

The Lasers Powering AI

Artificial Intelligence (AI) networks rely on vast amounts of data processed and transferred at incredible speeds to function effectively. This data-intensive nature requires robust infrastructure, with lasers playing a pivotal role. AI networks depend on two primary processes: AI training and AI inference. Training involves feeding large datasets into models to learn and make predictions, while inference uses these trained models to make real-time decisions.

Lasers are crucial in enhancing the efficiency and speed of these processes by enabling high-speed data transfer within data centers and across networks. This article explores the various ways lasers power AI networks, the specific requirements for data center connections, and their broader impact on AI infrastructure.

Requirements for Data Center Connections

The connectivity requirements for data centers supporting AI workloads are stringent. They must handle enormous volumes of data with minimal latency and high reliability. The primary requirements for lasers in these environments include:

High Bandwidth: AI applications, especially those involving large language models and real-time data processing, require interconnects that can support high data rates.

Low Latency: Minimizing latency is crucial for AI inference tasks that require real-time decision-making. Lasers enable faster data transfer compared to traditional electronic interconnects, significantly reducing the time it takes for data to travel between nodes.

Energy Efficiency: AI data centers consume vast amounts of power. Integrated photonics combines optical components on a single chip, reducing power consumption while maintaining high performance.

Scalability: As AI workloads grow, the infrastructure must scale accordingly. Lasers provide the scalability needed to expand data center capabilities without compromising performance.



Lasers Arrays in Data Center Interconnects

In 2022, [Intel Labs demonstrated an eight-wavelength laser array](#) fully integrated on a silicon wafer. These milestones are essential for optical transceivers because the laser arrays can allow for multi-channel transceivers that are more cost-effective when scaling up to higher speeds.

Let's say we need an intra-DCI link with 1.6 Terabits/s of capacity. There are three ways we could implement it:

Four modules of 400G: This solution uses existing off-the-shelf modules but has the largest footprint. It requires four slots in the router faceplate and an external multiplexer to merge these into a single 1.6T channel.

One module of 1.6T: This solution will not require the external multiplexer and occupies just one plug slot on the router faceplate. However, making a single-channel 1.6T device has the highest complexity and cost.

One module with four internal channels of 400G: A module with an array of four lasers (and thus four different 400G channels) will only require one plug slot on the faceplate while avoiding the complexity and cost of the single-channel 1.6T approach.

	4x400G	1x1.6T	1x4x400G
# of modules	4	1	1
Cost/complexity	Medium	High	Low
Faceplate density	Low	High	High
External MUX	Yes	No	No

Same spectral efficiency

Table 1: Comparison of different ways to implement a 1.6 Terabit/s link: four modules of 400G, a single-channel 1.6T module, and a module with four channels of 400G.

Multi-laser array and multi-channel solutions will become increasingly necessary to increase link capacity in coherent systems. They will not need more slots in the router faceplate while simultaneously avoiding the higher cost and complexity of increasing the speed with just a single channel.

Broader Impact on AI Infrastructure

Beyond data centers, lasers are transforming the broader AI infrastructure by enabling advanced applications and enhancing network efficiency. In the context of edge computing, where data is processed closer to the source, lasers facilitate rapid data transfer and low-latency processing. This is essential for applications like autonomous vehicles, smart cities, and real-time analytics, where immediate data processing is critical.

Lasers also play a significant role in the integration of AI with 5G and future 6G networks. The high-frequency demands of these networks require precise and high-speed optical interconnects, which lasers provide.

Conclusion

Lasers are at the core of modern AI networks, providing the high-speed, low-latency, and energy-efficient interconnects needed to support data-intensive AI workloads. From enhancing data center connectivity to enabling advanced edge computing and network integration, lasers play a pivotal role in powering AI. As AI continues to evolve and expand into new applications, the reliance on laser technology will only grow, driving further innovation and efficiency in AI infrastructure.

More Articles from EFFECT Photonics