TELEDYNE LECROY
Everywhereyoulook™

Solving the AI Bottleneck

AI, HPC and the Terabit Ethernet Revolution

WHITE PAPER

How Terabit Ethernet can meet the
high-bandwidth, zero-loss needs of AI & HPC

CONTENTS

Executive Summary

The rapid rise of Artificial Intelligence (AI) and High-Performance Computing (HPC) is transforming data center infrastructures. In just a few years, these applications have shifted from niche workloads to core drivers of traffic growth and architectural change. Unlike traditional cloud and enterprise services, AI and HPC demand dense, high-throughput communication between thousands of compute nodes. This is pushing Ethernet speeds to the limit.

To support this evolution, data centers are entering a new era of performance requirements. The transition to 800Gbps Ethernet (GE) is already underway, with 1.6Tbps Ethernet (TE) as the next milestone. Such high-speed connectivity is essential to meet the demands of massive GPU clusters used in large-scale model training and real-time AI inference.

This white paper explores the key factors driving the need for Terabit Ethernet, including the scale of AI traffic, data center infrastructures and the technological innovations enabling the next leap in networking speed. Teledyne LeCroy Xena's current range of Ethernet Traffic Generation (TG) solutions for 400GE, 800GE and 1.6TE are designed to deliver the broad flexibility needed by network equipment vendors, service providers and data center operators to develop the next generation of high-speed Ethernet devices and services.

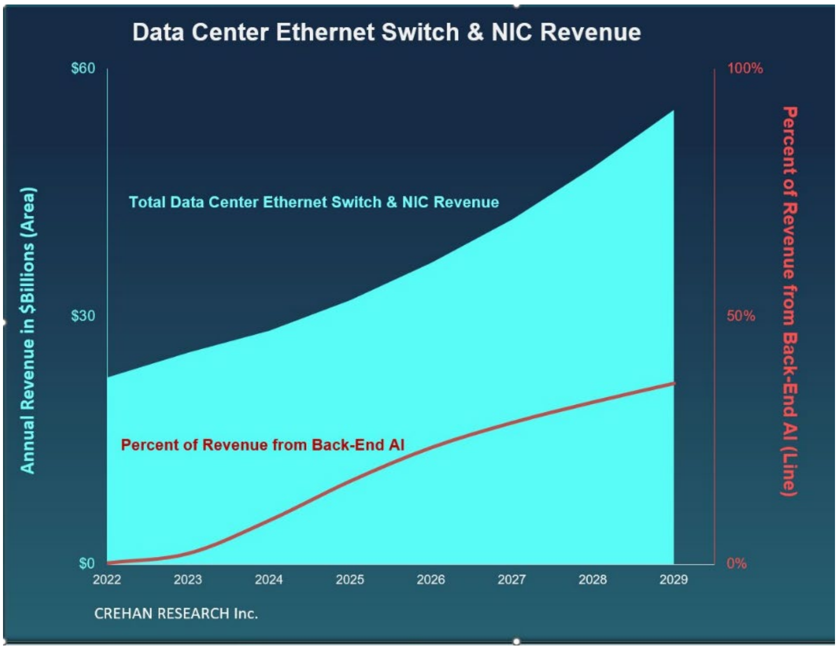
What's driving Terabit Ethernet?

Data centers are undergoing a dramatic transformation. What was once an infrastructure optimized primarily for general-purpose cloud computing and enterprise applications is now being reshaped by the explosive growth of Artificial Intelligence (AI) and High-Performance Computing (HPC). These new workloads not only consume vast amounts of compute and storage capacity—they also generate unprecedented volumes of east-west traffic within the data center, stressing the network in ways that were unimaginable just a few years ago.

Unlike traditional workloads, AI and HPC applications require tightly coupled communication across large clusters of accelerators. In training large language models, tens of thousands of GPUs must exchange massive data sets. This shift places extreme performance demands on the underlying network, where latency, bandwidth, and scalability are all critical to maintaining efficiency and minimizing job completion times.

Market trends underscore the scale of this shift. According to a January 2025 report from Crehan Research[1], combined revenues for data center switches and network interface cards (NICs) are expected to more than double, from approximately $20 billion in 2022 to over $50 billion by 2029. As depicted in Figure 1, the portion of that spend driven by AI back-end networking is forecast to grow from virtually zero in 2022 to more than $20 billion in 2029, making it the single largest driver of growth in the segment.

