

# 2026 Data Center eBook

**COMMSCOPE**<sup>®</sup>  
an Amphenol company

# Contents

- Introduction** ..... 3
- Chapter 1** | Cabling Strategies to Support Your Data Center’s Business Model ..... 4
- Chapter 2** | Data Center Interconnect Enables AI Data Centers to Scale Across ..... 11
- Chapter 3** | Practical Considerations for Designing a Future-Ready AI Data Center ..... 19
- Chapter 4** | Networking Technologies in AI Backend Networks: The Battle of InfiniBand™ vs. Ethernet ..... 24
- Chapter 5** | Reducing Time to “Go-Live” for GenAI Deployments ..... 29
- Chapter 6** | Supporting Tomorrow’s Data Rates Depends on Overcoming Fiber-Optic Dispersion ..... 36
- Biographies** ..... 41

## Introduction

### In 2026, two letters are on everyone's lips: AI.

It's part of every conversation, from boardroom strategy sessions to creating a dinner recipe that uses whatever you happen to have in your refrigerator that day.

You don't have to operate a hundred-acre generative AI (GenAI) factory to be part of this amazing revolution. The rapid evolution of the technologies, infrastructures and architecture of AI data centers is being felt downstream, as well—from AI-powered network optimization in the enterprise to AI-enabled applications in central offices and other edge deployments.

To keep up, every data center operator—be it enterprise, multitenant, central office or a dedicated AI factory—must be agile and discerning in how they plan, design, build and upgrade their facilities. The path of evolution isn't linear, and it isn't always obvious, leading to several critical questions about the future. You'll find smart guidance on all these topics in the following chapters:

- What kind of cabling strategies will help boost fiber density without letting complexity get out of hand?
- How can data center interconnect (DCI) architecture help data centers scale across and deal with the problem of electricity scarcity?
- Can changes to the physical layer of your data center help you manage cooling more effectively and efficiently?

- In the battle of network protocols, what are the advantages and limitations of Ethernet and InfiniBand in growing data centers?
- How can data centers shave days or weeks off deployment schedules, just by their choice of cabling infrastructure—and how does shuffle architecture figure in?
- As bandwidth keeps increasing, how close are we really getting to the physical limitations of fiber-optic cabling?

In data centers of all types, all over the world, the race to evolve has already begun. Only a smart blend of innovation, expertise and efficiency will help you pull ahead—and that's what we're bringing you here in CommScope's 2026 Data Center eBook.

# Chapter 1

## Cabling Strategies to Support Your Data Center's Business Model

There are several schools of thought when it comes to cabling networks, usually based on the type and purpose of the network. Decisions regarding how to cable these networks have changed over the years as networks have evolved. However, the recent rapid expansion of data centers to support AI applications, especially generative AI (GenAI), has significantly changed these models. Today's data center designs are highly dependent on the business model of the data center.

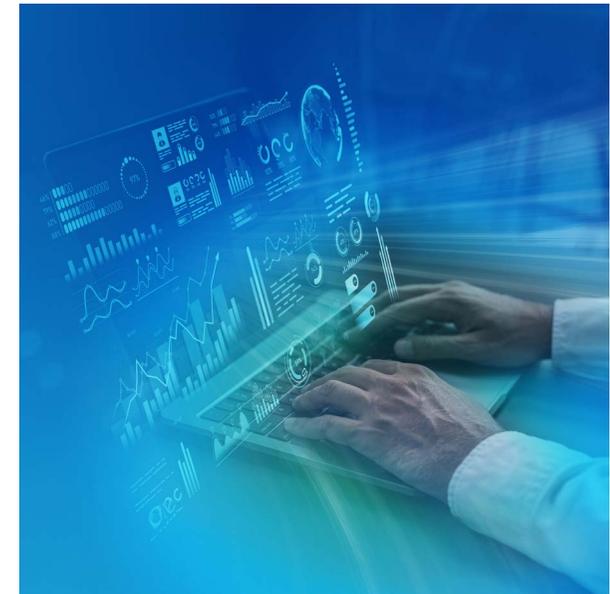
Key considerations include the density and volume of cables, as they have increased by 4x – 8x compared to traditional network cable counts; managing that density to and across cabinet rows; and costs, including not only the materials and labor to install, but also the time, labor and material associated with network life cycle and migration.

Data rates continue to increase, providing greater responsiveness and capacity for users. AI servers or nodes use graphics processing units (GPUs) to deliver greater capacity with each new generation of processor, released with a cadence of every year—or even sooner. Shorter refresh cycles may be justified—if the value outweighs the cost.

Power consumption has grown with newer AI processors, as have equipment cooling demands, requiring innovation in each infrastructure area to

support the network ecosystems. Pathways and spaces to manage connectivity within cabinet rows have not significantly grown, but the volume of cables inside and between cabinets has increased by at least 4x – 8x compared to previous networks. This chapter focuses on this growing cable density, as it directly impacts the flexibility and life cycle of the networks operating across the cabinet row.

With GenAI, there are several cabling options which can coexist within a network. At a high level, the options are either point-to-point or structured cabling, both of which are available in copper and fiber-optic variations. There are benefits and tradeoffs for each, determined by required data rate, distance, flexibility, power efficiency and cost, so the determining factors must depend upon the data center use case or business model.



## Cabling options

**Point-to-point:** Point-to-point (PTP) cabling has network connectors attached on each end with built-in transceivers. These cables are single speed, generally protocol-specific and either passive or active, depending on the application. The varieties of PTP are direct attach cable (DAC), which is a passive coaxial cable; active electrical cable (AEC), which is also a coaxial cable but draws power from the electronics; and lastly, active optical cable (AOC), which transmits over fiber-optic cable and draws power at each end as well.

These cables can be one-to-one connections or one-to-multiple for breakout applications. Figure 1 to the right show possible options. The smaller connectors on the ends are SFP+ for applications such as 10G or 25G. The larger connectors represent QSFP, QSFP-DD, or OSFP for applications 50 – 100G and above. The breakout designs can take 400G on one end and break-out to 2x 200G or 4x 100G, for example.

**Direct attach cables (DACs)** provide a simple point-to-point connection between network devices for short-reach applications. They are considered inexpensive connections within a cabinet, for example, from a top-of-rack (ToR) switch to servers within the rack. These coaxial copper cables are application-specific with SFP or QSFP interfaces at each end and they must be

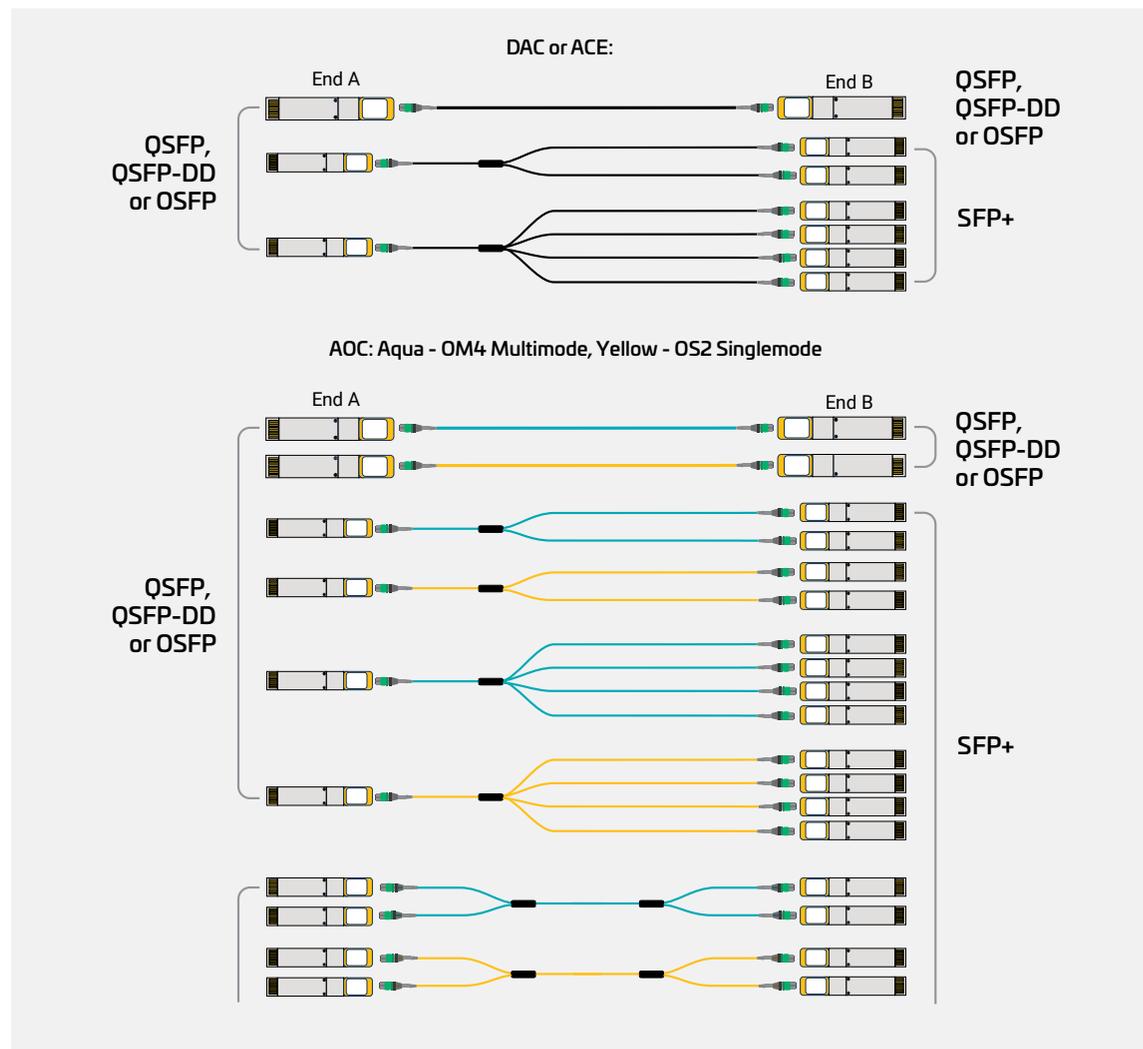
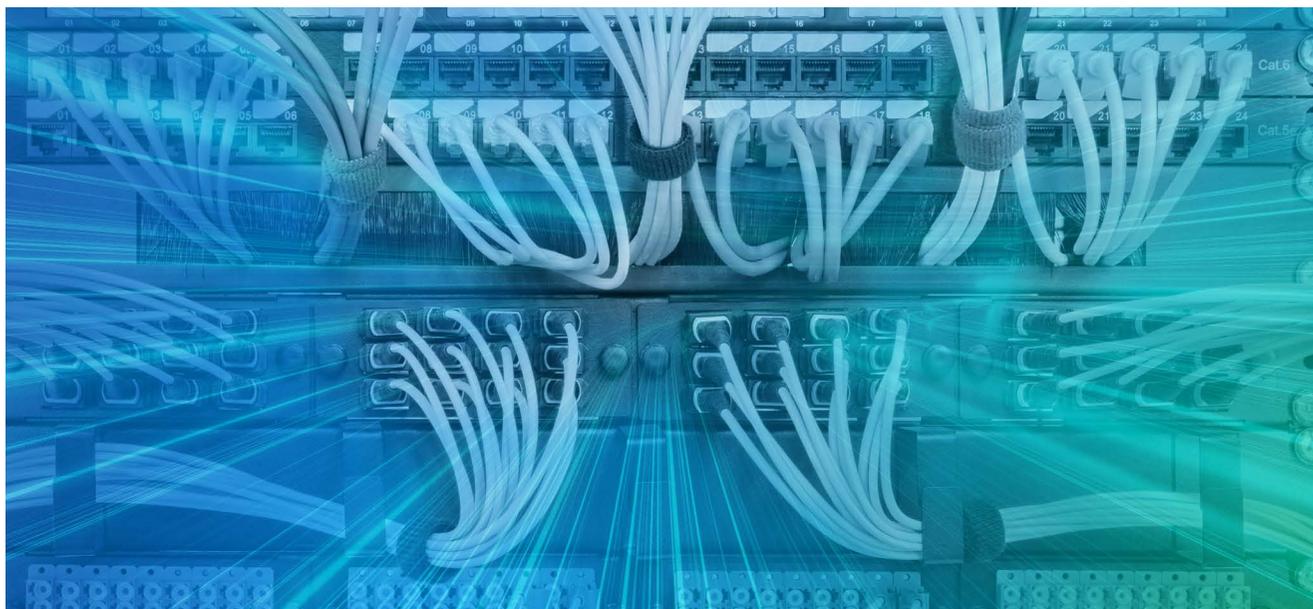


Figure 1. Point-to-point cabling options

removed and replaced as speeds change. They are typically distance-limited to 1 – 2 m or less; these length limitations get shorter as the data rates increase. DAC cables have larger diameters as lengths increase due to extra shielding required for application support, and they are more rigid with larger bend radii as a result. These cables are passive and do not directly draw power from the electronics but carry 0.15 watts per end from the electronics to transmit signals between devices.

