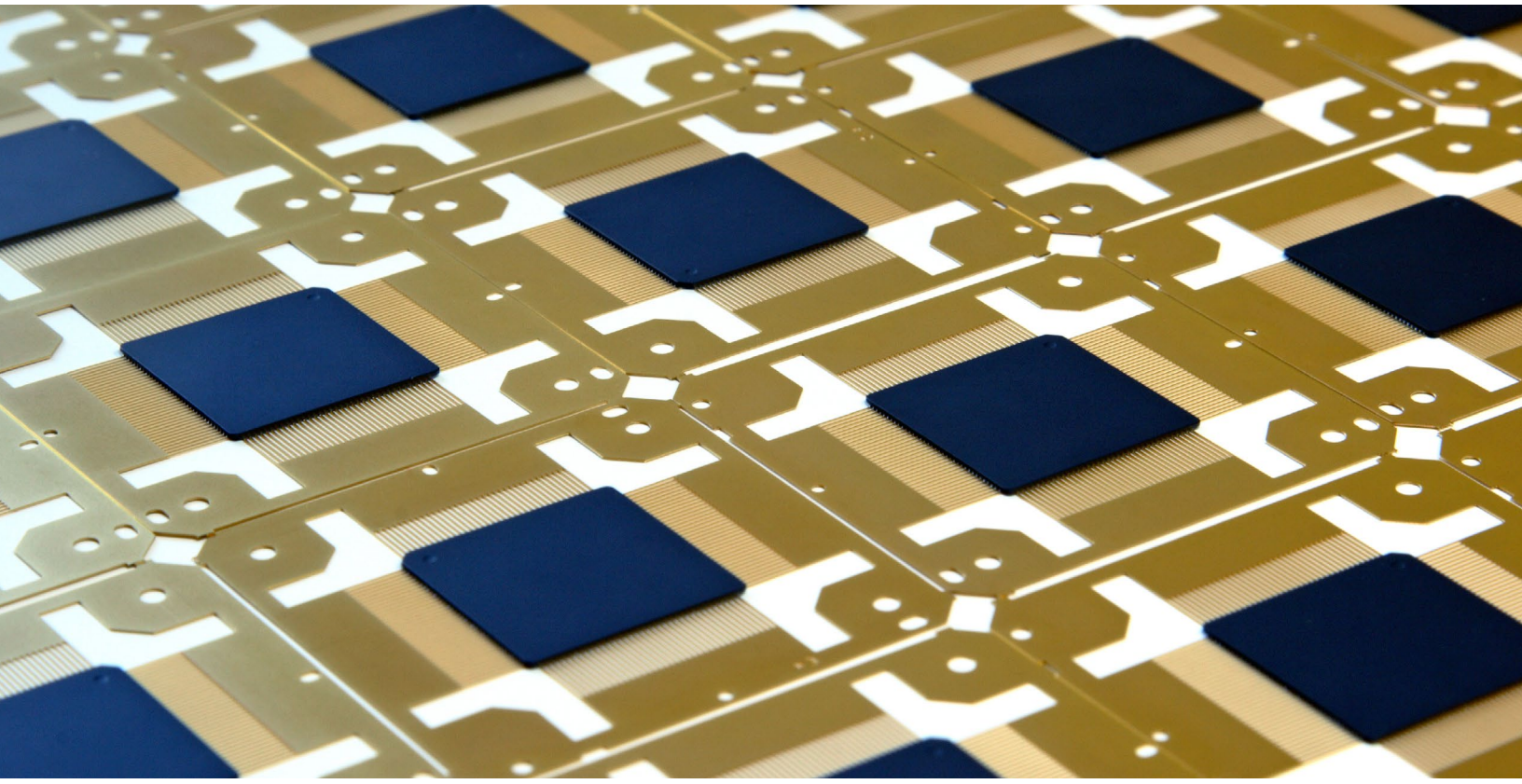
McKinsey & Company

McKinsey Direct

Opportunities in networking optics: Boosting supply for data centers

Potential shortfalls in networking optics supply could hinder data center and AI expansion. How can players boost supply and seize growth opportunities, especially in high-bandwidth transceivers?

This article is a collaborative effort by Bill Wiseman, Henry Marcil, Lari Hämäläinen, and Pankaj Sachdeva, with Wendy Zhu, representing views from McKinsey's Industrials & Electronics; Technology, Media & Telecommunications; and Semiconductor Practices.