This trend clearly signals a transition point in network infrastructure planning. To handle the scale and intensity of AI traffic, operators are rapidly adopting 800GE and preparing for the next leap to 1.6TE. Without this bandwidth evolution, AI training clusters will be bottlenecked by network limitations, underutilizing computer resources and driving up energy and infrastructure costs.



[1] Source: Data Center Ethernet Switch and NIC Market to Exceed $50

Figure 1: Annual revenue from Data Center Switches and NICs showing percentage of Back-end AI infrastructure. Source: Crehan Research

The AI and HPC bandwidth explosion

Modern AI models, especially Large Language Models (LLMs), are built on architectures that require parallelized training across thousands of AI accelerators. To scale efficiently, these models rely on two key techniques: data parallelism and model parallelism.

Data parallelism splits large data sets across many GPUs, each processing a different subset of the data. After each training step, all GPUs must synchronize their updates which require frequent so-called all-reduce updates across the entire cluster of GPUs.

Model parallelism divides the model itself across GPUs, with each device responsible for a portion of the forward and backwards passes. This requires communication between GPUs at every step in training.

Both methods depend on tight synchronization between GPUs and generate high volume east-west traffic. As model sizes grow in terms of parameters the amount of data exchanged between nodes increases dramatically. Any delay or bottleneck in the network results in GPU underutilization, lengthening of training time and increased energy and network costs.

In this environment, low latency is critical to reduce synchronization stalls, while high bandwidth ensures that parameter data can flow freely without congestion.

AI is affecting data center architectures

AI is reshaping network architectures in modern data centers, introducing new design priorities and traffic patterns. Traditional enterprise data centers were typically optimized for north-south traffic, where data flows in and out of the data center — from clients to servers and back. This model suited classic client-server applications and web hosting, where interactions primarily crossed the data center boundary.

In contrast, AI workloads — especially distributed training and large-scale inference — generate massive amounts of east-west traffic between servers inside the data center. These workloads involve intensive communication between clusters of GPUs or AI accelerators, often exchanging data in parallel across hundreds or thousands of nodes. To support this, networks must deliver low-latency, high-bandwidth connections to prevent bottlenecks that would slow down job completion. AI applications are also characterized by frequent elephant flows — large, long-lived data transfers that require consistent throughput and can quickly overwhelm traditional network designs if not properly handled.

To meet these new demands, many data centers have adopted spine-leaf architectures as shown on Figure 2, which differ significantly from older, hierarchical network topologies. Traditional designs often relied on multiple layers — core, aggregation, and access — creating oversubscription and bottlenecks as traffic moved upward and downward through the hierarchy. In contrast, a spine-leaf design flattens the network: each leaf switch connects directly to every spine switch, forming a highly scalable and non-blocking fabric. This ensures that any server connected to one leaf can communicate with any other server on a different leaf with consistent latency and bandwidth, regardless of traffic patterns. The design also simplifies scaling: adding more capacity simply means adding more spine or leaf switches, making it ideal for the horizontally scalable nature of AI infrastructure.