**Active electrical cables (AECs)** are also single-application coaxial cables with the SFP, QSFP, QSFP-DD or OSFP PC board interfaces on the ends, but, as the name implies, they are active. That means they draw power from the electronics to support higher data rates, but also to drive the signals further. They may draw 0.5 – 1.0W per end, up to approximately 50% of the power needed for transceiver applications end, but they can draw more for higher data rates or additional functionality. They are typically limited to  $\leq 7$  m reach and are generally used within or between adjacent cabinets. Similar to DACs, these cables are application- and (typically) protocol-specific, twin-axial copper cables and are thicker and less flexible as data rates or distances increase. Their rigidity can make them impractical for spanning



multiple cabinets in a row and a challenge to remove and replace. Considerations for migration to higher speeds should include repurposing or recycling the assemblies.

**Active optics cables (AOCs)** are the fiber-optic equivalent of DACs/AECs and are used primarily for higher-speed and longer-reach applications. They have higher data rate capabilities and can be built to support extended distances with QSFP, QSFP-DD or OSFP connectors on each end. As with DACs, they are normally application- and protocol-specific and must be removed and

replaced with a network upgrade. They are point-to-point between network devices but utilize smaller cable diameters because of their fiber construction. They can be used within a cabinet, across a cabinet row, or further, to network equipment using singlemode or multimode fiber. As application- and protocol-specific infrastructure, they also must be removed and replaced as upgrades occur. Considerations for their use should include refresh rates and the pathways utilized. Most materials within AOCs are not recyclable, so disposal upon refresh should

be properly coordinated. AOCs are used for AI applications at 400G and above. They typically draw 7 – 15W per end, generally in the 12 – 15W range, so proper planning for power and cooling needs is critical.

When running AOCs between cabinets or data center sections it is also important to recognize the weight and bulk of the connector ends which house the transceiver function. As they use small-diameter fiber-optic cables, care must be taken to minimize weight and stress at the connectors. This is especially a concern when entering or exiting cabinets where other cables or obstructions are present. Guidance from leading AI OEMs is to avoid using AOCs for data rates of 100G or higher due to weight and bend stress from the connector at the end of the smaller fiber cable assembly. For higher speeds, structured cabling with transceivers is recommended instead.

**Structured cabling systems (SCSs)** use passive cable and connectivity and are by nature application- and protocol-independent, supporting multiple network generations over multiple connection points between locations and equipment as needed. For fiber-optic cables,

passive connectors such as MPO12/8, MPO16 or LC duplex are terminated on each end for plugging into a transceiver at the network equipment. Fiber-optic transceivers draw similar power from the electronics at each end as AOCs, since they utilize power to convert electrical signals to optical. For applications of 400G and above, they typically draw 7 – 15W per end, and, like AOCs, generally in the 12 – 15W range per end.

With structured cabling, high fiber count, pre-terminated trunk cables are installed between network locations, typically via overhead fiber raceway. Those overhead trunk cables provide connectivity between locations within the data center architecture and also between locations within the cabinet row or POD. Patch panels are

utilized within or above the cabinets to provide transition or disconnect points between the trunk and equipment cables. That disconnect point provides flexibility for equipment patching and growth within the cabinet. Fiber patch cables from the panel connect to end devices within the cabinet, completing the circuit from the trunk cables above. If the architecture model can use preconfigured cabinets, then a network upgrade can be quickly accomplished by disconnecting trunk cables from the cabinet, rolling out the original cabinet, and replacing with the upgrade cabinet by reconnecting the trunk cables.

Figure 2 shows is a simple SCS channel using MPO trunk through adapters to breakout cables.



Figure 2. SCS channel using MPO trunk through adapters to breakout cables

Figure 3 depicts a representative baseline AI cabinet row including 32 GPU nodes within the eight outer cabinets and two middle-of-row management cabinets housing aggregated leaf-and-spine switches, compute and storage connections. The cabinet row cable counts are in the lower left of the graphic, recognizing 48 total MPO12/8 connections in each of the eight cabinets.

If cables exiting the cabinets are point-to-point, then the cabling effectively stitches the cabinets together and possibly to the overhead raceway. That may be acceptable for several years of operation in a static network, but when change is needed, those single-application cables must be removed, discarded or recycled and replaced. That will take time, affecting network availability. For dynamic networks like GenAI in particular, flexibility for change may be especially critical.

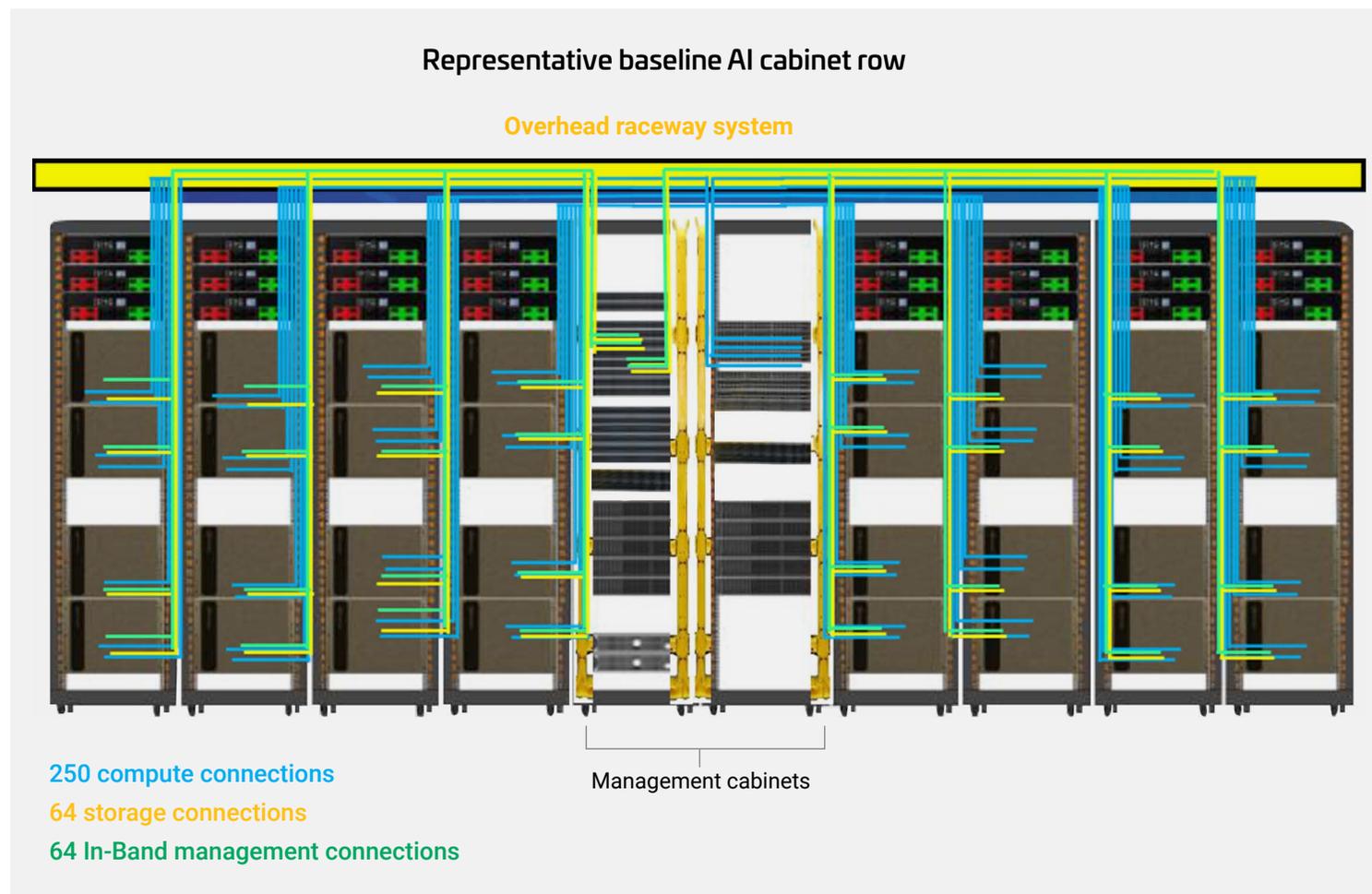


Figure 3. A representative baseline AI cabinet row

It should also be noted that Figure 3 is of an early generation scalable unit, with current options providing over 432 MPO12/8 connections to support 256 GPUs in the space of 5 – 6 cabinets (including switch cabinets).

## Relative time, reach and installed cost comparison:

| Media        | Application     | Reach at 100G lanes (400G or 800G ports) | Reach at 200G lanes (800 or 1.6T ports) | Cost per link | Installation time (1-5) | Applications       | Power   |
|--------------|-----------------|--|---|---------------|-------------------------|--------------------|---------|
| DAC (CR)     | Ethernet CR     | 2 m                                      | 1m                                      | <x            | 1                       | Server to ToR      | Low     |
| AOC MMF      | Ethernet VR, SR | 50 m, 100 m                              | 50 m                                    | 2x            | 4                       | Server to MoR, EoR | Medium  |
| AOC SMF      | Ethernet DR     | 500 m                                    | 500 m                                   | 2x            | 4                       | T1 to T2           | High    |
| Parallel MMF | Ethernet VR, SR | 50 m, 100 m                              | ~50 m                                   | x             | 2                       | Server to MoR      | Medium  |
| Parallel SMF | Ethernet DR     | 500 m                                    | 500 m, 2 km                             | 2x            | 4                       | T1 to T2           | High    |
| Duplex SMF   | Ethernet FR     | 2 km                                     | 500 m, 2 km                             | >2x           | 3                       | T2 to T3           | Highest |

## Conclusion

There are benefits and compromises for each cabling option, based upon data rate, distance, flexibility, power and cost. By design, GenAI networks repeat configurations across cabinet rows and switch layers to build and scale capacity. Those repeated configurations enable the use of pre-configured cabinets built offsite, then

rolled in and connected to trunk cables and other infrastructure systems to reduce time to network availability.

Regardless of the data center type, the cabling system used should both support the initial mission of the data center and provide a foundation for future applications and migrations.

*By Kenneth Hall, Solutions Architect,  
Data Center Solutions*

# Chapter 2

## Data Center Interconnect Enables AI Data Centers to Scale Across

In the context of this article, we will understand data center interconnect (DCI) to cover the physical infrastructure that connects two or more individual data centers together—allowing us to explore the use cases for DCI and methods of transmission. The geographic scope of DCI varies greatly, as it may include multiple data centers across a connected campus; however, it is more commonly used to describe clusters of data centers across several cities (also known as an “availability zone”). At even larger scale, data center networks spread across the world are connected to create a global DCI.

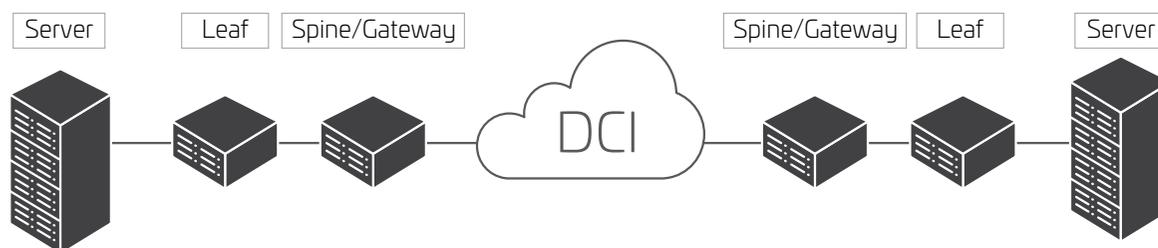


Figure 1: Basic architecture of a DCI

### Data center interconnect and the importance of scaling across for AI

Why has DCI become an important topic in today’s data center world? DCI has existed for a long time and has always been important; however, there’s been a significant recent change brought about by AI applications that has made it a critical parameter in the overall performance and economics of training large language models (LLMs)—and that change is centered on scaling.

To put this into context, a scale-up network is used to interconnect a large number of GPUs in a single cabinet with one another, forming a single block of compute. When the power, space or cooling to host more GPUs in that cabinet is exhausted, adjacent cabinets within the same data center or data hall are connected together to create a scale-out network. Scale-out networks lead to a significant increase of both compute

capability and the amount of electrical power required to keep the GPUs running. Power availability is a common logistical limiting factor to the amount of infrastructure that can be housed in a single data center building. This drives the need to build new data centers, either on a campus, across a city, or across a region to host the additional compute resources required for the training of large and hyperscale size models.

