Sunil Priyadarshi,
Andreas Bechtolsheim,
Brian Park and Vijay Vusirikala

Scaling AI Data Centers with XPO Optics

eXtra-dense Pluggable Optics are rewriting the rules of density, thermal management and reliability to future-proof artificial intelligence networks.



antoniokhr / Getty Images





The rapid expansion of AI and machine-learning workloads is reshaping the physical and logical architecture of modern data centers.

S. Sangsakawrat / Getty Images

The rapid expansion of artificial intelligence (AI) and machine-learning workloads is reshaping the physical and logical architecture of modern data centers. As large language models, neural networks and distributed training clusters grow in complexity and computational scale, they place increasing pressure on the underlying network infrastructure connecting accelerators, switches and storage systems. In particular, this demand is pushing existing optical interconnect technologies beyond their original design limits, exposing structural weaknesses in how the industry currently moves data.

In this new era of hyperscale computing, the performance of the network fabric has transitioned from a supporting role to a bottleneck governing overall system efficiency.

One emerging challenge is front-panel bandwidth density. Conventional octal small form-factor pluggable (OSFP) optics are approaching practical density limits for the highest-capacity AI network architectures, prompting the industry to explore new approaches. Extra-dense pluggable optics, or XPO, are designed to provide up to four times the front-panel bandwidth density of current pluggable solutions while retaining key advantages of the pluggable model, including configurability, manufacturability and serviceability. The approach also incorporates liquid cooling and is intended to accommodate a range of optical interface standards.

Development of the technology is being coordinated through the XPO Multi-Source Agreement (MSA), which includes more than 150 member organizations spanning optical-module and network-system vendors, connector and cooling suppliers, and other component manufacturers. The broad participation is intended to establish an interoperable ecosystem capable of supporting deployment as data center bandwidth requirements continue to grow.

The AI infrastructure bottleneck

The core challenge stems from a stark reality: AI data centers can require orders of magnitude more bandwidth than traditional cloud-computing environments, particularly for large-scale distributed training. For years, the industry has relied on established high-speed pluggable optical solutions to enable 400 Gb/s, 800 Gb/s and 1.6 Tb/s optical connections.

However, these traditional pluggable modules are fundamentally constrained by physical space and front-panel density. For example, one open rack unit (OU) switch configuration accommodating 32 OSFP modules at 1.6 Tb/s each provides 51.2 Tb/s of front-panel bandwidth. At that density, achieving a switching capacity of 102.4 Tb/s requires two full rack units. Reaching 204.8 Tb/s requires four OUs. In modern AI clusters requiring tens of thousands of continuous, high-bandwidth connections, this relationship between module size and faceplate area can create a bottleneck where the physical footprint of switch racks exceeds the number of AI compute racks.

Recognizing this ceiling, the industry actively explored alternative integration approaches, such as co-packaged optics (CPO) and near-packaged optics (NPO). In both cases, migrating the optical engines off the front panel and directly onto the host motherboard shortens the high-speed electrical path, which can reduce electrical losses and ease signal-integrity and power constraints. These architectures can also support high levels of bandwidth density.

CPO and NPO address density by moving optical engines closer to the switch ASIC, shortening the high-speed electrical path and reducing electrical loss. But they also shift more optical assembly and test into the system-manufacturing stage, where yield, qualification and repair can become more complex. They also tend to lock in the optical configuration earlier than pluggable approaches, reducing

1047-6938/26/11/34/8-$15.00 ©Optica

In modern AI clusters, the relationship between module size and faceplate area can create a bottleneck where the physical footprint of switch racks exceeds the number of AI compute racks.

late-binding flexibility across different reaches and interface standards.

The networking industry therefore faces a trade-off between the bandwidth density and efficiency enabled by greater integration and the modularity and field serviceability of front-panel optics. XPO is one approach to bridging that gap, retaining a hot-swappable pluggable module while increasing bandwidth density and integrating liquid cooling.

The five pillars of AI optical networking

To understand why a new optical paradigm is required, network architects must evaluate the specific technical demands of large-scale AI workloads. Design decisions made at the component level cascade directly into the performance, cost and scalability of the entire data center network.

