13 Oct 2026 04:33:03 ET | 14 pages

US Semiconductors and Hardware

OCP Day 1- Top Five Takeaways

CITI'S TAKE

We attended OCP Day 1 in San Jose CA and recap five key takeaways. We met with Marvell and believe the magnitude of the Microsoft Maia XPU ramp is not yet fully reflected in the company's revenue framework. We view NVIDIA co-sponsoring a pluggable standard with Cisco as a positive to neutral read for LITE & COHR. AMD's Helion extends its strategy of promoting open AI infrastructure positioning the company as an alternative to more vertically integrated platforms. We heard about hybrid copper-optical architectures and rack scale system level integration from server OEMs.

Atif Malik [AC]

Asiya Merchant, CFA

Papa Sylla

Kelsey Chia, CFA

Adrienne Colby

1) Maia Becomes Microsoft's Primary Inference Platform; Marvell the Likely Key Beneficiary — We believe the magnitude of the Microsoft ramp is not yet fully reflected in the company's revenue framework suggesting potential upside relative to existing custom silicon projections if those deployments materialize as expected.

2) NVIDIA/Cisco USFP Extends the Pluggable Runway — NVIDIA co-sponsoring a pluggable standard for the 400G-lane node is a meaningful signal that pluggables retain a central role for at least another generation, extending the module, a positive to neutral read for LITE & COHR.

3) AMD Helion Agentic AI Rack — Building on the company's **Helios** rack-scale architecture, Helion platform allows customers to tailor CPU performance, core density, power consumption, and GPU acceleration to the specific requirements of their AI agents, while adhering to an open, standards-based infrastructure approach. CEO Lisa Su in keynote mentioned that Helios is on the move with customer deliveries to major customers including Meta, MSFT, Oracle, and the AI labs and revealed 6th Gen AMD EPYC Verano, the next step in AMD's CPU roadmap.

4) **Hybrid Copper-Optical Architectures —** For larger-scale deployments, Broadcom expects networking architectures to increasingly optimize interconnect selection by distance, with copper deployed within racks for its cost, power, and latency advantages, and optical interconnects used between racks and clusters where reach and scalability become critical.

5) **AI is Rack Scale Systems Problem —** AI infrastructure is increasingly becoming a rack-scale systems problem rather than a server problem. HPE, Lenovo, Dell, and Supermicro all highlighted that rising power requirements, liquid cooling adoption, and growing deployment complexity are driving a shift towards treating the rack as the primary computing unit.

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliations.

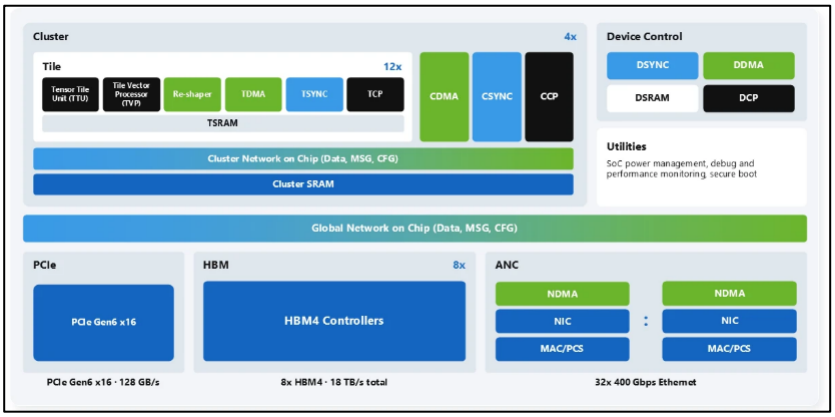
OCP Day 1 Highlights

1) Maia Becomes Microsoft's Primary Inference Platform; Marvell the Likely Key Beneficiary

At today's OCP Global Summit, MSFT (covered by Tyler Radke) said Maia will become the primary inference platform for its first-party products, with multi gigawatt deployments beginning in 2027. We view this as a meaningful positive for MRVL, MSFT's key XPU supplier. In fact, today's specifications show 288GB of HBM4 per chip and production beginning in 2027. The distinction from the prior generation matters. We believe Maia 300 is therefore the point at which Microsoft would shift from a peripheral relationship to a full XPU program for Marvell. The deployment numbers are large: 96 chips per rack (double Maia 200), clusters of 12,000+ chips. When combining these details by MSFT to the reported TSMC capacity talks for more than 300,000 units in 2027 and an ambition above one million over time (Reuters, August 10), these imply that MSFT could become another scaled custom-silicon franchise alongside AWS Trainium, which would further ease investor concerns about customer concentration, in our view.