For these model sizes that require distributed compute, geographically discrete data centers must be orchestrated to work together, and this is where DCI plays a crucial role. In fact, it's become so important that there's a new term that has evolved to describe this DCI network: scale-across.

Data centers that are used to train the largest LLMs frequently no longer exist as individual isolated islands but have become part of an archipelago of buildings that have formed a highly interconnected scale-across fabric, expanding horizontally to seamlessly enable increased flows of east-west traffic patterns. This scale-across network is collectively known as an AI-optimized DCI.

[According to the Dell'Oro Group](#), the annual capacity purchased for direct DCI is expected to show a threefold increase between 2025 and 2028, with the majority of interfaces shipped being 800G, 1.2T or 1.6T coherent optics, ZR and ZR+ transceiver standards.

### New DCI traffic types

DCI traffic has become extremely varied due to the demands of AI. For example, today's DCI networks can expect to transport:

- [AI model parameters \(e.g., all-reduce, all-to-all\)](#)
- [AI gradients](#)
- [GPU-GPU and node-node synchronization data](#)
- [Large latency-sensitive data flows](#)
- [Control and orchestration data for fault detection, scheduling, fabric telemetry, fault detection and recovery signaling.](#)

What's important to understand is that the majority of this data is intermediate compute data that keeps the fabrics and distributed workloads synchronized, rather than user data (such as images or documents). Although these DCI traffic patterns are driven mainly by training activity, inference AI can also require large amounts of data transported via DCI to synchronize its operations depending on the architectural model being deployed. At present, AI training (rather than inference) is the biggest driver of traffic;

however, this is expected to change in the future as the industry moves from primarily training to deploying more infrastructure to support inference.

### AI presents multi data center DCI challenges

With large AI models having 70 billion parameters or more, and hyperscale models like ChatGPT 4 utilizing over one trillion parameters, there are key infrastructure challenges that come into play when managing compute across multiple data centers. For example, there will be high inter-region latency (often 20 milliseconds or more, [as explained by NVIDIA in this technical blog](#)), which can introduce performance bottlenecks during large-scale LLM training.

With this and many other factors in mind, DCI network infrastructure must be designed with sufficient optical fiber capacity to support low-latency and high-bandwidth channels, ensuring fast and unconstrained communication flow between data centers to maximize training efficiencies.

### DCI transmission

There are a number of ways to transmit data between data centers over DCI. The technology can involve sending high-speed Ethernet via optical signals over dedicated fiber or wavelength services. In most cases, a DCI requires a high-speed WAN link, which could mean using MPLS, Ethernet, VPLS, Metro Ethernet, and so forth.

Selecting the right DCI infrastructure, architecture and topology depends on a wide range of variables, including the location of the data centers, the distance between data centers, bandwidth and availability requirements, the capabilities of local service providers, and security concerns.

There are just as many options and variables when you get down to the component level. Following are just a few of the solutions and developments that may figure into your decision.

### Border edge (gateway) devices

Leaf-and-spine switches are integral to creating a data center fabric and are used to ensure efficient east-west transmission of data. As a data center approaches cloud scale, the spine-layer switches are typically dedicated to facilitating the east-west data flow. A new group of switches, known as “border edge switches,” has been added to the

architecture to handle incoming and outgoing traffic at the edge of the data center fabric. In some cases, the job of handing off data to the DCI network can be performed at the leaf layer, using a border leaf switch.

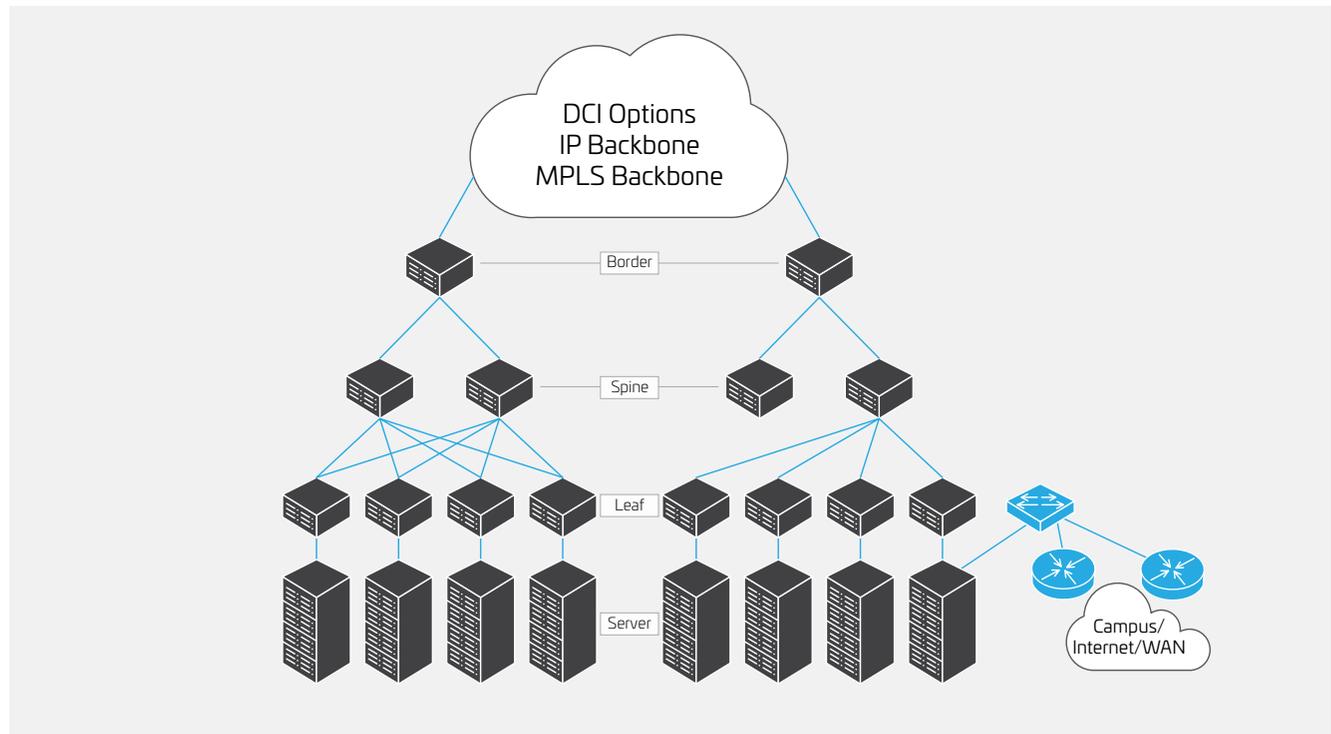


Figure 2: Example of a border edge switch facilitating DCI at the leaf layer

### Coherent optics

Coherent transmission has been used in long-haul and undersea links for about a decade, but we now see this technology migrating to the data center. Over time, the makers of coherent transceivers have reduced the size, power and cost of their optics to be more attractive for shorter and shorter links. IEEE Ethernet standardized the use of

100G and 400G coherent signaling on fiber links up to 80 km. A new project (802.3dj) will write standards for 800G coherent over just 10 km. The Optical Interworking Forum (OIF) has published the 400ZR and 800ZR standard specifying 800G transmission over a single wavelength with reaches of 80-120 km. Work on 1600ZR is currently underway at OIF.

As data centers continue to grow and become more interconnected, the bandwidth needs of the DCI network are also growing, reaching 100T over multiple wavelengths. This bandwidth demand is supported by faster data rates—400G and 800G per wavelength—and will only continue to escalate.

Coherent optics are typically used for ultra-high bandwidth applications ranging anywhere from 100G to 1T over very long distances. Powerful digital signal processing chips (DSPs) are embedded within these systems to mitigate linear effects caused by fiber impairments, including chromatic dispersion and polarization mode dispersion.

Coherent fiber optics utilize the natural properties of light to optimize digital modulation practices and fiber-optic carrying capacity in long-range applications. However, coherent transmission will change to suit DCI applications. Supported by a proposed new standard (IEEE802.3dj), this will attempt to enable 800G links greater than 10 km.



Figure 3: QSFP-DD coherent optic module

Figure 4: CFP2 coherent optic module

### Dense wave division multiplexing (DWDM)

DWDM enables multiple wavelengths of light to travel over the same fiber simultaneously, with each wavelength carrying a discrete signal. Tight wavelength spacing can enable up to 96 channels on a single fiber. DWDM is a versatile transmission technology able to support coherent optics and on-off keying (OOK) amplitude signaling. When combined with coherent modulation, individual channel bandwidth can expand to 400 or 800 gigabits (see Figure 5).

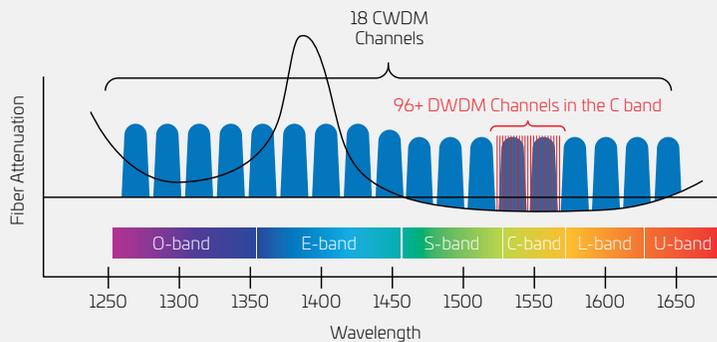


Figure 5: CWDM/DWDM band and channel configurations

### QPSK coherent coding vs. PAM4

Coherent systems are based around phase shift keying (PSK)—phase modulation techniques that allow multiple symbols per bit to be encoded based on four phase shift orientations (e.g., 0°, 90°, 180°, and 270°). Many systems use quadrature phase shift keying (QPSK) to encode two bits per symbol. Dual polarization QPSK (DP QPSK) uses horizontal and vertical polarization along with QPSK to represent twice as many bits (see Figure 6).

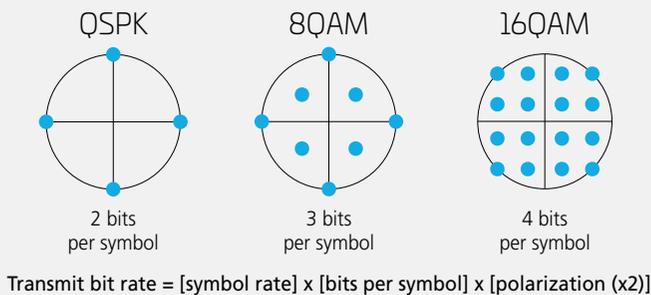


Figure 6: High-order modulation—Constellation diagrams

Pulse amplitude modulation (PAM4) is a four-level modulation scheme designed for short-haul fiber links. PAM4 (see Figure 7) uses four amplitude pulses, each containing two bits, to double the bandwidth of conventional binary signaling. The simplicity and low power requirements of PAM4 make it a popular option for 100G and 400G Ethernet applications.

Unlike coherent optics, PAM4 is highly susceptible to fiber impairments. This limits range to  $\leq 10$  km, whereas coherent optical fiber communication systems can potentially span thousands of kilometers on amplified links.

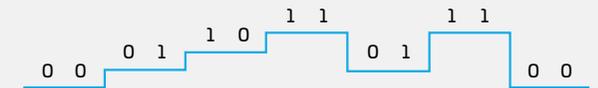


Figure 7: PAM4 signaling technology typically uses gray coding, reducing bit errors due to each bit only being one bit different than the adjacent levels.

### New designs driving higher fiber-count DCI networks

The expansion of east-west data flows and the move to a low-latency leaf-and-spine switching architecture has created a tsunami of data inside and among data centers, particularly AI data centers where training LLMs is the key function. This has pushed the development of data center campuses with multiple buildings into overdrive. This trend isn't isolated to a few hyperscalers in the U.S.; it also impacts cloud and multi-tenant data center (MTDC) providers globally.

### The physical layer

Once the external cable has been terminated inside the data center, it must be presented for splicing to an onward destination or connected to patching and cross-connection equipment. This is done using optical distribution frames (ODFs) located in the meet-me room (MMR) or main distribution area (MDA), where all network cabling comes together and is prepared for distribution.

To ensure this zone can support future growth and is manageable, patch cord hygiene should be mandatory. As noted earlier, the role of the DCI network across the campus is increasing; in the

near future, they will have to support throughput of 100T or more. This will require thousands of fibers—all converging at the MMR or MDA, underpinning the importance of good patch cord

management. So, too, ensuring the ODF has both fiber patch cord routing and slack management will be key to ensuring the MDA and MMR can support all future growth needs.