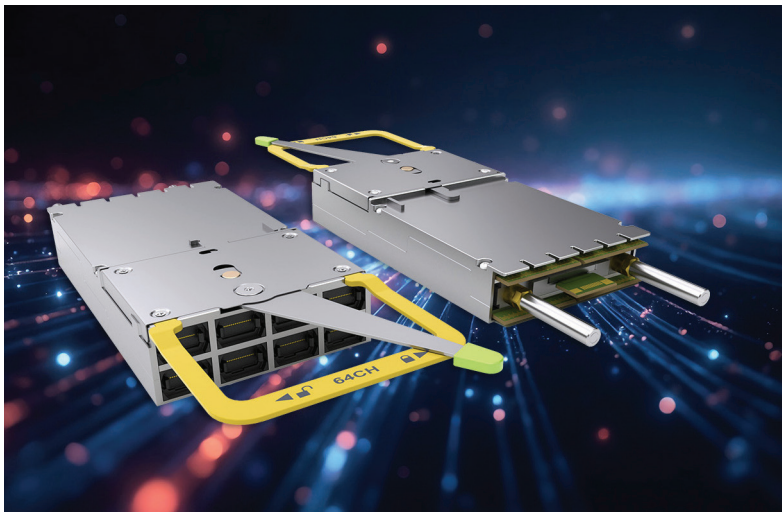
High bandwidth requirements: Bandwidth per accelerator and per switch ASIC is increasing rapidly. Scale-up fabrics can require tens of terabits per second per XPU, while scale-out links add multiple terabits per second. Therefore, the network needs large numbers of non-blocking, high-radix ports.

High port density: The limited front-panel density of current optics can force additional switch racks and more complex network topologies, increasing building footprint, cabling complexity and infrastructure cost.

Liquid-cooling integration: Large AI systems are increasingly adopting warm-water liquid cooling. Any new optical interconnect must integrate natively and efficiently with that cooling environment rather than depending on a separate air-cooling strategy.

High reliability: At hyperscale, networks can contain millions of optical links. Reliability, therefore, has to improve substantially to minimize both hard failures and performance-degrading soft failures that interrupt or fragment large training jobs.

Power efficiency: Every watt consumed by the optical interconnect is a watt unavailable to computation. Improving energy efficiency per transmitted bit is therefore a system-level requirement, not simply a component metric.



Example of XPO, designed to provide up to four times the front-panel bandwidth density of current pluggable solutions while retaining key advantages of the pluggable model.

Arista Networks

A paradigm shift in high-density pluggable optics

Developed by Arista Networks in collaboration with the MSA ecosystem, XPO represents a new class of high-density optical pluggable module designed specifically for AI data center fabrics. It targets the interdependent requirements of bandwidth, density, reliability, cooling and power efficiency while retaining the configuration and serviceability advantages of a pluggable form factor.

Using 64 independent electrical lanes, an XPO module is engineered to deliver 12.8 Tb/s of total throughput at 200 Gb/s per lane and is intended to scale to 25.6 Tb/s at 400 Gb/s per lane. Physically, the module is housed in a compact outer enclosure measuring approximately 60.8 mm in width, 111.8 mm in length and 21.1 mm in height.

This specific geometry makes the XPO module roughly 2.7 times wider than the standard OSFP form factor—16 XPO modules plus management ports can fit within a single OU. That supports 204.8 Tb/s of switching per OU at 200 Gb/s per lane, and a future 409.6 Tb/s per OU at 400 Gb/s per lane—up to a four-fold increase in front-panel bandwidth density relative to conventional OSFP-based designs.

Most importantly, XPO retains the configuration, serviceability and manufacturability advantages that have made pluggable optics the industry standard for decades. If an optical engine experiences failure or degradation, an operator can simply pull the individual module out and replace it. To manage the mating forces required by 64 high-speed electrical lanes, the module is secured by a top and bottom shell assembly featuring a sophisticated mechanical ejector. This integrated release pull tab provides 1:11 mechanical leverage, facilitating smooth insertion and extraction against high-density connector forces without needing to power down or service the host switch assembly.