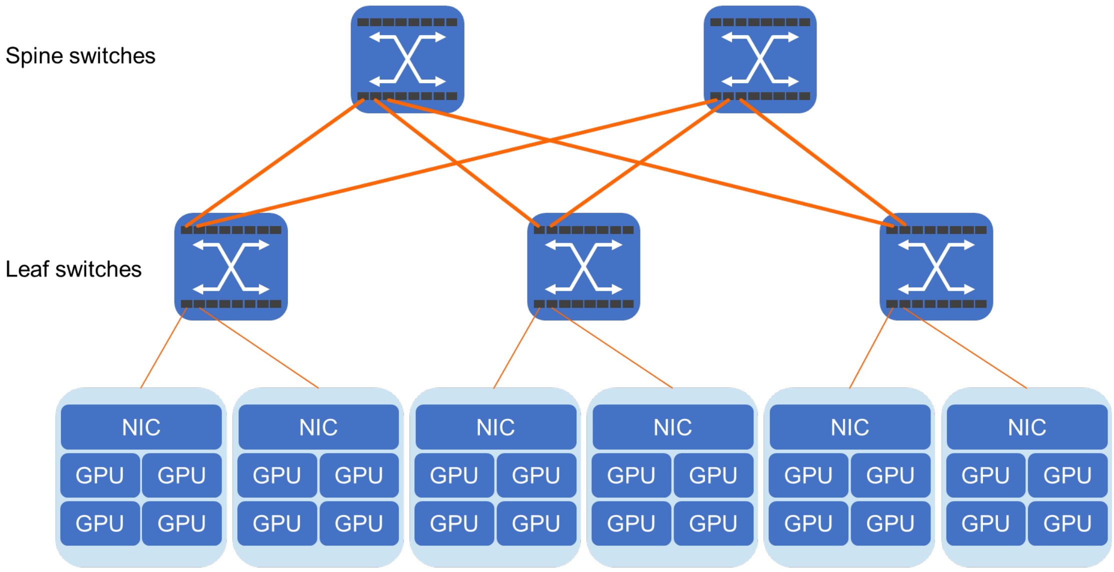


Figure 2: Typical Spine-Leaf network connecting GPU clusters in a data center.

Why 1.6TE per port?

A common data center rack architecture is illustrated on Figure 3. Each server rack houses one or two Top-of-the-Rack (ToR) Leaf switches that aggregate network traffic from the servers within that rack and connects upward to the spine switches in a separate rack. This layout minimizes intra-rack cabling complexity, simplifies management, and reduces latency between the servers and the rest of the network fabric.

Each ToR switch serves as the local network gateway for the servers it connects. In AI clusters, where GPU-equipped servers are densely packed and traffic patterns are dominated by east-west communication, the ToR switch must be capable of handling extremely high bandwidth. A single AI training job may generate traffic between dozens of servers in multiple racks, all requiring low-latency, high-throughput data exchange. If the ToR switch becomes a bottleneck, it can degrade the performance of the entire workload.

This is why port speed at the ToR level is critical, with the latest designs pushing toward 1.6TE per port. Each port must be capable of supporting high-speed links to servers, as well as uplinks to multiple spine switches. Given the number of GPUs per server and the high bandwidth requirements of AI workloads, slower ports simply cannot keep up. For example, just one server with eight GPUs may need hundreds of gigabits per second in sustained network throughput. Multiply that across an entire rack, and it's clear that older 100 or even 400GE links are not sufficient.

By deploying 800Gbps or 1.6Tbps-capable ports at the ToR level, data centers can ensure that server-to-server communication remains fast and predictable, even under peak loads. This high-speed connectivity is essential not only for performance but also for scalability — as AI models grow and workloads increase, the network must remain an enabler, not a constraint.

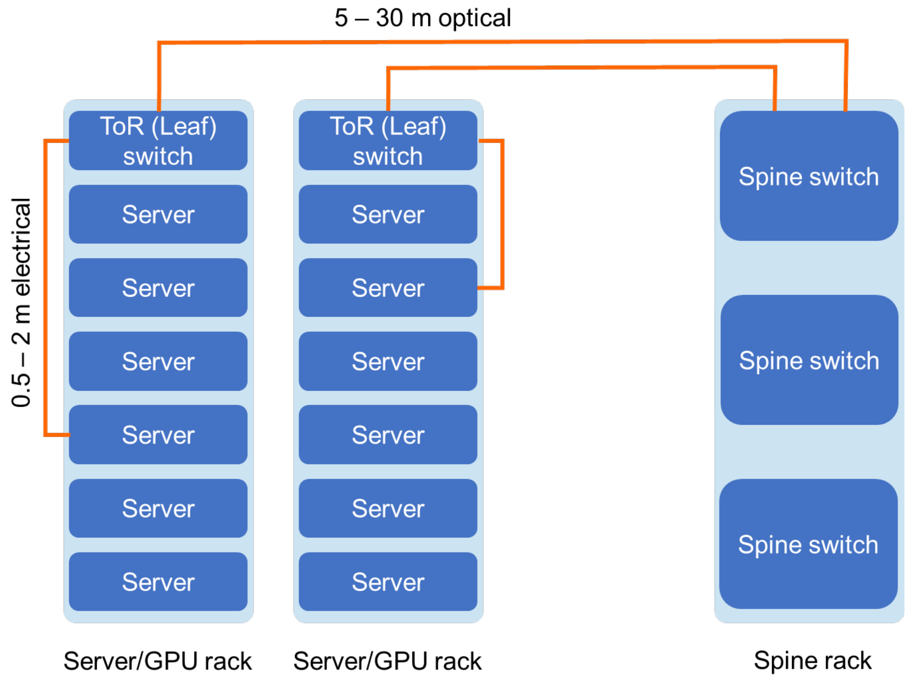


Figure 3: Typical rack architecture in data centers with Top of Rack Leaf switches in each server rack and Spine switches placed in a separate rack.