We met Marvell IR team at OFC today and they stated that 100% of the $12B+ FY30 custom revenue target is already tied to programs that have been won, implying the target does not require additional design wins to be achieved. On the scale of Microsoft's Maia opportunity, they restated prior comments that the program is larger than the two current custom XPU programs combined. On newer programs, management noted they are intentionally conservative in their outlook because there are multiple variables, particularly around supply-chain scaling and deployment timing, that remain under their control. We believe the magnitude of the MSFT ramp is not yet fully reflected in the company's revenue framework suggesting potential upside relative to existing custom silicon projections if those deployments materialize as expected.

Figure 1. Maia 300 SoC



Source: Microsoft

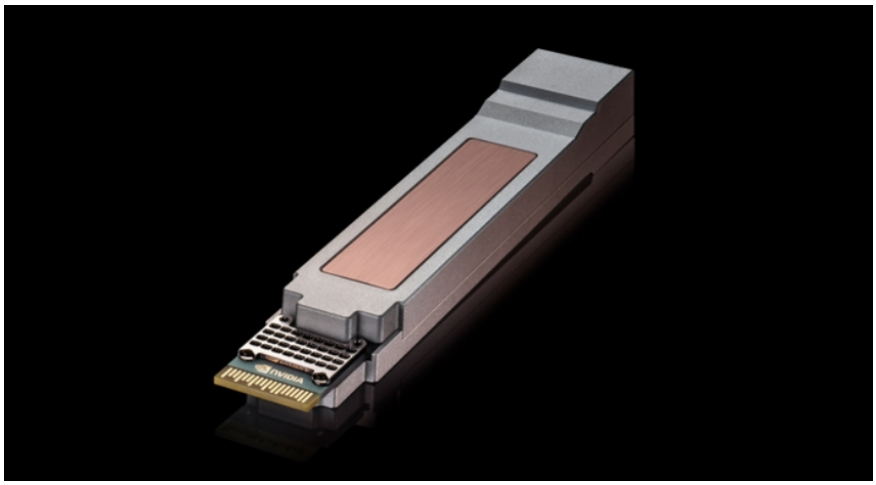
2) NVIDIA/Cisco USFP Extends the Pluggable Runway

NVDA and CSCO announced plans at the OCP Global Summit fo
Small Form Factor Pluggable (USFP). It is a new optical module f
400Gb/s-per-lane electrical signaling, with first implementatio
duplex lanes in an OSFP-sized footprint. The companies cite 4x
front-panel density and roughly 1.3x the power efficiency of curr
lane pluggables, up to 100W of dissipation without embedded n
and headline capacity of up to 12.8Tb/s. TEL and APH (both cove

人在草木间

Merchant) are the named cage, connector and module partners, and an open specification will follow through a community-led multi-source agreement. No availability timeline was disclosed. We note that 16 lanes at 400G implies 6.4Tb/s per direction, so we believe the 12.8Tb/s figure reflects aggregate bidirectional capacity or the future 32-lane variant, though we await more details from upcoming presentations.

Figure 2. NVDA & CSCO's Newly Introduced Pluggable Form Factor



Source: NVIDIA , Cisco

Read for CSCO: From Partner to Co-Architect

We see USFP as the latest step in a steadily deepening Cisco–NVIDIA relationship. It began with Cisco Silicon One becoming the only third-party silicon inside Spectrum-X in February 2025. It then extended to Cisco switches built on NVIDIA silicon, including the 102.4T N9100 on Spectrum-6 announced at GTC in March 2026. Most recently, in August, the Secure AI Factory was expanded with Supermicro compute for neocloud and sovereign cloud buyers. With USFP, CSCO moves from integrating NVIDIA technology to co-defining an industry interconnect standard alongside it. We believe this strengthens Cisco's claim to be NVIDIA's preferred Ethernet partner outside NVIDIA's own switching line, particularly in enterprise, neocloud and sovereign deployments where Cisco's channel and operating model are differentiators.

Read for COHR and LITE: Neutral to Positive

In our view, NVIDIA co-sponsoring a pluggable standard for the 400G-lane node is a meaningful signal that pluggables retain a central role for at least another generation, extending the module. LITE and COHR's strategic alignment with NVIDIA also reduces exclusion risk. We estimate that a 16-lane, 400G-per-lane module raises both laser count and laser performance requirements per port. We believe this plays to LITE's laser dominance and COHR's vertically integrated InP position anchored by the 6-inch Sherman fab, where supply remains the binding industry constraint.

We nevertheless see three areas that warrant monitoring. First, CSCO is itself a pluggable optics supplier through Acacia and its silicon photonics platform, and its optics organization co-led the specification, which could confer a time-to-market advantage on Cisco-attached ports. Second, a 4x step-up in bandwidth per slot

reduces unit volumes per bit deployed, placing greater weight on ASP scaling to protect revenue per port. Third, the release emphasizes optical designs with simplified signal processing, which could lower module bill-of-materials and pressure module pricing even though most module makers do not supply the DSP. No optical module vendor, including COHR or Lumentum, was named among launch partners. We regard early MSA membership as the key indicator.