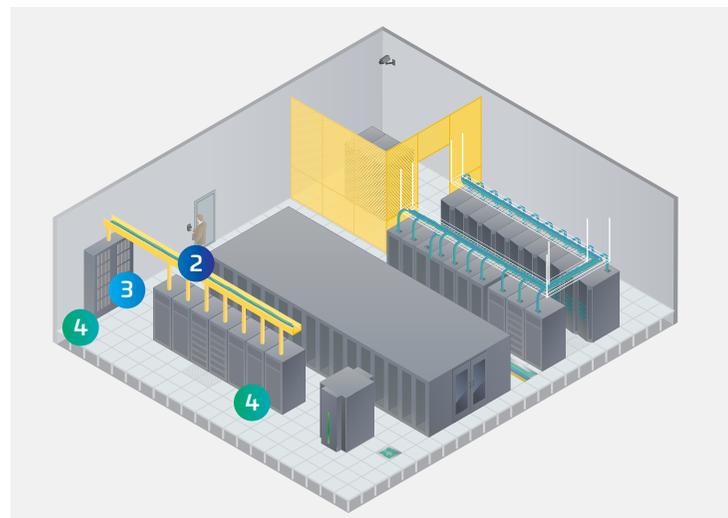


Figure 8: Entry into the data center building or MTDC meet-me room

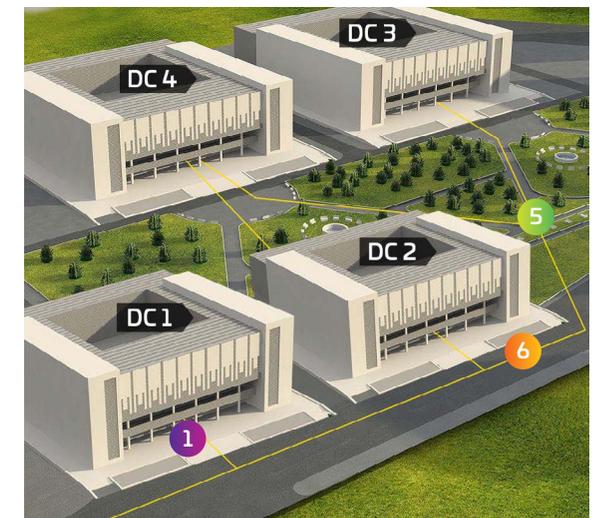


Figure 9: Infrastructure components for a typical campus-scale DCI

### To the future

Increasing deployment and expansion of DCI networks across the campus and into the cloud will continue unabated. As AI demands more scaling in networks, more and more data is created and shared by GPUs that never take a break or grow tired. As a result, data loads and scale-across DCI network requirements keep increasing. This presents physical challenges for the campus network designer, especially if the network must interconnect data centers across a metro area or land owned by others.

If the data is to traverse these geographies seamlessly, the data-carrying capacity of the fiber cables must increase. This will be done either by adding more fibers per cable—via size-reduced cladding—or introducing a commercially viable alternative glass technology, like hollow core fiber for example. But regardless of specific technology, DCI will continue to grow in importance to the growth and evolution of AI data centers.

*By Alastair Waite, Senior Manager,  
Global Data Center Market Development, CommScope*



# Chapter 3

## Practical Considerations for Designing a Future-Ready AI Data Center

Designing, implementing and managing data cabling in an AI data center involves unique challenges due to the introduction of high-performance processor chips to the servers (i.e., Accelerator Application Specific Integrated Circuits like GPUs and TPUs). These high-performance integrated circuits are capable of parallel processing huge amounts of data, and when combined with High Bandwidth Memory componentry in a rack-scale system (RSS) are capable of achieving 3.6 exaFLOPS of compute performance.

Today, some of the highest-performing AI superchips can process at **speeds of up to 40 petaFLOPS<sup>1</sup>** (floating point operations per second). For comparison, the performance of a traditional central processing unit (CPU), of the type commonly found in data center servers, has been found to provide theoretical peak performance of around **1,200 times slower than the latest GPU.<sup>2</sup>**

As mentioned previously, in an AI cluster, an Accelerator Application Specific Integrated Circuit (AASIC) does not operate on its own but rather works in parallel with multiple other AASICs and supporting hardware like High-Bandwidth Memory (HBM), scale-up and scale-out switching hardware, power supplies and cooling infrastructure. Often, an RSS will be configured with up to 72 GPUs working together in a chassis or rack, each GPU requiring huge amounts of data bandwidth flowing in and out via 400G, 800G and even 1.6T channels to maintain peak operating performance.

### The cooling challenge

With increased performance comes new cooling challenges that can impact the compute cabinet and, more specifically, the space available to introduce and manage fiber-optic cabling. Nothing comes for free, and the increased performance of these systems comes at the cost of higher power consumption that in turn creates heat as a byproduct. This heat must be evacuated from the servers and racks to keep the data center hardware running at optimum performance levels.

Previously, compute rack power amounted to 8 to 10 kW, and the resulting heat could be easily transferred away from active hardware using chilled air to maintain a constant ambient temperature of around 77°F (approximately 25°C). For the past 20 years, most data centers have been designed to support such air-cooled operations.

As higher-performing compute hardware with higher thermal densities have been introduced, rack power requirements have moved beyond 10 kW and are tracking upwards—as high as 80 or 100 kW—with some cloud providers forecasting rack-level power loads of 0.5 MW, levels that air cooling alone cannot efficiently or cost-effectively support. This trend has pushed designers to consider liquid cooling for their operations. Water and other fluids can be 50 to 100 times more efficient than cooling with air alone, which promises to address the thermal challenges of modern AI clusters. That fluid must be pumped around the data halls, into the rows and racks, and finally into the server chassis hosting the powerful accelerators.

It's at this point where the physical layer cabling design takes on enhanced importance, because we have to consider how to physically connect all those accelerators to one another in a coherent

way that doesn't impede the performance of the overall AI architecture. Plus, this task is complicated by the introduction of additional cooling hardware (in the form of fluid piping), alongside a dramatic increase in the number of fibers required to fit into a small footprint, whether that be a server chassis or an entire rack. **Figures 1 and 2 below illustrate the challenge facing designers when regarding cooling AI clusters.**

Additional DTC pipework infrastructure consumes real estate inside the rack that was once available for structured cabling. This further stresses the interior space's capacity as the number of fiber cables deployed in AI cabinets is also on the rise.

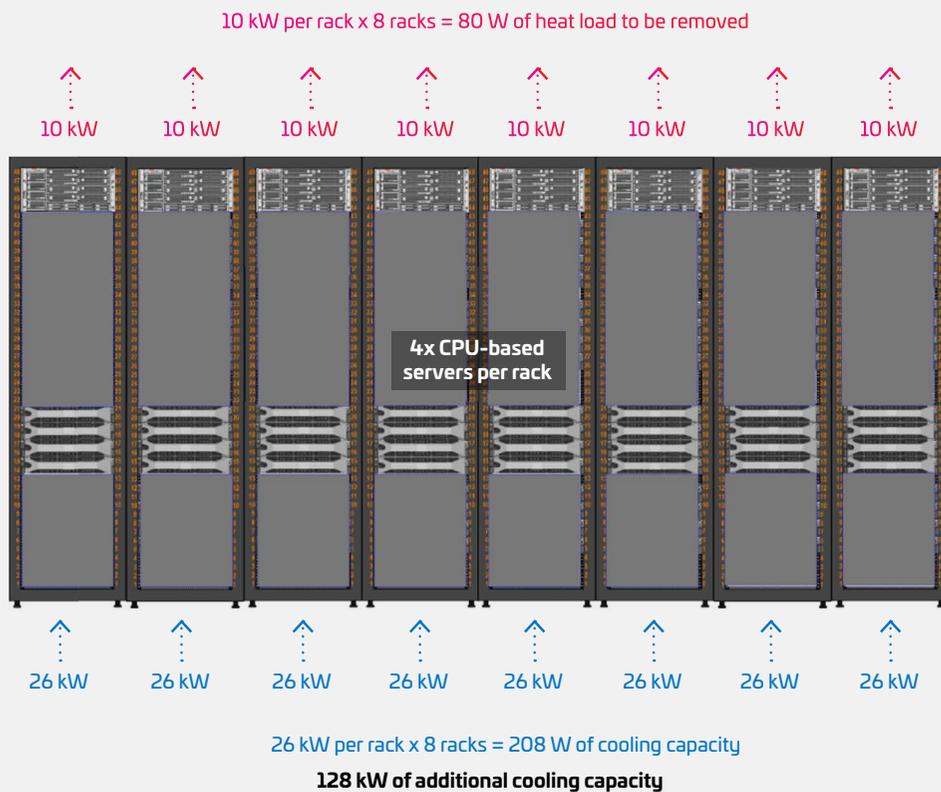


Figure 1: 4x CPU-based servers per rack: Air-cooled capacity in a data center using CPU-based servers. This scenario demonstrates surplus air handling capacity.

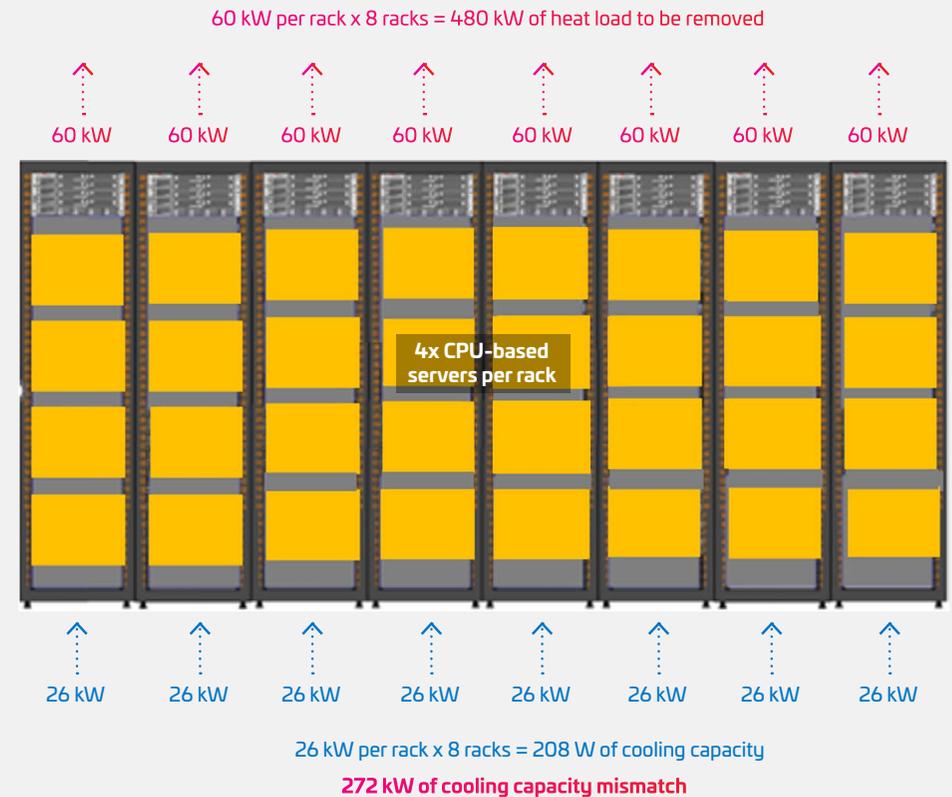


Figure 2: 4x GPU-enabled servers per rack: Air-cooling system is no longer sufficient and requires more cooling capacity, such as from liquid cooling in the form of direct-to-chip (DTC) and its additional pipework for each cabinet, rack and row.

## Key points in designing a physical layer to support AI clusters

### 1. Additional networks must be considered.

In a traditional data center design, we typically have the east-west network fabric cabling (e.g., server to T0 switch, and T1 to T2 switching layers), plus the out-of-band management and dedicated SAN links.