The AI power challenge

AI workloads are significantly increasing power demands in data centers. According to the International Energy Agency (IEA)[2], global data center electricity consumption reached approximately 415TWh in 2024, which accounted for roughly 1.5% of worldwide electricity use. AI-related compute is projected to rise around 30% annually, while conventional server usage grows by about 9%.

[2] IEA, Energy demand from AI – Energy and AI – Analysis - IEA

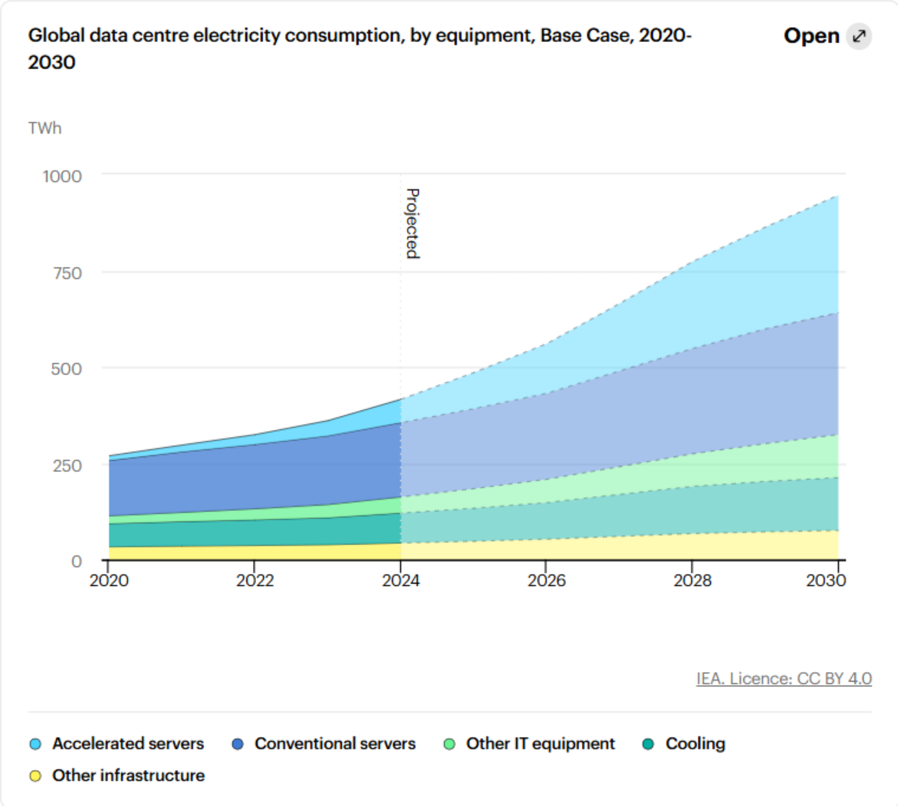


Figure 4: Power consumption of data centers per workload type. IEA License: CC BY 4.0

Estimates vary, but by 2030 data center electricity consumption is expected to roughly double to 945TWh, representing nearly 3% of global electricity demand. Each server rack equipped with GPUs for AI can consume 240 to 300 kW, whereas traditional workloads typically require only 30 – 50kW per rack.

The rapid increase in electricity consumption poses significant challenges beyond just power availability. Higher energy use translates directly into increased carbon emissions, especially in regions where the grid still relies on fossil fuels. In addition, cooling these high-density racks requires vast amounts of water, putting pressure on local water supplies which may already be a scarce resource. From a business perspective, the rising energy demand also escalates the operational costs for data center operators.

Therefore, the networking industry is taking actions to develop more energy-efficient equipment and network infrastructures while at the same time increase the bandwidth.

