ARISTA

White Paper

XPO: Redefining Pluggable Optics for AI Networking

Executive Summary

The explosive growth of Artificial Intelligence (AI) workloads is fundamentally reshaping the requirements for data center infrastructure. Next-generation AI clusters demand dramatically higher bandwidth density, improved thermal management, and greater system-level reliability than traditional cloud data centers were designed to support.

While the industry-standard OSFP (Octal Small Form-Factor Pluggable) module has successfully enabled 400Gbps, 800Gbps, and 1.6Tbps optical pluggable modules [1], it is limited to 32 modules per Rack Unit (RU), typically requiring 2 RUs to achieve 102.4Tbps and 4 RUs to reach 204.8Tbps of switching capacity. Although co-packaged optics (CPO) and on-board optics (OBO) have been proposed to increase bandwidth density, these approaches introduce significant challenges in field serviceability, scalability, and manufacturability, making them difficult to deploy widely in hyperscale environments [2].

To address these challenges, Arista Networks, together with an ecosystem of more than 45 industry partners, introduces eXtra-dense Pluggable Optics (XPO) [3]. XPO represents a new class of optical pluggable module designed specifically for next-generation AI data center fabrics. Each XPO module delivers 12.8Tbps of bandwidth using 64 electrical lanes and incorporates an integrated liquid-cooled cold plate capable of supporting 400W+ module power consumption.

The XPO module is approximately 2.7× wider than the OSFP form factor and delivers x8 higher bandwidth per module, enabling up to 204.8Tbps of switching capacity within the vertical space of a single Open Rack Unit (1OU)—a 4× increase in front-panel density compared to OSFP.

The XPO pluggable module preserves the advantages of field pluggability, enabling quick replacement or upgrades of optical modules without servicing the entire switch and minimizing downtime. It also supports a pay-as-you-grow deployment model and accommodates various optical architectures—including DR, FR, LR, SR, and ZR—to support diverse network configurations while providing greater operational flexibility.

AI Infrastructure Demand Growth

The exponential growth of artificial intelligence (AI) and machine learning workloads is fundamentally reshaping data center architecture. This transformation is placing unprecedented pressure on network infrastructure, pushing existing interconnect technologies far beyond their original design limits. As AI models continue to increase in complexity and scale, the performance of the underlying network becomes a critical factor in overall system efficiency. Consequently, the development of next-generation optical interconnect solutions is essential to prevent the network from becoming the primary bottleneck that limits AI innovation.

The core challenge stems from a simple reality: AI data centers require orders of magnitude more bandwidth than traditional cloud data centers. This dramatic increase in required throughput cannot be addressed through incremental improvements alone; it requires a fundamental re-evaluation of interconnect technologies. To support the distributed training clusters and massive data sets that power modern AI workloads, every component of the network fabric must be re-engineered to deliver higher performance, greater density, and improved efficiency. This paper outlines the new requirements imposed by this AI-driven transformation and introduces a purpose-built optical architecture designed to meet these challenges.

Requirements for Optical Interconnects for AI Networks

Understanding the specific technical demands that AI workloads place on networking infrastructure is essential. Design decisions made at the component level can have cascading impacts on the performance, cost, and scalability of the entire data center. This section outlines the five critical requirements that define the next generation of data center optics and examines why existing standards—originally developed for traditional cloud computing environments—are increasingly insufficient to meet the demands of large-scale AI workloads.

Extreme Bandwidth

A key requirement for AI networks is the ability to move vast amounts of data with minimal latency. Training large-scale AI models across distributed clusters of accelerators requires significantly higher network bandwidth than traditional data center workloads. This challenge extends beyond simply increasing link speeds; it requires building a non-blocking, high-radix network fabric capable of sustaining simultaneous communication among tens of thousands of accelerators.