However, an AI cluster deploying the new NVIDIA NVL72 RSS sees each rack typically having the following scale-out fiber network links:

- 72x compute back-end/GPU fabric
- 36x storage + in-band front-end/CPU fabric
- 4x out-of-band management (OOB), an out-of-band management copper link, using Category 6 or Cat 6a structured cabling
- 8x Cat 6 or Cat 6a copper links for the admin network
- 20x additional fiber cabling links to support redundancy

Adding these up, it's quickly apparent that the traditional cabling approach for compute cabinets used in non-AI environments will no longer be a sustainable strategy in an environment where

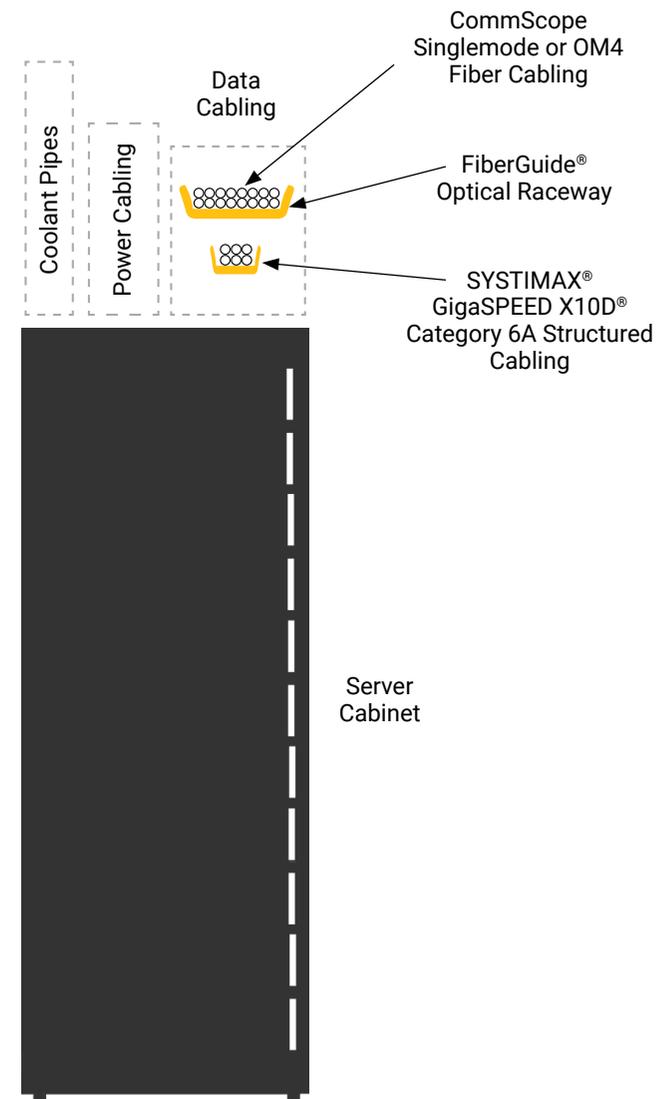
liquid cooling infrastructure is deployed alongside fiber cabling. There is only so much “give” in the network.

### 2. Infrastructure zoning and overhead pathway positioning.

Having individual zones above cabinet rows can prevent a conflict between data cabling, power feeds and the liquid piping used to transport liquid cooling fluid. Protecting the fiber cabling across the data center hall and having maximum run lengths for AI-enabled servers will help maximize performance. Introducing a fiber raceway like the FiberGuide® system is an effective platform for transporting cabling.

### 3. Cabinet selection and cable management.

Server racks come in different sizes: 600 mm and 800 mm widths. It might be tempting to select a smaller footprint to conserve space, but in the world of AI, the vertical riser space must be sufficiently large to enclose all the additional infrastructure that is now required. Selecting a wider—and deeper—cabinet is often the more prudent choice. However, this too brings challenges that make traditional cable patching from the rear of a cabinet very difficult, because of the lack of physical space to access the rear



of panels. In this instance, designing in-cabinet patching solutions with all-front access platforms, like the Rapid Fiber Connect™ platform, can significantly improve installation time and easier Day 2 moves, adds and changes (MACs).

**4. Maximizing GPU performance.** Essential to AI performance is the amount of low-latency, high-bandwidth data provided to the GPUs. High-quality data cabling and optical end-face performance are considered “table stakes” today.

As fabric bandwidths continue to increase, underpinned by higher data rates, singlemode, and OM4 optical fiber, structured cabling is essential. Many optical channels require so much bandwidth that parallel optics must be deployed. These channels require parallel connectivity like MPO16 and MPO12/8 supplied with APC polishing, preventing unwanted optical reflections that impact the performance of the link.

**5. Cable organization.** Use structured cabling to have easier MACs and to streamline future scalability for network growth, especially as the pace of evolution in AI hardware is so rapid.

Prioritizing and acting on these areas of focus early on in an AI data center build can help maintain efficient, reliable and scalable data cabling solutions to support their demanding computational workloads. As GPUs demand more resources to power their growing levels of performance, data centers must keep a future-focused view on their buildouts.

*By Alastair Waite, Senior Manager,  
Global Data Center Market Development, CommScope*

<sup>1</sup> <https://developer.nvidia.com/blog/inside-nvidia-blackwell-ultra-the-chip-powering-the-ai-factory-era/>

<sup>2</sup> <https://www.heise.de/en/news/AMD-Epyc-9005-server-CPU-First-test-confirms-high-performance-and-efficiency-9976883.html#:~:text=The%20Epyc%209755%20with%20the,AMD's%20Zen%205%20compact%20cores>

# Chapter 4

## Networking Technologies in AI Backend Networks: The Battle of InfiniBand™ vs. Ethernet

Ethernet protocol has traditionally dominated data center networks, in both cloud and enterprise environments. Its long-term evolution and robust ecosystem of suppliers and experts with decades of experience have given it a significant edge over other networking protocols. However, the rise of generative AI (GenAI), exemplified by large language models (LLMs) like ChatGPT, and the market dominance of leading graphics processing unit (GPU) suppliers like NVIDIA, have brought the InfiniBand networking protocol into the spotlight.

InfiniBand is a high-speed networking technology and industry-standard specification primarily designed for high-performance computing (HPC) applications. NVIDIA's acquisition of Mellanox (a leading InfiniBand supplier) and their subsequent promotion of the protocol for their AI platforms have propelled InfiniBand to become the leading networking protocol in AI backend networks.

That doesn't mean Ethernet is giving up without a fight, however. This analysis examines these two competing technologies in AI backend networks and provides insights into the future of AI networking applications.

### AI clusters: A fabric of intelligence

AI cloud data centers comprise two critical networks—frontend and backend. The frontend network, dominated by Ethernet, serves as the client-facing interface that connects AI to the

broader cloud infrastructure and ultimately to the outside world. The backend network, where InfiniBand has gained prominence, houses the AI

cluster, a fabric of interconnected GPUs that deliver AI functionality. Figure 1 illustrates the frontend and backend networks in an AI data center.

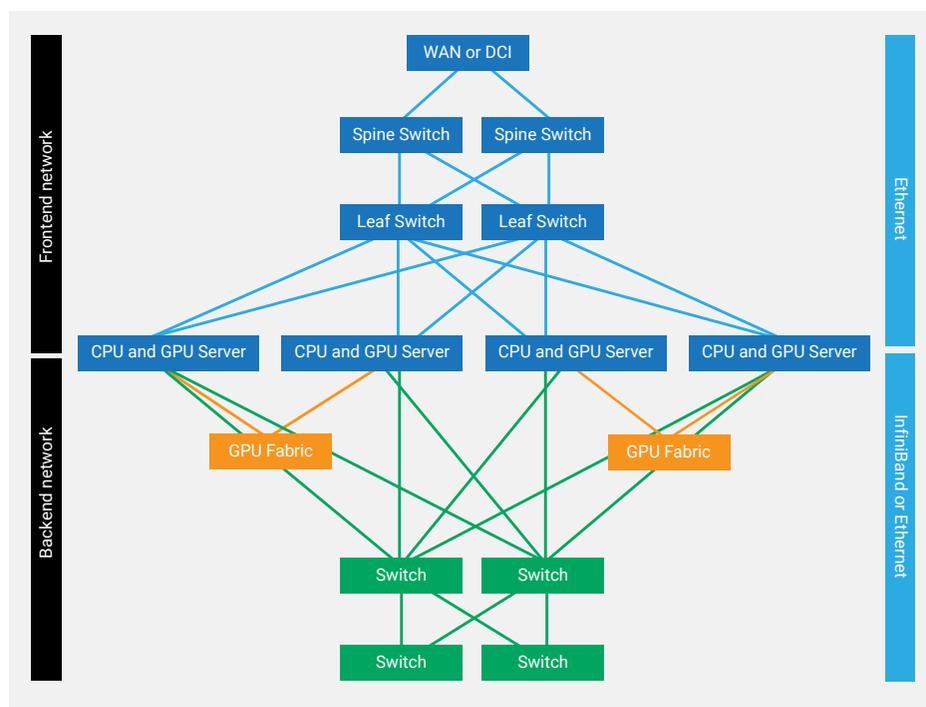


Figure 1: AI data centers include both frontend and backend networks.

AI functionality can be predictive AI or GenAI. Predictive AI utilizes data to forecast or infer likely predictions and outcomes. GenAI utilizes data to generate something new, such as text, music or graphics. Both types of AI functionality necessitate “on-the-job” training of AI models before they can provide services. This training process occurs within the GPU fabric of the AI backend network.

A prime example of AI training is a machine learning (ML) model that analyzes extensive datasets to learn parameters, enabling them to recognize patterns, make decisions or predict outcomes. In 2020, OpenAI released the Generative Pre-Trained Transformer 3 (GPT-3), a state-of-the-art LLM with 175 billion parameters and, while official numbers have not been made available, GPT-4 is rumored to include **1.8 trillion parameters**.<sup>1</sup> Meanwhile, Meta’s deep learning recommendation model (DLRM) boasts several trillion parameters and is poised for further growth. This training process is computationally intensive, requiring powerful processors, vast amounts of memory, and high-speed data storage and transfer capabilities.

The backend network plays a critical role in both the training and inference phases of AI

workloads. During training, the backend network facilitates the transfer of large volumes of data between storage, compute nodes, and other components. This requires high throughput, low latency, and the ability to handle data traffic bursts without congestion or loss. When the network underperforms, expensive AI clusters are underutilized. Therefore, the infrastructure interconnecting these compute resources must be as efficient and cost-effective as possible.

### The impact of network performance on AI

Primarily designed for high-performance computing (HPC) environments, the InfiniBand protocol offers exceptionally low latency and high bandwidth, making it ideal for applications like AI training that require a predictable and lossless network fabric. Compared to traditional Ethernet, InfiniBand exhibits lower latency and fewer congestion-related packet losses. In terms of bandwidth, Ethernet and InfiniBand offer comparable performance, with Ethernet enjoying an edge in market adoption. Let’s delve deeper into latency, packet losses, and bandwidth considerations.

### Latency

Latency, in simple terms, is the time it takes for data to transmit from one point to another on a network. In AI networks, latency occurs at two locations—at the network switch and the network interface card (NIC). Tail latency, the delay experienced by the slowest packets, is especially concerning as it can significantly impact AI training times.

Traditional Ethernet switches typically use a “store-and-forward” switching technology. This involves receiving and storing the entire data packet within the switch, checking it for errors, and transmitting it to the destination. Other switch types use the faster “cut through” technology where the switch starts forwarding the packet as soon as it reads the destination address, without waiting for the entire packet to be received. This means minimal error checking, but lower latency compared to “store-and-forward.” Many modern Ethernet AI switches can dynamically adjust between these modes based on network conditions and requirements, striking a balance between speed and reliability.

At the NIC level, InfiniBand uses Remote Direct Memory Access (RDMA) technology that allows two servers to read and/or write to each other's memory without involving the server's processor, cache or operating system (i.e., kernel). Network, storage, and compute applications benefit from this technology. By bypassing the networking kernel, RDMA frees up the CPU, lowers latency and provides much faster data transfer between two servers. Modern Ethernet NICs support AI implementations using RDMA over Converged Ethernet (RoCE). This enables direct memory exchange over an Ethernet network, effectively allowing Ethernet to achieve latency levels comparable to InfiniBand.

### **Packet loss**

Packet losses—and the subsequent re-transmission of data packets—significantly impact the overall performance of the AI backend network, making a lossless network protocol crucial for optimal operation. InfiniBand is designed as a lossless networking protocol, incorporating congestion management, buffer control, and quality of service (QoS) mechanisms.

While Ethernet is not inherently lossless, it can be configured to operate in a lossless manner



by implementing specific techniques and methods. These include priority flow control (PFC), acknowledgment congestion control (ACC) and enhanced transmission protocols like the aforementioned RoCE to bring Ethernet close to the lossless performance of InfiniBand.

### **Bandwidth**

Last, but not least, is the importance of network bandwidth. While both Ethernet and InfiniBand currently support 800G as their fastest data rate,

Ethernet is poised to achieve 1.6T by the beginning of 2027 through ongoing standards development, with the development of 3.2T also in sight. A key factor is the market adoption of transceivers supporting these faster speeds.

InfiniBand currently uses 800G with fiber-optic transceivers featuring 2x400G ports, enabling 400G bandwidth per GPU in an AI network. In contrast, the broader and more established Ethernet ecosystem already offers fiber-optic transceivers featuring a single native 800G port.