Understanding tradeoffs in data center architectures and switch designs

Before exploring how we can efficiently increase Ethernet speeds in a power-conscious manner, it is important to consider the trade-offs inherent in data center architectures and switch designs. This was captured in a presentation by Cisco Fellow Rakesh Chopra in his keynote presentation at the Ethernet Alliance TEF 21 event[3].

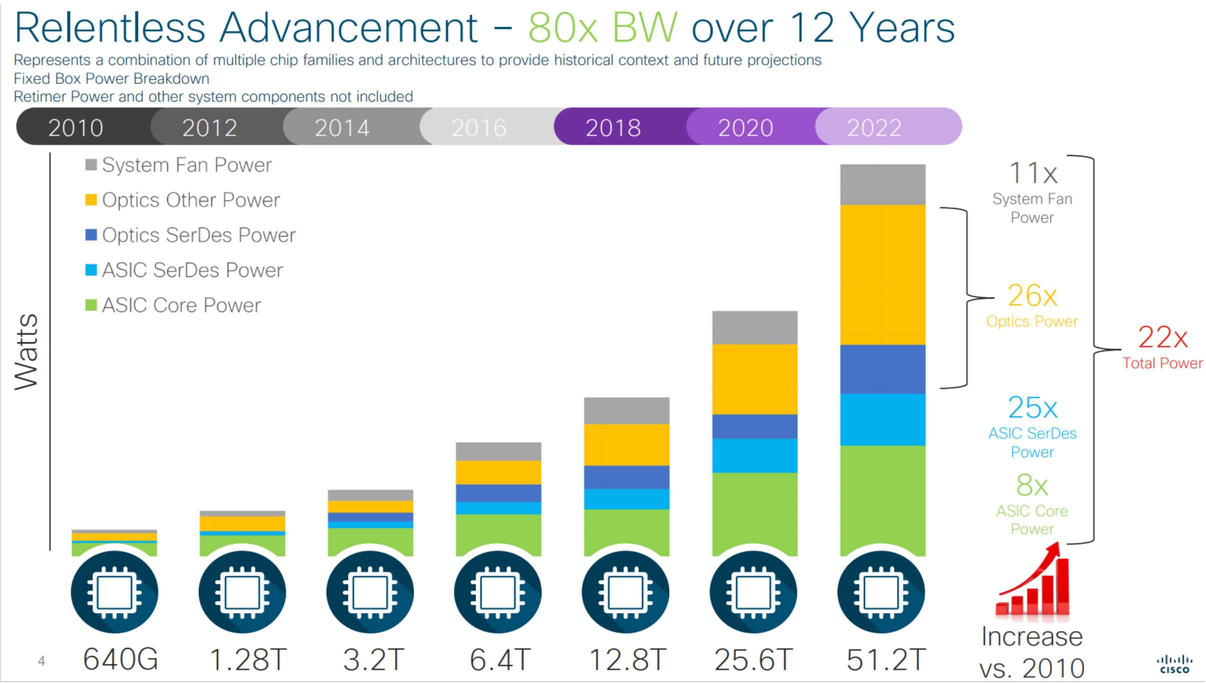


Figure 5: 80 times bandwidth increase for just 22 times power increase.

As illustrated out in Figure 5, power efficiency has improved - the bandwidth of switches has increased by a factor of 80 while power consumption has “only” increased by a factor of 22 during the last decade.

Each time the number of ports on a switch, or radix, is doubled, the link speed needs to be reduced by half to meet the same overall switch capacity budget. The switch can only deliver a fixed amount of bandwidth, which needs to be shared by all ports. So, scaling out by increasing the radix is power efficient, but inefficient from a link utilization perspective as it limits the amount of data that could have been delivered to a given port.

Similarly, scaling up by adding an additional network layer increases the cost and total power consumption. For example, the network power per server of a 5-tier network is 3 times higher than a 2-tier network.