Inside the XPO architecture

To achieve such dramatic leaps in both capacity and faceplate density, the engineers behind XPO had to abandon the traditional single-printed circuit board (PCB) layouts found in legacy modules. Attempting to route 64 high-speed 200 Gb/s lanes across a single flat plane while managing the massive thermal load of 12.8 Tb/s optics was a physical impossibility. Instead, the XPO architecture introduces a highly optimized "belly-to-belly" mechanical design.

Inside the rigid outer shell, the module incorporates two completely identical 32-channel paddle cards. These PCBs are arranged facing inward toward a shared, central element. The layout of the components on these cards is strategic. High-power, high-heat-generating components—including the 3-nm process digital signal processors (DSPs), transimpedance amplifiers (TIAs), continuous-wave (CW) laser diodes, and silicon photonics Mach-Zehnder modulators (MZMs)—are assembled on the inward-facing "hot" side of the PCBs. Conversely, lower-power circuitry, such as receivers, microcontrollers and general control logic, is mounted on the outward-facing "cold" side, looking toward the outer enclosure.

Native liquid cooling: overcoming the thermal barrier

Traditional air-cooled optical modules rely heavily on thermal interface materials (TIMs) and bulky, cage-mounted riding heatsinks. These legacy modules are forced to operate across wide temperature ranges—often from 0°C to 70°C. These large thermal excursions can induce severe mechanical stress on PCBs and delicate optical components, increasing the risk of material delamination, solder cracking and long-term reliability degradation. XPO fundamentally changes this paradigm by bringing active cooling directly inside the module shell, supporting the hyperscale shift toward liquid-cooled AI infrastructure.

	Conventional pluggable	XPO	NPO	CPO
Optical-engine location	Front-panel module	Front-panel module	Near ASIC	On/adjacent to ASIC package
Field replaceability	Hot-swappable	Hot-swappable	Factory/service replaceable	Limited; design-dependent
Electrical path	Long; PCB	Long; low-loss Twinax/CPC	Short	Shortest
Electrical loss	Highest	Low–moderate	Low	Lowest
Front-panel bandwidth density	32 × 1.6T = 51.2T/OU	16 × 12.8T = 204.8T/OU; future 409.6T/OU	32 × 6.4T = 204.8T/OU; future 409.6T/OU	32 × 6.4T = 204.8T/OU; future 409.6T/OU
Cooling	Air; IHS/RHS	Direct liquid	System air/liquid	Package/system liquid
Power delivery	3.3 V	48 V	12V	System/package power
Manufacturing	Mature pluggable ecosystem	Pluggable/module ecosystem	Optical-engine + system integration	Heterogeneous packaging
Serviceability	Front-panel replacement	Front-panel replacement	Requires system access	Most challenging
Integration complexity	Lowest	Moderate	Moderate–high	Highest
Primary advantage	Maturity + serviceability	Density + pluggability + cooling	Short path + modularity	Integration + lowest loss
Primary tradeoff	Loss + density	New form factor + liquid cooling	Fiber/cooling + integration	Packaging + serviceability

Comparison of pluggable optics for data centers.

To achieve dramatic leaps in both capacity and faceplate density, the engineers behind XPO had to abandon the traditional single-printed circuit board layouts found in legacy modules.

