Cabling Data Centers

Design Guide

DU-10438-001 | March 2023

Document History

DU-10438-001

Version	Date	Authors	Description of Change
001	2021-06-28	Brian Forbes, Craig Tierney, Dennis O'Brien, Jeremey Rodriguez, Robert Sohigian, Steven Hambruch, and NVIDIA Professional Services (NVPS)	Initial release
002	2021-08-27	Michael Balint and Robert Sohigian	Minor changes
003	2022-12-16	Brian Forbes and Robert Sohigian	NDR cables
004	2023-03-31	Robert Sohigian	New NVIDIA template

Contents

Chapter 1. Introduction

We are living in the age of data. Between millions of tweets and shares of photos, music, and video, the amount of global digital information created has grown sixfold in the past 5 years to nearly 6 zettabytes in 2014. And data consumption continues to grow every minute thanks to smartphones, social media, video streaming, and big data.

With this exponential growth of data and the increase of applications that can take advantage of real-time massive data processing the market demands faster and more efficient interconnect solutions. NVIDIA InfiniBand and Ethernet solutions provide an ideal solution for environments requiring high throughput and low latency.

This guide provides an overview of the different types of Ethernet and InfiniBand cables that exist today, suggests how to survey and understand connectivity requirements between systems in a data center to speed cable installation while maximizing maintainability, and discusses the methods and procedures used by NVIDIA teams around the world to successfully stand-up massive data centers.

1.1 Target Audience

The target audience spans a wide range of technologists including IT managers, system architects, network architects, network administrators, and technicians who must design, deploy, manage, or maintain data center networks.

1.2 Document Conventions

> **Note:** Identifies important information that contains helpful suggestions.

> ! **Caution:** Alerts to a risk of personal injury, system damage, or loss of data.

Warning: Warns that failure to take or avoid a specific action might result in personal injury or a malfunction of the hardware or software. Hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents before you work on any equipment.

1.3 Warnings

Warning: These warnings apply to InfiniBand connectors.

> Do not step on the cable or connectors.
> Do not drop the cable or connectors from any height.
> Do not drag the cable or connectors over any surface.

Warning: These warnings apply to cables.

> Do not use cable/zip ties or any other hard fastener.
> Do not staple the cables.
> Do not insert connectors into connector receptacles by pushing on the cable.
> Do not pull cables.
> Do not bend cables beyond the recommended minimum radius.
> Do not attempt to squeeze/force cables in a space that is too small.
> Do not route cables through pipes, holes, or unserviceable apertures.
> Do not drape cables for a length more than two feet (60 cm).
> Do not compromise on routing.

Common hazards to avoid when dealing with cables are shown in Figure 1.

Figure 1. Problems to avoid

Do not kink the cable | Do not over-bend the cable behind the connector | Do not twist the connector

Chapter 2. InfiniBand Cables Primer

The InfiniBand Trade Association formed in 1999 to solve two fundamental data center issues:

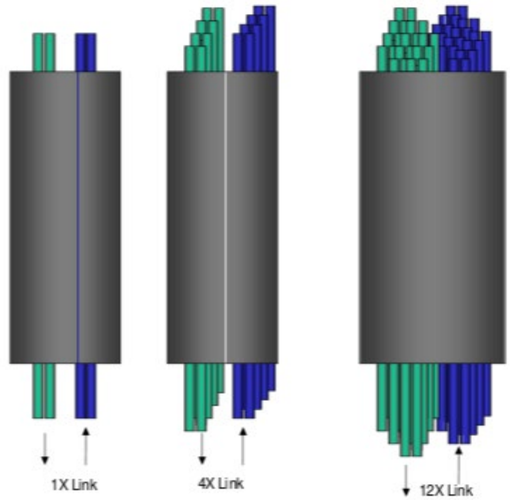
- The gap between CPU computing power and network speeds continued to widen—"no Moore's Law for I/O."
- Proliferation of network types—Ethernet, Fiber Channel, and proprietary High-performance Computing (HPC)—consumed many I/O slots per server and significant rack space for switching.

Pursuit of 'one network for messaging, storage, and HPC' that could fulfill both small and very large data center requirements led to these InfiniBand design points:

- Native support for server-to-server memory access—remote direct memory access (RDMA).
- Lossless switched network—no packet drops.
- Cut-through switching—a packet may begin transmission on an outbound switch port while it is still being received on an inbound port.
- Operating system (OS) bypass and zero memory copies within the server.
- Complete Transport Layer offload by the I/O adapter.
- Scalable Link Layer (cable) performance with backwards compatibility.
- A strategic vision for network performance growth.

Cables are a key element in InfiniBand performance, scalability, and future-proofing. To achieve scalable performance, InfiniBand implements a multilane cable architecture, striping a serial data stream across N parallel physical links running at the same signaling rate. Figure 2 shows three link widths of 1, 4, and 12 parallel lanes, referred to as *1X*, *4X*, and *12X*.

Figure 2. 1X, 4X, and 12X link widths



In copper cables, each lane uses four conductors—two differential pairs, one pair for transmitting and one for receiving. Optical cables use two fibers per lane, one for transmitting and the other for receiving.

Since the inception of InfiniBand, performance has increased by a factor of 25. From the original specification single data rate (SDR) to the current version (HDR), there has been a constant increase in performance over the years. Table 1 summarizes the past, present, and planned future of InfiniBand generations.

Table 1. InfiniBand generations

Name	Signaling Rate per Lane (Gbps)	Effective Bandwidth for 4X link (Gbps)[1]	Connector	Year
SDR	2.5	8	CX4	2003
DDR	5	16	CX4	2006
QDR	10	32	QSFP	2008
FDR10	10	39	QSFP	2013
FDR	14	54	QSFP	2013
EDR	25	97	QSFP28	2015
HDR	50	200	QSFP56	2019
NDR	100	400	OSFP	2021
XDR	250	800	TBD	2023[2]

1. Accounting for bit encoding overhead
2. XDR is planned after 2023.

Until the advent of 200 Gbps InfiniBand (HDR), 4X cables were the dominant width. HDR introduced a variant called HDR100 that uses two lanes at 50 Gbps to deliver a total of 100 Gbps. There are plans for a similar two link variant of NDR, NDR200. These two-lane formats effectively double switch density, reducing overall costs.

2.1 Mixing Widths and Rates

InfiniBand cable width (lane count) can change from hop to hop, permitting lower-cost implementations to coexist with higher-performance ones. Similarly, InfiniBand switches perform speed matching between ports running at different lane rates. Per-hop rate and width-matching are automatically negotiated between switch pairs, enabling interoperability between InfiniBand generations. Auto-negotiation can also maintain fabric continuity by connecting at a lower rate or width if cable signal integrity is degraded. For example, a damaged HDR cable may still be capable of 100 Gbps (EDR), or of running one lane at 2.5 Gbps (1X SDR).Bit Error Rate (BER)

In a lossless network running at high data rates, InfiniBand cables must have excellent signal integrity and a low BER. The InfiniBand specification requires less than one-bit error per 10^12 bits transmitted. NVIDIA cables for both InfiniBand and Ethernet are individually factory tested to less than one-bit error per 10^15—1,000 times better.

2.2 Cable Latency

With InfiniBand switch latencies on the order of 100 ns, speed of light delays in cables can have a significant effect on end-to-end latency. Although there are differences due to in-connector electronics and the physical medium (copper compared to glass), latency due to cable length is roughly 5 ns per meter, for both InfiniBand and Ethernet cables.

2.3 Connectors

InfiniBand uses Quad Small Form-factor Pluggable (QSFP) cable connectors for QDR, FDR, EDR, HDR, and HDR100 links. Examples of QSFP connectors are shown in Figure 3.

Figure 3. QFSP connectors

QSFP connectors are also used by Ethernet, in 100 GbE and 200 GbE for example. QSFP connectors have improved as speeds and thermal demands have increased but are compatible with previous generations. QSFP28 is the technical designation for lane speeds of 25 Gbps, and QSFP56 supports lane speeds of 50 Gbps.

2.4 Breakout Cables

A physical 4X HDR switch port can be split into two independent HDR100 ports. This is analogous to Ethernet port splitting, for example, converting a 4-lane 100 GbE port into four 25 GbE connections. HDR-to-dual-HDR100 breakout cables, or splitters, are available for this purpose, with QSFP connectors on all three legs (Figure 4).

Splitter cables are available for both optical and copper cables.

Figure 4. HDR to dual-HDR100 splitter

2.5 Port Power

All cable types except passive copper draw power from the QSFP port on the switch or host channel adapter (HCA), to boost the signal or drive optics such as lasers. The required port power increases with lane speed, the number of lanes, and with cable length. An HDR cable can require 6W per port.

2.6 Cable Technologies

InfiniBand cables use two media types: copper and optical fiber.

2.6.1 Copper

Copper InfiniBand cables have several advantages:

> Low cost.
> High reliability.

Their disadvantages:

> Short reach, especially at high signaling rates.
> Larger diameter than optical cables—in high-density cabling situations this can reduce airflow and complicate cable routing.

Due to their short reach, copper cables are mainly used within a rack.

2.6.1.1 Passive Copper

Passive copper cables are referred to as direct attach copper (DAC) cables. No port power is required. The maximum length for HDR is 2m. See Figure 5 for an example of a DAC cable.

Figure 5. Direct Attach Copper (DAC) cable

2.6.1.2 Active Copper

Active copper cable (ACC) products extend reach by introducing signal boosting electronics in the connector. They are more expensive than DACs, and thinner. For HDR, maximum length is 4m, and 3.5W are required from each InfiniBand port. An example of an ACC is shown in Figure 6.

Figure 6. Active Copper Cable (ACC)

2.6.2 Optical Fiber

Optical cables have several advantages:

- Longer reach.
- Smaller diameter than copper.

Disadvantages compared to passive copper:

- Higher cost.
- Require power from the InfiniBand ports.
- Less reliable due to electronics, and to additional physical interfaces in the case of separate transceivers.

InfiniBand optical link types:

- Active optical cables.
- Optical transceivers with a passive optical cable.

2.6.2.1 Active Optical Cables

An active optical cable (AOC) is a passive multilane fiber cable with QSFP optical transceivers integrated at both ends. InfiniBand was an early adopter of AOC cables due to these advantages over physically separate transceivers:

- The optical fibers can be perfectly aligned in the factory and their characteristics are well known, so mechanical and electronic costs can be optimized.
- Higher reliability due to the elimination of physical interfaces between transceivers and fiber cable.