June 2025

Networking optics are mission-critical components in advancing AI and data center infrastructure build-outs. To move the massive amounts of data required for AI training and inferencing, hyperscalers and other data center operators need increased access to optical transceivers capable of rates of 800 gigabits per second (Gbps) or even 1.6 terabits per second (Tbps) and higher.[1] But these crucial components could be in short supply in the near term: Production of 800-Gbps transceivers is expected to fall 40 to 60 percent short of demand through 2027, and shortfalls of 30 to 40 percent in the supply of 1.6-Tbps transceivers are likely to persist through 2029, according to McKinsey analysis.

Addressing these shortages will require industry standards, collaboration, capacity scaling, and significant investment of capital expenditure to ensure the necessary capacity by 2030.[2] The amount of capital expenditure needed depends on where the data center is located and the type of construction involved—retrofitting, new build, or expansion. This article explores current trends in networking optics technology as well as the market factors affecting production of higher-data-rate transceivers . It illustrates how industry leaders could act today to ease bottlenecks in component supplies and help enable the continuous build-out of the infrastructure needed to support growing AI adoption.

As AI market growth skyrockets and technology advances, industry dynamics are in flux

As the AI market continues to boom, compute power—the hardware, processors, memory, storage, and energy needed to operate data centers—is emerging as a critical global resource. By 2030, data centers will likely require $6.7 trillion worldwide to keep pace with the demand for compute power.[3]

While compute performance has continued to scale year over year, in part because of advances in GPUs and the proliferation of AI-specific application-specific integrated circuits, networking performance has not grown at a similar pace. But AI workloads require massive data movement to train models and deliver real-time inferences. Networking has thus become a pivotal differentiator for hyperscalers and all AI industry customers. Consequently, supply shortages and performance shortfalls in networking optics affect all data centers, regardless of their configuration (see sidebar, "Understanding networking optics and data center optics configurations").

Networking optics market growth by segment
The rapid growth in the networking optics pluggable-transceiver market is largely spurred by hyperscalers and telecom, especially as AI adoption increases.[4] According to McKinsey analysis, this market could expand at 14 percent CAGR through 2029 when, amid larger industry growth, the market for intra–data center (both front end and back end) optics supporting data rates greater than 1.6 Tbps could expand significantly, exceeding a 30 percent CAGR and likely reaching about $24 billion.[5] McKinsey analysis suggests that hyperscalers will shift about 87 percent of their back-end optics to 800 Gbps and greater by 2029, with 1.6-Tbps transceivers accounting for more than 40 percent of this demand (Exhibit 1).

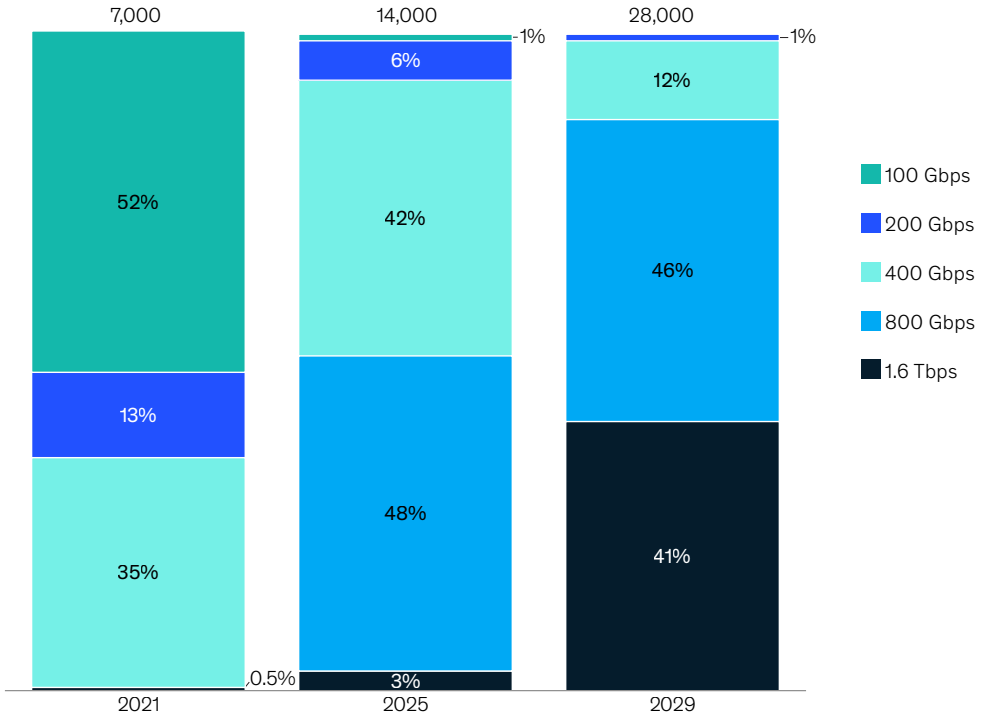
Additionally, data center interconnect (DCI), particularly front-end metro with inter–data center connectivity of 80 kilometers or less, is expected to see the greatest near-term growth. This is because high energy demands are challenging the feasibility of large-scale single data centers, leading to the emergence of multiple data centers working in tandem and driving DCI growth at a projected 34 percent CAGR through 2029.[6] Telecom companies are also decommissioning legacy networking technologies