3) AMD Helion Agentic AI Rack: AMD introduced **Helion**, an open blueprint for agentic AI infrastructure based on an Open Compute Project (OCP) specification. Building on the company's **Helios** rack-scale architecture, Helion is designed to give customers and partners greater flexibility in deploying AI agent workloads by enabling configurable combinations of CPUs and GPUs. The platform allows customers to tailor CPU performance, core density, power consumption, and GPU acceleration to the specific requirements of their AI agents, while adhering to an open, standards-based infrastructure approach.

We believe Helion extends AMD's strategy of promoting open AI infrastructure and ecosystem interoperability, positioning the company as an alternative to more vertically integrated AI platforms. By allowing customers to optimize compute and acceleration resources for emerging agentic AI workloads, AMD aims to broaden adoption of its CPU and GPU portfolio across next-generation AI deployments. Moreover, CEO Lisa Su in keynote mentioned that Helios is on the move with customer deliveries to major customers including Meta, MSFT, Oracle, and the AI labs. She also revealed 6th Gen AMD EPYC Verano, the next step in AMD's CPU roadmap.

4) **Hybrid Copper-Optical Architectures:** Discussions across the conference reinforced our view that the future of AI networking will be defined by hybrid copper-optical architectures. As AI clusters scale and cross-rack compute domains expand, networking requirements for bandwidth density, reach, latency, and energy efficiency continue to increase. However, both Broadcom and Amphenol emphasized that copper retains a longer technological runway than many investors anticipate. Advances in signaling technologies such as 400G PAM6, co-packaged copper, and improved system architectures could extend electrical connectivity to higher bandwidths and potentially beyond a single rack, supporting copper's role inside AI infrastructure for several more years.

Meta, on the other hand, discussed their optical backplane architecture for AI racks to address physical limitations with copper only architectures. Management described it as a way to achieve a significant increase in rack-level bandwidth while maintaining existing rack serviceability, cooling, and operational characteristics. The company emphasized that the optical backplane is intended to provide a step-function increase in bandwidth density without requiring major changes to how racks are deployed, serviced, or cooled. Meta noted that it already has a prototype optical backplane on display, suggesting the technology has progressed beyond a conceptual design stage. At a broader cluster level, Meta highlighted the enormous scale of fiber deployment within its AI infrastructure, noting that Prometheus (1 GW+ datacenter campus) contains approximately 3-4 million kilometers of fiber, reflecting the growing importance of optical connectivity as AI clusters span multiple buildings and require massive east-west bandwidth. We expect the Hyperion campus (5 GW+) to have even more significant optical content underscoring the strategic importance of Meta's $6 billion supply agreement with Corning.

For larger-scale deployments, Broadcom expects networking architectures to increasingly optimize interconnect selection by distance, with copper deployed

within racks for its cost, power, and latency advantages, and optical interconnects used between racks and clusters where reach and scalability become critical. To facilitate this transition, Broadcom introduced ICX packaging technology, which enables copper and optical connectivity to coexist within the same package, supporting flexible hybrid networking architectures. Looking further ahead, the company highlighted Optical Compute Interconnect (OCI) as a potentially important technology that could deliver the reach benefits of optics while maintaining power characteristics closer to copper, enabling larger and more efficient AI clusters. Combined with Broadcom's next-generation Tomahawk switching portfolio and ongoing industry standardization efforts, the roadmap aligns closely with the industry's broader evolution toward hybrid interconnects.

While optical connectivity (near-package optics (NPO) and co-packaged optics (CPO) adoption) will gain traction as AI infrastructure scales, adoption remains constrained by challenges related to serviceability, optical alignment, repairability, fiber management, manufacturing capacity, and overall system complexity. Industry participants emphasized that successful optical deployment requires innovation not only in optical engines themselves, but also in chip-to-fiber coupling, high-density connector ecosystems, and increasingly sophisticated fiber-routing architectures spanning servers, racks, and clusters. In this context, TE Connectivity highlighted Fusion-Fold manufacturing technology, which uses automated, high-precision optical assembly techniques to improve scalability and potentially address some of the manufacturing and reliability bottlenecks facing next-generation optical deployments.

Overall, the emerging industry consensus is that next-generation AI systems will deploy the right interconnect for the right distance: copper for short-reach scale-up connectivity within racks, and optics for longer-reach scale-out and scale-across deployments between racks and across clusters. Rather than a binary transition from copper to optics, the industry appears to be moving towards hybrid architectures that offer the most attractive combination of performance, cost, reliability, manufacturability, and scalability for large-scale AI infrastructure.