## The outlook ahead

New “Ethernet-boosting” technologies have been developed and are in the pipeline. The [Ultra Ethernet Consortium \(UEC\)](#), established in 2023 by leading Ethernet vendors and major users, is working to modernize RoCE to improve bandwidth, latency, tail latency and scalability. The Ultra Ethernet Transport (JET) protocol is central to the UEC's efforts. This new protocol optimizes next-generation AI and HPC networks through multiple innovations that increase network utilization and reduce tail latency, both critical factors for accelerating AI training times. Furthermore, the UEC is defining these specifications while preserving the following key Ethernet advantages, [as outlined in a UEC whitepaper](#)<sup>1</sup>:

- **A robust, multi-vendor ecosystem**, comprising interoperable Ethernet switches, NICs, cables, transceivers, optics, management tools and software from numerous participants.
- **Proven addressing and routing scale**, leveraging the scalability of IP networks for rack-scale, building-scale, and data center-scale deployments.

- **A comprehensive suite of tools for testing**, measuring, deploying, and efficiently operating Ethernet networks.
- **A proven track record of cost reduction** driven by a competitive ecosystem and economies of scale.
- **A demonstrated ability to rapidly advance**, illustrated by IEEE Ethernet standards which have quickly and regularly progressed across various physical and optical layers.

With these advancements, it appears that Ethernet will likely do in AI backend networks what it has done in frontend networks—dominate. Chip giants like Intel and AMD support Ethernet with their new-generation AI nodes, and major switch manufacturers (including NVIDIA) are implementing enhanced Ethernet into their systems. In fact, one of the world's largest AI clusters was recently built using Ethernet. Leading market research firms in the AI networking environment appear to have also already reached a consensus, predicting that Ethernet-based AI implementations will surpass those using InfiniBand by 2027.

*By Hans-Juergen Neithammer,  
Solutions Architect, Data Center Solutions*

<sup>1</sup> [Number of Parameters in GPT-4 \(Latest Data\)](#) by Josh Howarth, published on Exploding Topics, June 17, 2025

# Chapter 5

## Reducing Time to “Go-Live” for GenAI Deployments

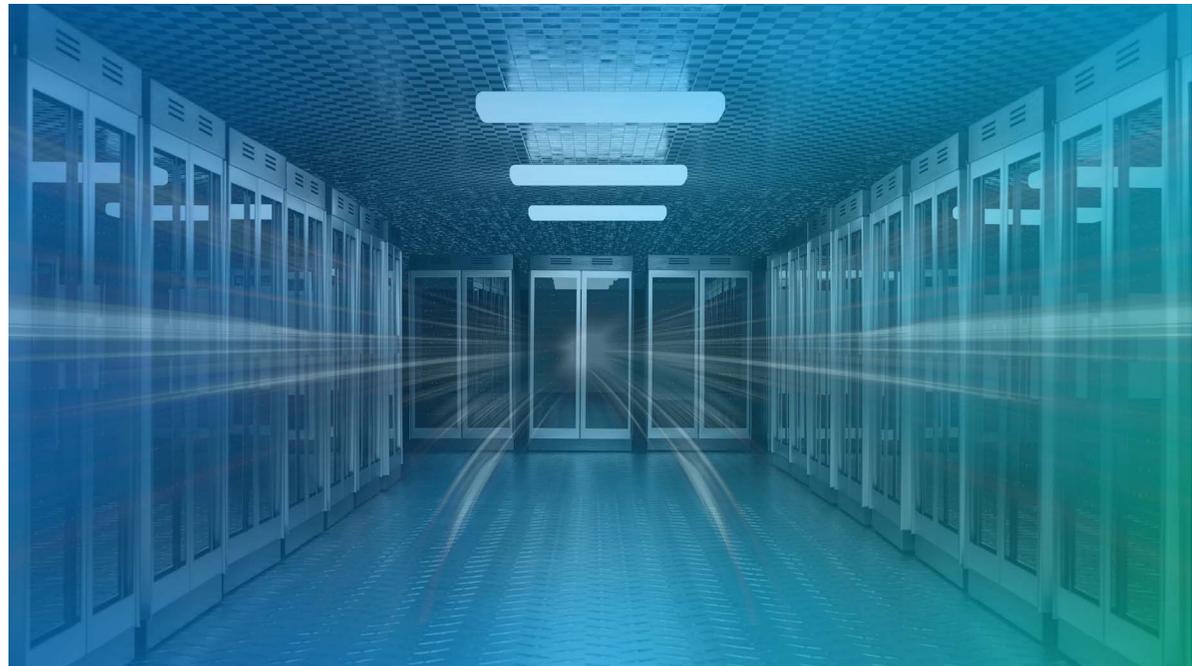
**Generative AI (GenAI) data centers need to build quickly to provide availability. Project schedules are getting compressed, from years to months or weeks, and deployed fiber counts have exploded from the thousands to hundreds of thousands.**

Generative AI (GenAI) data centers need to build quickly to provide availability. Project schedules are getting compressed, from years to months or weeks, and deployed fiber counts have exploded from the thousands to hundreds of thousands.

These two evolutionary trends have forced operators to look to their partners for solutions. Certainly, much goes into the planning, construction and delivery of these sites. As complex as these data centers are with their infrastructure power, cooling, and network architecture requirements, their need to scale and provide availability at a rapid pace requires constant deployment innovation. Time efficiencies onsite require a combination of infrastructure readiness, offsite staging and preparation of common components designed and packaged for modularity, coupled with localized product availability and processes to speed installation and deployment.

An important onsite challenge is managing the flow of the builds. While the scale of GenAI data centers can vary based on the site, the architectures are built upon logical, scalable building blocks to provide capacity across the

data center. That modular, repetitive approach—from data halls down to cabinet rows and purpose-built cabinets—provides the foundation for simplifying cabling structure and planning.



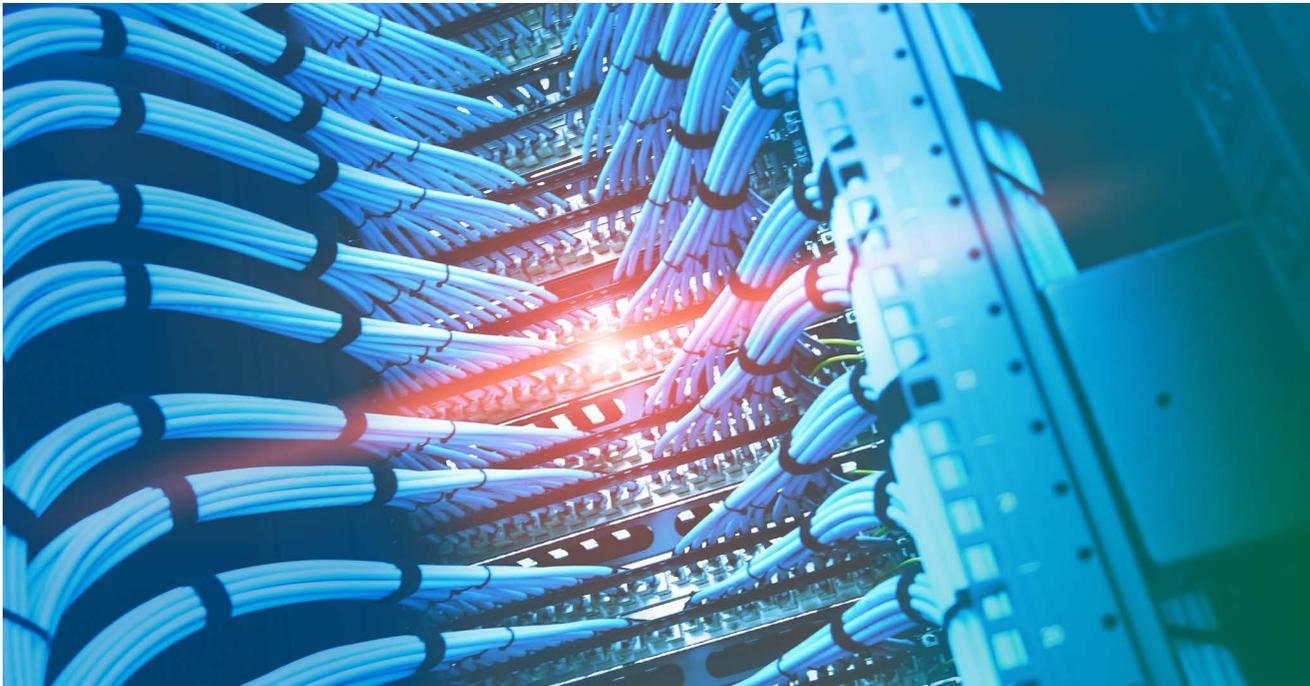
### Point-to-point: advantages and disadvantages

Consider that compute sections are grouped by cabinet row with common layouts across each multi-GPU compute cabinet. Within each cabinet is a predictable number of cable assemblies between devices in the cabinet; also, those exiting the cabinet to connect to a leaf switch in the management cabinets. As these cable assemblies are consistent throughout the compute cabinet architecture, they can be produced for quality

assurance, tested and verified offsite. Patch cords between the top-of-rack panel and node connections can be installed in cabinets offsite for roll-in and connect network scaling. Trunk cables can be strategically placed overhead in raceways above the cabinet row for quick connections between switch and server cabinets, saving valuable onsite time and resources. The same thought process can be implemented for other cabinet types.

That noted, some instances within a cabinet can utilize point-to-point cabling, such as direct attach cable (DAC), active electrical cable (AEC) and active optical cable (AOC) assemblies which can have initial cost and/or power benefits. As network speeds increase, however, those point-to-point cables will need to be replaced, discarded and upgraded. With copper DAC or AEC assemblies, the rigidity of their coaxial construction may make them challenging to remove, recycle or responsibly discard—only to replace them with more rigid higher-speed assemblies.

Point-to-point solutions are even more challenging when the cables need to exit the cabinet for routing elsewhere in the cabinet row. DAC assemblies have short distance limits, and additional shielding for extending this limited reach reduces their flexibility. AOCs are similarly challenging between cabinets, as the weight of high-speed transceivers on the cable connection increases the possibility of damage or breakage of the fiber cable. For these reasons, structured cabling is recommended instead, as it is application- and speed-agnostic and supports multiple network generations.



### Cable construction options

Cable construction is an important consideration. Not only is the routing plan critical, but pathway space for initial installation, along with flexibility for Day 2 additions for migration, must be considered. Individual bundles of 8- or 16-fiber cables in high quantities can quickly fill pathways, whether inside or overhead between cabinets. Should you prefer the granularity of individual cords, consider using 16-fiber assemblies, since they occupy less volume than comparable fiber counts with 8 fiber units. A more efficient cable footprint would come from higher fiber count assemblies with smaller bundle size with 8- or 16-fiber breakouts at appropriate locations, using MPO connectors to connect at the transceiver.

New smaller very small form factor (VSFF) multipair connectors are in early market deployments, which can initially bring reduced-size benefits for trunk and pathway applications. MMC16 or MMC24 and SN-MT16 connectors deliver the same fiber count as their traditional MPO16 or MPO24 counterparts in roughly one-third the space. Their smaller construction significantly reduces the area required in the vertical or horizontal pathway, making it easier to pull assemblies between locations in the data center.

The market is developing for these connectors, as well as for emerging ganged connectors which also leverage VSFF constructions to provide high fiber counts in a much smaller connector footprint. The expectation is that they will reduce the time needed to connect the fibers and reduce the risk of connectors being installed into the wrong port.

### Connections in evolving data center environments

Depending on the architecture, cable connections can occur on the front or rear of a cabinet (or both), based on their purpose. They may need to be routed around power devices or cooling hoses. When possible, plan to use larger or taller cabinets to maximize space for the systems installed.

Recent deployments scaling up and out of AI data center PODs have brought to light the challenges of collocated systems in cabinets, mainly GPU compute cabinets with limited vertical space on the front, along with CPUs for management, hoses for direct to chip (DTC) liquid cooling and power. These cabinets are generally 4 feet deep or deeper and, as such, are problematic to cable using traditional structured cabling methods.

Obstructions within the cabinets are only a part of the challenge. Human arms are generally shorter than 4 feet long, so accessing the panel connections from the rear is not practical. Solutions that are front-facing-only provide possible options using high-density VSFF connections providing the depth and height are factored into the design. In general, panel space is limited to 1 – 2 RU in a GPU cabinet due to the presence of other network devices necessary for functionality.

Bear in mind that direct connecting GPU cabinet transceiver ports to leaf switch ports can be accomplished as front-facing assemblies by using cable bundles. Managing the identification, routing, and protection of slack and spare cables, along with the extended breakout lengths over and between cabinets, is challenging for Day 1 installers. Day 2 becomes harder to access or change as the cabinets get stitched together with that point-to-point design. Running individual patch cables has challenges as well; cable bulk from the volume of several hundred 8- or 16-fiber cords between cabinets is difficult for installation as well as troubleshooting.