[1] "How data centers and the energy sector can sate AI's hunger for power," McKinsey, September 17, 2024.
[2] "The cost of compute: A $7 trillion race to scale data centers," *McKinsey Quarterly*, April 28, 2025.
[3] "The cost of compute: A $7 trillion race to scale data centers," *McKinsey Quarterly*, April 28, 2025.
[4] "AI infrastructure: A new growth avenue for telco operators," McKinsey, February 28, 2025; "How data centers and the energy sector can sate AI's hunger for power," McKinsey, September 17, 2024; Vladimir Kozlov, *April 2024 market forecast*, LightCounting, April 2024; *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024.
[5] "AI power: Expanding data center capacity to meet growing demand," McKinsey, October 29, 2024; demand for intra–data centers, metro inter–data centers, and long-haul data centers, projected; *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024; McKinsey analysis; *Optical transport report* (*front end*), Dell'Oro, 2024.
[6] *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024. Growth is calculated for 400-Gbps or faster transceivers. The weighted average is approximately 34 percent.

Exhibit 1

High-bandwidth transceiver demand will continue to grow and shift dramatically through 2029.

Hyperscaler transceiver demand projection for pluggable, by bandwidth, thousand units



Source: *April 2024 market forecast*, LightCounting, April 2024; *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024; McKinsey analysis

McKinsey & Company

such as time-division multiplexing, private internet protocol, and direct internet in favor of wave technology. This transition requires high-bandwidth transceivers that can accommodate traffic growth as well as traffic aggregation and technology changes.

The front-end metro optics market is undergoing rapid expansion fueled by the adoption of coherent zero dispersion shifted range (ZR/ZR+) transceivers with high data rates. Historically, coherent transceivers have been widely used in the telecommunications industry at data rates of 400 Gbps or less.[7] However, because AI training and inference require low-latency interconnects, data center operators are deploying high-speed coherent transceivers and investing in high-speed fiber infrastructure to connect data center clusters. About half of the metro segment uses 400-Gbps ZR and ZR+ transceivers; by 2029, the market could transition quickly, with 800-Gbps ZR and ZR+ transceivers and above capturing about 70 percent of market share.[8] Meanwhile, because of cost sensitivity, traditional telecom end users are more likely to use solutions with lower data rates rather than following data centers in adopting advanced digital signal processing (DSP) technology. As the front-end metro optics market continues to expand rapidly, new OEM players are entering the segment, led by laser suppliers' vertical integration.

[7] Bo Gowan, "Coherent optical turns 10: Here's how it was made," Ciena, March 9, 2018.
[8] *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024.

Understanding networking optics and data center optics configurations

Networking optics includes optical transceivers within data centers, coherent zero dispersion shifted range (ZR/ZR+) optical transceivers, and long-haul transponders. Optical transceivers convert electrical signals to optical signals and vice versa, enabling high-speed communication between networking devices over fiber-optic links.

Optical transceivers depend on a variety of components and technologies:

- *electro-absorption modulated laser:* integrated indium phosphide (InP) edge-emitting laser and electro-absorption modulator (EAM) on a single laser chip
- *silicon photonics (SiPh):* a technology that integrates all optical components, except laser diodes, onto one silicon chip; few companies are able to fully integrate optics with lasers in SiPh
- *digital signal processing (DSP):* a silicon-based chip that processes and improves incoming optical signals

Transceivers include the following types:

- *Pluggable.* These are modular transceivers that plug into networking devices. They are known for ease of implementation and serviceability.
- *Active optical cables*. A preassembled system consisting of two transceivers connected by a fiber optic cable, active optical cables typically are lower-cost and are used for shorter distances due to fixed-length cables.
- *Co-packaged optics (CPO).* This type of transceiver integrates the optical engine directly within the logic or networking chip, significantly reducing the distance data must travel—thereby improving energy efficiency and increasing data rates. Future versions of CPO aim to be completely integrated by adding a laser diode onto the chip. Current technologies rely on pluggable lasers for CPO.

