, 18x more NVLink bandwidth (TB/s), 45x more AI FLOPS (PF), with just 10x more energy consumption (to 100kW).
- Regarding robotics development, Nvidia is developing two ecosystems: warehouse and factory. Using Giant (9921 TT, NT$224.5, OP) as an example for the warehouse ecosystem, a system integrator may connect edge AI and robotics with the products and services of software or ODM firms through SDK and API. Underlying technical support will be provided by Nvidia. The factory ecosystem, exemplified by Foxconn Industrial Internet, (CN) uses digital twin technology and allows a system integrator to put together cutting-edge applications, equipment and data with Nvidia's Omniverse and AI technologies in order to optimize factory and production line layouts.

Figure 1: Nvidia AI chip roadmap

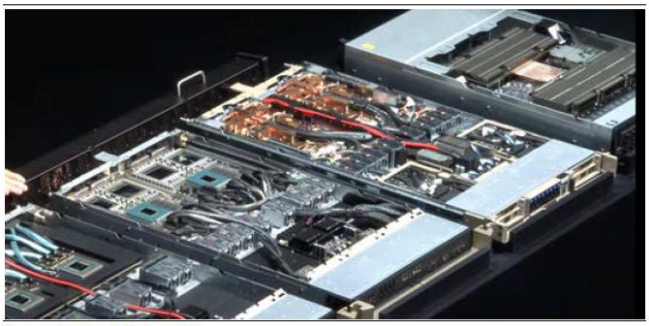
Source: Nvidia; KGI Research

Figure 2: Nvidia Spectrum-X is an ethernet platform specifically built for AI applications



Source: Nvidia; KGI Research

Figure 3: DGX Blackwell will adopt liquid cooling thermal solutions

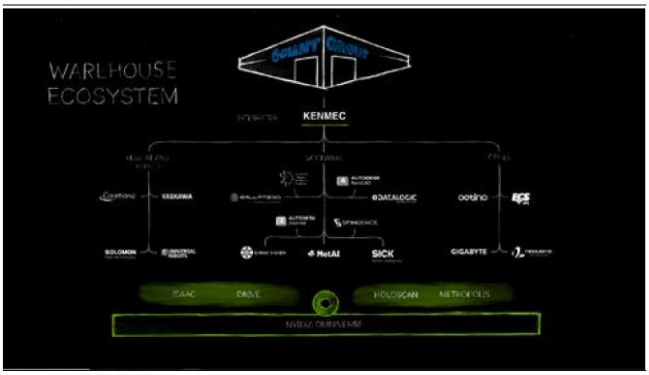


Source: Nvidia; KGI Research

Figure 4: Nvidia introduced several RTX AI computer models at its keynote speech

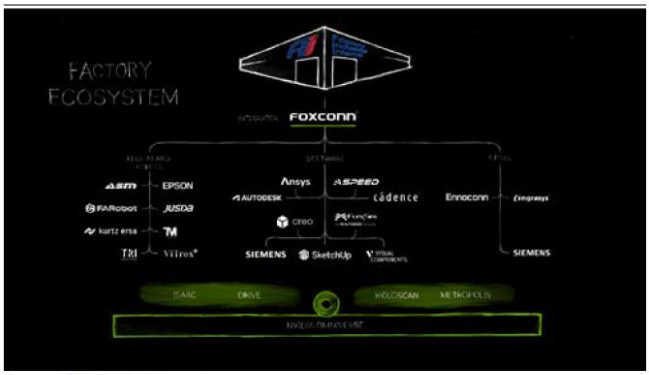
Source: Nvidia; KGI Research

Figure 5: Schematic of warehouse system using Giant as an example



Source: NVIDIA; KGI Research

Figure 6: Schematic of factory ecosystem using Foxconn Industrial Internet as an example



Source: NVIDIA; KGI Research

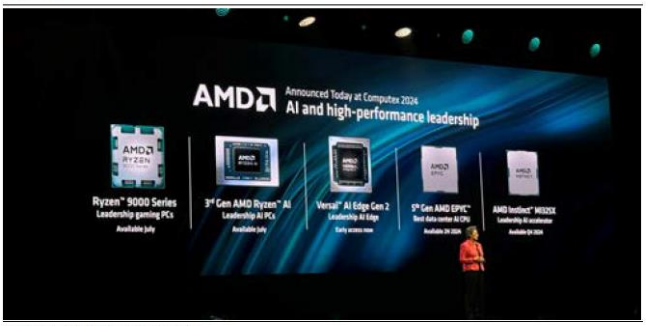
AMD (US) Keynote

- AMD unveiled Granite Ridge, a desktop processor in the Ryzen 9000 series. The series adopts the latest Zen 5 core and AM5 platform, and supports PCIe 5.0 and DDR5. AMD also introduced its flagship model Ryzen 9 9950X at the event. The Ryzen 9000 series will be launched officially in July 2024, and it will come with other models, including the Ryzen 9 9900X, Ryzen 7 9700X, and Ryzen 5 9600X.
- AMD also showcased its third-generation Ryzen AI NB processor AI 300, which is code-named Strix Point (9HX 370). The processor comes with the latest Zen 5 CPU, RDNA 35 graphics card, and XDNA 2 NPU to support AI applications. The XDNA 2 NPU provides 50 TOPS computing performance with 8-bit floating point, which is 5x better than its predecessor. In addition, the power efficiency is 2x better. XDNA 2 NPU is the first NPU to support Block FP16, which means it can maintain FP16 compute accuracy, while achieving the same performance as INT8. Many NB brands have announced new models that will come with the Ryzen AI 300. We estimate more than 100 such new models will hit the market after July 2024.
- AMD's EPYC CPU saw its server market share climb from a respective 27% and 31% in 2022-23, to 33% in 1Q24. At its keynote speech, AMD announced the fifth-generation EPCY Turin CPU (3nm). The series will use the same SP5 socket of Genoa, with up to 192 cores and 384 threads. The performance of Turin will be 3.9x better than Intel's (US) Xeon in summarizing text, 5.4x better in chat bot, and 2.5x better in translation applications. The official launch of Turin is scheduled for 2H24.
- Regarding the roadmap for server GPU, AMD will launch CDNA 3 MI325X in 4Q24, with the same architecture as MI300X, HBM3E, memory capacity of 288GB (2x H200) and bandwidth of 6TB/s (1.3x H200), and 1.3x higher peak theoretical FP16 and FP8

