ılıılı
CISCO

The Critical Role of High-Quality Optics in AI Networks

How Cisco Delivers Reliability and Performance

Contents

The rapid growth of Artificial Intelligence (AI) and Machine Learning (ML) workloads demands highly efficient and scalable network infrastructures to support massive data transfer and low-latency communication across Graphics Processing Unit (GPU) clusters. High-quality optics play a critical role in achieving the required performance by enabling high-bandwidth, low-latency connectivity and minimizing data loss across large-scale AI networks. This paper will look at some of the downsides of using low-quality optics in AI clusters and identifies what Cisco does to help ensure the delivery of high-quality optics, enabling high performance, reliability, and scalability and helping avoid suboptimal operations and potential financial losses due to GPU inactivity.

Designing an AI cluster network requires careful consideration to meet performance, scalability, and reliability demands, with optics playing a critical role. Key considerations include:

- **Bandwidth and latency:** Use high-speed, low-latency optics that match the bandwidth of GPUs, switches, and Network Interface Cards (NICs) to prevent bottlenecks.

- **Reach and distance:** Select optics (e.g., single-mode, multimode, copper, or Active Optical Cables [AOCs]) based on the data center's physical layout.

And above all:

- **Reliability:** Prioritize uptime and fault tolerance by choosing:

  ◦ Optics from reputable vendors with strong track records in reliability

  ◦ Optics that include features like Forward Error Correction (FEC), Digital Optical Monitoring (DOM), and high Mean Time Between Failures (MTBF)

  ◦ Vendors that deliver reliable optics through interoperability testing in multivendor environments

  ◦ Vendors offering optics that comply with IEEE and MSA standards

  ◦ Products leveraging advanced silicon photonics from experienced vendors

What is unique about AI networks?

Unlike traditional data centers, AI networks operate in environments with extremely high data throughput, low latency requirements, and distributed workloads that depend on reliable connections. AI networks require an infrastructure that can handle continuous high utilization and harsh thermal conditions as well as high density and frequent upgrade cycles. With all these challenges, there will be a growing demand for infrastructure that can handle these workloads without failure. A failure results in downtime or suboptimal performance in AI networks that will result in significant financial loss.

Why do optics fail more often in AI networks?

AI networks are tough on optics. AI workloads are very different from what traditional data centers handle. Training large models like GPT-4 and Llama 3 moves huge amounts of data across thousands of GPUs, requiring ultra-fast optics. Unlike traditional data centers, AI clusters push optics to their limits, demanding much higher speeds and lower latency to keep up with the workload. High speeds make transceivers more vulnerable to dust, heat, and mechanical stress because of tighter signal margins, increased heat generation, and greater sensitivity to physical disturbances. The compact, high-density design of AI clusters adds further mechanical stress, making optics more fragile and prone to failure.

AI networks also have unique traffic patterns. Instead of the usual structured data flow, AI workloads involve all-to-all traffic, meaning data is constantly being exchanged between GPUs. Some clusters have over 10,000 GPUs, leading to major congestion. Since AI training runs for weeks or even months, optics are under heavy stress.

AI workloads push buffers and links to their limits, exposing weaknesses that wouldn't normally show up. The high utilization exposes weaknesses in network hardware, including optics, where transceivers with better link margin will reduce the number of correctable errors and the likelihood of uncorrectable errors and improve network efficiency.

Heat is another big issue. AI racks can generate 50 to 100 kW of power, and GPU temperatures can hit 85°C (185°F) or more. Temperature shifts can mess with link margin and cause errors. AI compute appliances in back-end networks famously consume large amounts of power compared to front-end networks. Ports in many of these applications are often set up in portside exhaust configurations, placing the transceiver in a higher-temperature environment than if it was in a portside intake configuration. Higher operating temperatures will increase the failure rate of any transceiver, and inconsistent cooling airflow can also lead to unpredictable failures. Even turning systems on and off frequently (thermal cycling) wears down the optics faster. Using transceivers that can operate at these elevated temperatures reliably and have better field failure rates will improve the overall reliability and efficiency of these networks.