High Reliability

In large-scale AI fabrics comprising tens of thousands of optical links, component failures become statistically inevitable. However, the hard and soft failure rates of today's optical modules remain higher than desirable for the operational reliability required in such environments. A single failure can disrupt or halt a multi-million-dollar training job, resulting in wasted compute cycles and significant financial impact. Diagnosing and replacing a failed module within a fabric containing 50,000+ optical links presents a major operational challenge, often triggering cascading effects on job scheduling and leading to severe resource fragmentation across the network.

Liquid Cooling

The immense computational density of modern AI accelerators generates thermal loads that exceed the capabilities of traditional air cooling. As a result, hyperscale AI data centers are increasingly adopting liquid cooling as a fundamental infrastructure requirement. This shift places strict thermal design constraints on every component within the rack. Any optical interconnect solution that cannot integrate efficiently with a liquid-cooled environment is therefore fundamentally unsuitable for next-generation AI data center deployments.

Power Efficiency

Power consumption is a critical constraint in modern data center design. High-density racks operate within a finite power budget, and every watt consumed by the network is a watt unavailable for revenue-generating compute resources. Consequently, optical interconnects must deliver significantly lower power consumption per transmitted bit. High efficiency is essential not only for reducing operational costs but also for maximizing the computational density and overall performance of each rack.

Unprecedented Density

Physical space is a premium resource in modern data centers. To build the massive-scale fabrics required for AI workloads, network architects require solutions with significantly higher bandwidth density. Insufficient density forces the deployment of larger and more complex multi-tier network topologies with additional spine and super-spine layers, increasing latency, cost, and cabling complexity. The density limitations of current standards present a significant barrier; for example, OSFP supports only 32 modules per 1U, which is insufficient for the scale demanded by large AI clusters.

Based on these five fundamental requirements, it is evident that the widely adopted OSFP module is not optimally suited for the emerging demands of AI-driven data centers. This gap between existing optical technologies and the requirements of next-generation AI infrastructure highlights the need for new architectural approaches to optical interconnect design.

Solution Overview: The Arista XPO Pluggable Optics Module

The Arista XPO (eXtra-dense Pluggable Optics) module is a purpose-built solution designed from the ground up to address the unique challenges of hyperscale AI data centers. Its architecture directly targets the critical metrics of bandwidth, density, reliability, cooling, and power efficiency required for next-generation AI and machine learning workloads. By rethinking the optical module form factor and its integration with the host system, XPO delivers a significant advancement in network scalability and performance.

The key specifications and benefits of the XPO module directly correspond to the five critical requirements outlined previously:

- **Bandwidth:** Each XPO module delivers 12.8Tbps of bandwidth, configured as 64 channels at 200Gbps, enabling the construction of high-radix, non-blocking network fabrics capable of supporting the most demanding AI clusters.
- **Reliability:** Through optimization of the electrical channel, thermal management, and overall component architecture, XPO improves reliability per transmitted bit, reducing job interruptions and increasing overall system availability.
- **Cooling:** The XPO module incorporates an integrated liquid-cooled cold plate, providing an efficient thermal solution for liquid-cooled data center environments and enabling direct heat removal from high-power optical components.
- **Power:** The design leverages a high-quality linear interface channel, improving signal integrity and enabling lower power consumption compared to conventional optics that rely on complex and power-intensive signal conditioning.
- **Density:** XPO delivers 8× the bandwidth of OSFP, 4X front panel density of OSFP, enabling up to 204.8Tbps of switching throughput per Open Rack Unit (1OU) and establishing a new benchmark for pluggable optical module density.