Chopra showed the impact of these trade-offs when considering how a data center architecture can evolve with higher capacity switch configurations:

[3] Source: https://ethernetalliance.org/tef-2021-the-road-ahead-recordings-presentations/

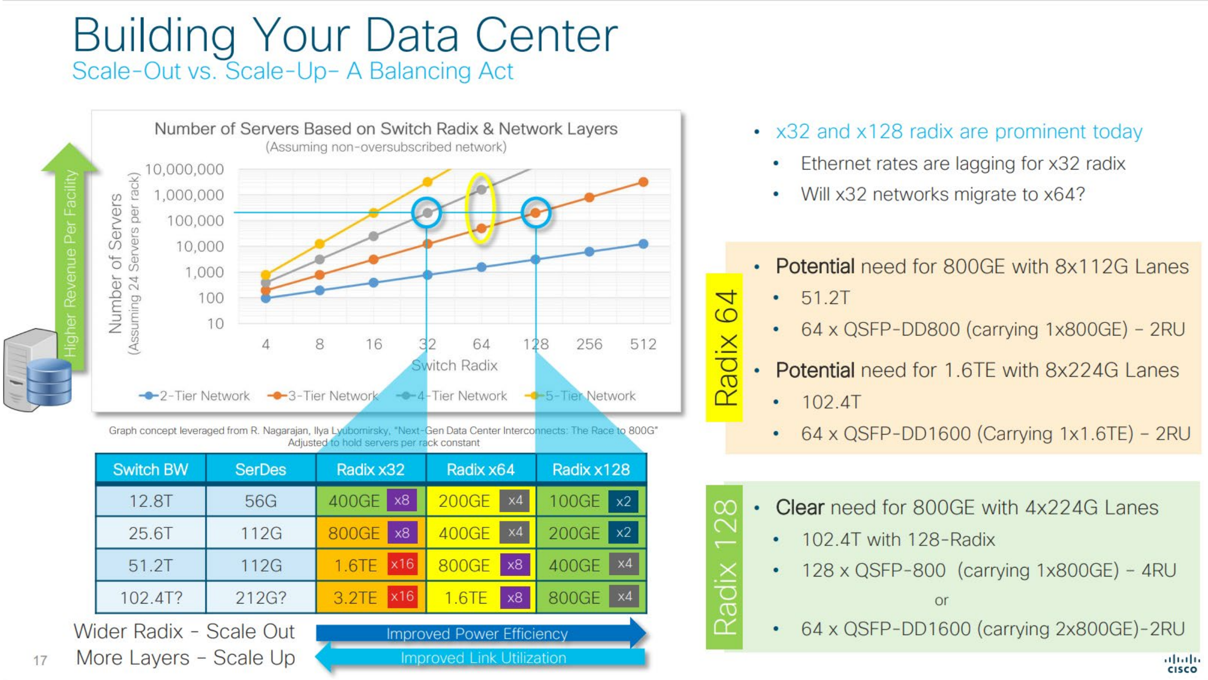


Figure 6: Outlining the need for 800GE and 1.6TE

As can be seen in Figure 6, increasing the radix of the switch to x32, x64 and x128 improves power efficiency but lowers link speeds. This means that there will be a need for 800GE connectivity based on 112G and 224G SerDes when trying to achieve better power efficiency in data centers.

How the power-speed balance is being addressed

To reduce power consumption in the core ASIC, switch designs are evolving as the industry moves from 800GE to 1.6TE. This transition requires a massive increase in total switching bandwidth towards 51.2Tbps or even 102.4Tbps. Achieving this without exceeding power and thermal limits depends heavily on advancements in semiconductor technology nodes. Moving from 16 nm and 7 nm down to 5 nm and 3 nm allows for more transistors per chip area, enabling higher SerDes density, faster processing, and larger buffers, all while improving energy efficiency.

As illustrated on Figure 5, optics account for an increasingly large portion of the overall power consumption. To address this, data center designs typically use electrical Direct Attach Cables (DACs) for short in-rack connections, as shown on Figure 3. DACs generally consume less power and cost less than optical cables but are limited to a few meters at 1.6TE. The reach can be slightly extended by using Active Electrical Cables (AECs) that include equalizers and optionally also drivers and re-timers in the cable.

High-speed connectivity between racks and spine switches are typically too long for electrical connections and requires optical cables. To address the power consumption issues of optical interfaces, Linear Pluggable Optics (LPO) modules have been introduced by several vendors. These modules omit the DSP functions retiming, equalization, and error correction, instead relying on similar functionality anyway built into the Switch or NIC ASIC, thereby reducing power consumption. LPOs are typically compatible with existing Small Form-factor Pluggables (SFPs) used for high-speed Ethernet like OSFP and QSFP-DD. Using LPOs rather than traditional pluggable transceivers typically reduce the power consumption of the optics by ~50% and by ~25% for the entire switch.