AI clusters also pack a ton of networking gear into a small space. Each GPU has multiple 400-Gbps links, so a 1000-GPU cluster can have over 4000 transceivers when connections between the compute and the leaf and spine nodes are factored in. The more optics there are in play, the higher the chance of failure. A single bad transceiver can slow down the entire network by causing data retransmissions.

What are the downsides of using low-quality optics in an AI network?

The downsides of using low-quality optics in an AI network are significant and can severely impact performance, reliability, and costs. Low-quality optics exhibit higher Bit Error Rates (BER), which leads to packet drop, higher latency, and lower throughput, all of which slow down an AI model training and inference task. Low-quality optics may also fail more frequently, causing unplanned network outages.

The frequent replacements, troubleshooting, and maintenance of unreliable optics drive up expenses due to financial losses incurred by GPU downtime.

Understanding GPU downtime costs

GPU downtime is costly in an AI infrastructure. A single GPU within a high-performance compute cluster can cost upward of $30,000, and these GPUs are typically operating in massive parallel clusters of hundreds or thousands of units. Every minute of downtime can result in thousands of dollars in lost productivity. Table 1 below provides a breakdown of the financial impact associated with GPU infrastructure and potential downtime due to optics failures.
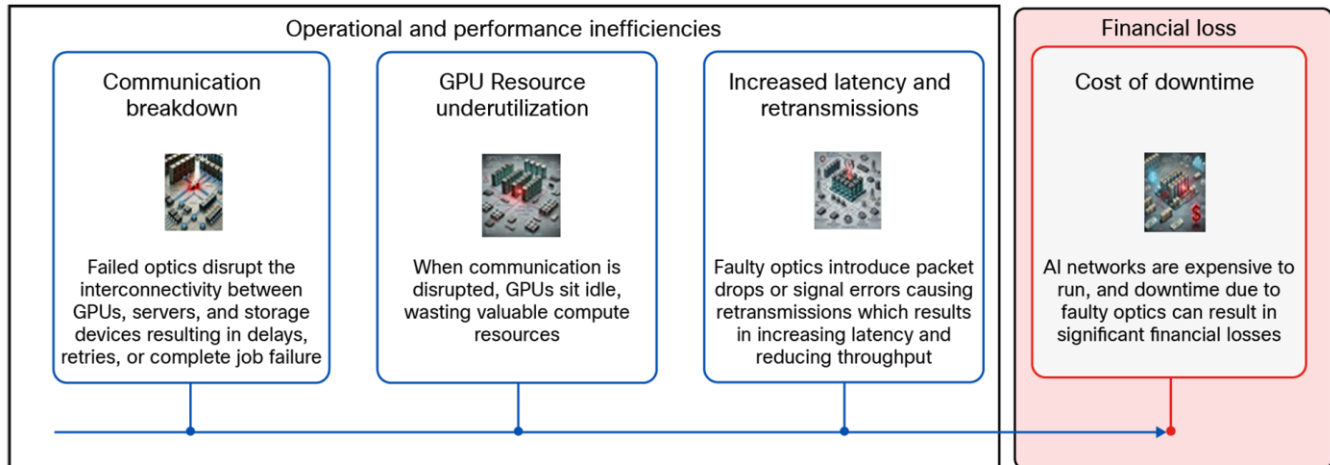
Table 1.   Impact on costs of GPU downtime

Example Scenario	Cost impact
Cost of a single AI server	$500,000+
Cost per minute of GPU downtime	$50 to $200+
Cost of downtime due to optics failure (per hour)	$2,000 to $12,000

The cost of a single AI server with eight GPUs typically exceeds $500,000. The high costs emphasize the need for a reliable and efficient infrastructure to maximize returns. The cost of GPU downtime ranges from $50 to $200 per minute, depending on the workload and the importance of the tasks being executed. Even a short interruption can result in substantial financial losses. Each hour of downtime due to optics failures could cause a loss of from $3,000 to $12,000. This is due to the cost of additional factors such as operational expenses, urgency, wasted electricity, delays in workload completion, or penalties for missing service-level agreements.