Optics technology has advanced rapidly since 2018 with the adoption of 400-gigabit-per-second (Gbps) transceivers, which is the baseline for high-speed data centers. Leading-edge tech is quickly becoming mainstream tech. In addition to 800-Gbps and 1.6-terabit-per-second (Tbps) bandwidth transceivers, advancements include the following:

- *Coherent ZR transceivers.* These transceivers enable high-capacity, power-efficient transmission at or above 400 Gbps through the use of coherent optics to maximize the performance of existing pluggable technology for longer distances.
- *Multiplexer (MUX) and demultiplexer (DEMUX) line systems*. These systems greatly increase the capacity and efficiency of long-haul line systems. A MUX aggregates multiple incoming data streams into one multiwavelength signal that can travel through one fiber. A DEMUX then disaggregates the multi-wavelength signal back into individual data streams.
- *Coherent CPO.* Coherent CPO significantly reduces latency and increases energy efficiency through the integration of optics with chips in a compact form factor. These elements are critical for large-scale data centers where energy consumption and overall density are a concern.
- *Coherent DSP.* Maximizing the performance of optical networks for longer distances, this technology combines DSP with coherent detection for increased accuracy and characterization of incoming signals, which enhances signal quality and capacity.
- *Coherent-lite optics.* This technology delivers cost-effective, lower-power coherent transmission for shorter reaches—such as data center interconnect and metro inter–data center optics (front end)—by simplifying coherent designs while retaining key performance benefits.

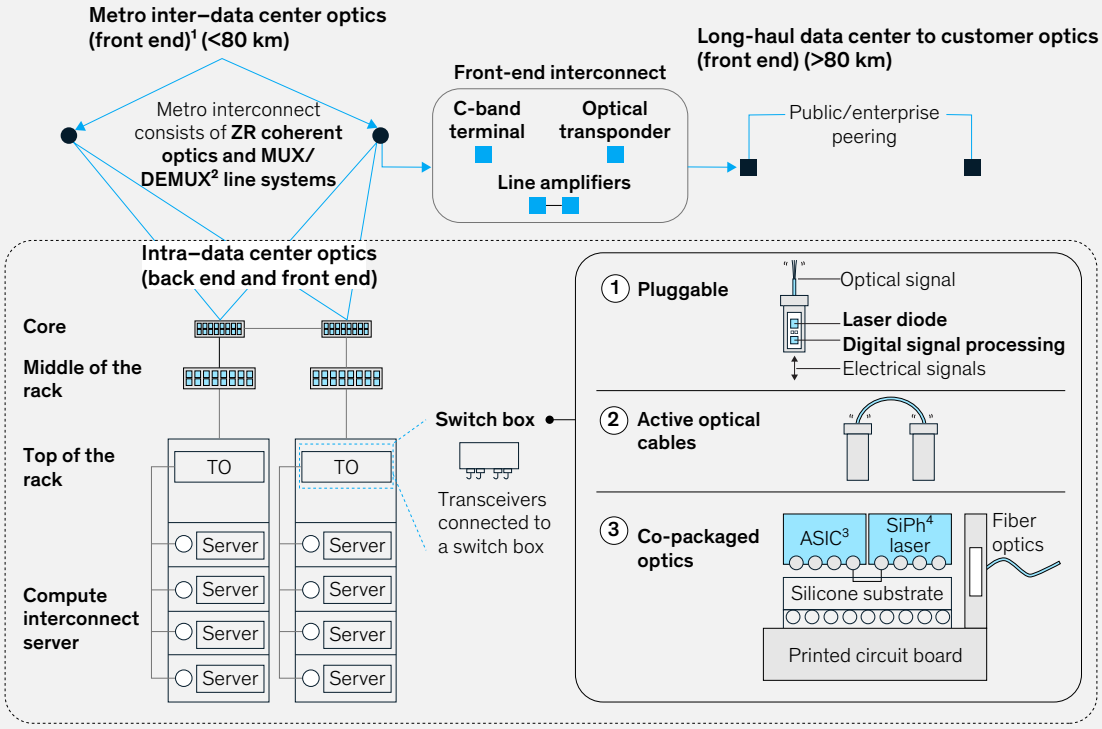
Data center optics configurations and functionality vary depending on the customers they serve, services they provide, and where they are located. Generally, these configurations can be divided into three types:

- *intra–data center optics (back end and front end):* optical links used within a single data center, including connections between servers, switches, and racks
- *metro inter–data center optics (front end):* optics connecting data centers located within 80 kilometers or less of each other
- *long-haul data center to customer optics (front end):* optics connecting data centers to end users or remote sites over distances greater than 80 kilometers

Optical transceivers and their various components are integral to supporting capacity and performance within various configurations for data center optics (exhibit).

Exhibit

Data center networking can generally be divided into three segments.



[1]Data center.
[2]Multiplexer/demultiplexer line systems.
[3]Application-specific integrated circuit.
[4]Silicon photonics.
Source: *Optical transceivers for data and telecom 2024*, Yole Group, 2024; McKinsey analysis

McKinsey & Company

Shortfalls in electro-absorption modulated lasers could be eased to some extent by ramping up silicon photonics and continuous wave laser production.

Potential bottlenecks in networking optics production

Electro-absorption modulated lasers built on indium phosphide (known as InP-based EML) are used widely in transceivers of all bandwidths. Production of InP-based EML for data center optics is concentrated among three to five players globally. Manufacturing processes are highly specialized and complex, with historically low yields for leading-edge and high-speed EML designs—particularly in early production. According to McKinsey analysis, yields can range from 50 percent to as low as 15 percent because of challenges in wafer-level fabrication, precise optical alignment, packaging complexity, and tight performance tolerances required at higher wavelengths. As demand for high-data-rate optics accelerates, InP-based EML supply is increasingly constrained. Shortfalls could reach 17 percent in 2026, and shortages are expected to continue into the second half of 2027 (Exhibit 2).[9]

Such shortfalls could be eased to some extent by ramping up silicon photonics (SiPh) and continuous-wave (CW) laser production to address growing demand for transceivers of 1.6 Tbps and above.

Exhibit 2

Supply dynamics of individual laser technologies will have the greatest impact on high-bandwidth optical transceiver supply.

View of pluggable transceiver supply sufficiency, 2024–29, thousand units

Undersupply ◄ ► Oversupply

By laser technology

	2024	2025	2026	2027	2028	2029
VCSELs[1]	1,200	–2,500	–2,000	–900	–1,200	–3,500
EML[2]	500	–400	–2,000	–1,500	1,000	2,200
DML[3]	400	2,500	4,000	6,500	7,500	9,000
SiPh[4]	1,200	3,300	3,300	4,000	3,500	5,000
Total	3,300	2,900	3,300	8,100	10,800	12,700

By bandwidth

	2024	2025	2026	2027	2028	2029
≤200 Gbps	1,500	1,300	3,200	5,000	3,500	1,000
400 Gbps	5,500	7,000	7,500	9,000	11,000	12,000
800 Gbps	–3,500	–5,000	–6,000	800	4,500	10,000
1.6 Tbps	–10	–200	–1,500	–7,000	–8,000	–8,000
Total	3,490	3,100	3,200	7,800	11,000	15,000

[1]Vertical-cavity surface-emitting laser.
[2]Electro-absorption modulated laser.
[3]Directly modulated laser.
[4]Silicon photonics.
Source: McKinsey proprietary Networking Optics Model; World Fab Forecast, SEMI, 2021–24; "Silicon Photonics 2024 - focus on SOI, SiN, and LNOI platforms," Yole Group, November 2024; McKinsey analysis

McKinsey & Company

[9] McKinsey proprietary Networking Optics Model; World Fab Forecast, SEMI, 2021–24; "Silicon Photonics 2024 - focus on SOI, SiN, and LNOI platforms," Yole Group, November 2024; *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024; McKinsey analysis.