Disadvantages:

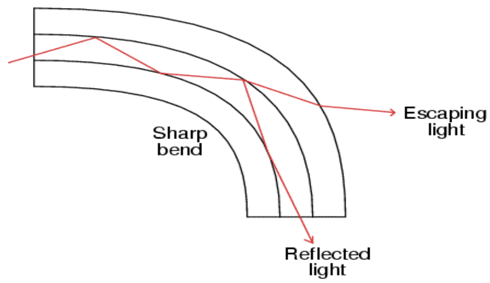
- Fixed lengths (although custom lengths are possible).
- Not compatible with optical patch panels.

A typical AOC is shown in Figure 7. The maximum AOC length for HDR is 150m.

Figure 7. Active Optical Cable (AOC)

An AOC fiber cable is thinner and more flexible than a copper cable. However, a tight bend can cause signal degradation if light passing through the fiber reflects and refracts abnormally (Figure 8).

Figure 8. Signal degradation due to fiber bend radius



The amount of signal degradation depends on the materials and construction of the cable itself. Some cables are capable of tighter bend radiuses than others, and manufacturer recommendations should always be observed (Table 2).

Table 2. Bend radius of typical AOC cable

Cable Type	Single Cable Diameter	General Minimum Bend Radius Guidance[1]
QSFP 4 Channel AOC	3 mm ± 0.1 mm	15 × diameter (45 mm) dynamic 10 × diameter (30 mm) static Assembly (cable plus connector): 56 mm (2.2 in)
1. Check manufacturer recommendations for specific guidance		

2.6.2.2 Optical Transceivers with Passive Fiber

Two HDR InfiniBand optical transceivers are available: the MMA1T00-HS for short range and the MMS1W50-HM for up to 2 kilometers. Both have a QSFP56 connector on the InfiniBand port side.

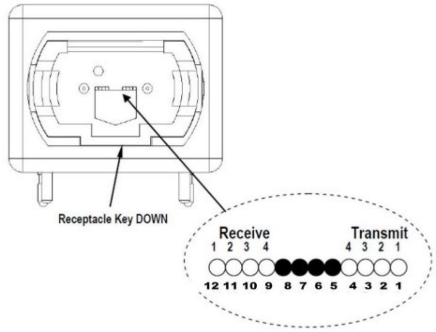
2.6.2.2.1 MMA1T00-HS Optical Transceiver

The MMA1T00-HS has an MPO-12 (Multifiber Push-On) receptacle for a passive 8-fiber multi-mode cable. Its reach is 100m, and it requires up to 5w from the InfiniBand port. An MMA1T0-HS module is shown in Figure 9.

Figure 9. MMA1T0-HS optical transceiver

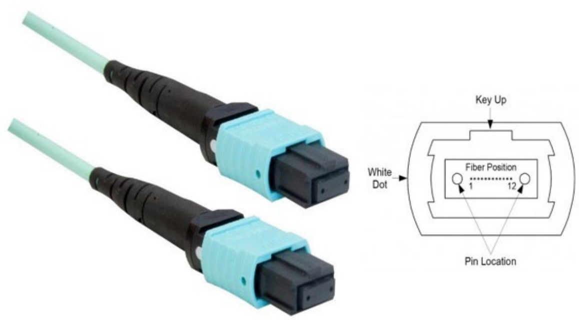
The MPO receptacle on an MMA1T0-HS transceiver is depicted in Figure 10.

Figure 10. MMA1T0-HS MPO transceiver receptacle



Point-to-point (P2P) connections between MMA1T0-HS modules use a Type B MPO jumper cable (also known as a reversed cable or crossover cable). The middle four fibers of the 12 possible fibers are unpopulated. An example of an MPO Type B jumper is shown in Figure 11.

Figure 11. MPO-12 Type B jumper



2.6.2.2.2 MMS1W50-HM optical transceiver

The MMS1W50-HM has a duplex LC connector for a passive dual-fiber single mode cable. Its reach is 2 kilometers, and it requires 5W from the InfiniBand port. Data streams from the four InfiniBand lanes are combined within a single fiber in each direction using coarse wavelength division multiplexing (CWDM). An MMS1W50-HM module is shown in Figure 12.

Figure 12. MMS1W50-HM optical transceiver

Chapter 3. NDR InfiniBand

The introduction of 400 Gbps (NDR) InfiniBand doubles network performance compared to HDR, and the increase from HDR's 40 switch ports to 64 ports greatly reduces the amount of equipment needed to implement a customer fabric. NDR also offers excellent backwards compatibility with existing HDR and EDR deployments.

This chapter assumes familiarity with InfiniBand concepts covered in Chapter 2.

3.1 NDR Overview

Compared to HDR, NDR introduces several new capabilities:

> Link speeds of 400 Gbps (four lanes of 100 Gbps) and 200 Gbps (two lanes of 100 Gbps). The latter is referred to as NDR200 and differs from HDR, which is four lanes of 50 Gbps.
> Octal Small Form-factor Pluggable (OSFP) connectors, which enable higher port densities than QSFP by supporting two InfiniBand links per physical connector.
> Twin cables and transceivers—pairs of cables or transceivers that share an OSFP connector.
> QSFP112 connectors, which are used on DPUs where space is at a premium.

Due to higher speed and increased power requirements, the mix of NDR cables and transceivers has changed from HDR:

> The reach of DAC cables, which are the least expensive and lowest power cables, is shorter than for HDR. A range of NDR ACC cables provides an intermediate step in distance and power draw between copper and optical connectivity.
> Active optical cable (AOC) technology, in which the transceivers and fiber are a single assembly, is not practical for NDR. Distances greater than a few meters are managed by NDR optical transceivers and passive multi-strand MPO fiber cables.
> A new family of MPO cables, called MPO/APC (Angled Physical Contact, or Angle-Polished Connector), increases link reliability and distance. These cables are not compatible with HDR or EDR transceivers.
> Splitting NDR optical links into two NDR200 links introduce the need for MPO fiber splitters.

> Backwards compatibility with HDR, HDR 100, and EDR involves a new family of cables having OSFP connectors at the switch end.
> The OSFP twin-port approach somewhat constrains the usage of adjacent ports. For example, cables from the InfiniBand port pair must be the same media type.

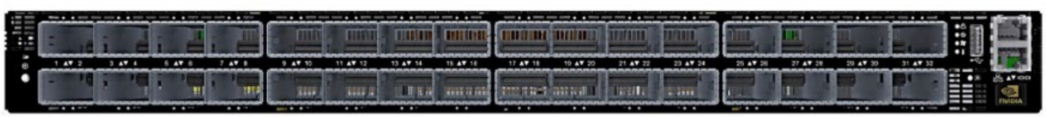
3.2 Components

NDR introduces new connector types for switches, cables, and adapters. It also introduces a new family of optical transceivers and MPO fiber cables.

3.2.1 NDR Switch Connectors (OSFP)

Figure 13 shows the connector side of a QM9700, which is an air-cooled 64-port NDR switch.

Figure 13. QM9700 NDR switch

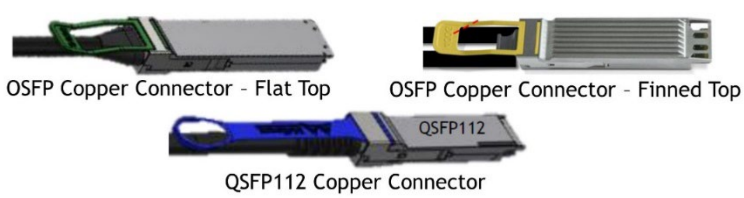


To fit 64 connections into one rack unit (1U, or 1RU), it uses 32 OSFP connectors and each physical connector provides dual NDR ports. The OSFP cages only accept ‘finned’ OSFP male connectors, to help cool such a dense array of ports.

3.2.2 NDR Cable Connectors

NDR introduces two new connector form factors to InfiniBand: two variants of OSFP and QSFP112 (Quad Small Form Factor Pluggable, rated for 112 Gbps per lane), as shown in Figure 14.

Figure 14. NDR cable connectors



The figure shows connectors on passive and active copper cables. Optical transceiver versions, which have MPO/APC connectors on the outboard end, are discussed in the next section.

The flat OSFP connector is single-port and is used for adapter connections. The finned OSFP version has ‘twin’ ports and is physically taller and is only used in switch ports.

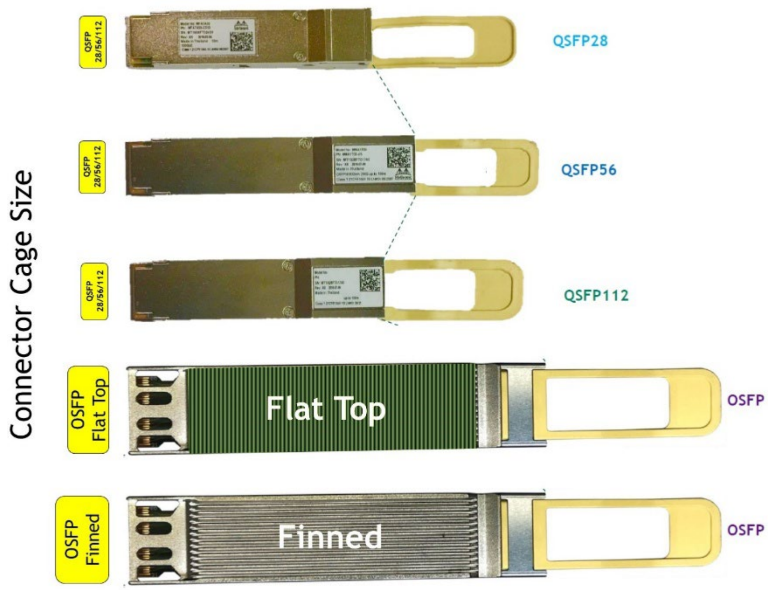
QSFP112 connectors are single-port and are used for DPU connections where card edge space is limited.

The QSFP112 connector technology is also used for 400 GbE.

3.2.3 NDR Optical Transceivers

Figure 15 compares the physical sizes and power requirements of InfiniBand transceivers. Note the height difference between the flat OSFP connector used for adapter ports and the taller finned OSFP connector used for NDR switch ports.

Figure 15. InfiniBand transceiver sizes



OSFP transceivers can support two standard MPO connections, with a pair of standard MPO connectors turned 90 degrees and mounted side by side, as shown in Figure 16. Flat OSFP transceivers, intended for NDR and NDR200 adapter ports, are single-port with the second port physically blocked.