What happens to a GPU cluster during an optics failure?

A failed optic will disrupt communication between GPUs, cutting off the flow of critical data. A single optic failure may also isolate an entire AI node. This can result in stalled compute resources, as GPUs need high-bandwidth connectivity to perform these operations for training jobs. Even a single node failure can force the entire cluster to sit idle while waiting for data retransmission recovery. Optics failures can cause timeouts or the isolation of a node, requiring the network to reroute the traffic. This will add significant latency and reduce the overall throughput using a suboptimal path. Furthermore, troubleshooting and repair of optics can take hours, especially when failures are intermittent and it is difficult to identify the root cause. Replacing optics usually requires human intervention, leading to additional downtime.

Operational and performance inefficiencies			Financial loss
Communication breakdown	**GPU Resource underutilization**	**Increased latency and retransmissions**	**Cost of downtime**
Failed optics disrupt the interconnectivity between GPUs, servers, and storage devices resulting in delays, retries, or complete job failure	When communication is disrupted, GPUs sit idle, wasting valuable compute resources	Faulty optics introduce packet drops or signal errors causing retransmissions which results in increasing latency and reducing throughput	AI networks are expensive to run, and downtime due to faulty optics can result in significant financial losses

Optics failures in AI data centers can have severe operational and financial impacts

Figure 1.
Impacts of optics failures in AI data centers

In high-performance AI environments, the savings from reduced downtime significantly outweigh the investment in premium optics. For instance, consider a 1000-GPU AI cluster training a large language model over one month. If low-quality optics with a 2% failure rate are used, the result could be approximately 20 hours of downtime. With GPU costs averaging $10,000 per hour, the downtime would lead to a loss of $200,000 in the 20 hours of downtime. Investing in high-quality optics at a premium over low-quality optics mitigates such failures, helping ensure seamless operations. The avoided downtime leads to substantial savings.

Is the cost of premium optics worth it?

While the premium paid for high-quality optics might appear significant upfront, it represents a relatively small fraction of the total cost of building and operating an AI infrastructure. In a typical AI compute node, optics are 3% to 5% of the total cost. With the bulk of the expenses going to GPUs, High-Bandwidth Memory (HBM), and advanced cooling systems, the cost of reliable optics is minimal by comparison. However, the impact of low-quality optics on the overall system can be disproportionately high, making this small investment in premium optics a smart move for maximizing returns on the entire infrastructure. While the use of third-party optics may come at a lower cost, the disadvantages are significant due to unpredictable quality and lack of support when it comes to troubleshooting an end-to-end network.

What do Cisco optics offer where many vendors fall short?

Cisco tests optics rigorously for compatibility and performance. AI networks are predominantly multivendor environments. In a typical AI data center, the network and compute elements are multivendor, which makes interoperability very challenging. Within a compute node, the components could be multivendor, with the server and the NICs sourced through different vendors, adding even more complexity when it comes to interoperability. Ensuring interoperability within a multivendor environment is critical for delivering reliable optics in AI networks.

Today, Cisco stands out as one of the few vendors offering a complete portfolio that encompasses networking, compute, and storage, along with the optics necessary to seamlessly interconnect the entire ecosystem.

Having a complete portfolio allows Cisco to do comprehensive testing to help ensure compatibility and reliability across diverse environments, a capability that sets it apart from single-product vendors.

This testing includes:

- **Independent testing in Cisco environments using Cisco® network and compute nodes:** Verifies seamless integration and performance for Cisco's core platforms.
- **Testing in third-party environments:** Helps ensure consistent performance across mixed-vendor infrastructures.