To further reduce the power consumption of optics, Co-Packaged Optics (CPO) is emerging as a promising solution. In CPO technology, the optics and the ASIC are integrated on the same substrate, which reduces losses, latency, and power consumption even further. However, since the ASIC and optics are now integrated, they do not fit into a traditional transceiver cage like an OSFP but use specialized fiber optic connectors. CPOs can reduce power consumption of optics by ~60-70% and by ~30-35% for switch. See Figure 7 for an overview.

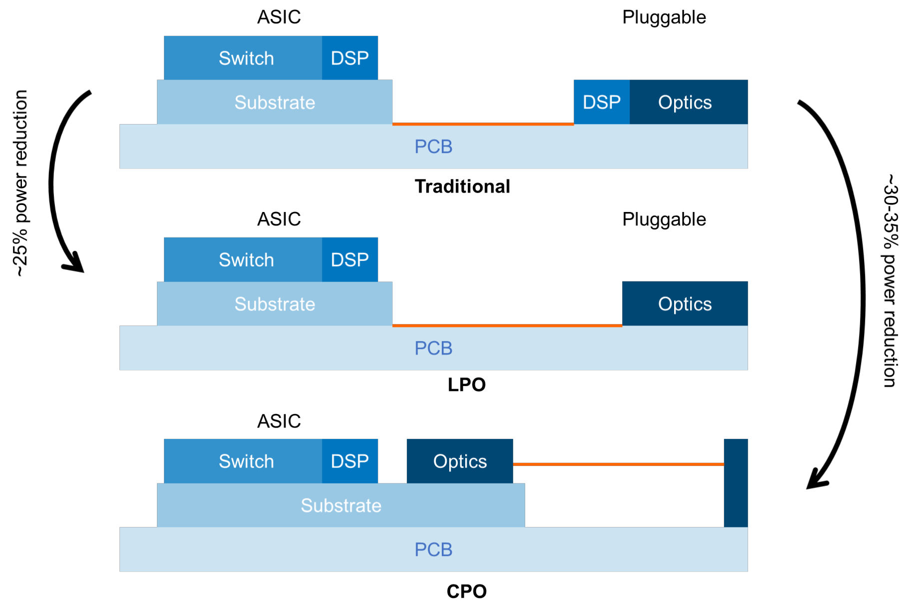


Figure 7: Comparison of switch board with traditional transceiver, an LPO and a CPO.

Terabit Ethernet technical challenges and testing requirements

Moving beyond 100GE required new approaches and technologies that now must be considered when developing and testing Terabit Ethernet. This includes new modulation schemes and even new optical approaches that lead to a broad variety of potential Terabit Ethernet options.

100GE was considered the end-of-the-road for single-lane solutions based on Non-Return-to-Zero (NRZ) modulation schemes leading to the adoption of 4-level Pulse Amplitude Modulation (PAM-4) as a means of increasing the effective bit rate. With NRZ, only one can be represented per symbol period, as NRZ is based on just two voltage levels representing a “1” and a “0”. PAM-4, on the other hand, uses 4 voltage levels and can thus represent two bits per symbol period, as shown in Figure 8. Now, double the number of bits sent with each clock cycle leading to an effective doubling of the number of bits transmitted.

When this option is combined with the speed of the lane (112G or 224G) and the number of parallel lanes used, several different paths to 800GE and 1.6TE emerge.

From a product development and design perspective, the challenge becomes how to implement and test PAM-4 ensuring low Bit-Error-Rates (BERs) as the Signal-to-Noise Ratio (SNR) is now reduced. This is important both in achieving good performance, but also in ensuring interoperability with other systems and is a major challenge when moving to Terabit Ethernet speeds.