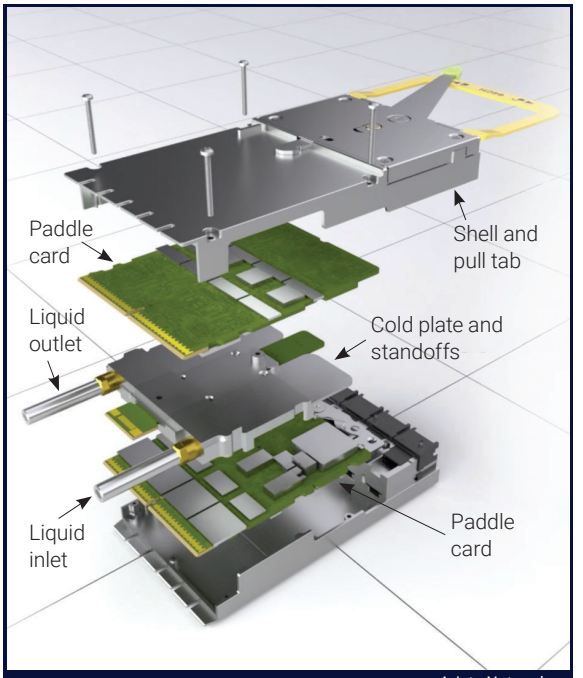
A central liquid cold plate is physically sandwiched between the two paddle cards, allowing it to simultaneously draw heat away from the inward-facing "hot" sides of both PCBs. This integrated design efficiently removes thermal energy from high-power module configurations up to 500 W. By utilizing standard 40°C to 45°C warm-water liquid cooling using either deionized water or a PG25 coolant mixture, the XPO architecture can keep crucial component temperatures 20°C to 25°C below those of comparable air-cooled designs. The XPO configuration has a maximum module power envelope of 500 W, and the narrower operating-temperature range afforded by liquid cooling is also intended to reduce thermally induced mechanical stress and improve long-term reliability.

Linear channels and high-power efficiency

The electrical design of the XPO module is segregated to improve both signal integrity and power efficiency. High-speed transmit and receive signals are separated onto opposite sides of the internal paddle cards to minimize crosstalk. This physical separation, combined with a low-loss Twinax copper cable assembly and an enhanced XPO connector, creates a clean linear channel. The resulting low insertion loss—it maintains end-to-end ASIC-to-optics loss below 18 dB—makes the platform ideal for power-efficient linear-drive pluggable optics (LPO) implementations. The high-speed paths use separate connectors and fly-over copper to the switch ASIC, while a separate low-speed connector handles control and power.

Traditional pluggable modules rely largely on a 3.3 V power input, which limits practical module power to roughly 40 W. XPO's much larger thermal envelope requires a different approach: a dedicated power-and-control card-edge connector that accepts 46 V to 53 V DC input (nominally about 48 V) directly from the rack's primary bus bar.

On-board point-of-load voltage regulators located directly on the module's paddle cards then step down the 48 V supply to 3.3 V and other required logic levels. This approach eliminates the need for bulky, worst-case provisioned voltage converters on the host switch



Arista Networks

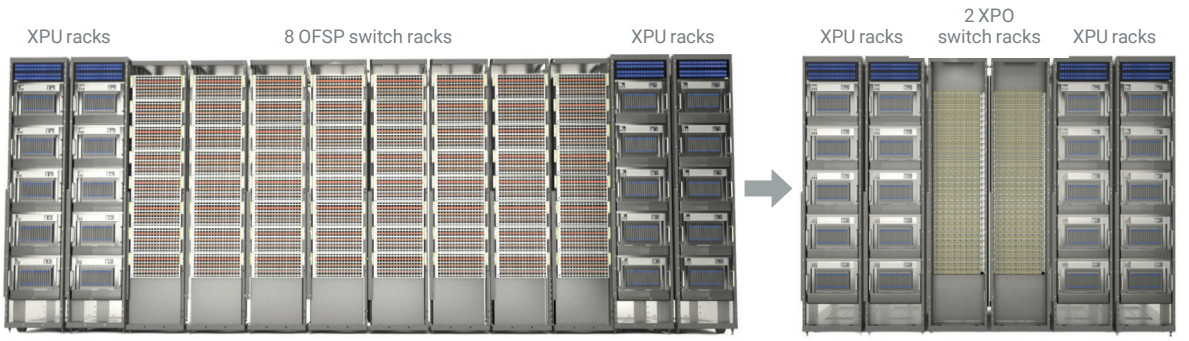
Key elements of the XPO architecture

Paddle cards: Two printed circuit boards are arranged on either side of the module's central cooling structure. In the XPO architecture, the inward-facing sides accommodate higher-power components so that heat can be transferred efficiently toward the cold plate.

Central liquid cold plate: Positioned between the two paddle cards, the cold plate provides a direct thermal path for removing heat from components on both sides of the module.

