Spirent™



WHITE PAPER

Turning Ideas into Action for 1.6T Ethernet

Turning Ideas into Action for 1.6T Ethernet

Ushering in the Next Ethernet Rate

The explosion of Edge computing, virtual reality (VR) and augmented reality (AR), and AI applications are set to significantly increase demand from their networks as new requirements accelerate the adoption of higher speed Ethernet connectivity solutions to satisfy these data needs. This trend is disrupting traditional timelines for the development and availability of connectivity technologies. This will have a major impact on the ecosystem powering these new technologies, including Network Equipment Manufacturers, Data Center operators, interconnect manufacturers and system integrators.

Key Requirements of Next-Gen Applications

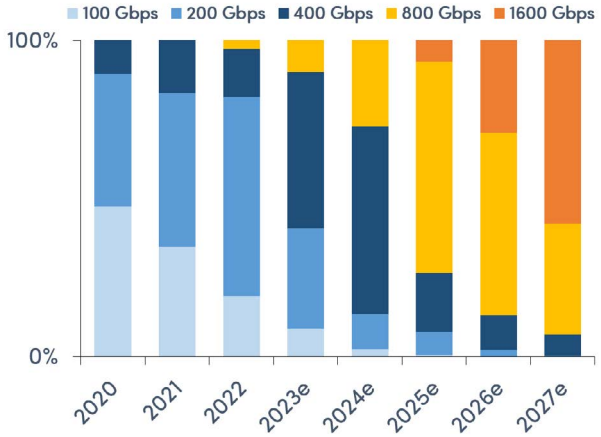
For AI/ML training and inference, there is a high degree of sensitivity to bandwidth and latency. The job completion times (JCT) are gated by tail latency which negatively affects the ROI on the expensive AI infrastructure because the GPUs will now have to sit idle waiting for the delayed packets to arrive in order to process them while other jobs in queue suffer.

The IEEE's 802.3dj group is formulating the upcoming Ethernet standard, which includes physical layers and management parameters for speeds up to 1.6 Terabits per second. Since the AI traffic workloads are characterized by the presence of many elephant flows, the greater the available bandwidth, the better the large-scale data transfers can take place which will help reduce job completion times.

In addition, the cost per bit, which is a major consideration for large deployments, will decrease due to more efficient optical module designs expected for 1.6T, resulting in a decreased power consumption per bit.

> In our latest "AI Networks for AI Workloads" Advanced Research Report, we predict that by 2025, most ports in AI networks will operate at 800 Gbps, with a swift shift to 1600 Gbps by 2027. This demonstrates a rapid uptake of the highest speeds available in the market. This pace of migration is nearly twice as fast as what we have typically observed in traditional front-end networks connecting general-purpose servers.
>
> **Sameh Boujelbene,**
> **Vice President at Dell'Oro Group**

Migration to High-Speed Networks in AI Clusters (Back-End)



*Includes both Ethernet and InfiniBand
* Source: Dell'Oro Group "AI Networks for AI Workloads" Report