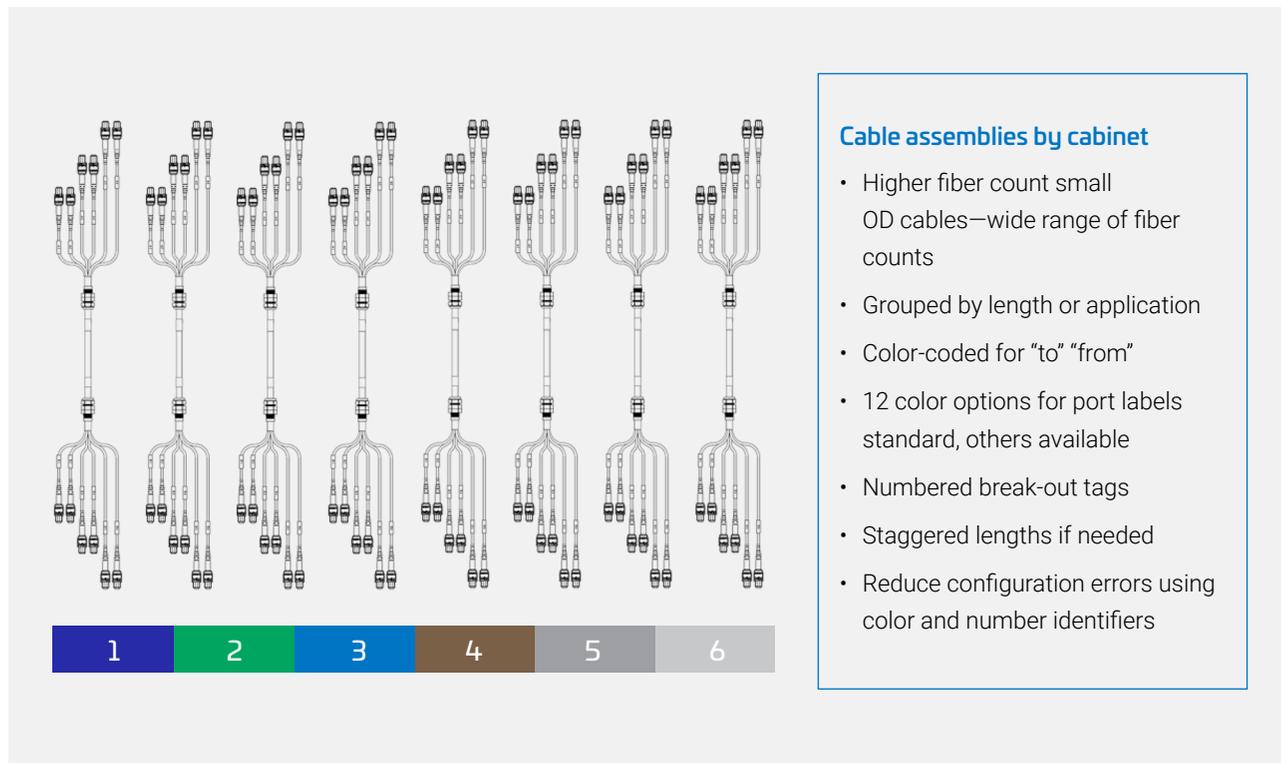
Accommodating and routing cable through available pathways relies on a well-planned approach. Pre-measuring known cable paths—building in a manageable slack factor—is important for installation and change management. Once length is verified, one must determine timing and volumes required. It’s important to plan this as far in advance of deployment as possible, as custom assemblies purchased in volume may require additional lead time.

In addition, as GenAI deployments require groups of MPO assemblies at each GPU to deliver higher fiber count connections back to the management cabinets, vertical and horizontal pathways can quickly become congested. Planning for that cable mass is critical to Day 2 operations when cables may need to be added for next-generation applications. In a short period of time, we have seen the migration from 400G using eight fibers to 800G using 16 fibers (2x8f); soon, we’ll see 1.6 Tbps becoming mainstream for the larger AI sites, possibly doubling the fiber count again. For backward compatibility, using 100G lanes would double the fiber count. With 200G lanes on the horizon, there will be options to maintain the current fiber counts with 2x speed. Either way,

planning cable assembly configurations can simplify installation and enable flexibility for operations teams.

For example, within spine-leaf switch and management cabinets, high fiber count array assemblies color-coded with numbered labels can simplify and speed up onsite installation.

Since these assemblies are replicated through each cabinet row, they can be pre-measured for pathways and placement. These can be made in various port and fiber counts and break-out lengths to provide efficiencies upon installation. Assemblies with staggered lengths can simplify proper placement.



To save time onsite, these assemblies—along with the required network equipment—can be pre-installed in a cabinet offsite by a value-add partner. Offsite or off-data-center-floor preparation can reduce effort expended by simply dropping trunk cables from overhead into the top of the rolled-in cabinets and connecting to the modules at the top-of-rack fiber panel.

If fit-out were done onsite, it would require unpackaging, cleaning and inspecting the assemblies, determining which assemblies are positioned where within the cabinet, neatly routing and managing the cables in the pathways (ensuring to and from connections are complete for dozens of cables in each cabinet), and then connecting trunk cables to the panel at the top of the cabinet.

Advance steps can be taken to improve efficiency when materials get to the site. From a fiber connectivity perspective, overhead pathways should be installed before the cabinets are brought into the data center. Factory-terminated and tested trunk cables should be provisioned above the planned cabinet locations. Labeled trunk cables provide fiber routing between endpoints in the

cabinet row or across the data center. All fiber patching can be completed within this cabinet to the front of a fiber panel at the top of the cabinet. Once arriving onsite, the cabinet is positioned, trunk cables are dropped and routed into the cabinet and are then connected to the installed equipment cords via the fiber panel.

Aside from simplifying and speeding up the installation process, the building block approach provides the foundation for future rapid migrations via disconnecting the cabinet trunks, rolling the cabinet out for next steps and then rolling in and reconnecting trunks to the next-generation cabinet.

### Steps to simplify deployment

Whether the cabinets are configured and cabled offsite or onsite, cable assemblies can be built, organized and packaged to simplify the process and reduce time required onsite. For example, all of the cable assemblies for a single cabinet could be factory built, tested, color-coded and number labeled, packaged, and marked with “to” and “from” labels for each of the requirements in that cabinet. QR-coded assemblies can provide all factory test data and documentation for each assembly.



Another option would be to package up all similar purpose assemblies for the cabinet row, with similar labeling. Examples would be node-to-fiber panel at top of the cabinet, or spine-to-leaf switch connections in the management cabinet. Installers onsite could plug in the assemblies as they work down the row. The idea is to set the workflow and follow the simplified sequences to be more efficient onsite.

### Taking a long-term view

Consider the need to have redundancy and error-free transmissions to reduce the chance that the network fails. In recent years, “mesh” or “shuffle” network cabling provided a way to keep the network operating when a port or switch failed on a duplex fiber level. This is more critical for multi-pair applications.

For example, 400G applications would split into 4x100G ports along with other similar ports in the network. Should a 400G device fail, a small, connected portion of the network would continue operating at a slightly reduced speed (300G), but the network would not go down. Service could be scheduled but would not require an immediate truck roll. GenAI cluster uptime requirements are better supported by incorporating a shuffle element in some critical areas. As most of these cabinet row connections use parallel pairs of fibers for speeds of 200G, 400G and above, redundancy can be included using common configurations. The higher the pair count in connections, the greater the benefit of shuffle design.



### Making the best choice

There are benefits and tradeoffs for each cabling option, with dependencies based upon data rate, distance, flexibility, power and cost. By design, GenAI networks repeat configurations across cabinet rows and switch layers to build and scale capacity. Those repeated configurations enable the use of pre-configured cabinets built offsite to be rolled in and connected to trunk cables and other infrastructure systems to reduce time to network availability.

Regardless of the data center type, the cabling system(s) used should support the initial mission of the data center with a vision for future applications and migrations.

*By Kenneth Hall, Solutions Architect,  
Data Center Solutions*

# Chapter 6

Supporting Tomorrow's Data Rates

Depends on Overcoming Fiber-Optic Dispersion

The demand for bandwidth continues to increase unabated, driven most recently by artificial intelligence (AI) and machine learning (ML) applications. AI clusters are the first to use high-speed transceivers; they have more high-speed inter-rack cabling and consume more fiber cable than traditional data center networks<sup>1</sup>. New 800G and 1.6T transceivers with 200G lanes will be standardized in the IEEE 802.3dj task force, with publication expected in 2026.

The 802.3dj standard will include parallel singlemode fiber applications with intensity modulation and direct detection transceivers (IMDD) like 800G-DR4, 800G-DR4-2, 1.6T-DR8 and 1.6T-DR8-2. Duplex singlemode fiber applications that use wavelength division multiplexing (WDM) like 800G-FR4-500, 800G-FR4 and 800G-LR4 will also be included.

These future data rates will become available to your data center soon, and fiber parameters must be re-evaluated for these higher baud rates. Fiber impairments must be considered with these high-speed transceivers, and chromatic dispersion is the largest concern. This article will discuss chromatic dispersion and what steps the standards bodies and CommScope have taken for these fibers to support these new data rates.

### What is fiber-optic dispersion, and why is it important in fiber-optic communication systems?

Dispersion is anything that causes some parts of an optical signal to travel faster or slower than other parts. This leads to signal distortion and limits the bandwidth and reach of fiber-optic links.

There are different types of dispersion, including modal, chromatic and polarization mode dispersion. In multimode fibers, each mode travels at a different speed; we call this "modal dispersion." With polarization mode dispersion, the two polarizations of an optical signal travel at slightly different speeds. Singlemode fibers are primarily concerned with chromatic dispersion where each wavelength in a signal sees slightly different glass properties and consequently travels at a different speed. Each of these types of dispersion will close an eye diagram and increase the bit error ratio (BER).

Fiber dispersion contributed to the adoption of PAM4 modulation. Each of these types of dispersion has a larger penalty at higher speeds or baud rates. With higher modulation formats like PAM4, you can double the bit rate compared to non-return-to-zero (NRZ) modulation while maintaining the same baud rate or symbol rate. Data throughput increases without increasing the dispersion penalty. This is particularly important with chromatic dispersion, since that penalty quadruples with each doubling of the baud rate.

### How does fiber-optic dispersion influence the design and deployment of long-distance communication networks?

In the past, chromatic dispersion was a major area of concern in long-haul and undersea networks. These networks typically would follow a span of singlemode fiber with a span of special fiber

with negative dispersion to cancel out chromatic dispersion. About a decade ago, coherent transceivers were introduced with powerful digital signal processors (DSPs) that can mathematically remove the impact of dispersion.

Unlike these long-haul networks, data center links are much shorter and operate near 1310 nm, where chromatic dispersion is negligible and hasn't been a major concern. However, as we get to 200G lanes, the impact of chromatic dispersion can't be ignored.

Two parameters are used to determine the chromatic dispersion at a given wavelength: the zero-dispersion wavelength and the slope at zero-dispersion. The three-term Sellmeier equation uses these two parameters to calculate the dispersion at an operating wavelength.

The ITU-T in standards G.652<sup>2</sup> and G.657<sup>3</sup> sets the range of dispersion values allowed in G.652.D, G.657.A1 and G.657.A2 fibers. The permitted zero-dispersion wavelength is from 1300 – 1324 nm and the slope at zero dispersion is in the range 0.073 – 0.092 ps/(nm<sup>2</sup>\*km).

In the past, Ethernet standards used both extreme zero-dispersion wavelengths (1300 or 1324 nm)

and maximum slope (0.092 ps/nm<sup>2</sup>\*km) to determine the range of dispersion values for testing wavelength channels. The shortest operating wavelength would assume 1324 nm zero dispersion wavelength, and the longest would assume 1300 nm.

The dispersion at a given wavelength will vary from fiber to fiber. Figure 1 graphs the dispersion

in ps/(nm\*km) for two different extreme fibers, one with a zero-dispersion wavelength of 1300 nm and slope of 0.073 ps/(nm<sup>2</sup>\*km), and another with zero-dispersion wavelength of 1324 nm and slope of 0.092 ps/(nm<sup>2</sup>\*km). Of these two extreme fibers, Fiber 1 has nearly double the dispersion at 1264.5 nm and half the dispersion at 1337.5 nm compared to Fiber 2.