Liquid inlet and outlet: Coolant enters and exits through dedicated connections integrated into the module, enabling direct liquid cooling.

Shell and pull tab: The outer shell encloses the internal assemblies, while the pull-tab mechanism supports insertion and removal of the pluggable module.



	OSFP-based rack	XPO-based rack	Delta	Calculation / basis
Bandwidth per rack	1.6 Pbps	6.5 Pbps	4×	204.8T switch in 4 OU vs. 1 OU
Number of 204.8T switches	8	32	4×	Total capacity 32 OU
Power per rack	40 kW	160 kW	4×	Switch ~3.5 kW + linear optics ~1.5 kW

Rack-level infrastructure utilization and economics of an OSFP-based rack versus an XPO-based rack.

motherboard. This reduces overall system complexity, improves host system reliability and helps the module achieve an overall power conversion efficiency of approximately 92%, according to the XPO specifications. At the system level, moving this conversion onto the module can remove several kilowatts of worst-case power conversion capacity from the host motherboard; the authors estimate roughly 3.2 kW for a 204.8 Tb/s switch populated with 16 fully retimed 200 W XPO modules.

The hyperscale footprint advantage

The value of an architectural innovation is only fully realized when component-level specifications translate into systemic economic efficiencies across an entire data center. The 4× density advantage of XPO can alter the math of hyperscale facility design, enabling reductions in physical footprint, with potential implications for capital expenditure and operational complexity.

Consider a scale-up AI cluster containing 512 high-performance accelerators (XPUs), each connected at 25.6 Tb/s. That implies 13.1 Pb/s of switching bandwidth, equivalent to 64 switches, each providing 204.8 Tb/s and 8,192 1.6-Tb/s ports. If 128 XPUs fit in a compute rack, the cluster occupies four XPU racks.

Using OSFP technology, a switch rack accommodating 1,024 1.6-Tb/s ports would require eight switch racks for that cluster. With XPO, the same rack can accommodate 4,096 1.6-Tb/s ports while reducing the requirement to two switch racks. The networking footprint falls by 75%, while the combined compute-plus-network rack count drops from 12 to 6—a 50% reduction under these illustrative assumptions.

At gigawatt scale, the arithmetic becomes more dramatic. The authors model a 256,000-XPU facility with 2,000 XPU racks and about 10 million 1.6-Tb/s optical interfaces across scale-out and scale-up fabrics. At 1,024 OSFP interfaces per switch rack, that would imply roughly 10,000 switch racks, or 12,000 racks overall. At 4,096 1.6-Tb/s interfaces per XPO switch rack, the requirement falls to about 2,500 switch racks and 4,500 total racks—a 62.5% reduction in overall rack count. These figures are scenario calculations rather than deployment forecasts, but they illustrate how optical-module density can translate into building-scale consequences.

Modern liquid-cooled Open Rack v3 (ORv3) infrastructure can support rack-level cooling capacities of 160 kW or more. An OSFP-based rack at about 40 kW uses only a fraction of this capacity; a four-times-denser XPO-based rack at about 160 KW can use it much more fully. Fewer switch racks for the same total bandwidth can mean fewer racks, manifolds, bus bars, power shelves and plumbing runs, as well as shorter fiber paths. Those system-level effects, rather than module density alone, are where the potential infrastructure savings arise.

A unified foundation for AI data centers

AI data center networks must span multiple physical reaches within one architecture—from very short scale-up connections inside a cluster, through data center and campus links, to metro and longer inter–data center connections. The XPO form factor is intended to support optical and non-optical interfaces across that range rather than tying the switch to a single transmission technology.

The XPO platform is designed to handle the complete spectrum of AI connectivity requirements. It can power short-reach (100 m) scale-up connections utilizing high-efficiency DR (data center reach) standards

The value of an architectural innovation is only fully realized when component-level specifications translate into systemic economic efficiencies across an entire data center.

for scale-up links within localized XPU SuperPods. It can also accommodate medium-reach (2 km) FR (fiber reach) and LR (long reach) connections bridging vast spine switch fabrics. And crucially, it supports the power requirements of advanced coherent-lite and ZR technologies, enabling high-bandwidth, long-haul (80 km+) scale-across links connecting geographically separated data centers. In this terminology, ZR and ZR+ refer to coherent pluggable interfaces used for metro and longer-reach connections.