Testing optics in Cisco environments using Cisco network and compute nodes

One of the key approaches used by Cisco is to test Cisco optics with third-party NICs deployed in any server. The testing methodology includes:

- **Chamber test case:** Evaluation under varying environmental conditions, such as high/low temperatures and voltage (HTHV, HTLV, LTHV, LTLV).
- **Optics verification:** Validation of supplier details, serial numbers, and specifications (product ID, virtual ID, etc.).
- **Power cycle test:** Repeated cycling of server, NIC, ports, and software to assess resilience and consistency, including Electrically Erasable Programmable ROM (EEPROM) and DOM status checks.
- **OIR (online insertion and removal):** Testing hot-swappable components to help ensure seamless infrastructure optimization without downtime.
- **Traffic test:** Generation of network traffic through the NIC connecting upstream to a switch to validate performance under real-world conditions. With the successful completion of a traffic test, Cisco publishes comprehensive compatibility and performance data for connectivity solutions through the Cisco Optics-to-Device Compatibility Matrix, offering customers valuable tools and insights to make well-informed decisions. The matrix includes details on the optics, the NIC, and the minimum firmware/software required, as seen in the screen shot below:

Mellanox (Nvidia)													
		Transceiver Description										Software Release	
Network Device Product ID	Transceiver Product ID	Data Rate	Form Factor	Max. Reach	Cable Type	Media	Connector Type	Transceiver Type	Case Temp	DOM HW Capable		Minimum	DOM SW
Mellanox (Nvidia) Bluefield-3 B3140H 1 x 400G2E QSFP112 Network Interface Card	QSFP-400G-DR4	400 Gbps	QSFP112	500m	Ribbon Fiber	SMF	MPO-12	Optic	0 to 75C	Y			

Testing in third-party environments

Beyond the Cisco ecosystem, independent testing is also conducted in third-party environments through joint effort with third-party vendors. This broader scope helps ensure that Cisco's solutions perform consistently across heterogeneous infrastructures, offering flexibility for customers operating mixed-vendor environments. Upon completion, Cisco collaborates with third-party vendors to publish joint product briefs that detail the testing results, offering transparency and valuable insights for actionable decision-making.

These combined efforts help ensure that Cisco continues to deliver high-quality, reliable solutions for a wide range of IT environments.

Comprehensive compatibility testing by Cisco (Avalanche)

Cisco has developed a methodology called Avalanche to simulate many third-party switches and routers, allowing for rigorous testing of optics under stress conditions. By helping ensure that Cisco optics remain system vendor agnostic, Avalanche provides a significant advantage over optics from other system vendors. Cisco delivers superior compatibility and reliability for its optics through an extensive testing framework that comprises three key components:

- **Optics-level testing**

  ◦ Each Cisco optic undergoes rigorous optical, electrical, and mechanical design verification to meet high standards.

  ◦ Transceivers are tested against multiple suppliers to help ensure robust interoperability.

- **Avalanche testing**

  ◦ Focused on closing compatibility gaps in third-party environments, Avalanche testing accelerates the validation of optics for diverse hosts. It includes:

    ◦ **Stress testing** to push optics to their limits, identifying potential failure points.

    ◦ **Issue exposure** that proactively uncovers hard-to-find problems before deployment.

    ◦ **Model ports** for accurate modeling of Cisco and third-party interfaces.

- **Platform-level testing**

  ◦ While many vendors validate their optics only on their own platforms, Cisco conducts transceiver testing at the platform level, helping ensure seamless link operation across different environments.

Cisco's comprehensive optics testing methodology, combining optics-level testing, platform-level testing, and Avalanche—a system designed to emulate third-party environments as shown in Figure 2 below, helps ensure robust and vendor-agnostic optics performance.
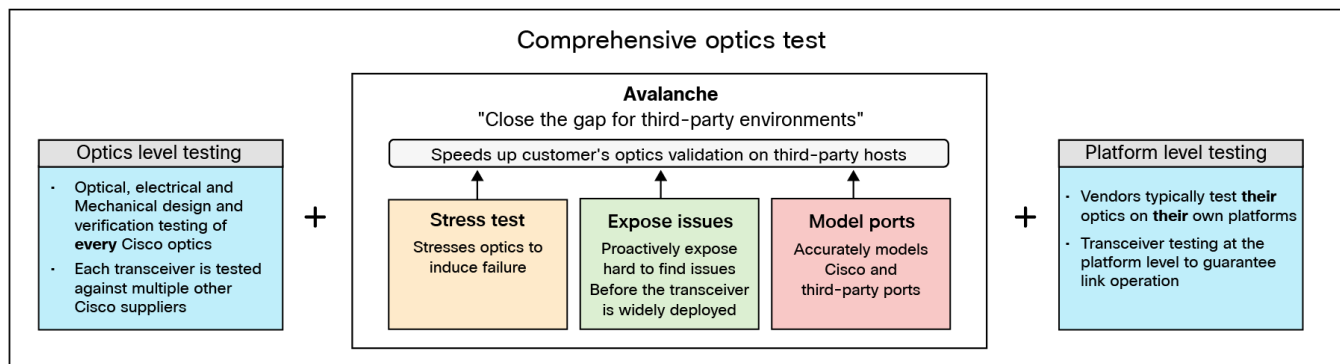
Figure 2.
Avalanche, Cisco's comprehensive optics testing