When used in SiPh, CW lasers can replace multiple discrete InP lasers; in fact, one CW laser can power two to eight (or more) lanes, depending on the architecture. Reducing the number of required components by combining SiPh and CW lasers could help ease the shortage. However, because the SiPh ecosystem has yet to mature and SiPh is currently more expensive to manufacture than InP-based EML, end customers' spending on optics could increase. Alternatively, addressing shortages by 2029 by expanding existing InP-based EML fabs or building new facilities could take at least five to seven years and require an investment of approximately $500 million to $2 billion, depending on which approach to capacity expansion is chosen.[10]

Vertical consolidation and global shifts in the networking optics supply chain

In the back-end intra–data center segment, competition among suppliers has heated up as AI adoption rises and more OEMs in the industry vertically integrate (via M&A and strategic partnerships) to secure critical laser and DSP supplies. For example, because leading US-based OEMs are acquiring laser fabrication facilities to enable faster time to market, only one or two Japan-based laser suppliers with mass production capability remain available to other OEMs for potential outsourcing.

The OEM market is concentrated, with about 80 percent of market share divided among 11 OEMs.[11] Most laser suppliers are located in the United States or Japan, though China-based suppliers have increased their share of the back-end pluggable market rapidly, growing by about 20 percentage points between 2017 and 2023 to roughly 60 percent. Notably, most OEMs in China still rely heavily on laser and DSP supplies, and their in-house capabilities are largely in assembly and packaging.

Despite profoundly shifting market dynamics, the industry's geopolitical landscape appears relatively balanced. Many critical component manufacturing facilities have expanded capacity in Southeast Asia or Europe, and all major vendors in China have at least one assembly and packaging site outside of China. For example, about 75 percent of fabrication sites are based in the United States and Europe, and assembly and packaging operations are concentrated in Southeast Asia, with one-third of all sites located in China.

Nonetheless, given existing and potential future changes to tariffs, customers and players throughout the value chain should consider how to optimize their supply chain footprint to ensure cost efficiency and sustained competitiveness in global markets. Additionally, major back-end intra–data center OEMs have engaged in horizontal integration and expanded their portfolios to include the front-end metro market segment, where they are poised to increase their market share by 20 percentage points between 2024 and 2027. Price competition has heightened as more OEMs have entered the market segment and expanded available product options.

The following three factors have helped more players enter the front-end metro market segment:

- The industry's standardization of ZR transceivers (provided via standards introduced by the Optical Internetworking Forum and Multi-Source Agreement) expanded the front-end market segment beyond proprietary solutions.
- Hyperscalers have supported their back-end suppliers (particularly those with laser supply capabilities) in expanding their offerings to include front-end metro services and have benefited from reduced time to market and sufficient supply.
- When optical systems integrators made critical subcomponents, such as coherent DSPs, available on the open market, new entrants were able to access previously inaccessible technology. The resulting influx of new players has, in turn, fostered increased competition and innovation.

As the optical supply chain continues to evolve, vertical consolidation, regional diversification, and new market entrants will reshape competitive dynamics, driving further shifts in the industry landscape.

[10] Calculated based on data for current fab capacity and size; World Fab Forecast, SEMI, 2021; company press releases to determine fab expansion and capacity expansion and cost. Cost per three-inch-equivalent wafer for EML lasers determined from World Fab Forecast and press releases multiplied by average EML three-inch wafer gap from McKinsey proprietary Networking Optics Model.

[11] *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024.

AI networking's impact on demand for pluggable versus co-packaged optics

Pluggables will continue to meet AI-driven demand over the next five years because they offer competitive performance and total cost of ownership benefits. However, emerging technologies such as co-packaged optics (CPO) are gaining momentum to disrupt the market as power efficiency and performance requirements intensify. Some projections suggest that CPO could lower data center energy consumption by up to 30 percent while enabling bandwidths of 3.2 Tbps and beyond, supporting a broad range of next-generation optical products in the next ten years.[12]

