

IT Hardware

Strong demand for Blackwell despite several adjustments

Key message

1. Recent hot discussions in the AI server supply chain include: (1) Nvidia's (US) cancellation of dual NVL36 rack designs; (2) the cancellation of the B200A GPU and plans for the B300A Ultra GPU to launch by end-2025F, (3) socket design to begin for Blackwell Ultra GPUs in 1H25F.
2. NVL36x2 design discontinuation will not affect the assembly industry landscape; B300A and B300 Ultra introductions will be neutral for baseboard and thermal suppliers.
3. GPU socket designs present new opportunities for CPU socket plays.

Event

Recent hot discussion of Blackwell AI servers include: (1) Nvidia (US) will stop producing dual NVL36 racks, and will produce NVL72 for 2025F projects; (2) Nvidia cancelled development of the B200A GPU, and intends to launch next generation AI GPU B300 Ultra (CoWoS-L) and B300A Ultra (CoWoS-S) GPUs in late 2Q25F; (3) socket design will likely start for Blackwell Ultra GPUs in 1H25F.

Impact

GB200 ramp-up on track; NVL36 dual configuration to be discontinued. Nvidia recently ceased developing NVL36 dual racks, and the firm is encouraging users to adopt NVL72 racks, given their higher efficiency. The dual rack NVL36 configuration (NVL36x2) was adopted mainly by Amazon Web Services (AWS; US), with other NVL36 designs being mainly for Meta (US) and Google (US). We expect this change will apply to new projects in 2H25-2026F, and will not impact the supply chain. Hon Hai Precision Industry (2317 TT, NT\$199, Restricted) and Quanta Computer (2382 TT, NT\$283.5, OP) will continue assembling GB200 racks. We maintain our forecast for shipment of 55k GB200 racks in 2025F, and our estimate of a NVL36/72 weighting of 55-60%/ 40-45% in 2025F, while we expect the original design of NVL72 for Microsoft (US) will have higher priority.

B200A GPUs cancelled; new B300 Ultra to be launched in late 2Q25F. Nvidia cancelled development of the B200A GPU, and may launch B300 Ultra (CoWoS-L) and B300A Ultra (CoWoS-S) GPUs in late 2Q25F, followed by the Rubin series in 1H26F. The roadmap is likely to become clearer in 1Q25F. The supply chain expects a new GPU (codenamed B102 Ultra) to enter production in late 2Q25F as a replacement for the B200A, and we think this will likely be called the B300A Ultra. The B300A Ultra is a single die chip with four high bandwidth memory (HBM) modules, like the B200A, and uses TSMC's (2330 TT, NT\$1,045, OP) CoWoS-S process for packaging. Blackwell B300 Ultra will succeed the B200, with an HBM density increase. B300A Ultra NVL36 racks will use air cooling solutions, while the B300 Ultra is likely to be paired with liquid cooling solutions. We don't expect any significant impact on baseboard suppliers like Wistron (3231 TT, NT\$109, OP) or Hon Hai, as they supply baseboards for B200, B300A Ultra, B300 Ultra or Hopper GPUs. Should B200 shipments rise due to the cancellation of the B200A, it will benefit these firms on higher content value per unit. Despite frequent launch schedule adjustments for several products, Blackwell GPU demand remains very strong.

Socket design to begin for Blackwell Ultra GPU in mid-2025F. Nvidia will likely adopt socket designs for GPUs for several reasons. While GPUs did not previously use socket designs, being directly mounted on boards due to transmission performance requirements, failure costs have risen for GB200 units, which have two GPUs and one CPU. Metal thermal interface material (TIM), such as indium, is also able to dissipate heat better than film TIM, but the lower melting point creates challenges for direct mounting. Socket designs could partially solve these problems, and we may receive Nvidia's final decision by the end of 2024F. We expect GPU sockets to be first adopted for the B300 Ultra generation in mid-2025F, at the earliest. CPU socket suppliers FIT (HK, a connector subsidiary of the Hon Hai group), Molex (US), Amphenol (US), TE (US) and Lotes (3533 TT, NT\$1,670, OP) all are aggressive about gaining this GPU socket opportunity, and recently sent samples to Nvidia. This will bring positive impacts for FIT/ Hon Hai and Lotes in the long.

Stocks for Action

On booming AI server demand, investors should revisit the supply chain on currently low valuations, with our top picks being Hon Hai, Quanta, Asia Vital Components (3017 TT, NT\$647, OP) and Delta Electronics (2308 TT, NT\$387.5, OP).

Risks

CSPs' AI server capex plan; AI server GPU supply constraints.