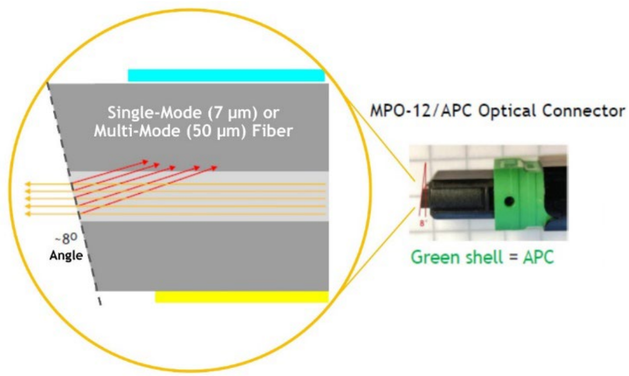
Figure 16. OSFP twin-port transceiver

3.2.4 NDR MPO/APC Fiber

The NDR transceiver family supports both single-mode and multi-mode passive fiber to provide a range of reach, cost, and power requirements. In most regards, NDR fiber cables are the same as those for HDR as described in Section 2.6. Unlike HDR, NDR fiber cables use MPO/APC (Angled Physical Contact or Angle Polish Connector) connectors, which provide superior signal integrity but do not interoperate with previous InfiniBand MPO cables.

Figure 17 illustrates the MPO/APC concept.

Figure 17. MPO APC concept



Fiber used in other connectors such as MPO/UPC have a polished end face that is perpendicular to the fiber axis. This end face acts as a mirror and reflects a portion of the laser beam back toward the transmitter (back scatter). This can reduce signal integrity and degrade the laser. MPO/APC uses fibers with a slightly angled face, so the reflected light is sent off-axis out of the fiber.

3.2.5 Minimum Bend Radius for NDR Cables

As with previous InfiniBand media, NDR cables can be damaged or have their signal integrity degraded by bending them too sharply:

- For DAC or ACC cables, the minimum bend radius at the connector itself is 70 mm (2.75 inches). Elsewhere in the cable the minimum radius is smaller: 34.5 mm (1.4 inches).
- For fiber, whether single-mode or multi-mode, the minimum bend radius is 30 mm (1.2 inches).

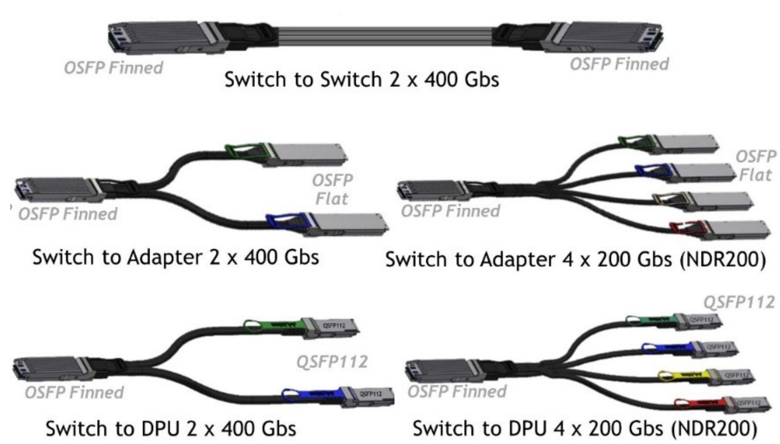
3.3 Connecting NDR Components Together

This section discusses the options for interconnecting NDR switches and endpoints.

3.3.1 NDR Copper Connectivity

The NDR copper cable family is shown in Figure 18. It includes both DAC and ACC. DAC requires no port power but has a limited reach; ACC draws port power to amplify the signal—approximately 1.5 Watts per connector-- and extend the reach. Copper cables are mainly used within a rack, or between adjacent racks.

Figure 18. Copper connectivity



Switch-to-switch cable assemblies have a finned OSFP connector at each end with a pair of copper cables between them. The two cables are logically independent InfiniBand links.

Cable assemblies to connect a switch to NDR adapters have a finned OSFP connector at the switch end and two copper cables, each ending in a flat OSFP connector. In the case of DPUs, where space is at a premium, the two copper cables end in QSFP112 connectors.

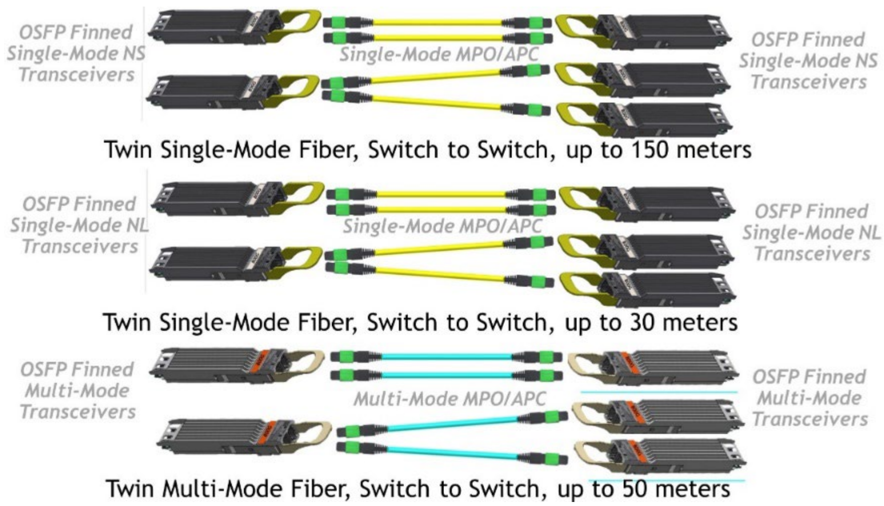
When the adapters or DPUs have ND200 ports (200 Gbps), each OSFP connector at the switch end splits into dual copper cables that use two data lanes instead of four. Four adapters or four DPUs can be connected. For DPUs, the cables end in QSFP112 connectors, otherwise they end in flat OSFP connectors.

Copper cables to connect NDR switches to HDR and EDR are discussed in Section 3.3.4.

3.3.2 NDR Switch-to-Switch Fiber Connectivity

Three options to interconnect switches using MPO/APC fiber are shown in Figure 19.

Figure 19. Switch-to-switch fiber



For single-mode fiber up to 30m, pairs of 'NL' OSFP transceivers can be used. Up to 100m can be achieved using 'NS' transceivers, which cost more and draw more power than 'NS' models. An 'NL' transceiver port can be connected to an 'NS' port but the maximum distance will be 30m.

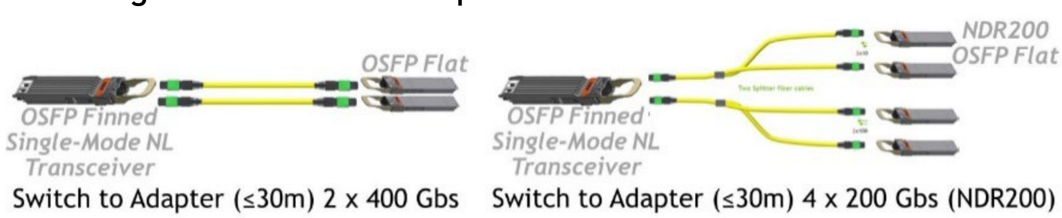
Multi-mode fiber and multi-mode OSFP transceivers can connect switches up to 50m apart. This is the least expensive fiber option and requires the least port power.

Fibers in a pair can have different lengths. Cables that are not MPO/APC, such as those used for HDR and EDR, will not work. Multi-mode and single-mode transceivers will not interoperate.

3.3.3 NDR Switch-to-Endpoint Fiber Connectivity

There are multiple fiber options to connect switches to adapters and DPUs. Available configurations differ between single-mode and multi-mode. Single-mode is available for adapter connections reaching up to 30m, but not DPUs, as shown in Figure 20.

Figure 20. Single-mode fiber to adapters

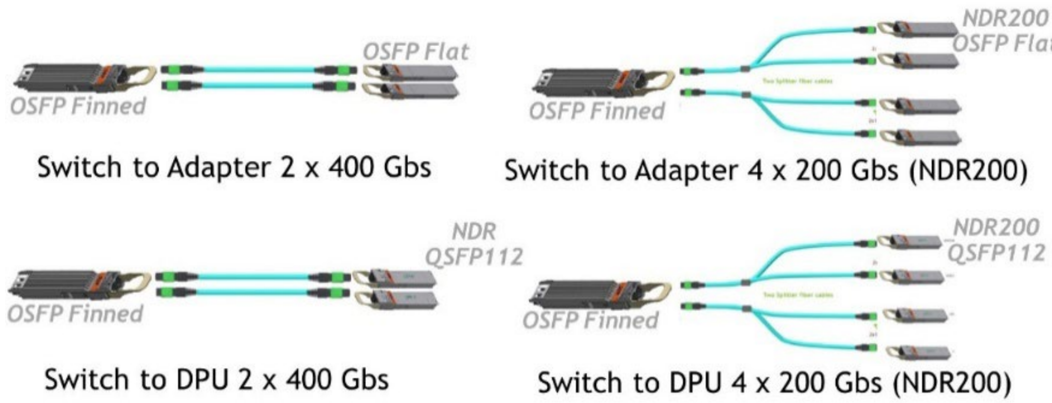


NDR adapters are connected using OSFP finned twin 'NL' transceivers on the switch end, up to 30m of single-mode MPO/APC fiber, and OSFP flat single-mode transceivers at the adapter ports.

For NDR200 adapters, twin two-way single-mode splitters are used with 200 Gbps OSFP flat single-mode transceivers at the adapter. Each of the two splitters can be a different length. As usual, the physical fiber split occurs near the shared MPO connector.

Multi-mode fiber offers up to 50m connections to adapters and to DPUs. See Figure 21.

Figure 21. Multi-mode fiber to adapters and DPUs



Multi-mode options for NDR and NDR200 adapters are like single-mode. Transceiver form factors at the switch and at the adapter ports are the same, using multi-mode versions. The maximum distance is 50m and less port power is needed.