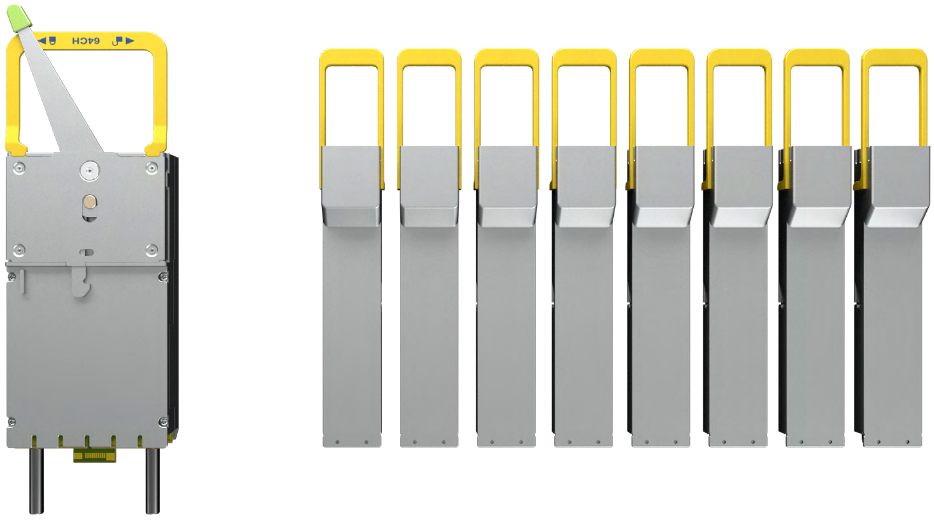

Figure 1: Comparison of OSFP to XPO highlighting the density improvement of XPO compared to 8X OSFP

Mechanical Architecture

The "Belly-to-Belly" Design To achieve an 8x increase in bandwidth per module compared to standard 1.6Tbps OSFP, XPO abandons traditional single-PCB layouts.

1. **Dimensions & Pitch:** The module features a compact mechanical form factor with dimensions of 60.8 mm (width) × 111.8 mm (length) × 21.3 mm (height), enabling high front-panel density.
2. **Dual Paddle Cards:** Inside the module shell, the XPO architecture contains two separate 32-channel printed circuit boards (PCBs), referred to as paddle cards.
3. **Belly-to-Belly Layout:** These two identical cards (Card 1 and Card 2) are arranged in a "belly-to-belly" configuration, facing inward toward a shared central element. High-power, heat-generating components (such as transmit circuitry and laser drivers) are assembled on the inward-facing "hot" side of the PCBs, while lower-power components (like receive circuitry and control logic) are mounted on the outward-facing "cold" side.
4. **Ejector Mechanism:** Due to the high number of electrical contacts, insertion and removal require significant force. The XPO module utilizes a mechanical ejector with a release pull tab that provides a 1:11 mechanical leverage to assist operators during insertion, engaging smoothly with the host cage.

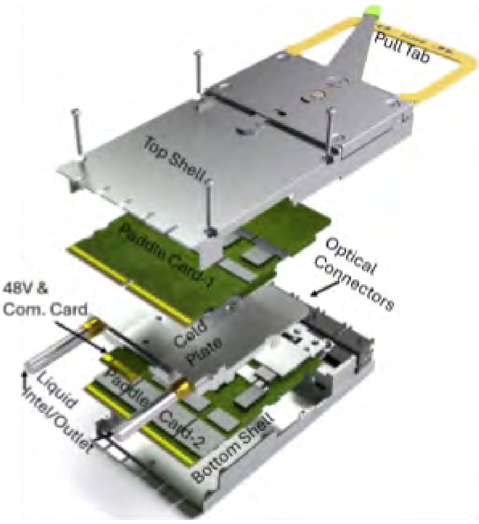

Figure 2: Exploded view of XPO

Integrated Liquid Cooling

A defining innovation of the XPO module is its native liquid cooling. Instead of relying on thermal interface materials (TIM) and cage-mounted riding heatsinks, XPO brings the cooling directly inside the module.

1. **Central Cold Plate:** A liquid cold plate is physically sandwiched between the "hot" sides of the two paddle cards, cooling both circuit boards simultaneously.
2. **Thermal Capability:** This design efficiently removes heat from high-power modules exceeding 400W, easily supporting intense applications like 8 of the 1.6Tbps ZR optics inside the XPO module. By utilizing 40-45°C warm-water liquid cooling, XPO keeps component temperatures 20°C to 25°C lower than air-cooled equivalents.
3. **Fluid Interface:** The module features integrated fluid channels and connects to the host system's cooling manifold via blind-mate, quick-disconnect liquid connectors. These dripless connectors are rated for 500 mating cycles and support dynamic flow rates ranging from 0.25 Liters Per Minute (LPM) for lower-power modules to 0.7 LPM for high-power modules.