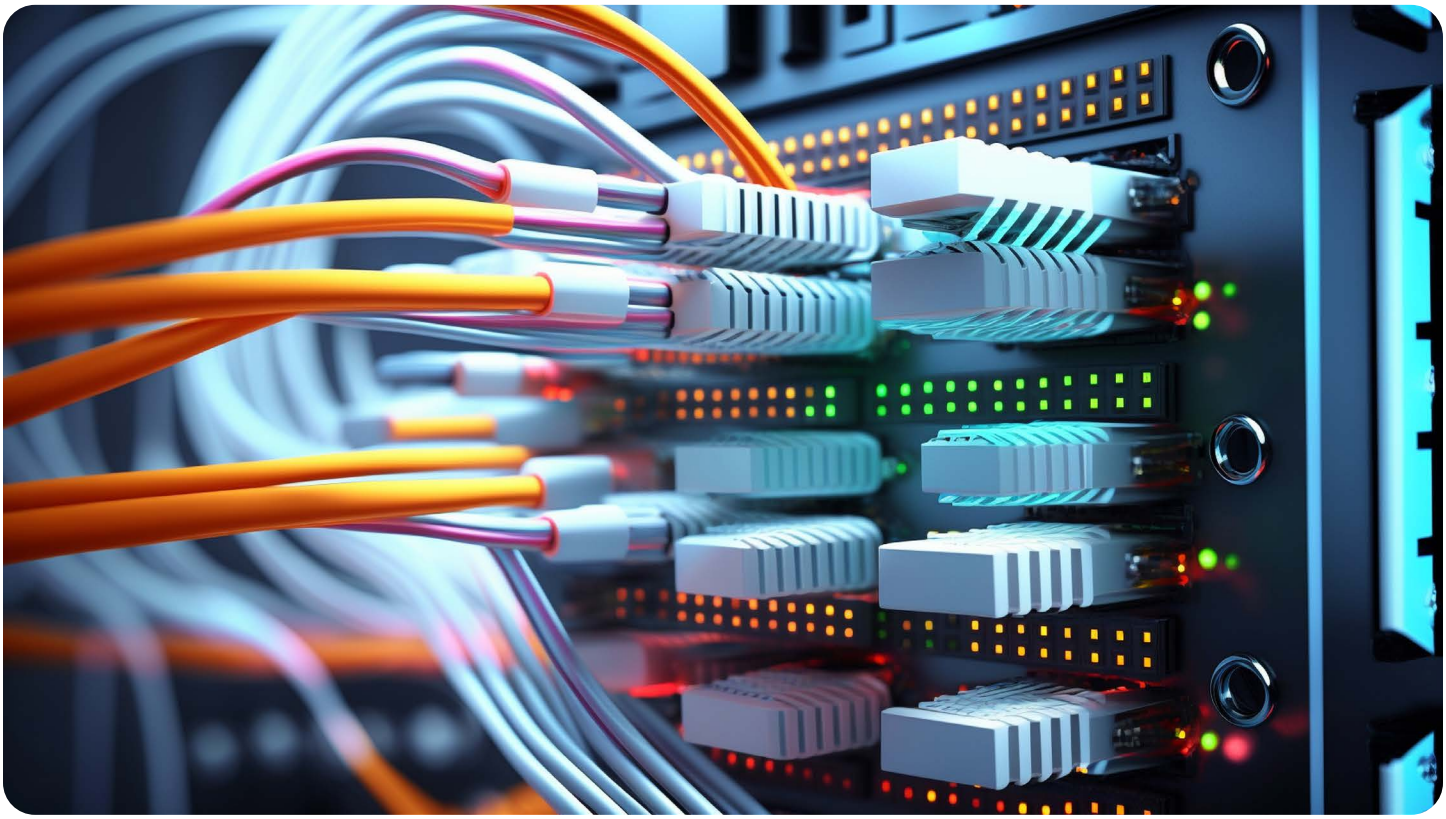
Technology Challenges on the Path to 1.6T Ethernet

As each new Ethernet generation increased the link speed, there was a corresponding increase in the rate of the signals carrying the data. In some cases, industry has been able to keep the line rate the same between successive generations to a certain degree by utilizing methods such as parallel lanes and PAM4 modulation, but continued progress inexorably calls for faster and faster bit transitions.

IEEE 802.3dj gets to 1.6T Ethernet by doubling the physical layer interface bit rate from 800G technology; 112 Gbps up to 224 Gbps. Since PAM4 will continue to be used, the doubling of bit rate inherently doubles the spectral content of the physical layer signal. Doubling the symbol rate of the signal requires a combination of improved analog channel characteristics and more sophisticated SERDES transmitter and receiver technologies.

Electrical channels like PCB traces, connectors, and copper cables will be challenged to overcome insertion losses and reach useful distances. The channel between the ASIC and front panel can be improved by using new PCB materials, adding more retimers, using low-loss flyover cables, or even using co-packaged optics. In addition, new forward error correction and link training algorithms are under consideration to improve overall link fidelity budget. The channels between Ethernet ports will similarly need to adopt new technologies and techniques to offset the ever-increasing power consumption inherent in faster symbol rates.

As with many communication system advances, carrying 1.6T information across a physical medium turns into an engineering challenge of how to best achieve the performance requirements while mitigating complexity. Dollar costs, power consumption, and latency all need to be factored into the complexity equation for 1.6T. To achieve the right balance for multiple applications, a wide variety of solutions will likely be required.



The Evolving Connectivity Landscape

Connections between Ethernet ports have historically been made through copper cables for shorter distances and optics for longer reaches. However, starting with 800G the growing disparity in reach, cost, and power between DACs and optics created a need for new technologies to fill the gap.

Passive Direct Attach Cables (DACs) are a tried-and-true Ethernet connectivity technology. Being relatively simple in design, DACs are inherently reliable, relatively inexpensive, have low latency and consume essentially no power. In a datacenter environment, they are the popular choice for connections within a rack. Link training is typically used to optimize the transmitters and receivers at both end points. However, with increased data rates to 112 Gbps per lane and now up to 224 Gbps per lane for 1.6T, the electrical signal losses limit how far even advanced SERDES can reach. Losses can be reduced by using thicker copper wire, but this approach has its drawbacks. Thicker cables are heavier, harder to manage, and tend to block off the airflow to the faceplate, which can be problematic for high density deployments. As a result, for 1.6T, the longest DACs may no longer be practical or even capable of reaching from the top of rack to the bottom.