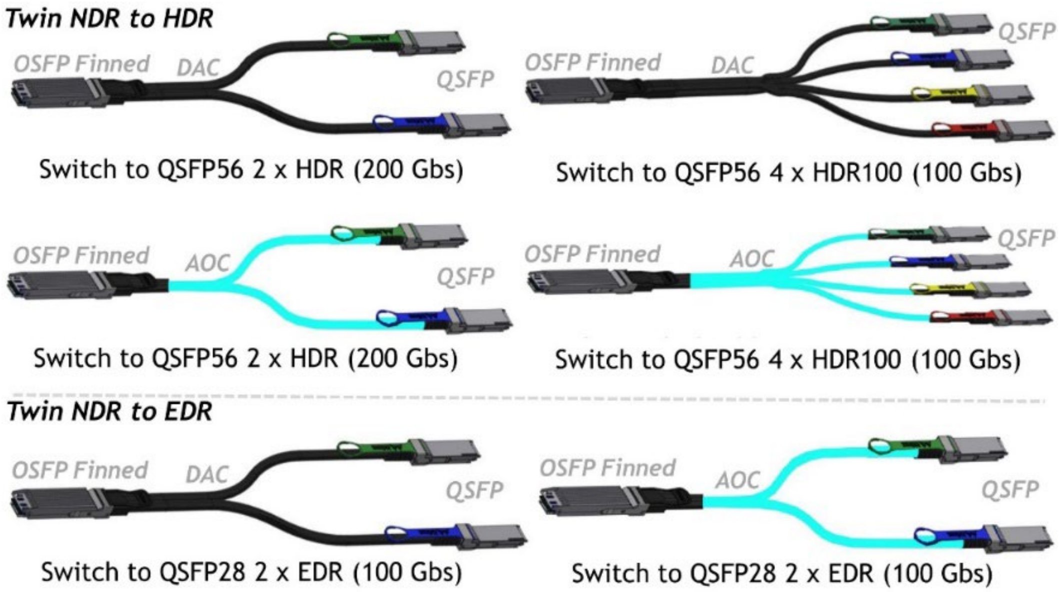
Multi-mode also offers DPU connectivity, where the DPU transceiver form factor is QSFP112. NDR and NDR200 dual-splitter connections can be up to 50m long.

Each of the two splitters can be a different length. The multi-mode splitter family is the same for DPU and adapter links.

3.3.4 NDR Connections to HDR and EDR

Cables for interoperability with HDR and EDR are shown in Figure 22. DAC cables are used for distances up to 2m and AOC cables provide up to 30m. These cables can be used to connect an NDR fabric to previous generation switches and adapters and can be mixed on the same NDR switch with NDR copper and optical cables.

Figure 22. NDR to HDR and EDR



The upper portion of Figure 22 shows solutions for HDR interoperability. All options use an OSFP twin connector at the switch end and the appropriate variant of QSFP for the other legs. The physical split is located close to the OSFP.

The lower portion of the Figure covers interoperability with EDR. EDR cannot be split so the options are fewer. Recall that 100 Gbps EDR and 100 Gbps HDR100 are not the same: EDR is four lanes of 25 Gbps and HDR100 is two lanes of 50 Gbps.

Note that there are no ACC cables, and no non-split cables. There is no easy way to use fiber to connect NDR to HDR or EDR because previous InfiniBand optics and fiber cables do not support MPO/APC.

3.4 Maintaining NDR Connectors and Cables

The expanded use of optical links in NDR requires thorough and consistent maintenance procedures.

3.4.1 Optics Need Cleaning

All optical fiber cables as well as the transceivers ship with their connector faces protected by dust caps.

During installation it is imperative that you clean the MPO connector as well as the transceiver receptacle before cable insertion. This must be done every time that a cable is plugged in.

Should a cable be inadvertently detached, clean again before insertion. The active optical connection is only 9 µm in diameter and the signal can easily be degraded by dust.

Plugging the cable without cleaning may cause degradation of the transmission, thereby resulting in the loss of data.

The type of fiber cleaner needed for MPO connectors is shown in Figure 23.

Figure 23. Fiber cleaner for MPO connectors

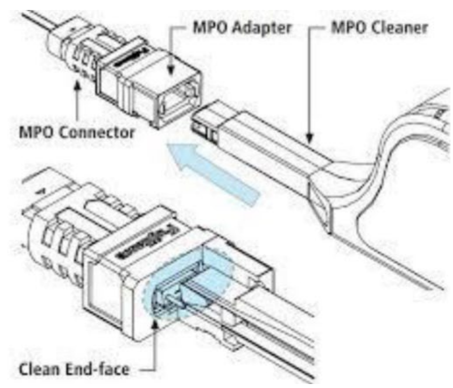
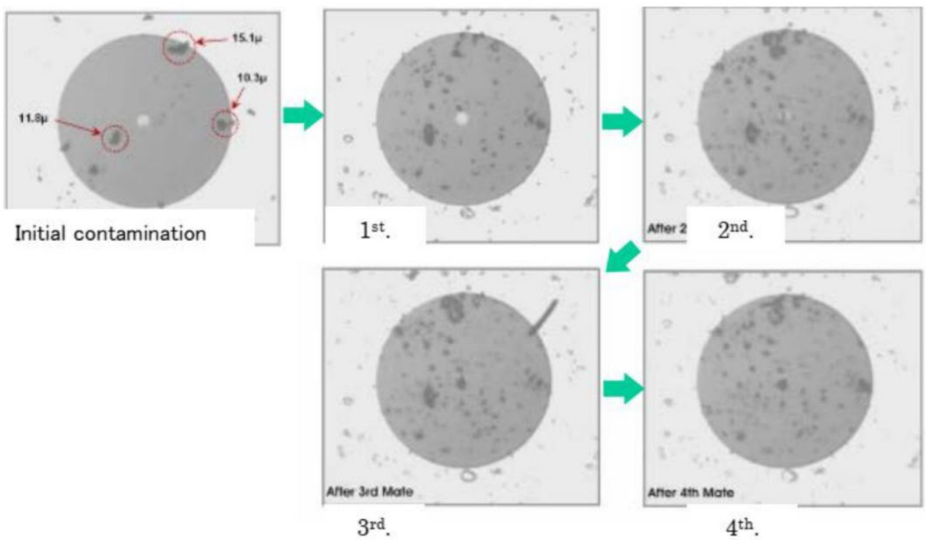


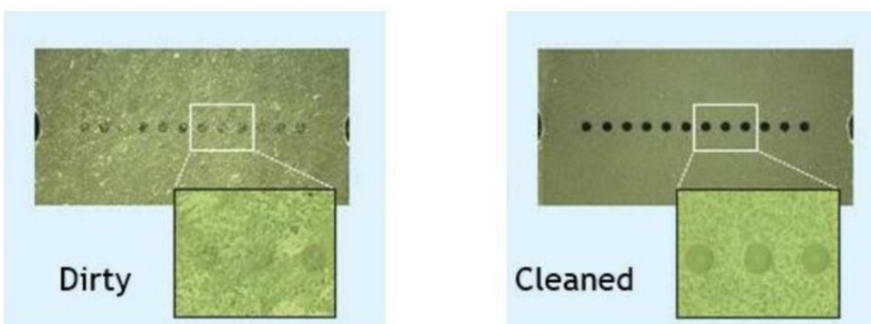
Figure 24 shows examples of a connector getting contaminated if not cleaned for each insertion. The little bright spot in the center is the active area for a single mode fiber, which is only 9 µm across. It got covered after the second insertion.

Figure 24. Examples of connector contamination if not cleaned



The difference between how an MPO connector could look, but should look, is illustrated in Figure 25.

Figure 25. How an MPO connector could look and should look



Chapter 4. Ethernet Cables Primer

Ethernet in today's data centers consists of a family of multilane link types with various signal rates per lane (Table 3).

Table 3. Ethernet speeds

Name	Lanes	Signaling Rate per Lane (Gbps)	Effective Bandwidth (Gbps)	Connector	Deployed
GbE	1	1	1	SFP, RJ45	
10 GbE	1	10	10	SFP+	2002
25 GbE	1	25	25	SFP28	2016
40 GbE	4	10	40	QSFP	2010
50 GbE	2	25	50	QSFP28	2016
50 GbE	1	50	50	QSFP56	2016
100 GbE	4	25	100	QSFP28	2010
200 GbE	4	50	200	QSFP56	2018

Ethernet support of single-lane links provides a wider range of speeds than InfiniBand:

- There are two versions of 50 GbE: a single lane at 50 Gbps, and dual lanes at 25 Gbps. This document will refer to the latter as "50 GbE (2x25)."
- There are two versions of 100 GbE: using four lanes at 25 Gbps, and dual lanes of 50 Gbps; the latter will be referred to as "100 GbE (2x50)."

4.1 Commonalities with InfiniBand

As previously noted, Ethernet and InfiniBand cable and transceiver technologies have nearly converged but are not interchangeable due to minor standards differences. Descriptions in Chapter 2 of these InfiniBand technologies also apply to Ethernet except where noted:

- QSFP connectors.
- DAC, ACC, and AOC cables.
- MPO-12 passive optical cables for multilane transceivers.

This section covers additional Ethernet cabling components and considerations.

4.2 Additional Cable Latency

In addition to speed of light delay—roughly 5 nanoseconds per meter depending on the cable technology—copper Ethernet links may require Forward Error Correction (FEC) to achieve the necessary BER.

FEC techniques introduce additional link delays of up to 120 nanoseconds because they add redundant error correcting codes to every data frame. In latency sensitive applications, use of low-loss DAC cables can eliminate the need for FEC. Such cables implement the IEEE CA-N or CA-L standards.

4.2.1 SFP Connectors

In addition to the QSFP connectors that are common between Ethernet and InfiniBand, Ethernet uses small form-factor pluggable (SFP) connectors for 10 GbE (SFP+), 25 GbE (SFP28), and 50 GbE (SFP56) cables. SFP connectors (Figure 26) are used for cables and for transceivers.

Figure 26. SFP connector

4.2.2 1G BASE-T Connectivity

Connections to 1G BASE-T cables can be achieved using an adapter that has an SFP+ connector on the port side and an RJ45 jack on the cable side (Figure 27).

Figure 27. SFP to 1G BASE-T (RJ45) adapter

4.3 Ethernet Switch Ports

Ethernet switches, especially top of rack (ToR) switches, often must support a variety of link speeds on the same switch, for example, 200 GbE uplinks with a mix of 50 GbE and 10 GbE downlinks within the rack. A variety of cable types is required to provide this flexibility.

4.3.1 Mixing Port Speeds

Enabling the necessary mix of uplinks and downlinks, including backwards compatibility to previous Ethernet generations, involves a combination of per-port speeds and port splitting. Switch vendors try to simplify this by offering models with different mixes of port speeds, but cables provide another layer of customization. For example, if a customer still has a few legacy GbE-connected nodes it may make sense to connect them to 10 GbE-capable ports instead of keeping the legacy GbE switches.