compared to H200. A server with MI325X will be able to process a model with 1.0tn parameters, which is also 2x H200.

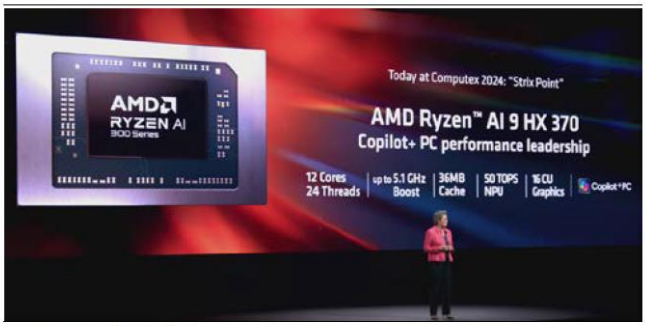
- In 2025, AMD intends to launch CDNA 4 MI350X, which will be built via 3nm, with 288GB HBM3E memory for low floating point (FP4 and FP6) computing. Using the same architecture as MI300X, the inference performance of MI350X will be 35x better than CDNA 3. For 2026, AMD plans to roll out CDNA Next MI400X, which will come with new architecture to deliver better GPU performance and efficiency.

Figure 7: AI products were highlight of AMD's keynote speech



Source: AMD; KGI Research

Figure 8: AMD introduced Ryzen AI 300 series processor for AI PC



Source: AMD; KGI Research

Figure 9: AMD unveiled fifth-generation EPYC server CPU, code-named Turin

Source: AMD; KGI Research

Figure 10: AMD's server GPU roadmap; new chip will be launched every year

Source: AMD; KGI Research

Figure 11: AMD MI300X provides better performance than Nvidia's H100 in large AI accelerator model

Source: AMD; KGI Research

Figure 12: AMD MI325X boasts better memory capacity, bandwidth & parameter processing capacity than Nvidia's H200 GPU

Source: AMD; KGI Research

Intel (US) Keynote

- At the Computex 2024 keynote, Intel CEO Pat Gelsinger unveiled advancements in the company's product lineup and AI capabilities. Intel introduced the Xeon 6 processor, designed to modernize data centers with exceptional performance per watt and energy efficiency. With 144 cores, Xeon 6 significantly outperforms its predecessor, Xeon 2, which has only 28 cores. This enhancement allows a single Xeon 6 rack to replace multiple Xeon 2 racks, offering substantial space savings and the ability to transcode 144 videos simultaneously, a 4.2x performance gain. Additionally, upgrading to Xeon 6 can save substantial energy, exemplified by a potential 80,000 MWh annual electricity use reduction for a 200-rack data center. Future iterations, like the second generation of Xeon 6, with 288 cores, promise even greater performance. Customer feedback from companies like eBay and SAP underscored the performance and energy efficiency improvements offered by Xeon 6.

- In the realm of AI, Intel's Xeon processors, combined with AI accelerators like Gaudi, offer powerful and cost-effective solutions for generative AI training and inferencing. The Gaudi 2 and Gaudi 3 AI kits are designed to provide high performance at competitive prices. Gaudi 3, in particular, showcases 40% faster training times and 2.3x performance per dollar in throughput compared to competitors like the H100. These accelerators support scalable open industry standards and open-source frameworks, making them versatile for a range of applications, including medical imaging and real-time database environments. Gelsinger stated that customers like Naver (KR), Bharti Airtel (IN), Bosch (DE), and Infosys (IN) are adopting Gaudi.

- Intel's Lunar Lake processors are set to revolutionize the AI PC market. By 2028, 80% of PCs are expected to be AI-capable, and Lunar Lake processors are at the forefront of this transformation. Featuring advanced security, extended battery life, and up to 48 TOPS of NPU performance, Lunar Lake processors will power over 80 SKUs of new AI-enhanced PCs, and will start shipping in volume in 3Q24F. The Lunar Lake processor, developed in collaboration with TSMC, promises significant improvements in performance and power efficiency. Lunar Lake's architecture delivers up to 40% lower power consumption than its predecessor, Meteor Lake, and includes enhanced capabilities for on-device AI experiences.

- Intel's commitment to innovation continues with the upcoming Arrow Lake and Panther Lake processors. Arrow Lake, set for release in 4Q24F, aims to bring AI capabilities across all PC categories, starting with desktops. Panther Lake, anticipated to debut at Computex 2025, represents Intel's next leap in technology, being the first on the 18A process node. This marks a return to process leadership for Intel, promising significant advancements in performance and efficiency.