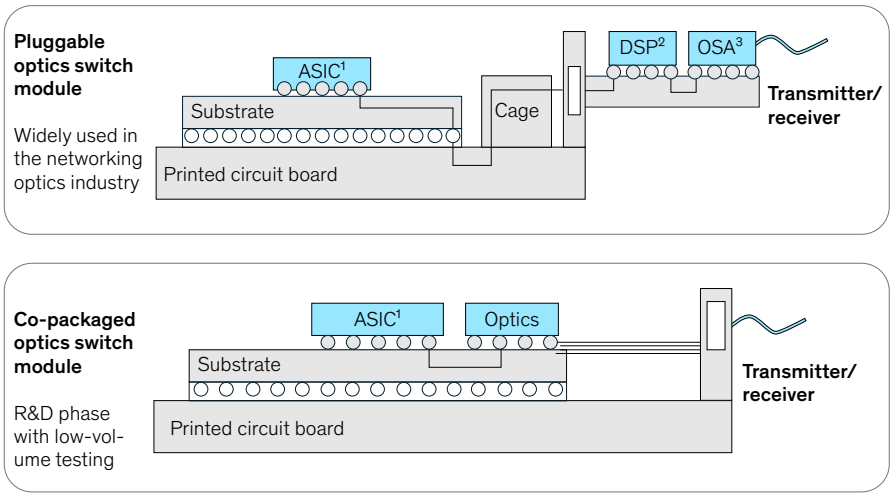
As the industry shifts, a wave of new entrants—incumbents and start-ups—will likely compete in the CPO space. Nonetheless, CPO faces challenges in reliability, thermal management, and a lack of interoperability standards. And if an integrated optical module fails, replacing the entire switch is costly (Exhibit 3).

Additionally, because integrated optics and electronics can generate heat at levels that traditional cooling methods may not be able to dissipate, CPO faces manufacturing hurdles—particularly in packaging and assembly—that must be overcome to ensure reliability, consistent performance under fluctuating thermal conditions, stable connectivity, and compatibility among components integrated into a system. Setting industry standards for system-level integration of CPO is essential to increase CPO device yield and accelerate CPO adoption.

Exhibit 3

Because optical engines are integrated in co-packaged optics, failing modules may need to be replaced rather than repaired.

Pluggable vs co-packaged optics switch modules, illustrative



[1]Application-specific integrated circuit.
[2]Digital signal processor.
[3]Optical subassembly.
Source: Broadcom; *Optical transceivers for data and telecom 2024*, Yole Group, 2024; McKinsey analysis

McKinsey & Company

[12] Broadcom; *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024; "Delta showcases advanced AI data center networking solutions with co-packaged optics (CPO) Ethernet switch and next-generation switch design at COMPUTEX 2025," Delta, May 15, 2025; "Micas Networks announces industry's first 51.2T co-packaged optics network switch system now in volume production," Micas Networks, March 17, 2025.

Various solutions are emerging, with companies focusing on optimizing specific CPO components or improving interconnectivity between critical elements. For example, several start-ups are focusing on improving overall power efficiency through integrated quantum dots and vertical-cavity surface-emitting lasers, while others are working on addressing yield challenges through novel alignment and manufacturing improvements.[13] But while large-scale CPO adoption is unlikely until 2035, the CPO supplier segment could still see a CAGR of 20 percent through 2029.[14] Data centers began testing 1.6-Tbps CPO as a proof of concept in 2024, and acceleration in adoption hinges on improvements in yield, reliability, and cost.

Emerging trends in data center networking and operations

As hyperscalers and data center operators adapt and expand to accommodate rapid growth in AI adoption while improving their efficiency and reducing their total cost of ownership, clear trends are emerging in terms of their top priorities, strategies, and use of particular systems and solutions.

— *Interoperability as a strategic priority*. Ecosystem diversification has renewed data centers' focus on interoperability, which has received strong backing from industry consortiums such as the Optical Internetworking Forum and the Institute of Electrical and Electronics Engineers. Multivendor demos are on the rise, showcasing seamless interoperability across CPO interfaces, pluggable modules, and host systems in 800-Gbps and 1.6-Tbps modules.

— *Hyperscalers embrace optical circuit switching*. Optical-native, cost-efficient architectures are taking center stage. Optical circuit switching is gaining significant traction as a scalable solution for DCI that offers ultralow latency and high throughput.

— *Low-power optics and DSP-free designs are gaining ground*. For short-reach applications, low-power optics and DSP-free designs offer compelling energy savings and easier interoperability. These approaches are becoming strong alternatives to CPO, though CPO could still become the leading design solution if its integration and yield challenges are resolved.

— *High-radix switching enables network flattening*. Simplified network layers—enabled by CPO—help reduce power and latency, with fewer switch ports delivering the same bandwidth. This supports growing demand for scale and efficiency.

— *Shifting from InfiniBand to Ethernet to lower total cost of ownership*. Data centers are increasingly focusing on maximizing their hardware utilization to reduce their operational costs. Many hyperscalers, for example, are transitioning from InfiniBand to Ethernet to enhance scalability and cost efficiency, though adoption of the latter is hindered by the high cost of switching.