Tailoring switch port speed depends on:

- Native capabilities of the switch hardware, and switch vendor feature decisions
- Fundamental 'link format' considerations: lane rates and number of lanes

The following discussion deals with the fundamentals, which may be further restricted by a given switch model.

Table 3 showed the fundamental properties of each Ethernet speed class and how its aggregate throughput maps to a physical link. Often it is also possible to reduce a port's lane rate for backwards compatibility, for example to use a 100 GbE-capable port for 40 GbE, but all lanes of the port must run at the same rate.

4.3.2 Splitting Switch Ports

For multilane ports, it is often possible to logically split the lanes into subsets and use a physical splitter (breakout) cable.

Each 'leg' of a splitter must conform to a row in Table 3. For example, it is not possible to derive a 4 GbE link by running a 100 GbE port's lanes at 1 Gbps, nor is it possible to create a 20 GbE link by running a 50 GbE (2x25) port at 10 Gbps per lane. Some legacy implementations cannot be accommodated, for example 100 GbE using 10 lanes of 10 Gbps each.

4.3.3 Splitter Cables

Table 4 shows the valid combinations of splitters and lane rates.

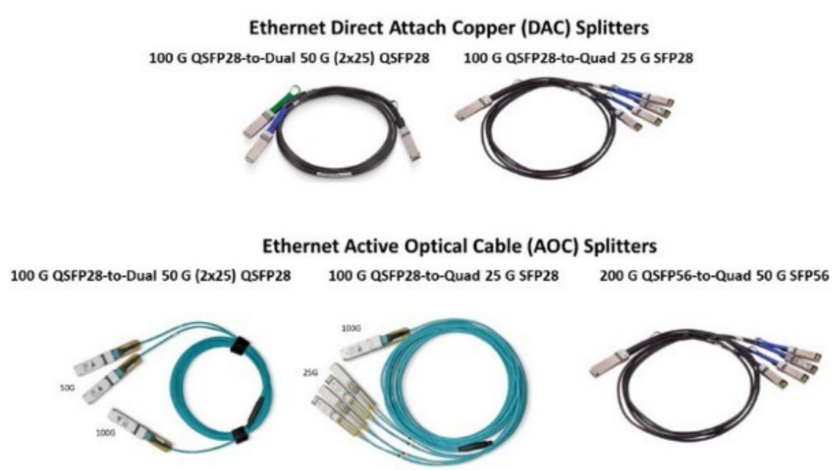
Table 4. Ethernet speed and port splitting combinations

Switch Port (GbE)	Lanes	Switch Connector	Port Mode	Lanes per Split	Rates per Lane (GbE)			
					50	25	10	1
10	1	SFP+	Unsplit	1			10[3]	1
25	1	SFP28	Unsplit	1		25	10	1
40	4	QSFP	Unsplit	4			40	
			4-way	1			4x10	4x1
50	2	QSFP28	Unsplit	2		50 (2x25)		
	1	SFP56	Unsplit	1	50			
100	4	QSFP28	Unsplit	4		100	40	
			2-way	2		2x50[1]		
			4-way	1		4x25	4x10	4x1
200	4	QSFP56	Unsplit	4	200	100	40	
			2-way	2	2x100[2]	2x50[1]		
			4-way	1	4x50	4x25	4x10	4x1

1. Each leg is two lanes of 25 Gbps (2x25)
2. Each leg is two lanes of 50 Gbps (2x50)
3. Blue = SFP+/SFP28/SFP56 as appropriate

For example, a quad splitter for a 100 GbE port can provide four links of 25 GbE, 10 GbE, or 1 GbE and will have a QSFP on the port end and SFP connectors on the other four legs. Again, specific switch models may not support all combinations shown. Several examples of Ethernet splitter cables are shown in Figure 28.

Figure 28. Examples of Ethernet splitter cables



4.3.4 Connecting SFP to QSFP

If two ports must be connected but one is SFP and one is QSFP, for example to connect a 25 GbE HCA to a QSFP port capable of both 100 GbE and 25 GbE, there are three solutions:

- Use a four-way splitter and do not use three of its legs.
- Hybrid cables.
- QSFP to SFP Adapters (QSAs).

A hybrid cable has a male SFP connector on one end and a male QSFP connector on the other. One specific lane from the QSFP side is connected to the single lane of the SFP. Though technically simple, there are practical issues:

- Multiple lengths are needed, with a mix of copper and optical versions.
- Signal integrity must match the application. For example, a hybrid DAC cable designed for 10 GbE will be constructed differently than a 25 GbE DAC.

To simplify connections between QSFP and SFP ports, NVIDIA introduced the QSA in 2011.

4.3.5 Quad to Single Adapters (QSAs)

A QSA is a simple passive module with a QSFP male connector on the port side and an SFP female connector on the other. It resembles a transceiver but is essentially a zero-length hybrid cable that needs no port power. Any compatible standard cable can then be used to complete the link. Examples of QSA modules are shown in Figure 29.

Figure 29. QSFP to SFP (QSA) adapters

To optimize cost based on speed and thermal requirements, there are multiple QSA types:

- 40 GbE to 10 GbE.
- 100 GbE to 25 GbE.
- 200 GbE to 50 GbE.
- Extended temperature versions for hotter environments.

4.3.6 Ethernet Transceivers

Transceiver costs, and power drawn from an Ethernet port, are a function of speed and distance (reach).

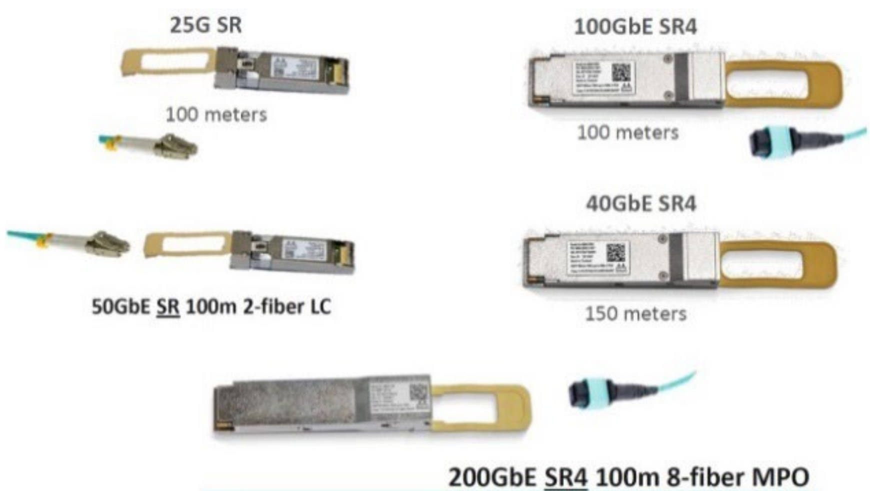
4.4 Short Reach (SR) Transceivers

For distances <200m, transceivers using multi-mode fiber are cost-effective. There are two categories:

- SR: these transceivers have an SFP connector on one end and a duplex LC connector for the optical cable. Typical reach is 100m for 25 GbE and 50 GbE.
- SR4: these transceivers have a QSFP connector on one end and an MPO-12 receptacle for the optical cable. Typical reach is 150m for 40 GbE, and 100m for 100 GbE and 200 GbE. MPO-12 cables are discussed in InfiniBand Cables Primer.

See Figure 30 for examples of short reach transceivers.

Figure 30. QSFP to SFP (QSA) adapters



4.4.1 Long Reach (LR) Transceivers

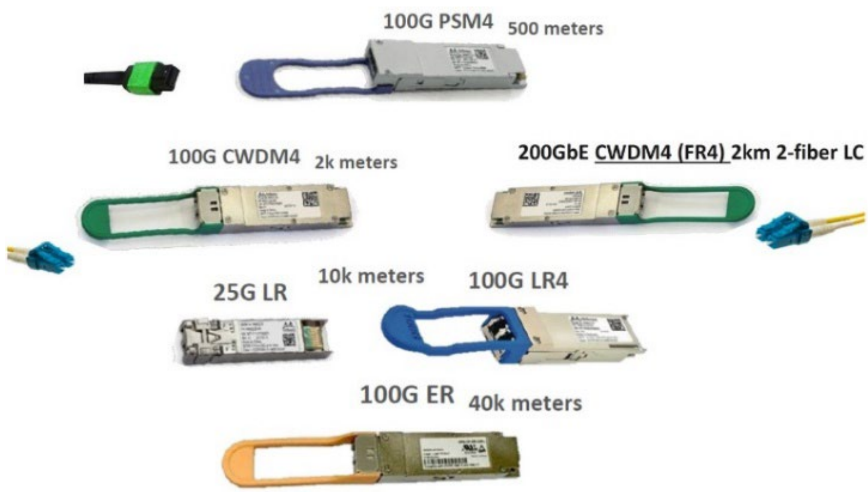
Transceivers for longer distances use single-mode fiber. Multilane signaling is accomplished using Wavelength division multiplexing (WDM) over a single fiber. Table 5 shows several types that available for different speeds and distances.

Table 5. LR transceiver specifications

Type	Speed	Maximum Distance
PSM4	100 GbE	500m using eight optical fibers in an MPO-12 cable
CWDM4	100 GbE	2 km using a pair of single-mode fibers with duplex LC connectors at each end
LR	25 GbE	10 km using two single-mode fibers with duplex LC connectors at each end
LR4	100 GbE	10 km over a pair of single-mode fibers with duplex LC connectors at each end
ER	100 GbE	40 km using two single mode fibers with duplex LC connectors at each end

Examples of LR transceivers are shown in Figure 31.

Figure 31. Ethernet LR transceivers



Chapter 5. Cable Deployment Planning

As with any well-planned network, those employing NVIDIA InfiniBand or Ethernet solutions should begin with a logical network design and a physical network design. This guide, being aimed at cable deployment, will speak to elements of the physical network design. It should be noted however that since logical design influences physical it should be completed first, or in tandem.

The logical network design must account for: servers, Ethernet switches, InfiniBand switches, storage, and any other cluster components. The physical network design will detail how these components are laid out and should include: floorplans, rack elevations, and infrastructure details. And most importantly, from the perspective of this guide, the physical network design will detail cable information including physical cable routes, lengths, and spares.

5.1 Considerations

Developing a physical network design involves a great many considerations and following the recommendations and guidelines in the following sections will ensure a smooth and maintainable cable deployment (Figure 32).