- Intel's keynote at Computex 2024 highlighted its leading role in driving AI and computing advancements. From the high-performance Xeon 6 processors to the AI-optimized Gaudi accelerators and the next-generation Lunar Lake, Arrow Lake, and Panther Lake processors, Intel is poised to shape the future of computing with innovative, energy-efficient, and AI-capable technologies.

Figure 13: Intel expects over 40 million units of CPUs with NPUs will be shipped by end-2024

Source: Intel; KGI Research

Figure 14: Performance of Lunar Lake processor

Source: Intel; KGI Research

Figure 15: Over 80 SKUs of AI PC designed by 20+ OEMs are expected to be in mass production in 3Q24F

Source: Intel; KGI Research

Figure 16: Intel's PC product road map

Source: Intel; KGI Research

Supermicro (US) Keynote

- DLC (Direct Liquid Cooling) can operate with room temperature water, saving up to 3% on initial capex and up to US$60mn in opex over 5 years, while reducing CO2 emissions and allowing smaller air conditioners.
- Supermicro aims for 15% DLC market share in new data center deployments in 1 year versus less than 1% currently, and DLC penetration may reach 30% in the following year. Thus, the firm thinks current DLC production of 1,000 per month will soon expand to 2.000 per month.
- DLC lead time will be around in 2-4 weeks to the clients, offering higher performance, solid uptime, and no premium on infrastructure and hardware acquisition costs.
- Switching from air cooling to DLC can save 33% power and up to 40% energy, while doubling computing density and supporting the highest TDP GPU and rack power densities
- Currently, Supermicro can ship 50 DLC racks daily. With DLC capacity of 1,000 per month, it expects to ship over 10k racks in 2024F at an ASP price of US$3mn per rack.
- Malaysia capacity for rack assemblies is 5,000 racks per month; to be expanded on strong AI server demand outlook.

Figure 17: DLC provides up to 0-3% net savings on initial capex

Source: Supermicro; KGI Research

Figure 18: DLC will have lower data center power requirements and CO2 emission, saving up to US$60mn in opex over 5 years

Source: Supermicro; KGI Research

Figure 19: Supermicro expects DLC market share will surge in the coming year

TSource: Supermicro; KGI Research

Figure 20: Supermicro's DLC solutions overcame bottlenecks

Source: Supermicro; KGI Research

Computex – Cloud AI

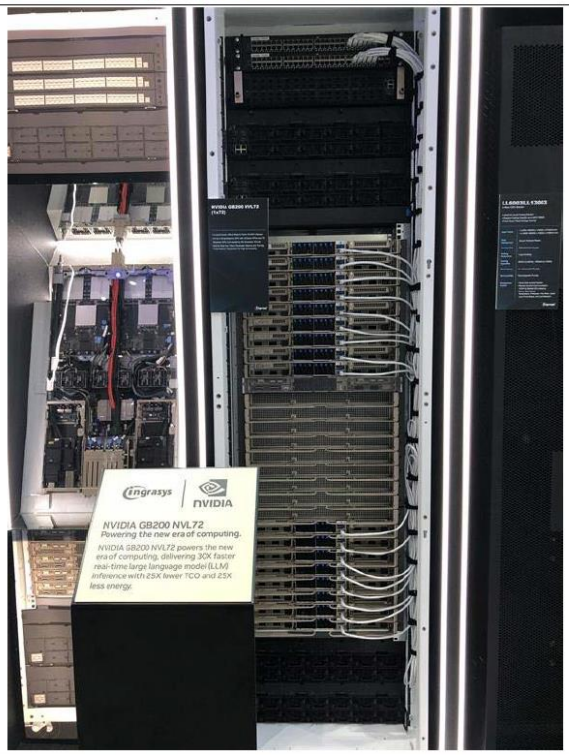
Ingrasys (subsidiary of Hon Hai (2317 TT, NT$177.5, OP))

- Ingrasys showcased AI servers, including Nvidia's GB200 NVL36 and 72 models, an H200 MGX design, a HGX design, and AI servers on AMD's MI300X and Intel's Gaudi 3 platforms.
- The firm's GB200 NVL36 is a 2U design and the NVL72 is in 1U design, both with liquid cooling solutions. The NVL36 has a TDP of 70kw while the NVL72 has a TDP of 143kw. Ingrasys expects this design to begin shipping by end of 2024F.
- The firm demonstrated its liquid cooling solutions for AI server racks, with sidecars for liquid-to-air solutions with RPU, and in-row CDUs for liquid-to-liquid solutions. It also showcased an in-rack CDU within a GB200 NVL72. Manifold and universal QD (UQD) were also visible on the back of the GB200 server racks.
- The liquid-to-air sidecar has 16 cooling fans on the fan door with RDHx, which can dissipate more than 70kw of heat in an NVL36 rack. The in-row CDU has a liquid-to-liquid cooling design supporting 8-10 GB200 NVL72 racks, with total heat dissipation capability up of 1,300kw, with a smaller 600kw unit designed for four NVL72 racks.
- GH200/GB200 MGX designs were also highlighted, with single and doubled Hopper/Blackwell GPUs, and Grace CPUs in 1U and 2U designs. These modular designs are more attractive to enterprises. 1U designs have nine cooling fans and 2U designs have six, offered in partnership with Foxconn Tech (2354 TT, NT$68.5, NR). Ingrasys also showcased an MGX 4U HGX design with better flexibility, as it is a PCIe rather than a universal baseboard (UBB) design.