More broadly, the platform can accommodate RF/microwave and copper links over a few meters; intensity-modulation/direct-detection (IM/DD) optics over data center reaches; Coherent Lite for campus-scale links; and ZR/ZR+ coherent optics for 100 km and beyond. The two large paddle cards provide PCB area for 8-, 16- and 32-channel optical components; the liquid-cooling and power envelope can accommodate higher-power coherent DSPs; and the module size allows the fiber routing and bend radius needed for optical amplification.

The same architecture also provides a scaling path. At 200 Gb/s per electrical lane, XPO supports 12.8-Tb/s modules; at 400 Gb/s per lane, the target rises to 25.6 Tb/s. The proposed roadmap calls for production-volume XPO deployments with 102.4-Tb/s switches in 2027 and 204.8-Tb/s switches in 2028. Because the form factor remains pluggable, future optical technologies could be introduced without fixing the transmission choice at the time the switch is manufactured.

Proponents also argue that XPO could benefit from an established silicon photonics manufacturing base. The underlying production infrastructure has already been used to manufacture hundreds of millions of 400 G, 800 G and 1.6 T transceivers every year, accumulating billions of device-hours of operation. That manufacturing maturity could help shorten the path to higher-volume deployment compared with architectures that require substantially different packaging or qualification approaches.

XPO represents one proposed route to the next generation of AI optical networking, increasing front-panel bandwidth density while retaining a pluggable operating model. Its 64-lane architecture, native liquid cooling, higher-voltage power delivery and support for multiple interface types are intended to address not only bandwidth, but also thermal management, reliability and serviceability. At 200 Gb/s per lane, 16 modules could provide 204.8 Tb/s per OU; a 400-Gb/s-per-lane generation would double that figure to 409.6 Tb/s per OU.

Whether XPO, CPO or a combination of architectures ultimately serves different parts of future AI fabrics will depend on system-level trade-offs in power, cooling, cost, reliability and maintainability. What is clear is that as the bandwidth in AI data centers grows, optical interconnects will need to deliver substantially more bandwidth within increasingly constrained physical and power envelopes. OPN

Sunil Priyadarshi (spriyadarshi@arista.com) is Senior Director of Photonics Engineering and AI & Cloud Network Architect, Andreas Bechtolsheim is Chief Architect and Cofounder, Brian Park is Technical Director and Vijay Vusirikala is Distinguished Lead, AI Systems and Networks, all at Arista Networks.

References and Resources

- XPO Multi-Source Agreement (MSA). "eXtra-dense pluggable optics (XPO) specification." xpomsa.com
- S. Priyadarshi. "Unlocking the potential of copackaged optics in AI and HPC: Opportunities and challenges," IEEE Communications Magazine, **64**, 24 (2026). doi:10.1109/MCOM.001.2500504
- OSFP Multi-Source Agreement. "Specification for OSFP-XD octal small form factor pluggable module," Rev. 1.11, 23 Nov 2025. osfpmsa.org
- IEEE Standard Association. "IEEE draft standard for ethernet amendment 13: Media access control parameters for 1.6Tb/s and physical layers and management parameters for 200 Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s operation," IEEE P802.3dj (2024).
- Open Compute Project (OCP). "Open rack v3 specifications and designs." www.opencompute.org/wiki/Open_Rack/SpecsAndDesign
- S. Priyadarshi et al. "Industry's first 12.8T 8×DR8 high-density liquid-cooled 64-channel pluggable XPO optical module for scale-up," ECOC 2026, Sept. 20–24, 2026.
- S. Priyadarshi et al. "Demonstration of the 12.8T 8×DR8 high-density liquid-cooled 64-channel pluggable XPO LRO module and XPO emulation platform," IEEE Hot Interconnects Conference, Technical Paper Session B: Hot Topic Presentations, Aug. 19, 2026.
- GIGABYTE. "DCA3-ST0." Accessed Sept. 24, 2026.