Figure 32. Cleanly deployed cables

5.1.1 Physical Cable Routes

Whether a switch exists at the ToR, intermediate/aggregate, or core/director tier in the fabric, cables destined for the same switch at the same rack position must be split in a 50/50 manner. That is, half the cables will be routed along the left side of the rack and half the cables will be routed along the right and then to their destination port. This division allows the great volume of the many InfiniBand cables to be balanced equally and thus reduces the total number of cables on any one side of the rack. Additionally, it allows for greater serviceability of all equipment within the rack ensuring no cable passes the midpoint of the rack impeding the insertion or removal of equipment.

If cables are run out the bottom and top of the rack, run them in quarters. This will reduce the congestion on the sides of the components where the cables run.

5.1.2 Selecting Cables

Generally fiber cables are employed for inter-rack runs due to their ability to traverse greater distances, smaller diameter, and better performance. However, when price is a determining factor, intra-rack cables will most often be passive copper cables as they are less costly and ideally suited to shorter runs. Active copper cables offer increased length compared to passive copper though not as long as fiber, and their price falls between the two. Active copper cables then are suited to both inter-rack and intra-rack runs.

Cable lengths should be chosen to minimize slack both inside the rack and in the cable management. Depending on equipment proximity, space, or cost constraints it may be advisable to mix copper passive, copper active, and fiber cables. While planning cable lengths consult the physical network design, pay particular heed to the floorplans and rack elevations, these documents will assist in selecting the proper lengths.

> Note:
>
> - Cables do impart some minor latency therefore it is best to keep all cluster components close to each other.
> - Active copper cables are usually thinner than passive copper cables and will fit easier within the rack.
> - Fiber cables are more expensive but take less room and weigh less than copper.
> - Use both passive and active Copper cables to minimize cost.
> - Use both Copper and Fiber cables if possible, to minimize cost.

Failure rates for InfiniBand cables are very low. However, like any hardware component, they do occur. Plan for one extra cable per 200 end points. Spares must be of varying lengths matching the lengths of the cables within its proximity. Alternatively, if spares are run, be sure that the cable has enough slack to reach additional ports.

Note: NVIDIA offers a unique option with cables, color-coded pull tabs. These tabs can assist the installer in identifying cables during the implementation.

5.1.3 Selecting Racks

For large multilayer fabrics built using high-density/director switches, it is best practice to house the switch chassis in its own rack. The rack should also meet the following guidelines:

> Source extra wide racks.

> Have 4-6 inches (10-15 cm) between the vertical supports and the sides of the rack. This is important when using copper cables due to the greater individual cable diameter.

> Do not remove front doors, rear doors, or side panels.

> Leave around 4U at the top and bottom of the rack. This will facilitate manipulating cables at the time of deployment.

5.1.4 Selecting Cable Management

Cable management comes in different sizes; review manufactures specifications and source large enough systems to account for the great number of cables and diameters. Where possible oversize cable management such that cable slack can be contained outside the rack.

5.2 Document Network Connectivity

Cable documentation will be part of a comprehensive physical network design and should include a P2P cable matrix/schedule. The P2P matrix defines the logical connections between any two ports as well as the physical distance between the port pairs. This will ensure that all interconnects are properly accounted for and will facilitate an efficient and error free deployment.

Note: Document the planned and later the actual topology of the configuration. Be sure to keep the documentation updated reflecting any changes to the planned topology.

5.2.1 Point-to-Point Cable Map

The P2P cable map is a comprehensive document consisting of a list of all cables to be implemented, a source and destination and bundling matrix describing every cable end point, and labeling information. There should be a map that expresses for each server or switch, the rack in which it is located and its exact position in the rack. Also, for each server type a map should be included explicitly showing the location of the port. This is particularly important for hosts with multiple network interfaces. From this map, cable lengths and bundling strategy can be determined.

It is important to review this map thoroughly to ensure it is correct. Incorrect cable lengths can lead to issues in system deployment and lead to long delays if cables need to be reordered. System owners should consider consulting with network experts to ensure their P2P map is correct.

5.2.2 Constructing Cable BOM

While planning the cable implementation various intricacies of the logical and physical design will dictate the sourcing of many different cables of different lengths and types. Key cable features including length, cable type, and cable connector must be determined. This step should be given the proper time and attention to ensure it is done correctly. Determining physical cable lengths from the P2P map can be difficult because of any number of factors that affect how cables are installed. Ordering cables that are too short will cause a delay in system deployment to reorder the right cables. While conservative estimates of cable length can prevent this issue, managing the extra cable can be challenging and lead to long-term cable maintenance issues.

5.2.3 Labeling Information

Many different naming conventions exist, and large cable deployments rely heavily on good practices. It is critical that all cluster elements have unique names to ensure all are connected per the design.

The cable labeling then should reflect at least:

- Designations representing origin components and destination components.
- Including the port the cable should connect to.
- Cable type and length details.
- And the BSN (see: Approach and Methodology).

Chapter 6. Cable Deployment Methodology

Following planning and documenting the cable requirements, the next phase is the physical implementation of the cable plant. The following represents a comprehensive outline of the cabling best practices that NVIDIA recommends.

6.1 Assess the Environment

Space is needed to prepare the cables before installing; be sure that there is an area large enough to work in before beginning.

6.2 Survey Cable Management

Cable management can take many forms and most data centers maintain a plan for the type of trays, baskets, and conduits used. The planning phase should have accounted for these variations. However, before running any cables, it is advisable to review paths the cable management system follows, looking for: obstructions, tight corners, and missing "cascades."

6.3 Inventory

Before any deployment work begins, inventory the cables. Sort the cables by type and by length. Compare the total count of each cable type against the planning documentation.

6.4 Stage

Stage cables before laying cables into their final location. In the previous step, the boxes should have been grouped by type and length increasing the efficiency of the staging process.

6.4.1 Un-Boxing and Un-Bagging

The first step in the staging process is to remove the cables from their boxes and antistatic bags. It is most efficient to unbox and unbag all cables at one time, provided space allows. Leave any cable wraps or connector covers in place at this point; these will be removed later.

Warning: Since this is a repetitive task, be sure to maintain standards, and focus to prevent damage to the cables.

Note: Cables should remain grouped by type and length.

6.4.2 Labeling and Grouping

In the previous steps, the cables have been grouped by type and length. However, it is now time to begin labeling and grouping the cables in preparation for bundling.

Note: It is impractical to print out labels onsite, proper documentation is critical to print labels before arrival.

Labeling and grouping are semi concurrent tasks; groupings are simply unbundled collections of cables that will become bundles in the next steps. Though properly formatted labels should make it unnecessary, the cable deployment documentation can be consulted to determine the cable groups.

To begin, review the sheets of preprinted cable labels; per best practice the labels should have Source, Destination, cable type/length details and a Bundling Sequence Number (BSN). The cable type/length details with the BSN will allow for grouping the cables without directly consulting the cable deployment documentation, though referencing is advisable to ensure that no mistakes were made.

- Using the cable type/length information from the label select a cable of the correct length.
- Apply the label 3-4 inches from the connector head.
- Attempt to orient all labels the same way by starting them in the same position.
- Group the cables by their sequence number.

Warning: This is a repetitive task. Maintain standards and focus to prevent damage to the cables.

Warning: While grouping the cables, lay them down away from foot traffic or rolling loads.

Note: Ideally all cables will be labeled and grouped before the next step. However, space constraint may dictate that this operation be completed piecemeal.

6.4.3 Unrolling and Bundling

With the cables sorted and grouped by type, length, and BSN they can now be unrolled and bundled making them ready for deployment. The BSNs should be ordered with the lowest number being the first bundle to be deployed and the highest the last.

6.4.3.1 Unrolling

Note: Proceed with the following steps one cable at a time.

Warning: Stage the cables away from foot traffic or rolling loads.

1. Find the cable grouping with the lowest BSN and move these cables to the area designated for staging the cables.
2. Divide the group of cables into two halves, placing each half on either side of the staging area.

   This prevents excessive walking—unrolling one cable and then going back and unrolling the second cable from the other side of the room.
3. Select a single cable and remove any cable wraps leave connector covers in place.
4. Begin unrolling the first cable.
   a. Free one end of the cable and in a handover hand fashion lower the connector to the floor.
   b. Continue the handover hand unrolling slowly laying the cable down in a straight line for as long as space is available. If the cable is long and there is enough space, turn the cable back leaving a large radius bend, and continue this pattern until the cable is fully unrolled.
5. Select the next cable and repeat step 2 this time laying the cable closely beside the first.
6. Continue for all cables in the grouping.

Warning: Cables should be unrolled rather than uncoiled.
Do not uncoil or stretch cables this can lead to kinks and damage.
Do not twist the cable to open a kink.
Do not pull the cable through any openings or around corners.

6.4.3.2 Bundling

This is the procedure for bundling cables:

1. Prepare short strips of hook and loop wraps.

   The strip should be able to surround the diameter of the bundle with extra 2 cm. Too long wraps will create fat bulges along the bundles that can cause issues during layout of bundles.
2. Starting at the connectors at one end of the bundle place a hook-and-loop wrap around the bundle approximately one foot from the end.
3. Progressing down the length of the bundle at approximately 2.5 foot intervals, place a hook-and-loop wrap around the bundle.
4. After all the hook-and-loop wraps are in place, wrap the connectors in plastic to prevent them from being damaged during the installation process.

Warning: Wraps must not be so tight as to distort the jacket of the cable. They are only used to prevent unnecessary movement of the cable; snug is tight enough.

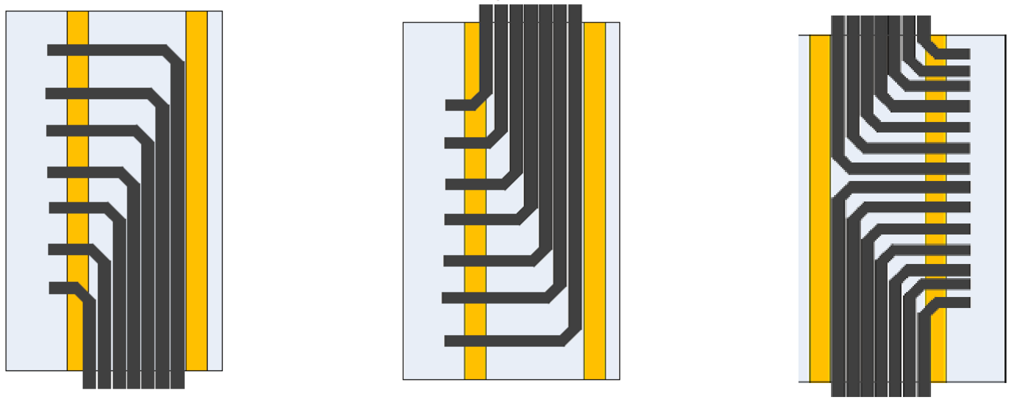
6.5 Deploying the Bundles

Cables entering a network rack can be said to enter in one of three ways:

> Top feed.
> Bottom feed.
> Top/bottom feed.