Figure 21: Ingrasys' booth with complete GB200 NVL 36/ 72 AI server demonstrations in DLC designs

Source: Ingrasys; KGI Research

Figure 22: Ingrasys' GB200 NVL72 rack



Source: Ingrasys; KGI Research

Figure 23: Ingrasys' GB200 with cold plate

Source: Ingrasys; KGI Research

Figure 24: Ingrasys's side car solution (front)

Source: Ingrasys; KGI Research

Figure 25: Ingrasys's side car solution (rear)

Source: Ingrasys; KGI Research

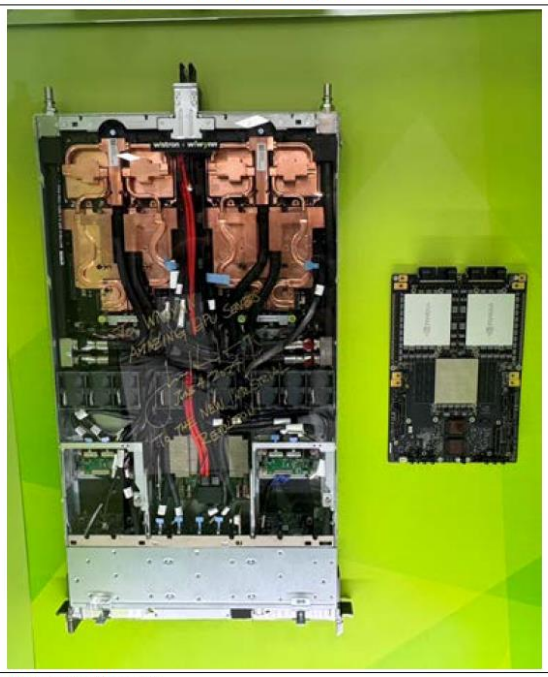
Wiwynn (6669 TT, NT$2,620, OP)

- Wiwynn had several AI server demonstrations in its booth, including GB200 NVL72 server racks, liquid-to-air direct liquid cooling (DLC) solutions, two-phase immersion cooling tanks, and air cooling AI server designs with H200, MI300X and Gaudi 3 GPUs.
- Wiwynn's GB200 server rack has completed the design phase, and the firm is cooperating with parent company Wistron. Wistron provides motherboards to Wiwynn, and Wiwynn ships server racks to the CSP clients.
- The firm's DLC solution was showcased, with Zutacore (US) as the cold plate design partner, Parker and Staubli (US) as the QD and tube supplier, and Lite-On IT as the power shelf and busbar design partner. The DLC solution aims to deliver 30x faster real-time large language model (LLM) inferencing with 25x lower total cost of ownership (TCO), while using 25x less energy.
- HGX H100 AI server designs were also shown at the booth, with an eight GPU design for H100, H200, B100 & B200 in DLC and air-cooled solutions. AMD's MI300x GPU and Intel's Gaudi 3 GPU were also shown in 8U server designs.
- Wiwynn highlighted single-phase and two-phase immersion cooling tanks with heat exchangers. The dielectric liquid for the single-phase solution is mineral oil, while the two-phase tank uses fluorinated liquid.

Figure 26: Wiwynn's GB200 rack

Source: Wiwynn; KGI Research

Figure 27: Wiwynn's GB200 compute board



Source: Wiwynn; KGI Research

Figure 28: Wiwynn's 2-PDLC (two-phase liquid cooling) cold plate

Source: Wiwynn; KGI Research

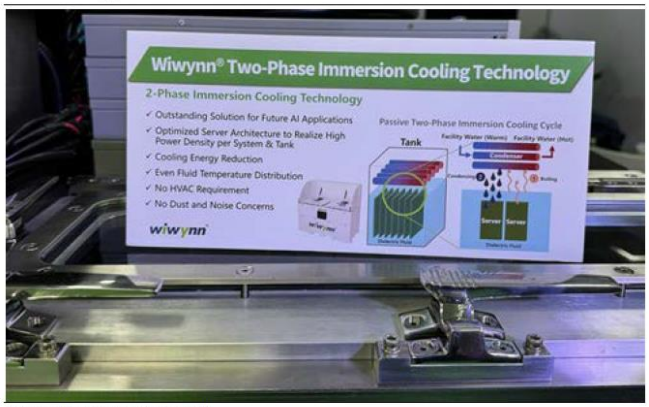
Figure 29: Wiwynn's HGX H100 solution

Source: Wiwynn; KGI Research

Figure 30: The server power density and chip thermal design power matrix of Wiwynn's solutions

Source: Wiwynn; KGI Research

Figure 31: Wiwynn's two-phase immersion cooling technology



Source: Wiwynn; KGI Research

Figure 32: Wiwynn showcased Intel's Gaudi 3 GPU in a server build

Source: Wiwynn; KGI Research

Figure 33: Wiwynn's AMD MI300x GPU server

Source: Wiwynn; KGI Research

Figure 34: Wiwynn's two-phase immersion tank

Source: Wiwynn; KGI Research

Figure 35: Wiwynn's single-phase immersion tank

Source: Wiwynn; KGI Research

Gigabyte (2376 TT, NT$317.5, OP)

- Gigabyte showcased AI server product lines, including designs with Nvidia B100 / B200 GPUs, the GB200 NVL72, as well as AMD's MI300X GPUs.
- The firm's GB200 solution will be two GB200 NVL36 racks with liquid cooling designs. Total power consumption will be 66kw per NVL36 rack, with internal busbars.
- The firm showcased busbar power supply designs for AI server HGX B200 racks. The firm's liquid cooling solution was co-designed with several vendors, including CoolIT System (CAN) and Motivair (US) for CDU and manifolds, and nVent (US) for sidecars and RDHx.
- The firm's single-phase immersion cooling solution can dissipate 40Kw of heat with a 12U design and coolant oil from Shell (US). European and US clients are more interested in adopting immersion cooling solutions.