Electrical Interface and 50V Power Delivery

The XPO module utilizes 64 high-speed electrical lanes to deliver 12.8Tbps using 200Gbps PAM4 signaling, with a roadmap to support 25.6Tbps via 400Gbps lanes. To maximize signal integrity and power efficiency, the electrical design is strictly segregated:

1. **Clean Linear Channel:** High-speed Transmit (Tx) and Receive (Rx) signals are separated onto opposite sides of the paddle cards to minimize crosstalk, providing an optimized linear channel ideal for Linear drive Pluggable Optics (LPO).
2. **Dedicated Power/Control Connector:** To prevent power supply noise from coupling into high-speed data lanes, power and low-speed control signals (such as I2C/I3C, Reset, and Interrupts) are routed through a completely separate, dedicated card-edge connector located in the center of the module.
3. **50V DC Architecture:** Traditional pluggable rely on a 3.3V DC input, which draws massive current for high-power optics. XPO introduces a 46V to 53V DC input (nominally 48V or 50V) directly from the rack's bus bar. This high-voltage input significantly reduces the required current rating and the physical size of the power connectors.

4. **Motherboard Simplification:** By utilizing on-board 48V-to-3.3V voltage regulators located directly on the module's paddle cards, XPO eliminates the need for bulky, worst-case provisioned voltage regulators on the switch motherboard, maximizing overall system reliability.

Comparative Analysis: Quantifying the XPO Advantage Over OSFP

To fully appreciate the architectural impact of the XPO module, a direct, data-driven comparison with the incumbent OSFP standard is essential. This analysis evaluates the performance differences at both the individual module level and the rack level, illustrating how innovations at the component level translate into system-wide efficiency gains.

At the most fundamental level, XPO delivers a significant increase in front-panel bandwidth density. To achieve a total switching throughput of 204.8Tbps, an XPO-based system requires only one-quarter of the rack space compared to an equivalent OSFP-based deployment. This represents a clear 4× density improvement over OSFP, enabling network architects to build substantially more powerful fabrics within the same physical footprint.

Figure 3: Illustration of 204.8Tbps switch with XPO (top) and OSFP (bottom) showing the 4X density improvement

When this density advantage is applied at the system level within a standard ORv3 (HPR) liquid-cooled rack [4], the benefits become even more significant. The following table compares a fully populated rack based on each optical standard. This comparison reveals an important insight into total cost of ownership (TCO). Liquid cooling infrastructure represents a substantial capital investment, and to justify this expense, rack deployments must typically target 120 kW or higher power density.

An OSFP-based rack, with a maximum power draw of approximately 32kW, significantly underutilizes the available cooling infrastructure. In contrast, an XPO-based rack, operating at approximately 128kW, fully leverages the rack's liquid-cooling capability. This allows the cooling and power delivery infrastructure to be efficiently amortized across a much larger computational payload. This is summarized in Table 1.

Table 1: Rack level comparison between OSFP and XPO

Metric	OSFP-based Rack	XPO-based Rack	Delta
Bandwidth per Rack	1.6 Pbps	6.5 Pbps	4X
Number of 404.8T Switches	8	32	4X
Power per Rack	32kW	128kW	4X

As a result, XPO enables data center operators to deliver 4× greater network capacity within the same infrastructure footprint. By dramatically increasing usable payload within the rack, the relative cost of supporting infrastructure—such as power delivery, cooling systems, and rack space—is effectively reduced. This improvement extends beyond component-level efficiency and begins to reshape the overall economics of large-scale AI data center deployments.