Top and bottom feeds are routed into a rack in much the same way, with the largest difference being that for top feed the first bundle to be run should correspond to the topmost switch or switch line card and for bottom feed the first bundle should be the bottommost switch or switch line card. Top/bottom divides the rack and switches within into upper half and lower half and runs the bundles respectively. Figure 33 shows how the cable bundles will be organized within the rack.

Figure 33. Side views of cable feeds



Bundles should always be plugged in from bottom to top. This is because when plugged in from top to bottom, the top cables will slow down plugging the next bundle under.

Additionally, as was detailed in Section 4.1.1, cables and thus bundles destined for the same switch at the same U must be split in a 50/50 manner. This allows half the cables to be routed along the left side of the rack and half the cables along the right and then to their destination port. This division allows the great volume of the many InfiniBand cables to be balanced equally and thus reduces the total number of cables on any one side of the rack. Additionally, it allows for greater serviceability of all equipment within the rack ensuring no cable passes the midpoint of the rack impeding the insertion or removal of equipment (Figure 34).

Figure 34. Top view of cables on the connector side

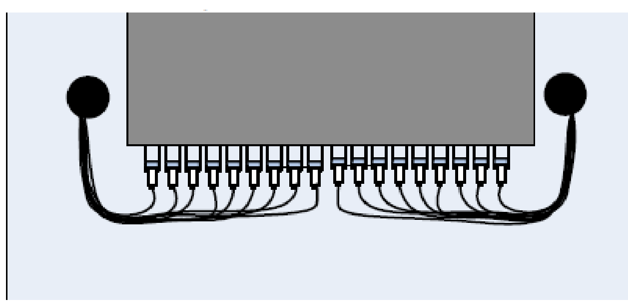
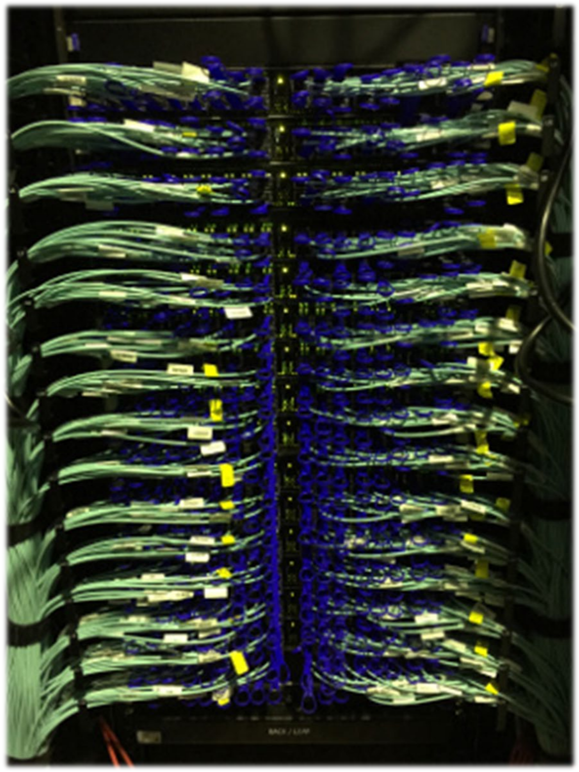


Figure 35 shows a front view of a properly cabled chassis-based switch.

Figure 35. Front view chassis-based switch



6.5.1 Install

Good cabling begins with good preparation. In previous steps, the environment and cable management have been surveyed for readiness. The cables are, by this point, labeled and bundled and thus are ready to be laid in their final position.

Typically, the BSN will be sequential and reflect the order in which to deploy the bundles. However, this is not always true, therefore it is recommended to first consult the cable deployment documentation to determine the order in which the bundles are to be deployed. In general, the inter-rack cable bundles will be run first, and the intra-rack cable bundles run last.

6.5.1.1 Installing the inter-rack cable bundles

Even the best cable management infrastructure can have obstructions that prevent directly placing the bundle within. It will therefore in some cases be necessary to coil the bundle as a whole and unroll it as it is placed in its final location.

> **Note:** The cable bundles must be supported in some fashion while being placed into the cable management. This may entail temporary wrappings and strappings or any supporting methods to ensure that the bundle does not kink or stretch while being worked with.

> **Note:** Slack should be positioned nearer the TOR switches. The core and intermediate/aggregate switches will have numerous cables running to them and therefore slack above these will cause congestion within the cable management.

> **Warning:** Never route cables through the middle of the rack. This results in air flow restrictions and makes servicing all equipment blocked by the cables difficult or impossible without disconnecting the cables.

Move the prepared cable bundles from the staging area to the area immediately near, in front of, the core switch/network rack and the bundle deployment.

1. Place the first eight feet or so of the cable bundle into the cable management +/- a rough estimate of the required free cable to reach the U for which it is destined.
2. Carefully fish the free cable through the designated cable openings and ensure that the connectors will reach the midpoint of the switch or line card.

   Do not connect the cable yet. Tie this end in place with hook-and-loop and continue with the bundle.
3. Moving down the length of the bundle place it into the cable management, carefully working around support structures, when making corners take the widest radius possible.

4. Upon reaching the destination rack, review the remaining length of the bundle and if considerable slack is present determine the best manner to address:
   a. Cable management comes in different sizes and ideally it was sized large enough to contain cable slack.
   b. Cable slack is best managed by creating large loops with large bend radius that double back on themselves.
5. As with the first end of the bundle, carefully fish the free cable through the designated cable openings and ensure that the connectors will reach the midpoint of the switch or line card.

   Do not connect the cable yet. Tie this end in place with hook-and-loop and continue with the bundle.
6. Repeat this process for the remaining bundles.

   Continue to Section 6.5.2 and then return to the next section.

6.5.1.2 Installing the intra-rack cable bundles

Cable management within a rack takes many forms. The key items to keep in mind are bend radius, cable support, and individual cable path management. Cable bend radius should never at any time exceed specifications. Proper cable support should be implemented to ensure longevity. Once cables are separated from the bundle, they must be managed to prevent the blocking of other equipment, restricting air flow, and maintaining equipment serviceability.

> **Warning:** Never route cables through the middle of the rack. This results in air flow restrictions and makes servicing equipment blocked by the cables difficult or impossible without disconnecting the cables.

Generally, intra-rack cabling is not bundled before installation and given the variables of rack cable management it must be left to the installer to determine whether to begin at the switch end of the cable or node end.

1. Whether starting at the switch or node plug the connector into its source port and fish the cable to the destination port, taking special care not to twist the cable near the connector.

> **Note:** If it is necessary to reorient the connector, begin the adjustment further down the cable, at least 75 cm away from the connector, so the twist occurs gradually along the cable instead of abruptly near the connector.

2. Use the cable management to orderly contain any slack, maintaining awareness of bend radius limitations.
3. Repeat this process for the remaining bundles.
4. Continue to Section 6.5.2 with attention paid to dressing.

6.5.2 Connecting and Dressing Cables/Bundles

Connecting and dressing cables/bundles are discussed together to ensure a clean appearance and an easily maintained deployment. Both must occur concurrently. Simply plugging the connectors in and running to the side will result in tangles and disorder that will progressively worsen as the number of cables increases.

> **Note:** Be sure to provide strain relief, cables must be supported to prevent stress. In general, a cable should be supported at least every two feet (60 cm).

> **Warning:** Straps must not be so tight as to distort the jacket of the cable. They are only used to prevent unnecessary movement of the cable; snug is tight enough.

6.5.2.1 Connecting

1. Work on the left side or the right side of the rack first.

   Begin by removing any plastic wrap that was used during the installation process. Additionally, remove any protective caps that might remain on the cable.
2. If cables enter through the top of the rack, begin with bundles with destinations at the top.
3. Using the label information, fully seat the connector into the correct port. It is possible that other cables may need to be unplugged and plugged back in again to avoid tangles and resolve twists caused by the cable wanting to return to its original coiled shape.
   a. Because cables routed to the left or right most ports will be shorter than those to the middle variations will result in slack, route this into the in rack cable management.
   b. Progress one U at a time and alternate between connecting ports and dressing cables.
4. Continue to the next section.

6.5.2.2 Dressing for Equipment

1. Having connected all the cables for the left or right side of a single U, begin near the middle and gently guide the cables into shape.
   a. Be careful with bend radiuses, and do not pull on cables.
   b. Keep the plane of the cables in line with the level of the U that they connect to.
2. Using hook-and-loop, wrap at intermediate points and then within the cable management.

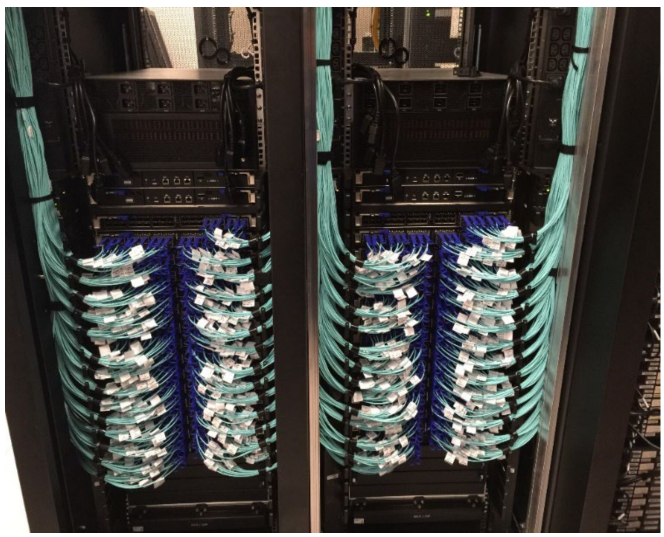
   In the early stages of the connection dressing process, keep the wraps quite loose as they may need to be adjusted.

3. Once reaching the last cable, add extra wraps as needed to any slack, wraps, or bends.
4. When everything looks right, tighten any wraps to just snug, and then close-up any cable management.