Figure 36: Gigabyte showcased a GB200 server design

Source: Gigabyte; KGI Research

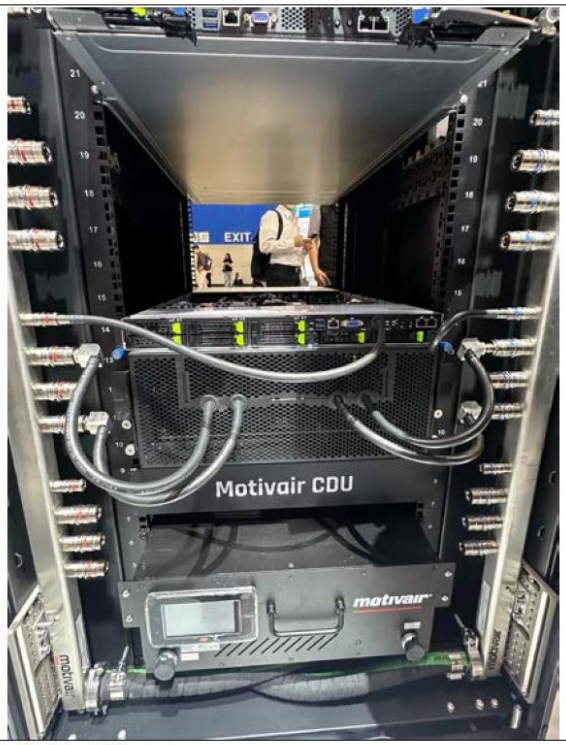
Figure 37: Gigabyte partnered with CoolIT and Motivair in direct liquid cooling solutions

Source: Gigabyte; KGI Research

Figure 38: Gigabyte's fan door design for sidecars

Source: Gigabyte; KGI Research

Figure 39: Gigabyte adopted Motivair's CDU as a liquid cooling solution



Source: Gigabyte; KGI Research

Figure 40: Gigabyte showcased an ORv3 rack design

Source: Gigabyte; KGI Research

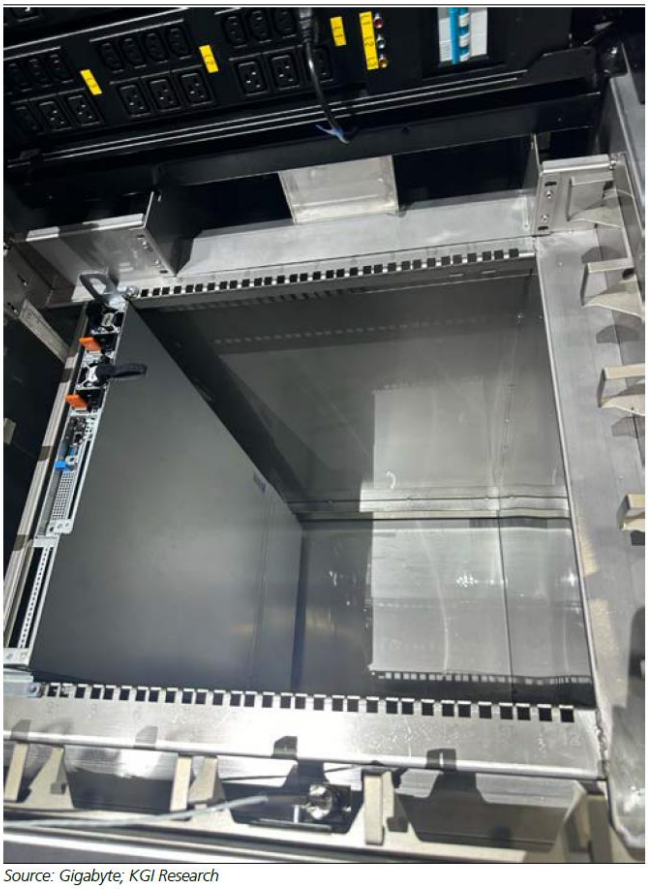
Figure 41: Gigabyte's AMD MI300X AI server

Source: Gigabyte; KGI Research

Figure 42: Gigabyte's immersion cooling tank

Source: Gigabyte; KGI Research

Figure 43: Immersion cooling tank interior



Source: Gigabyte; KGI Research

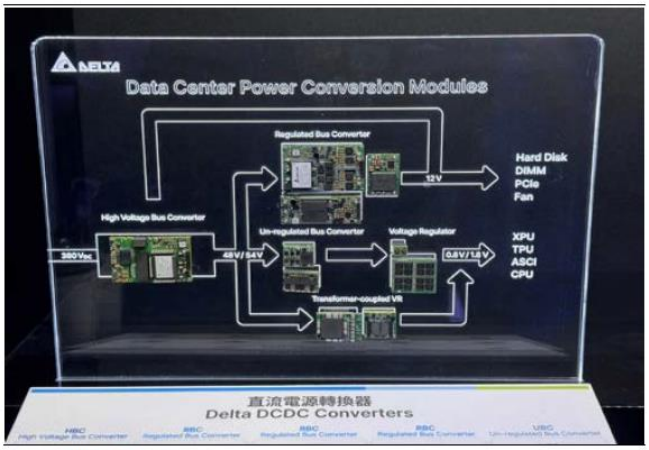
Delta (2308 TT, NT$345, OP)