5) AI is a Rack-Scale Systems Problem: AI infrastructure is increasingly becoming a rack-scale systems problem rather than a server problem. HPE, Lenovo, Dell, and Supermicro all highlighted that rising power requirements, liquid cooling adoption, and growing deployment complexity are driving a shift towards treating the rack as the primary computing unit. As AI clusters scale to gigafactory-like deployments, success will depend less on individual servers and more on integrated solutions spanning power delivery, cooling, storage, networking, deployment services, and lifecycle management. Industry participants also emphasized the growing importance of open standards, operational automation, and ecosystem collaboration to reduce deployment timelines, maximize GPU utilization, and manage increasingly complex AI environments.

Hewlett Packard Enterprise (HPE) discussed how AI workloads, higher power consumption, and liquid-cooling requirements are fundamentally changing datacenter architecture, leading to a shift from viewing individual servers as the primary computing unit to treating the entire rack as the computer. As infrastructure becomes more disaggregated, with components such as power shelves and cooling distribution units operating independently from the server, HPE highlighted new operational challenges around communication between devices, topology discovery, power sequencing, telemetry management, and lifecycle management. The company emphasized the need for vendor-neutral industry standards through the Open Compute Project (OCP), noting that future rack-scale infrastructure will require consistent management frameworks that

work across a wide range of server, rack, and management architectures rather than proprietary, closed systems.

Lenovo highlighted the practical challenges of deploying and scaling large AI infrastructure environments, emphasizing that the key metric for operators is reducing "time to first token," or the delay between hardware investment and productive AI output. The company noted that AI clusters are reaching gigafactory scale, where operational complexity grows exponentially due to the large number of interconnected components and where uptime becomes critical. Lenovo described four core areas required for successful deployments: designing optimized architectures, securing and building complex infrastructure, validating systems during deployment, and providing long-term operational support. The company also stressed that modern AI datacenters are increasingly designed around power delivery and cooling capabilities rather than physical floor space, and that rapid shifts in AI workloads, including the rise of agentic AI, are placing significant pressure on supply chains and deployment timelines.

Dell Technologies and Supermicro both emphasized that AI infrastructure is becoming increasingly complex and data-intensive, requiring integrated, rack-scale solutions that optimize performance, deployment speed, and operational efficiency. Relative to its Storage Solutions & ExaScale offering, Dell focused on the growing importance of storage architecture, KV-cache management, and disaggregated infrastructure to maximize GPU utilization and accelerate AI deployments; while Supermicro highlighted its end-to-end building-block approach spanning compute, networking, cooling, and datacenter deployment services, highlighting the need for scalable, pre-validated infrastructure solutions that can help customers rapidly deploy and operate large AI clusters while managing the growing power, cooling, and operational demands of next-generation AI workloads.

Companies Mentioned:

Advanced Micro Devices (AMD.O; US$594.24; 1; 12 Oct 26; 16:00) | Amazon.com, Inc. (AMZN.O; US$259.53; 1; 12 Oct 26; 16:00) | Amphenol Corp (APH.N; US$86.19; 1; 12 Oct 26; 16:00) | Broadcom Inc (AVGO.O; US$354.73; 1; 12 Oct 26; 16:00) | Cisco Systems (CSCO.O; US$115.34; 1; 12 Oct 26; 16:00) | Coherent Corp (COHR.K; US$288.02; 1; 12 Oct 26; 16:00) | Corning Incorporated (GLW.N; US$147.28; 1; 12 Oct 26; 16:00) | Dell Technologies (DELL.N; US$576.76; 1; 12 Oct 26; 16:00) | Hewlett Packard Enterprise (HPE.N; US$70.33; 1; 12 Oct 26; 16:00) | Lenovo Group (0992.HK; HK$32.34; 1; 12 Oct 26; 16:10) | Lumentum Holdings Inc (LITE.O; US$995.41; 1; 12 Oct 26; 16:00) | Marvell Technology Inc (MRVL.O; US$265.86; 1; 12 Oct 26; 16:00) | Meta Platforms Inc (META.O; US$742.96; 1; 12 Oct 26; 16:00) | Microsoft Corp. (MSFT.O; US$541.15; 1; 12 Oct 26; 16:00) | NVIDIA Corp (NVDA.O; US$226.32; 1; 12 Oct 26; 16:00) | Oracle Corporation (ORCL.K; US$137.39; 1; 12 Oct 26; 16:00) | Super Micro Computer, Inc. (SMCI.O; US$39.98; 2H; 12 Oct 26; 16:00) | TE Connectivity Ltd. (TEL.N; US$217.28; 1; 12 Oct 26; 16:00) | TSMC (2330.TW; NT$2540.0; 1; 12 Oct 26; 13:30)