Through this comprehensive approach, Cisco differentiates itself by providing reliable, extensively validated optics that help ensure seamless operation across Cisco and third-party platforms. This is beneficial to customers because it speeds up their optics validation on third-party hosts, reducing the risk of interoperability issues between Cisco optics and third-party hosts.

Silicon Photonics

One of the key areas that differentiates Cisco from many vendors in the optics industry is the innovative use of Cisco Silicon Photonics, which enables networks to rapidly scale to the next higher speed, far outpacing traditional optics technologies—from 800-Gbps optics today to 1.6-Tbps optics tomorrow.

By integrating multiple optics functions on an integrated circuit, Cisco Silicon Photonics offers the following benefits:

- **Higher reliability and scalability:** Silicon photonics integrates key optical components into a single chip, reducing failure points and improving network resilience.
- **Lower power consumption:** Efficient light-based communication reduces energy use and thermal output, minimizing overheating risks.
- **Stronger signal integrity:** Optical signals experience minimal loss over long distances, helping ensure stable interconnects for AI clusters.
- **Compact and durable design:** Silicon photonics chips are more durable and compact, making them well-suited for high-density AI networks.
- **Vendor-backed quality assurance:** The chips are rigorously tested by top vendors, helping ensure high performance and consistency across deployments.
- **Enhanced thermal stability:** Silicon photonics chips operate efficiently across wide temperature ranges, maintaining reliability in dense data centers.

Cisco, with the use of silicon photonics, can offer better transceiver design, smarter cooling, and improved monitoring tools, all of which are critical in preventing failures and keeping the AI infrastructure running smoothly.

By rigorously validating optics in real-world conditions, Cisco helps ensure that AI clusters achieve high availability, optimal throughput, and stable connectivity, reducing data loss, link flaps, and inconsistent performance that can lead to costly downtime.

Cisco also implements rigorous pre-deployment testing of optics to help ensure that they meet performance standards. This reduces failure rates and improves reliability.

Conclusion

AI networks present unique challenges, requiring an infrastructure that can handle high data throughput, low latency, and continuous reliability under extreme workloads. Cisco's approach stands out by combining advanced technologies, a complete portfolio, and rigorous testing methodologies to address these demands. Unlike many vendors, Cisco provides seamless interoperability across multivendor environments, delivering optics that are tested for both Cisco and third-party platforms.

With innovations like Silicon Photonics, Cisco offers higher reliability, improved thermal efficiency, and enhanced scalability, enabling networks to achieve next-generation speeds with minimal failure risk. Cisco's comprehensive testing and compatibility tools further provide customers with confidence in their investments, reducing downtime, preventing costly disruptions, and maximizing the return on AI infrastructure. By choosing Cisco, organizations benefit from a proven, end-to-end solution designed to support the most demanding AI workloads and help ensure long-term success.

Learn more

Cisco Optics-to-Device Compatibility Matrix
https://tmgmatrix.cisco.com/

Cisco Optics Product Information
https://copi.cisco.com/

Cisco Optics-to-Optics Interoperability Matrix
https://tmgmatrix.cisco.com/iop

Printed in USA   C11-4974751-00   02/25