Emerging Interconnect Technologies

Short range optics and Active Optical Cables based on SR technology are commonly used for intra-rack connections up to 100 meters but are arguably overkill for the 4 to 5 meters needed for inter-rack connections.

Two new copper cable types have emerged for 800G to slot between DACs and AOCs and are expected to continue for 1.6T:

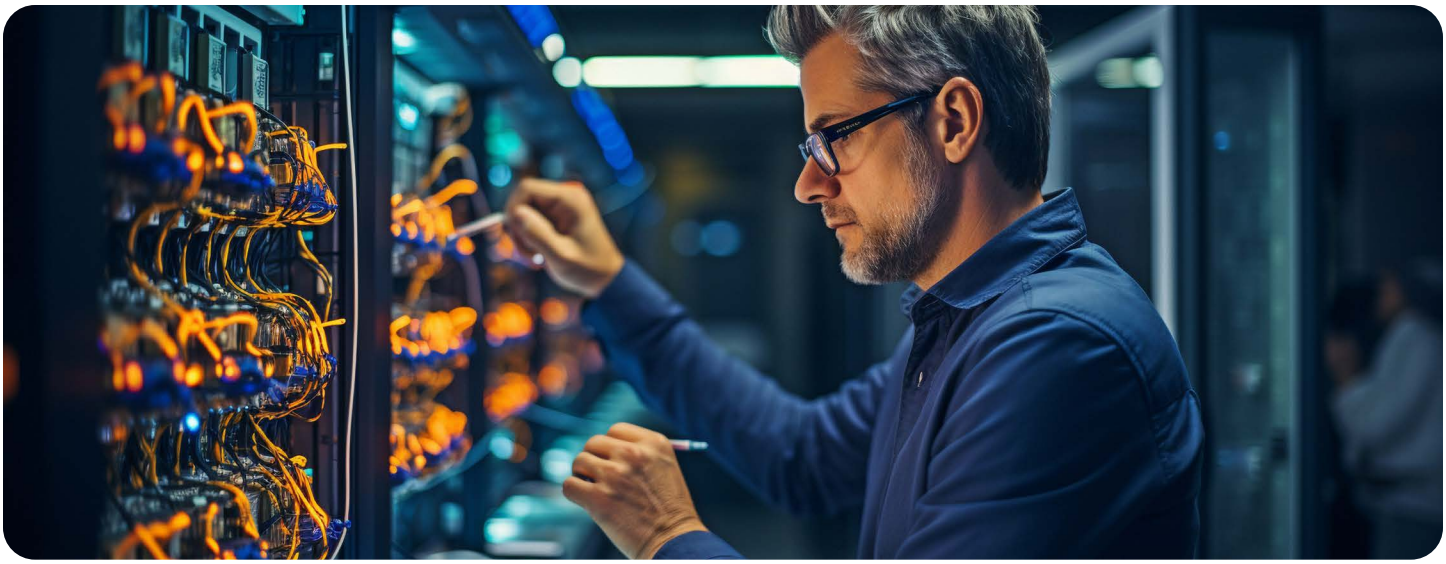
- **Active Copper Cables (ACC)** utilize a linear amplifier at the receiving end of the cable to boost the signal and extend the possible reach at the expense of increased cost and power.
- **Active Electrical Cables (AEC)** use re-timers at both ends of the cable to recondition the signal on the way in and the way out. By removing rather than simply amplifying noise, AECs promise better performance and reach than ACCs, but with yet higher cost and power.

Both active cable types have thinner gauge wire and less insulation than DACs, so they can offset their power consumption somewhat by allowing for better cable management and airflow, resulting in better cooling for high-density rack systems.

Optics remain the only viable solution beyond 10 meters and offer a variety of traditional reaches from SR up through DR, FR, and LR. Concerns about ever increasing power consumption and subsequent heat generation at higher data rates have led to the development of Linear Drive, also known as **Linear Pluggable Optics (LPO)**. LPOs aim to reduce the power and latency in the optics by removing the DSP in the module. The tradeoff is the increased burden on the host SERDES or PHY to handle the signal conditioning that was previously done in the module. LPOs have started to roll out for 800G systems and if system integrators can overcome the technical challenges, we may see LPOs become commonplace for 1.6T.

Another emerging optical technology is the use of **coherent optical modules** in the datacenter. Coherent modulation has long been used for long-haul networks where its ability to carry more bandwidth using less optical power than traditional optics made them attractive despite the large package size. Recent advances have allowed coherent to move into pluggable form factors such as QSFP-DD and OSFP, making them viable for high density applications in the datacenter.

New standards for 400G ZR, 800G ZR, and 1600G ZR are being developed by the OIF to streamline adoption. The challenge is that coherent modules still burn more power than other interconnect options, so the host needs to be able to supply the higher current and provide enough airflow to cool the module.



