Morgan Stanley | RESEARCH

UPDATE

October 13, 2026 04:27 PM GMT

Greater China Semiconductors | Global

OCP kick off – Optical, Packaging, Memory and Power in focus

Al infrastructure innovation is shifting toward rack-level co-optimization, creating opportunities in optical interconnects, advanced packaging, memory architectures, and power delivery as agentic Al drives inference demand.

Key Takeaways

- Optical interconnect: CPO is moving toward rack-scale scale-up, while pluggables retain a scale-out advantage.
- Custom ASICs and packaging: MediaTek sees 3.5D gaining traction as 2.5D approaches physical limits
- Memory architecture: SK hynix targets inference capacity and KV-cache bottlenecks through MARS arid HBF.
- Power and cooling: Racks above 200kW accelerate 800V DC, vertical power delivery and liquid cooling.

CPO: Optical scale-up is getting closer to reality. A 64-XPU CPO-enabled compute-rack demonstration highlighted the potential for direct optical links between accelerators. As 200G/lane electrical channels approach physical limitations, optical scale-up will become more important, particularly for customized XPUs. However, pluggagle optics should retain scale-out volume leadership given advantages in serviceability and cost. Packaging yield, fiber alignment, and reliability remain key commercialization challenges.

Advanced packaging: Moving beyond 2.5D. MediaTek highlighted custom ASIC designs approaching 10x reticle size, arguing that 3.5D integration could deliver approximately 10x better die-to-die energy efficiency versus conventional 2.5D interconnects. Management expects increasing adoption over the next 3-5 years. MediaTek also emphasized custom HBM, proprietary die-to-die interconnects, and DTCO optimization. We see structural opportunities for advanced packaging and custom ASIC design services as system complexity increases (link).

Memory: Capacity matters alongside bandwidth. SK hynix introduced its MARS disaggregated memory architecture, demonstrating a 3-5x improvement in time-to-first-token through shared KV-cache access versus recomputation. Meanwhile, HBF targets approximately half of HBM bandwidth but 14x higher capacity, with samples expected in early 2028. Inference workloads will increasingly require heterogeneous memory architectures encompassing HBM, DRAM, and NAND, potentially expanding memory content beyond traditional GPU-attached HBM (link).

Daniel Yen, CFA
Equity Analyst
Daniel.Yen@morganstanley.com +886 2 2730-2863

Charlie Chan
Equity Analyst
Charlie.Chan@morganstanley.com +886 2 2730-1725

Daisy Dai, CFA
Equity Analyst
Daisy.Dai@morganstanley.com +852 2848-7310

Lucas Wang
Equity Analyst
Lucas.Wang@morganstanley.com +886 2 2730-2875

Henry Zhao
Equity Analyst
Henry.Zhao@morganstanley.com +852 2239-7731

Ethan Jia
Research Associate
Ethan.Jia@morganstanley.com +852 3963-2287

GREATER CHINA TECHNOLOGY SEMICONDUCTORS
Asia Pacific
Industry View **Attractive**

investors should be aware that the firm may have a conflict of

Research as only a single factor i
decision.
**For analyst certification and oth
refer to the Disclosure Section, l
report.**
+= Analysts employed by non-U.S
with FINRA, may not be associate
and may not be subject to FINRA
communications with a subject c
and trading securities held by a re



Power and thermal: The next system-level constraints. Demonstrations of 6kw-class double-sided cold plates vertical power delivery, and advanced microchannel cooling highlighted intensifying thermal challenges. The transition toward 800V DC should reduce conversion losses and improve power density. Meanwhile, AMD's Helion agentic Al rack architecture reinforces incremental CPU demand alongside GPU infrastructure. Power delivery, advanced cooling, and system integration capabilities to become increasingly important differentiators.