Redefining Data Center Scale: Footprint, Cost, and Efficiency Gains

The true value of an architectural innovation such as XPO is realized when component-level improvements translate into system-level benefits across the data center. The 4× density advantage fundamentally alters data center design by enabling significant reductions in physical footprint, capital expenditure, deployment time, and operational complexity.

Consider an AI cluster of 512 XPUs (e.g., GPUs or other accelerators) connected in a scale-up domain. Assuming 25.6Tbps of bandwidth per XPU, this scale-up domain would require 64 switches with 204.8Tbps capacity each. Using the incumbent OSFP technology, which provides 204.8Tbps per 4 rack units, the network would require eight switch racks to deliver the necessary connectivity. In contrast, an XPO-based architecture can support the same cluster using only two switch racks, since XPO enables 204.8Tbps switching capacity within a single rack unit, delivering a 4× density improvement.

At hyperscale, this efficiency gain has profound implications. Consider a 400 MW AI data center supporting 128,000 XPUs. In this scenario, we assume a scale-up network with 12.8Tbps per XPU and a scale-out network with 1.6Tbps per XPU, with the accelerators interconnected through a three-tier Clos topology. Under these conditions, the switching capacity per rack is approximately 1.64Pbps with OSFP compared to 6.55Pbps with XPO, reflecting the substantial density advantage provided by the XPO architecture.

Network Tier	OSFP Switch Racks	XPO Switch Rack	Racks Saved
Scale-Up	1024	256	768
Scale-Out	384	96	288
Total	**1408**	**352**	**1056**

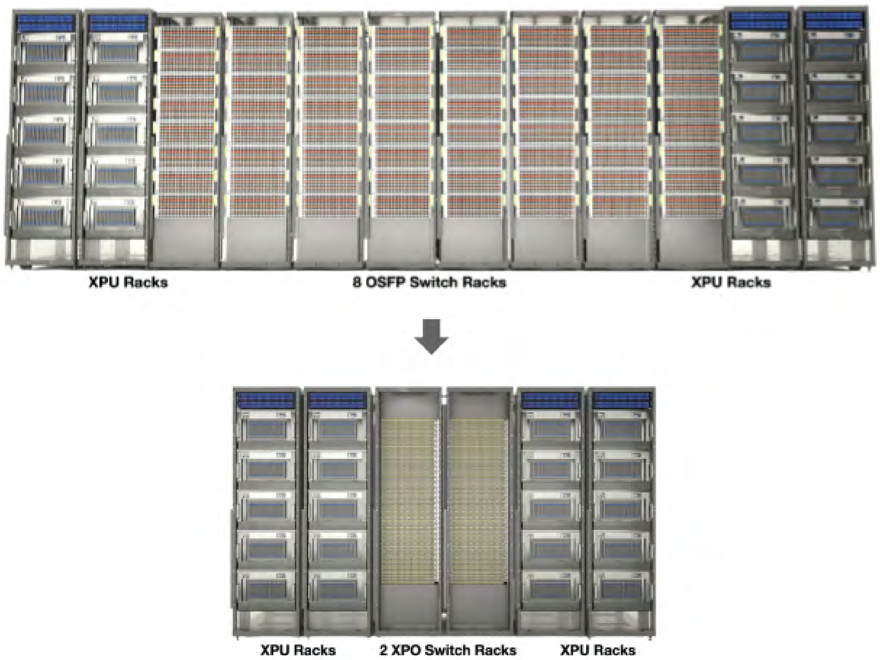

Figure 4: Illustration of the reduction in the switch rack footprint by 75% between OSFP and XPO

This analysis indicates approximately a 75% reduction in required switch rack footprint, along with corresponding reductions in electrical infrastructure, cooling capacity, and plumbing requirements. The impact on capital expenditure can therefore be substantial. For hyperscale AI facilities requiring billions of dollars in construction costs, such improvements could significantly reduce the number of required buildings or enable far greater compute capacity within the same facility footprint.