- A NVL72 rack generally adopts 4 power shelves, each at 1U height and equipped with six 5.5kW AC-DC PSUs with power efficiency of 97.5%+.
- Next generation PSUs could be up to 8kW per unit (for Nvidia's Rubin platform), with a power efficiency requirement of up to 98%.
- To achieve higher power density and efficiency, Delta has been adopting wide-bandgap (WBG) materials. SiC penetration is higher than GaN for now.
- Due to limits of power efficiency improvement, Delta is developing liquid cooling technologies for PSU.
- Other than CPUs and GPUs, switches, optical transceivers and HBM are also large power consumers.
- With AI servers upgrading from 12V to 48V systems (especially Blackwell GPUs), spec upgrades for DC-DC power modules and power chokes are expected.
- For thermal solutions, Delta offers one-stop shopping services, with a comprehensive product portfolio including 3D VC, cold plates, manifolds and cooling distribution units (CDU) (both in-rack and in-row), all of which leverage the firm's in-house production capabilities.
- The firm will introduce a 1.35MW in-row CDU, capable of dealing with heat generated by ten or more NVL72 racks.

Figure 44: Delta exhibits power shelves for AI servers

Source: Delta; KGI Research

Figure 45: Power conversion modules for data centers



Source: Delta; KGI Research

Figure 46: DC-DC power converters and chokes are upgraded on AI servers migrating from 12V to 48V

Source: Delta; KGI Research

Figure 47: Delta offers cold plate and thermal fan solutions for GB200 servers

Source: Delta; KGI Research

Figure 48: The firm is developing liquid cooling technologies for PSUs

Source: Delta; KGI Research

Kaori (8996 TT, NT$479.5, OP)

- Demo products: plate heat exchangers, manifolds, CDU (in-rack and in-row).
- Kaori believes its core technology vacuum brazing process to be much more advanced for manifold manufacturing than traditional gas tungsten arc welding (GTAW), due to better control of solder and welding beads, helping the firm stand out against the competition.
- The firm shipped CDU to Alibaba (CN) in 2019 with total volume of 13MW. Both in-rack and in-row CDU are under customer qualification, with in-row CDU shipments expected in 4Q24F.
- For CDU, Kaori believes its advantage is in software control capabilities.
- Management targets end-2024 annualized capacity of 50k manifold units, 50k radiator sets, and 2k in-row CDU units.
- Target customers are server brands and EMSs. In order to better service EMS customers, Kaori will be expanding its R&D and sales teams.
- The firm's liquid cooling revenue weighting is close to 10%, and will likely be closer to 20% by end-2024F.

Figure 49: The vacuum brazing (right) process offers smoother welding surfaces, leading to considerably lower likelihood of fluid leakage compared to traditional GTAW (left)

Source: Kaori; KGI Research

Figure 50: Kaori exhibits a 800kW in-row CDU

Source: Kaori; KGI Research

Bizlink (3665 TT, NT$275.5, OP)

- HPC revenue of approximately US$200mn in 2023, of which data and power cables accounted for 60-65% and 35-40%, respectively.
- Management guides HPC revenue to pick up noticeably from 4Q24F onwards.
- AEC cable: partnering with chip supplier Credo (US), with offering adopted by CSPs including but not limited to Microsoft (US) Maia platform.
- Bizlink has been awarded a reference design for power cables and busbar in Nvidia's GB200. The firm believes there are two other qualified vendors, neither of which is from Taiwan. Management cites that Nvidia puts much more emphasis on power in the GB platform designs, relative to the GH platform.
- Nvidia emphasizes product performance, reliability and service capabilities. Therefore newcomers are unlikely to penetrate the market on price considerations alone.
- The firm targets top-tier US CSPs, but also looks to sell to leading server brands.
- CPO: targets revenue contribution in 2026-27F.
- The firm is expanding capacity in Malaysia and Indonesia to fulfill customer demand.

Chenbro (8210 TT, NT$308.5, OP)

- Chenbro is a reference design partner for Nvidia's GB200 2U chassis. During Computex, Chenbro showcased two types of products for MGX 2U server chassis, one shorter at 76.6cm and the other longer at 90cm. The firm's MGX 4U chassis is under certification now, and the spec has yet to be finalized.
- Both versions of the MGX 2U chassis are modular designs with liquid and air cooling designs. The short version is more popular among CSP clients, and requires a busbar power connection. Enterprise clients prefer the 90cm version, which integrates a power supply within the chassis.
- Chenbro can help client to buy quick connector (QD) through a collaboration with Lotes. Chenbro can also manufacture its own quick connectors which are under certification with a CSP client.
- The firm also has new products, such as military-spec chassis for Hyve (US), which is a 2U 19" chassis with modular storage, power and networking configurations.

Figure 51: Chenbro's MGX 4U product

Source: Chenbro; KGI Research

Figure 52: Chenbro's military standard chassis

Source: Chenbro; KGI Research

Figure 53: Chenbro's MGX 2U chassis – short

Source: Chenbro; KGI Research

Figure 54: Chenbro's MGX 2U chassis – long

Source: Chenbro; KGI Research

Chenming Electronic Technology (3013 TT, NT$91.1, NR)

- Chenming showcased a 42U liquid cooling rack, developed in cooperation with several component manufacturers, including Lotes for quick connectors (QD), Kaori for manifolds and TaiSol (3338 TT, NT$84.4, NR) on cold plates. In addition to supplying chassis and racks, the company hopes to provide customers with total solutions.
- Chenming exhibited OCP (Open Compute Project) 3rd generation (ORv3) rack. The rack is 44OU high, and compatible with 47RU, the future MGX rack, and is currently under certification with clients.
- Single-phase immersion liquid cooling products exhibited by Chenming are 4U 7kw and 24U 80kw respectively. The firm provides CDU and chassis.