This new focus on modulation schemes means that Terabit Ethernet testing cannot only focus on Layer 2 but also needs to consider Layer 1 issues. This becomes even more important when we

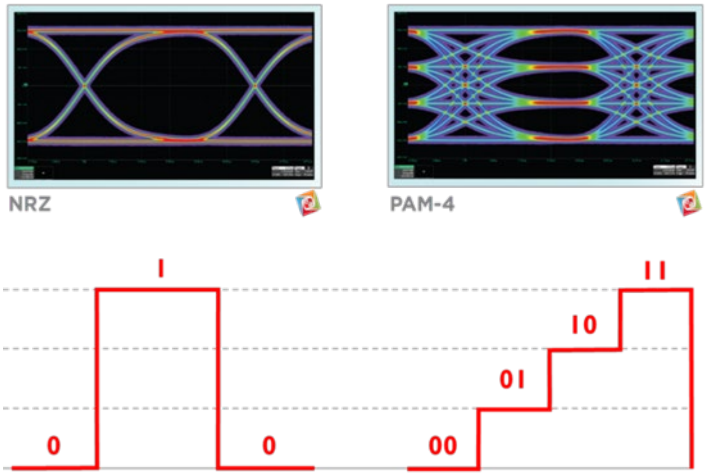


Figure 8: NRZ and PAM-4 encoding.

consider the new approaches using Co-Packaged Optics (CPO) where optical and electrical components are packaged together to achieve better cost-per-bit and power-consumption-per-bit.

Terabit Ethernet testing will thus require a broader view encompassing Layer 1 and Layer 2 but will also require a great deal of flexibility in accommodating various combinations of baud rates, modulation schemes and parallelization delivered in a variety of form factors, such as QSFP-DD and OSPF. Flexibility will thus become a key criterion for Terabit Ethernet testing equipment going forward.

Teledyne LeCroy solutions for Terabit Ethernet Testing

Test & measurement solutions need to be at the forefront of technology to help first movers develop new products – in this instance terabit switch, line-card, transceiver and active cable solutions. Teledyne LeCroy Xena's Ethernet test solutions are designed to give engineers all the options they need in terms of speeds, modulation schemes and form factors.

Teledyne LeCroy offers a range of Ethernet Traffic Generator and Analyzers covering all Ethernet speeds from 10Mbps to 1.6Tbps. All solutions are conveniently managed through the same XenaManager 3 (XM3) software. XM3 is the new version of XenaManager loaded with improvements to the user interface and Layer 1 to make testing of high-speed Ethernet even easier.

Our XOA open-source Python environment lets you integrate Ethernet testing with any of our traffic generators and analyzers into your own test environment. This makes it easy to migrate your existing test cases to new, higher-speed test modules.

Our Teledyne LeCroy Xena Z1608 Edun module supports 100GE, 200GE, 400GE, 800GE and 1.6TE using wither 112G or 224G SerDes and a single a OSFP port. Figure 9 shows a picture of the Xena Z1608 Edun.

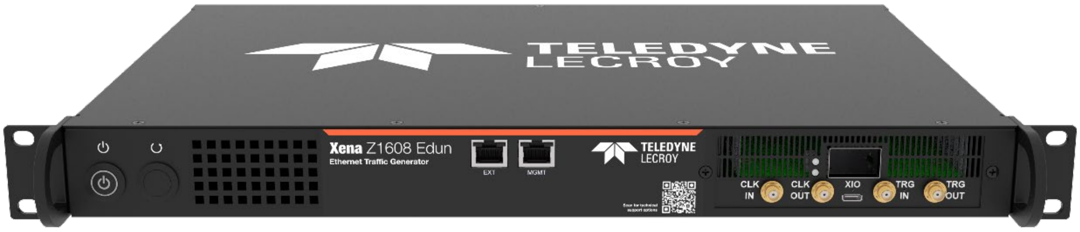


Figure 9: Teledyne LeCroy Xena Z1608 Edun 1.6Tbps Ethernet Traffic Generator and Analyzer.

Z1608 Edun use the same unified software solutions that supports all our traffic generators, ensuring a consistent user experience regardless of hardware configuration. The XenaManager software is an easy-to-use tool to configure ports, set-up traffic flows and analyze traffic properties. Furthermore, the XOA open-source Python environment integrates with your own testing environment for automation and scripting.

The high-level feature set of Z1608 Edun is as follows:

- 5-speeds: 1.6TE, 800GE, 400GE, 200GE & 100GE
- 1 x OSFP cage
- Supports 224G & 112G SerDes (PAM4)
- Test with optics, Active cables and DACs
- Extensive L1, L2 and L3 test features

For more information about Z1608 Edun please visit: https://xenanetworks.com/product/z1608-edun/