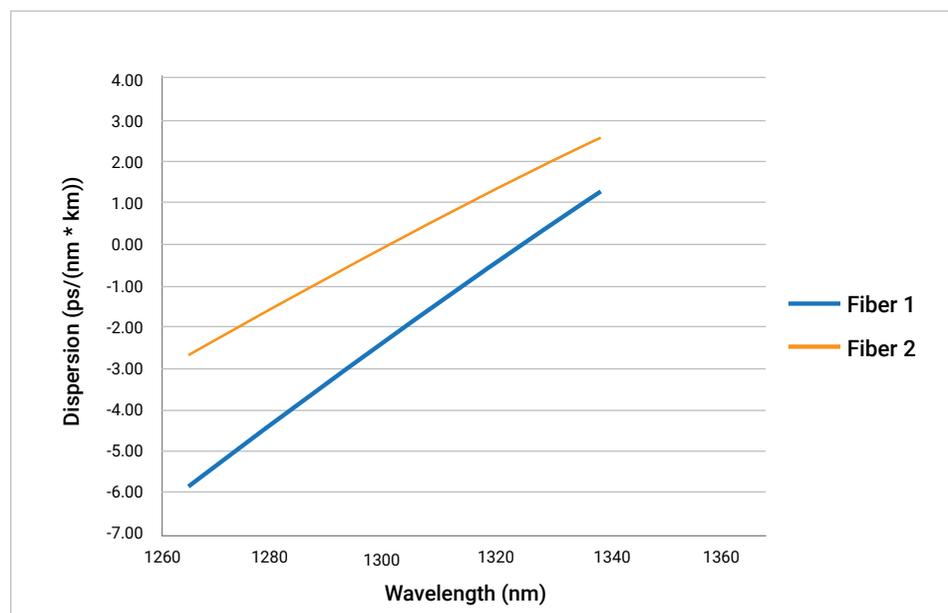


Figure 1: Dispersion vs. wavelength for two extreme fibers. Fiber 1 has the longest zero-dispersion wavelength and maximum slope; Fiber 2 has the shortest zero-dispersion wavelength and minimum slope.

This approach to dispersion effectively requires each transmitter to tolerate the most extreme fibers allowed by ITU-T standards at maximum reach. For 800G-FR4, this would lead to a dispersion range of -11.7 to +6.6 ps/nm over 2 km. For 800G-LR4, the range would be -28.1 ps/nm to +9.3 ps/nm over 10 km. A statistical approach to chromatic dispersion has the potential to require transceivers to comply with less dispersion while representing realistic fiber links.

### What steps have been taken to address chromatic dispersion for 200G lanes?

The IEEE 802.3dj task force has taken a statistical approach to fiber chromatic dispersion for 200G lane applications like 800G-FR4 and 800G-LR4. In the past, each transceiver was tested with the worst-case fiber. The ITU-T conducted a study with data from eight fiber manufacturers to understand the statistical distributions of fiber dispersion for G.652 and G.657 fibers. The results were published in the 2024 version of G.652. This is useful data that IEEE can reference.

CommScope also compiled a data set using millions of fibers in CommScope cables over the past 10 years. This CommScope data set—which

I compiled and graphed myself in Figure 2—showed good agreement with the ITU-T data set, confirming that dispersion has remained consistent over time.

Knowing the likelihood of encountering a fiber with both an extreme zero-dispersion wavelength and maximum slope is key to developing a statistical model for link dispersion. Figure 2 plots a bivariate histogram of the data set for zero-dispersion wavelength and slope. Fibers with the most

dispersion would be located in the top two corners of the graph (1300, 0.092) and (1324, 0.092). Nearly all the fibers are within a small distribution with moderate wavelength and slope, and very few fibers have both an extreme zero-dispersion wavelength and slope. This indicates that the approach taken with previous generations requires transmitters to be designed and tested to work with fibers displaying dispersion characteristics that are rarely seen in practice.

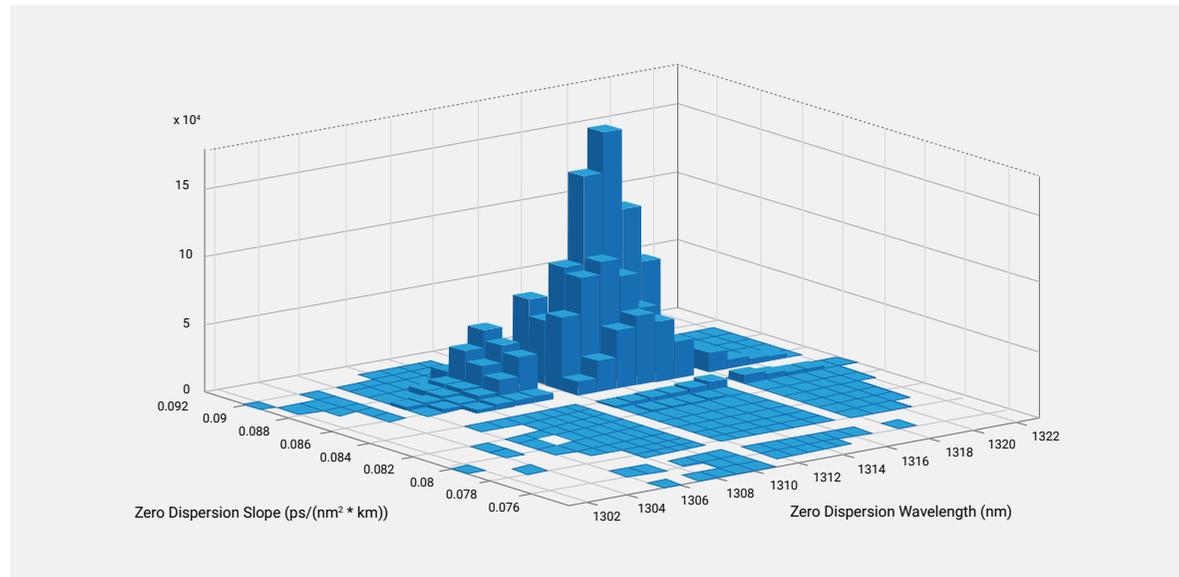


Figure 2: Bivariate histogram of fibers in the CommScope data set for zero-dispersion wavelength and slope.

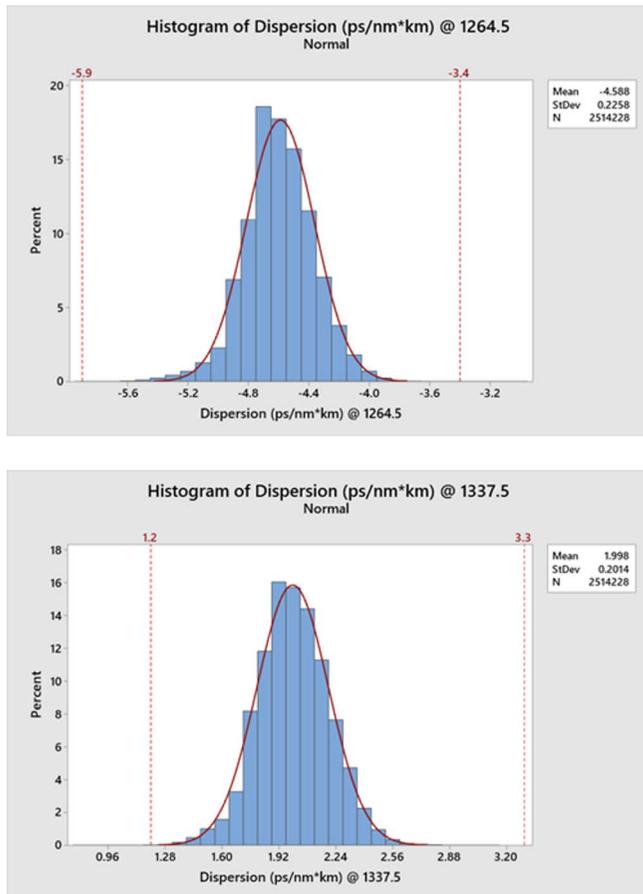


Figure 3: Histogram of per km dispersion for 800G-FR4 wavelengths at (a) 1264.5 nm and (b) 1337.5 nm. Worst-case dispersion is marked by vertical red lines.

Figure 3 includes histograms for per-km fiber dispersion in ps/(nm\*km) for the two extreme 800G-FR4 wavelengths: (a) 1264.5 nm and (b) 1337.5 nm. The vertical red lines indicate the most extreme fiber dispersion allowed by ITU-T standards. The vast majority of the fibers in the data set have much less positive dispersion at 1337.5 nm and less negative dispersion at 1264.5 nm. Since very few fibers in the data set have both an extreme zero-dispersion wavelength and maximum slope, it is very rare to see a fiber that would have worst-case dispersion at these wavelengths.

### What's in store for the future?

The results from the CommScope data set and the ITU-T study mean that a new singlemode fiber standard is not needed for IMDD WDM transceivers with 200G lanes. The next step will be to determine what is needed for 400G lanes and 1.6T transceivers like 1.6T-FR4 and 1.6T-LR4. The New Ethernet Applications Ad Hoc Ethernet for AI Assessment in IEEE 802.3 hosts a consensus-building activity where participants

discuss whether tighter chromatic dispersion specifications are needed for 400G signaling.

### Why does CommScope pay attention to transceiver standards and specifications?

We understand that our customers use the passive components they buy from us with active equipment to build their networks. It's very important that we understand these standards and specifications to enable our [Fiber Performance Calculator](#) tool and our [SYSTMIX® Application Assurance™ warranty](#). We want to be a trusted partner for our customers as they design and build their networks.

*By Earl Parsons, Director,  
Data Center Architecture Evolution, CommScope*

<sup>1</sup> E. Parsons, "Cabling considerations of AI data centers," CommScope white paper, 2023.

<sup>2</sup> G.652, "Characteristics of a single-mode optical fibre and cable", ITU-T, 2024.

<sup>3</sup> G.657 "Characteristics of a bending-loss insensitive single-mode optical fibre and cable", ITU-T, 2016.

### **Dr. Earl Parsons, Director of Data Center Architecture Evolution, CommScope**

Dr. Earl Parsons joined CommScope in 2014 as a principal optical engineer. Prior to joining CommScope, Earl received an MS and Ph.D. in optical sciences from the University of Arizona and was a senior member of technical staff at TE SubCom, formerly known as Tyco Telecommunications.

Dr. Parsons also served as an editor of the IEEE 802.3db-2022 standard. His interests include multimode and singlemode fiber-optic systems to enable artificial intelligence data centers.

### **Alastair Waite, Senior Manager Market Development Data Center, CommScope**

Alastair Waite is a seasoned professional at CommScope, having joined the company in September 2003 as a product manager for the Enterprise Fiber Optic division. Over the years, he has held several key positions, including head of Enterprise Product Management for EMEA, head of Market Management, and Data Center Business leader in EMEA.

In his current role, he has a focus on developing innovative infrastructure solutions and novel market approaches to meet the evolving demands

of AI in data centers. Alastair is also an active contributor to industry discussions, sharing insights on AI and data center infrastructure. Notably, he has authored articles such as “Bringing more order to AI data center infrastructure orders,” which provides guidance on navigating the complexities of structured cabling solutions in AI environments.

### **Hans-Jürgen Niethammer, Market Development, Strategic Cloud Business EMEA/APAC, CommScope**

Hans-Jürgen joined CommScope’s cabling division in July 1994 and has held several key positions in product management, technical services and marketing, including director of Program Management in EMEA, director of Marketing in EMEA, and director of Technical Services and sales operations in EMEA.

Since January 2013, he has been responsible for CommScope’s data center market development in EMEA, ensuring that CommScope’s solutions enable customers’ data center infrastructures to be more agile, flexible and scalable in order to meet the requirements of their dynamic market segment today and in the future.

Hans-Jürgen is an international expert in data centers, fiber optics and AIM systems. He is a member of several ISO/IEC and CENELEC standardization committees and he is editor of several international standards. He also holds a chartered engineer degree in electronic engineering and is a state-certified business economist.

### **Ken Hall, Data Center Architect NAR, CommScope**

Ken is responsible for technology and thought leadership for global scale, AI and cloud data centers. In this role he has been instrumental in the development and release of high-speed, ultra-low loss fiber-optic solutions to efficiently enable network migration for data center operators. Focused on cabling efficiencies for AI design and implementation, he continues to drive solutions for these rapidly evolving networks.

Ken has nine patents to date for fiber-optic connectors and infrastructure management systems. He is a registered Communication Distribution Designer (RCDD) and Network Technology Systems Designer (NTS).

CommScope pushes the boundaries of communications technology with game-changing ideas and ground-breaking discoveries that spark profound human achievement. We collaborate with our customers and partners to design, create and build the world's most advanced networks. It is our passion and commitment to identify the next opportunity and realize a better tomorrow. Discover more at [commscope.com](https://commscope.com).



[commscope.com](https://commscope.com)

Visit our website or contact your local CommScope representative for more information.

©2026 CommScope Technologies LLC, an Amphenol company. All rights reserved. CommScope and the CommScope logo are registered trademarks of CommScope and/or its affiliates in the U.S. and other countries. For additional trademark information see <https://www.commscope.com/trademarks>. All product names, trademarks and registered trademarks are property of their respective owners

EB-117988.3-EN (03/26)