Figure 55: Chenming's 42U liquid cooling rack was developed with several component manufacturers

Source: Chenming; KGI Research

Figure 56: Chenming single-phase immersion liquid cooling product

Source: Chenming; KGI Research

Computex – Edge AI / AI PC

ASUS (2357 TT, NT$493, OP)

- ASUS showcased the new ExpertBook P5, featuring the newly-launched Intel Lunar Lake CPU, with performance of over 100 TOPS, nearly three times Meteor Lake's 34 TOPS. Lunar Lake's design integrates memory and the CPU for improved speed, although it lacks expandability, which may be seen in future models such as Arrow Lake. ExpertBook P5 includes ASUS's unique AI features, such as voiceprint recognition for better audio management in meetings, and is set to launch in September 2024. AI PC models will carry a US$100-150 ASP premium compared to normal PCs.
- Microsoft Copilot+ application updates for x86 systems are scheduled for release in 4Q24. Customers who buy ExpertBook P5s can upgrade to newest Copilot application after its launch.
- The ASUS 2024 commercial NB roadmap will prioritize Intel processors, with potential consideration for AMD or Qualcomm based on future customer requirements.
- Another spotlight in the ASUS booth is the newly released ESC AI POD for Nvidia GB200 NVL72 AI servers, capable of inferencing with trillion-level parameters. The showcased cooling solution is a liquid-to-air sidecar that can dissipate 75kW of heat.
- Each ESC AI POD with Nvidia GB200 NVL72 AI server requires two sidecars when using this cooling solution. The sidecars, manifolds, cold plates, and quick connectors (UQC) are all supplied by Cooler Master.
- The ESC AI POD for Nvidia GB200 NVL72 AI servers features a centralized power supply design. It can be paired with a Battery Backup Unit (BBU) (possibly one per rack or shared among several racks) or an Uninterruptible Power Supply (UPS). ASUS is involved in the design of mechanical parts.
- A comprehensive range of AI notebooks powered by AMD Ryzen AI 300s were showcased, including ZenBook S16, VivoBook S14/S15/S16, ProArt P16/PX13, TUF Gaming A14/A16, and ROG's Zephyrus G16.

Figure 57: The compute tray of ASUS' ESC AI POD for the NVIDIA GB200 NVL72

Source: ASUS; KGI Research

Figure 58: ASUS ESC AI POD for the NVIDIA GB200 NVL72

Source: ASUS; KGI Research

Figure 59: ASUS revealed an AMD AI PC product line

Source: ASUS; KGI Research

Figure 60: ASUS showcased the flagship commercial NB ExpertBook P5, featuring the Intel Lunar Lake processor

Source: ASUS; KGI Research

Acer (2353 TT, NT$50.6, NR)

- Acer's AI PC "Swift 14 AI", which is equipped with a Snapdragon X Elite, is aimed at consumers, and will be available in Europe in June at price of US$1,099-1,199.
- Copilot requires an internet connection for functionality, but the "Recall" function is available without the Internet, and content won't be uploaded to the cloud because of privacy concerns.
- Microsoft Paint has been equipped with a "Co-creator" function, which can generate pictures through commands, though it requires an internet connection to function. Live Captions is able to translate audio between 44 languages.
- For AI PCs, Acer developed its own AI software "AcerSense", which has three performance modes for selection, including silent, normal, and performance. In addition, the Acer PurifiedView tab is an edge application, which lets users adjust camera settings effortlessly, with automatic framing, eye contact, and background blur functions.
- AI PCs equipped with Intel's Lunar Lake platform will be available in 2H24F.
- Nvidia GeForce RTX 4090 GPUs in Acer's notebooks, which already perform up to 686 AI TOPS, are able to use ChatRTX to run LLMs on edge devices.
- Altos, an Acer group company, is focused on solutions for keeping data on servers to support edge computing. Altos' aiWorks platform can support Nvidia Multi-Instance GPU (MIG) to divide a GPU into several instances, and each of which is completely independent for different users.

- Acer Medical, an Acer group company, has developed two AI medical devices. VeriSee DR is an AI-powered diabetic retinopathy diagnostic system. VeriOsteo OP is designed to screen for abnormal bone mineral density. This is Taiwan's first intelligent medical device approved by the Food and Drug Administration (FDA).

Figure 61: Acer has launched a new AI PC model "Swift 14 AI", which is equipped with a Snapdragon X Elite

Source: Acer; KGI Research

Figure 62: Acer Medical has developed "VeriSee DR" & "VeriOsteo OP" AI related medical devices

Source: Acer; KGI Research

MSI (2377 TT, NT$186.5, OP)

- MSI showcased AI PC models equipped with Intel Lunar Lake CPUs (48TOPS NPU), which may be launched in 3Q24F.
- MSI's AMR AI Copot-Pro can hold up to 150kg, equipped with Techman Robots' robotics arm. Major applications for MSI's robots are on industrial automation, educational and medical robots.