Related Technologies

Link training has long been used to improve performance across long copper channels by autonomously tuning transmitters and receivers to each other. Due to the challenges of 224 Gbps signaling even over short distances, proposals are being made in the IEEE and OIF to add link training to new segments, such as chip to chip, chip to module, and module to module. Active interconnects with multiple segments between end points would then benefit from increased channel margin at the expense of increased cost, power consumption, and latency.

Another area being studied is Forward Error Correction. When the industry transitioned from NRZ to PAM4 signaling, FEC between hosts endpoints became mandatory to offset the signal-to-noise penalty of the higher modulation. The coding gains and long burst error tolerance provided by FEC allow lower bit error rate targets that are easier to achieve.

In order to get to 224 Gbps per lane, new FEC architectures that add FEC between link nodes are being considered. Once again, additional FEC will inevitably add latency which may not be suitable for AI/ML applications.

OSFP vs QSFP-DD

Initially it was expected that pluggable modules for 1.6T would be deployed only using the OSFP form factor due to better heat dissipation and room for components allowed by the larger package. However, recent feedback from industry partners suggests a push to transition to the smaller QSFP-DD form factor soon after.



About Spirent

Spirent Communications (LSE: SPT) is a global leader with deep expertise and decades of experience in testing, assurance, analytics and security, serving developers, service providers, and enterprise networks. We help bring clarity to increasingly complex technological and business challenges. Spirent's customers have made a promise to their customers to deliver superior performance. Spirent assures that those promises are fulfilled.

For more information visit: www.spirent.com

Turning Ideas into Action for 1.6T Ethernet

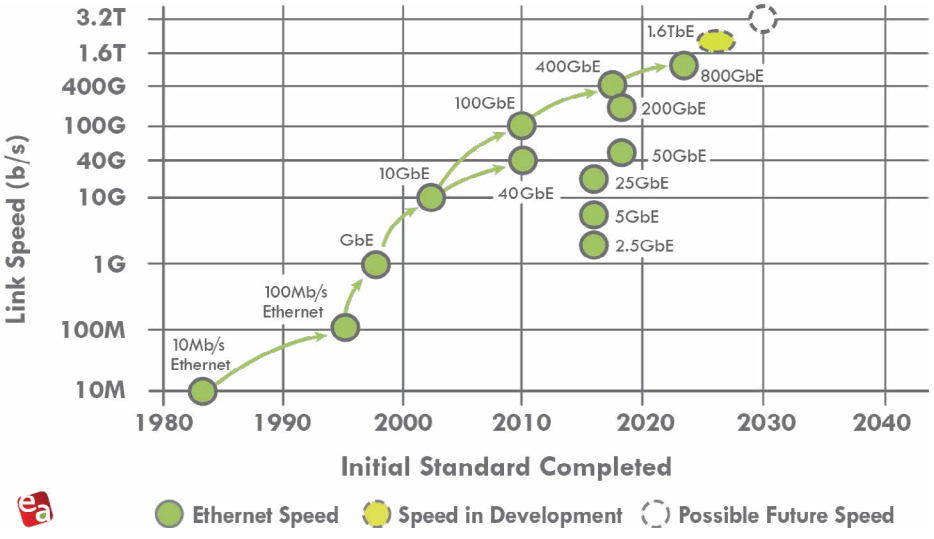
It is important to note that the observations presented in this paper represent the current working view of a dynamically evolving interconnect ecosystem.

It is unclear which interconnect technologies will be more prevalent in the coming years as new connectivity options are now being launched, with early industry demonstrations just starting to validate the interoperability, reliability and high-performance of these solutions.

The industry is closely monitoring the latest innovations in this space as there is still a lot to learn when it comes to Ethernet's next rates. The various interconnect technologies serve different use cases and budget considerations, while DACs may offer a cost-effective option for intra-rack communication.

Regardless of connectivity strategy, reliable performance and interoperability needs to be verified for the next-gen networking. When it comes to accelerating time to market for the interconnect solutions of tomorrow, it is essential to work with a trusted partner who brings decades of experience in Ethernet testing and offers vendor neutral support.

As a leading test & measurement solution provider, Spirent is well-placed to enable the industry to rise to this new technology challenge and turn ideas into action when it comes to 1.6T.



Source: Ethernet Alliance

Americas 1-800-SPIRENT
+1-800-774-7368 | sales@spirent.com

Europe and the Middle East
+44 (0) 1293 767979 | emeainfo@spirent.com

Asia and the Pacific
+86-10-8518-2539 | salesasia@spirent.com

© 2024 Spirent Communications, Inc. All of the company names and/or brand names and/or product names and/or logos referred to in this document, in particular the name "Spirent" and its logo device, are either registered trademarks or trademarks pending registration in accordance with relevant national laws. All rights reserved. Specifications subject to change without notice.

Rev A | 03/24 | www.spirent.com