— *Hyperscalers are hyper focused on optimizing costs throughout the networking stack*. Hyperscalers are doubling down on realizing efficiencies in their total cost of ownership. Intensifying price negotiations with OEM and subtier suppliers could lower hyperscalers' spend on networking hardware by 10 to 30 percent, for example. Hyperscalers may also replace OEMs with original-design manufacturers (ODMs) to lower their capital expenditure, seek lower-cost vendors for subcomponents, or redesign their optics using lower-cost reference designs.

Moving forward, the trajectory of AI-driven data center expansion could be determined by the optics industry's efforts to address supply chain gaps and make technological breakthroughs today.

[13] Holly, "Silicon photonics for AI: Balancing cost, power, and reliability," NADDOD, September 6, 2024.
[14] *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024.

Technology evolution in networking optics will require extensive ecosystem collaboration among OEMs, component suppliers, and industry-wide standardization bodies.

Looking ahead, networking optics needs to keep pace with data center demands

The increasing power demands and need to reduce latency per data center for next-generation AI training models could prompt a fundamental shift in data center architecture.[15] As securing large-scale power supplies in a single location becomes increasingly difficult, hyperscalers turn to distributed, smaller-scale data center clusters or composable infrastructure architecture to optimize resource utilization and power efficiency through dynamic workload allocation. Public telecom networks, too, are increasingly focused on data centers. The feasibility of such transitions could hinge on striking a delicate balance between the total cost of ownership (namely, potentially prohibitive costs to upgrade ZR transceivers and full front-end metro infrastructure) and the urgency of mitigating power shortages.

Importantly, technology evolution in networking optics will require extensive ecosystem collaboration among OEMs, component suppliers, and industry-wide standardization bodies. As optics technology advances beyond 1.6 Tbps, multivendor modules for transceiver components will likely become the norm. Thus, enabling the seamless integration of different components will be essential.

The networking optics industry must work quickly and collaboratively to recalibrate its ability to address imminent supply shortfalls and untangle manufacturing challenges to avoid becoming a bottleneck in the AI-driven expansion of data centers.

As hyperscalers intensify their focus on lowering the total cost of ownership, opportunities may emerge for ODMs, providers of electronics manufacturing services, and long-tail OEMs to compete on module-level assembly and packaging and for smaller subcomponent vendors to engage directly with hyperscalers. This could help bridge supply gaps while driving down costs. Hyperscalers should also support the standardization of CPO and next-gen optics formats.

To meet rapidly shifting data center needs, the networking optics industry will need the agility to innovate, which in turn will require supply chain stability and manufacturing efficiency. This agility could be enabled to some extent by public sector investment and incentives, but financial backing from the private sector and long-term industry commitments are imperative to support the $500 million to $2 billion that could be needed to close the capacity gap in components.[16] In the United States, the CHIPS and Science Act of 2022 has supported investments in networking optics, for example.[17] At the same time, China has provided incentives to expand its domestic and international capacity, and Chinese market players are steadily gaining global market share and shifting market dynamics.[18]

[16] The amount needed depends on the approach used to expand capacity: expanding existing InP-based EML fabs or building new facilities.
[17] "$36 billion+ in CHIPS for America funding announced," National Institute of Standards and Technology, updated January 23, 2025.
[18] *Optical transceivers for datacom and telecom 2024*, Yole Group, May 2024.

Because any returns on investments in expanding optics capacity will take at least three to five years to materialize, making substantial investments today is urgent. Robust growth in data center demand, potential competition from other players, and innovations that improve manufacturing yields could prompt players in various regions to accelerate and augment their capacity via increased capital expenditure investment. This, in turn, could cause global competition for market share and talent to heat up and affect market prices.

Despite hyperscalers' increasing focus on reducing the total cost of ownership, the future of high-speed networking to support existing and future AI-driven demands will not be determined by cost efficiencies alone. The success of this evolving ecosystem will require a prompt and collective effort by all players to ensure that innovation, scalability, capacity, and interoperability move in lockstep.

Bill Wiseman and **Lari Hämäläinen** are senior partners in McKinsey's Seattle office; **Henry Marcil** is a partner in the Bay Area office; **Pankaj Sachdeva** is a senior partner in the Philadelphia office; and **Wendy Zhu** is an associate partner in the Denver office.

The authors wish to thank Alexander Shoflick, Alison Fong, Pranjali Kumar, and Shi-Wei Toh for their contributions to this article.