Figure 63: MSI showcased Prestige 14 AI+ EVO, with Intel Lunar Lake CPU inside

Source: MSI; KGI Research

Figure 64: MSI's AMR (Autonomous Mobile Robot) robot

Source: MSI; KGI Research

Figure 44: CPU & GPU roadmap

	2022	2023	2024	2025	2026
			CPU		
intel	Eagle Stream (Sapphire Rapids CPU, Intel 7)	Eagle Stream (Emerald Rapids CPU, Intel 7)	Birch Stream (Granite Rapids/ Sierra Forest CPU, Intel 3)	Birch Stream (Clearwater Forest CPU, Intel 18A)	
AMD	Zen 4 (Genoa CPU, 5nm)	Zen 4c (Bergamo CPU, 5nm)	Zen 5 (Turin CPU, 3nm)	Zen 6 (Venice CPU, 3nm)	
NVIDIA	BlueField-3 DPU	Grace CPU (ARM-based)	BlueField-4 DPU		Vera CPU
arm	Neoverse V2 (Ampere One-1 CPU, 5nm)	Poseidon (Neoverse V3) (Ampere Siryn CPU, 5nm)	Neoverse N3 (TBA)		
			GPU		
intel	Ponte Vecchio (Xe HPC GPU)	Gaudi 2 (GPU)	Gaudi 3 (GPU)	Falcon Shores (GPU)	
AMD		Instinct MI300 GPU	Instinct MI325X GPU	Instinct MI350X GPU	Instinct MI400X GPU
NVIDIA	Hopper H100 GPU		Blackwell B100/ GB200 GPU	Blackwell Ultra GPU	Rubin GPU

Source: Company data; KGI Research

Figure 65: AI PCs to be aggressively marketed by CPU & NPU giants

Platform	Intel Meteor Lake	Intel Lunar Lake	Intel Arrow Lake	Intel Panther Lake	AMD 7040 Phoenix	AMD 8040 Hawk Point	AMD AI300 Strix Point	AMD 9000 Strix Halo	Qualcomm Snapdragon X Elite
Microarchitecture	Redwood cove	Lion cove	Lion cove	Cougar cove	Zen 4	Zen 4	Zen 5	Zen5	Oryon
Launch Date	4Q23	2H24F	2H24F	2025F	2Q23	1Q24	2H24F	2025F	1H24F
Process	Intel 4	Intel 18A	Intel 20A	Intel 18A	TSMC N4	TSMC N4	TSMC N4	TSMC N4	TSMC N4
AI architecture	Movidius VPU	TBD	TBD	TBD	Ryzen AI	Ryzen AI	Ryzen AI	Ryzen AI	Hexagon
TOPS	34 TOPS	120 TOPS	TBD	TBD	33 TOPS	39 TOPS	TBD	>70 TOPS	45 TOPS
NPU	11 TOPS	48 TOPS	50 TOPS	120 TOPS	10 TOPS	16 TOPS	50 TOPS	50-60 TOPS	45 TOPS
CPU	5 TOPS	5 TOPS							
GPU	18 TOPS	67 TOPS							

Source: Company data; KGI Research

Figure 66: Intel & AMD desktop CPU roadmap

	Intel					AMD			
	Raptor Lake	Raptor Lake refresh	Meteor Lake	Arrow Lake	Panther Lake	Ryzen 4000 (Renoir)	Ryzen 5000 (Vermeer)	Ryzen 7000 (Raphael)	Ryzen 9000 (Granite Ridge)
Time for launch	4Q22	3Q23	2024F	2024F	2025F	1Q20	4Q20	3Q22	2024F
Process (node)	Intel 7 (10nm)	Intel 7 (10nm)	Intel 4 (7nm)	Intel 20A	Intel 18A	TSMC N7	TSMC N7+	TSMC N5	TSMC N3
Microarchitecture (P-Core)	Raptor Cove	Raptor Cove	Redwood Cove	Lion Cove	Cougar Cove	Zen 2	Zen 3	Zen 4	Zen 5
CPU sockets (desktop)	LGA 1700	LGA 1700	LGA 1851	LGA 1851	LGA 1851	AM4 (LGA 1331)	AM4 (LGA 1331)	AM5 (LGA1718)	AM5 (LGA1718)
DRAM	DDR4 / DDR5	DDR4 / DDR5	DDR5 LPDDR5X	DDR5	DDR5	DDR4	DDR4	DDR5	DDR5
PCIe	Gen 5	Gen 5	Gen 5	Gen 5	Gen 5	Gen 4	Gen 3	Gen 5	Gen 5

Source: Company data; KGI Research

Figure 67: Intel & AMD NB CPU roadmap

	Intel						AMD				
	Alder Lake	Raptor Lake	Meteor Lake	Arrow Lake	Lunar Lake	Panther Lake	Ryzen 6000 (Rembrandt)	Ryzen 7000 (Phoenix)	Ryzen 8040 (Hawk Point)	Ryzen AI300 (Strix Point)	Ryzen 9000 (Strix Halo)
Time for launch	1H22	1H23	4Q23	2024F	2H24F	2025F	1Q22	1Q23	4Q23	2024F	2025F
Process (node)	Intel 7 (10nm)	Intel 7 (10nm)	Intel 4 (7nm)	Intel 20A	Intel 18A	Intel 18A	TSMC N6	TSMC N4	TSMC N4	TSMC N4	TSMC N4
Microarchitecture (P-Core)	Golden Cove	Raptor Cove	Redwood Cove	Lion Cove	Lion Cove	Cougar Cove	Zen 3+	Zen 4	Zen 5	Zen 5	Zen 5
DRAM	DDR4 / DDR5	DDR4 / DDR5	DDR5	DDR5	DDR5	DDR5	DDR5	DDR5	DDR5	DDR5	DDR5
PCIe	Gen 5	Gen 5	Gen 5	Gen 5	Gen 5	Gen 5	Gen 4	Gen 5	Gen 5	Gen 5	Gen 5

Source: Company data; 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.