Alternatively, operators with existing data centers can leverage this density advantage to increase accelerator density per building, maximizing the utilization of existing infrastructure and real estate. In addition, the higher-radix switches enabled by XPO allow for simpler scale-out network topologies with fewer tiers and lower round-trip latency, directly improving the performance and efficiency of large-scale AI training workloads.

Core Innovations and Platform Versatility

XPO's value stems from a series of engineering decisions that optimize its mechanical, thermal, and electrical design for the specific demands of hyperscale AI data centers, while maintaining broad compatibility with existing and future industry standards.

1. **Use of Existing Technology:** XPO supports increased capacity per module by leveraging existing photonic and silicon chipset technologies. This approach reduces adoption risk and allows the ecosystem to build upon mature, reliable, and cost-effective manufacturing processes.
2. **Integrated Cold Plate:** XPO incorporates native liquid cooling through a cold plate embedded between two paddle cards arranged in a belly-to-belly configuration. This design enables highly efficient heat transfer from the optical components and DSPs directly to the liquid cooling system.
3. **Clean Linear Channel:** Superior signal integrity is achieved using CPC fly-over cables and an optimized edge-connector pin-out. This clean, low-loss electrical channel reduces the need for power-intensive digital signal processing (DSP), contributing to lower overall power consumption.
4. **Power Efficiency:** In addition to enabling a linear channel architecture, XPO improves power delivery efficiency by utilizing the 50V DC bus-bar voltage directly as the module supply, minimizing power conversion losses within the system.
5. **Improved Reliability:** Reliability is enhanced through a combination of factors, including reduced component count, lower operating temperatures enabled by the integrated cold plate, minimized temperature variations, and improved signal integrity through the optimized electrical channel.
6. **High Density:** The XPO module density is achieved by optimizing the module's physical dimensions for maximum optical density using MPO-16 connectors. This configuration also aligns with the highest density available in high-speed electrical system connectors, enabling efficient routing and packaging. This pragmatic physical design is a key factor in achieving the 4× density improvement over OSFP.

In addition to these core innovations, the XPO platform was designed for maximum flexibility, allowing it to adapt to evolving optical technologies and future industry standards.

Conclusion: Enabling the Next Generation of AI Networking

The rapid expansion of artificial intelligence workloads is redefining the performance requirements for modern data center networks. AI clusters demand unprecedented bandwidth, higher reliability, efficient liquid cooling integration, improved power efficiency, and significantly greater front-panel density than traditional optical interconnect technologies were designed to deliver. These emerging requirements are difficult to meet using conventional pluggable solutions such as OSFP.

The XPO architecture addresses these challenges through a purpose-built pluggable module optimized for hyperscale AI infrastructure. By combining a dual-paddle mechanical architecture, integrated liquid-cooling cold plate, clean linear electrical channel, and high-voltage power delivery, XPO dramatically increases optical density while maintaining the operational flexibility and serviceability advantages of pluggable optics.

With 12.8Tbps per module and up to 204.8Tbps per 1OU switch, XPO delivers a 4× improvement in front-panel density compared to OSFP, enabling approximately 75% reduction in switch rack footprint while significantly lowering infrastructure cost and network complexity. At hyperscale, these improvements translate into substantial gains in capital efficiency, operational simplicity, and overall system performance.

By rethinking optical module architecture from the ground up, XPO provides a scalable foundation for next-generation AI networking infrastructure, enabling data center operators to build higher-capacity, more efficient, and more reliable networks capable of supporting the rapidly growing demands of artificial intelligence.

References

[1] https://osfpmsa.org/

[2] Unlocking the Potential of Co-Packaged Optics in AI and HPC: Opportunities and Challenges, Sunil Priyadarshi, IEEE Communication Magazine, February Edition, 2026, https://ieeexplore.ieee.org/document/11303304

[3] https://xpomsa.com

[4] https://www.opencompute.org/wiki/Open_Rack/SpecsAndDesigns

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office
1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062