6.6 Cable Management

Due to the volume of cables involved large configurations, the size of the SFP and QSFP connectors, as well as the bend radius of the various cable types, additional installation preparation may be required. Ensure that proper cable management standards and practices are followed (Figure 36).

Figure 36. Properly managed cabling in a rack



6.7 In-Rack Cable Management

Many accessory items are available to help manage cabling ingress, egress, and in-rack cable management. These include vertical and horizontal cable ducts, strain reliefs, and bend radius managers (Figure 37).

Figure 37. Various rackmounted cable management devices

6.8 Cable Management Requirements

Cable management requirements are to:

- Bundle cables together in groups of relevance (for example, ISL cables and uplinks to core devices) to ease management and troubleshooting.
- Use cables of correct length. Leave only a little slack at each end.
- Keep cable runs less than 90% of the max distance supported for each media type, as specified in the applicable standard.
- Keep copper and fiber runs separated.
- Install spare cables in advance for future replacement of damaged cables.
- Use color coding of the cable ties. The colors should indicate the endpoints. Place labels at both ends, as well as along the run.

6.9 Cable Management Best Practices

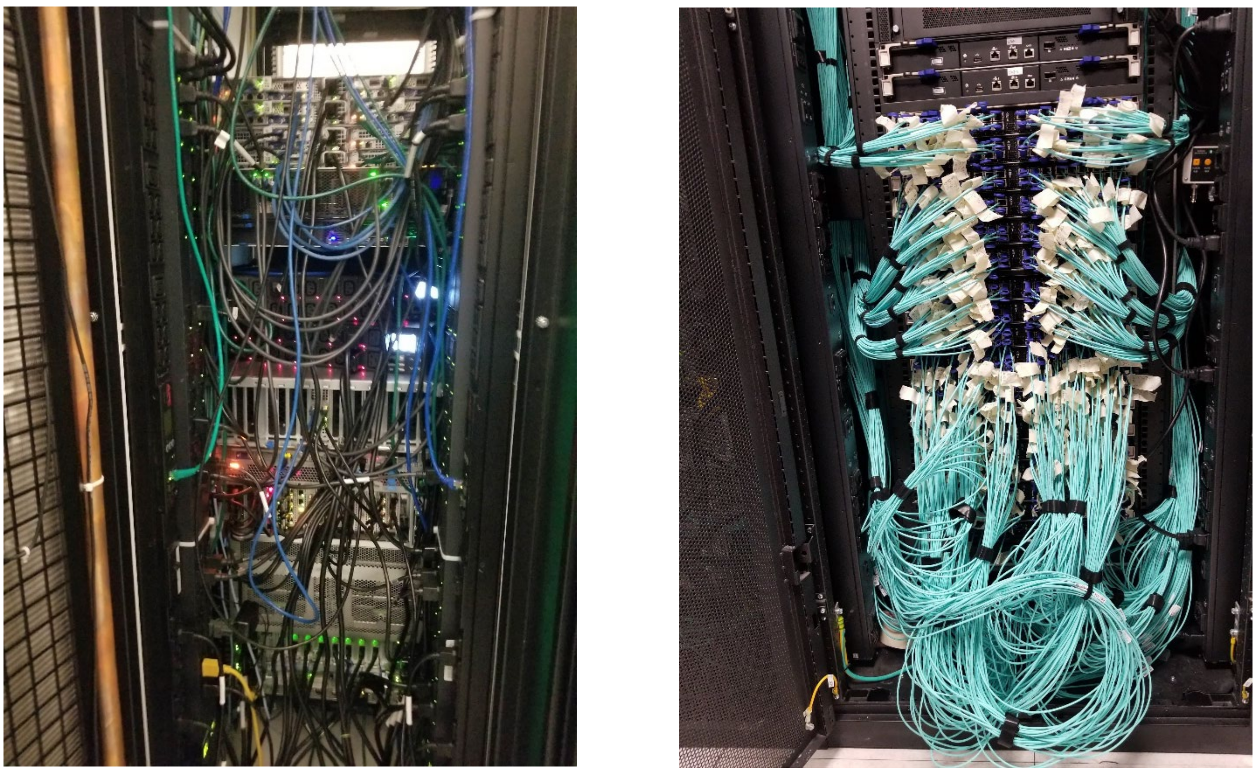
Ensure that:

- Transceivers and system units (for example, fan units) in the switch can be easily replaced.
- The rack is wide enough to place the cables between the switch and the rack side walls.
- The cables do not block air flow.
- The cables do not block transceiver or system unit extraction.
- The cables are tied to the rack structure to remove strain and tension on the connectors.
- The weight of the cables is supported by the cable management system and the rack floor. Refer to Cable Weight to plan weight loads for structured cabling.
- Cables are supported every 2m or placed in trays.
- Cables and bundles do not block other equipment.
- Cables are not routed through pipes and holes. This may impact serviceability as well limit additional future cable runs.
- Soft hook and loop fastener cable ties are used instead of plastic.
- Cables are in trays whenever possible.
- Abandoned cables that can restrict air flow, and could cause overheating or shutoff, are removed.
- Cables are disconnected by gently pulling by the tag of the connector.

6.10 Poor Cable Management

Figure 38 shows two examples of poor cable management. On the left is a rack with sparse cabling. Although there are few connections, the lack of labeling would make it hard to troubleshoot connections. In addition, cables running behind a device that is being removed for service might need to be removed, causing unnecessary interrupts to other elements of the configuration. Although the cables in the rack on the right are tagged and bundled, the slack at the lower part of the cabinet could cause equipment to overheat or shutoff.

Figure 38. Poor cable management



Chapter 7. Cable Support Systems

Cable support systems should provide the strength and working load capacity sufficient to meet the load requirements of the cable tray system and any future expansion. All work should be reviewed by a qualified structural engineer before implementing any installation.

All cable management and pathway apparatus should be selected, installed, and maintained in accordance with applicable industry standards including ANSI/TIA, BICSI, and NEMA. They should also be installed in compliance with all locally applicable fire safety and electrical codes, such as National Electric Code (NEC) in the United States and Mexico, The Canadian Electric Code in Canada, The International (IEC) in Europe, The British Standard (BS) 7671 in the United Kingdom, and NF C 15-100 in France.

7.1 Planning Considerations

There are many factors to consider when calculating cable support system capacities.

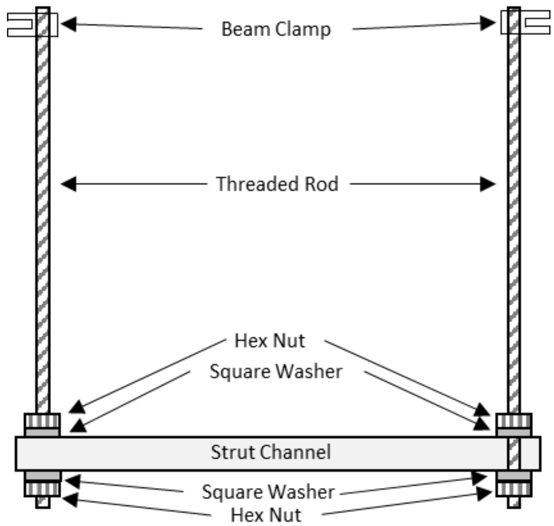
These include:

- Building overhead support structures.
- Trapeze capacities.
- Tray type and construction.
- Stacking multiple trays
- Installing the fiber optic tray system.
- Installing power busway system.
- The types of cables.
- The planned amount of cable runs.

7.2 Cable Support Mounting

Cable support systems are often hung from a ceiling support system called a "Trapeze" (Figure 39). There are several engineered solutions. The following illustration shows a basic trapeze assembly.

Figure 39. Trapeze cable support



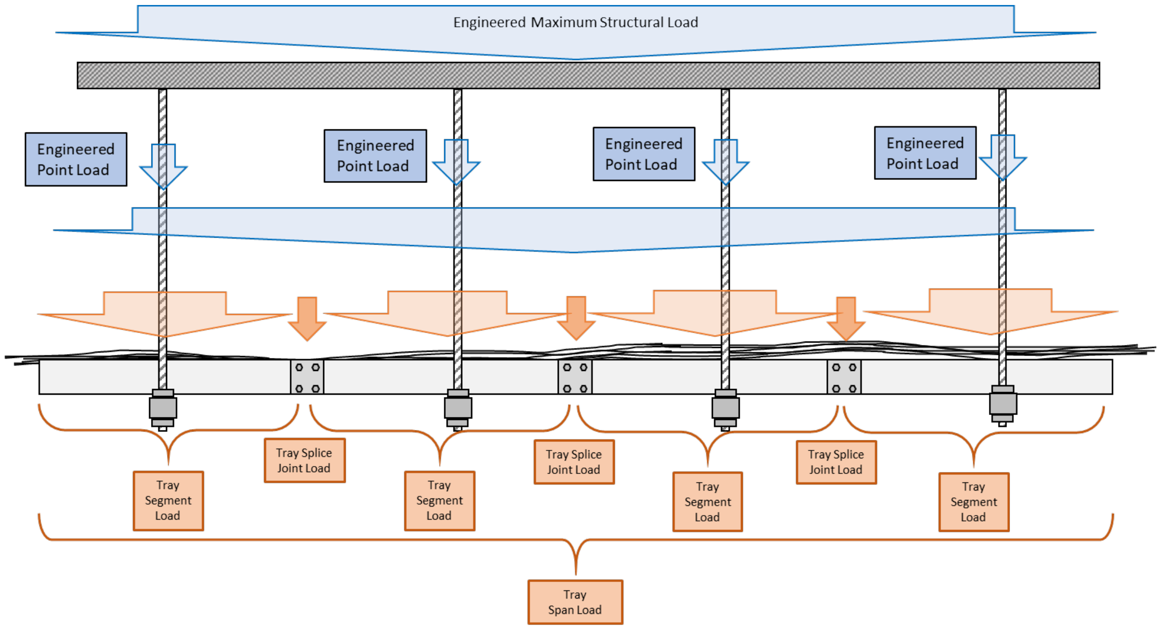
7.3 Cable Support Load Distribution

Key calculation points:

- Structural load. The engineered maximum load that the ceiling support structure can carry.
- Engineered point load. The overhead connection point that the trapeze is hung from.
- Tray segment load. The manufacturers specification how much weight can be supported per segment.
- Tray splice joint load. The manufactures specification on how much weight the mechanical splice can support.
- Tray span load. The manufactures specification on how much weight can be supported across multiple spans.

Figure 40 shows how the cable support system distributes the weight load.

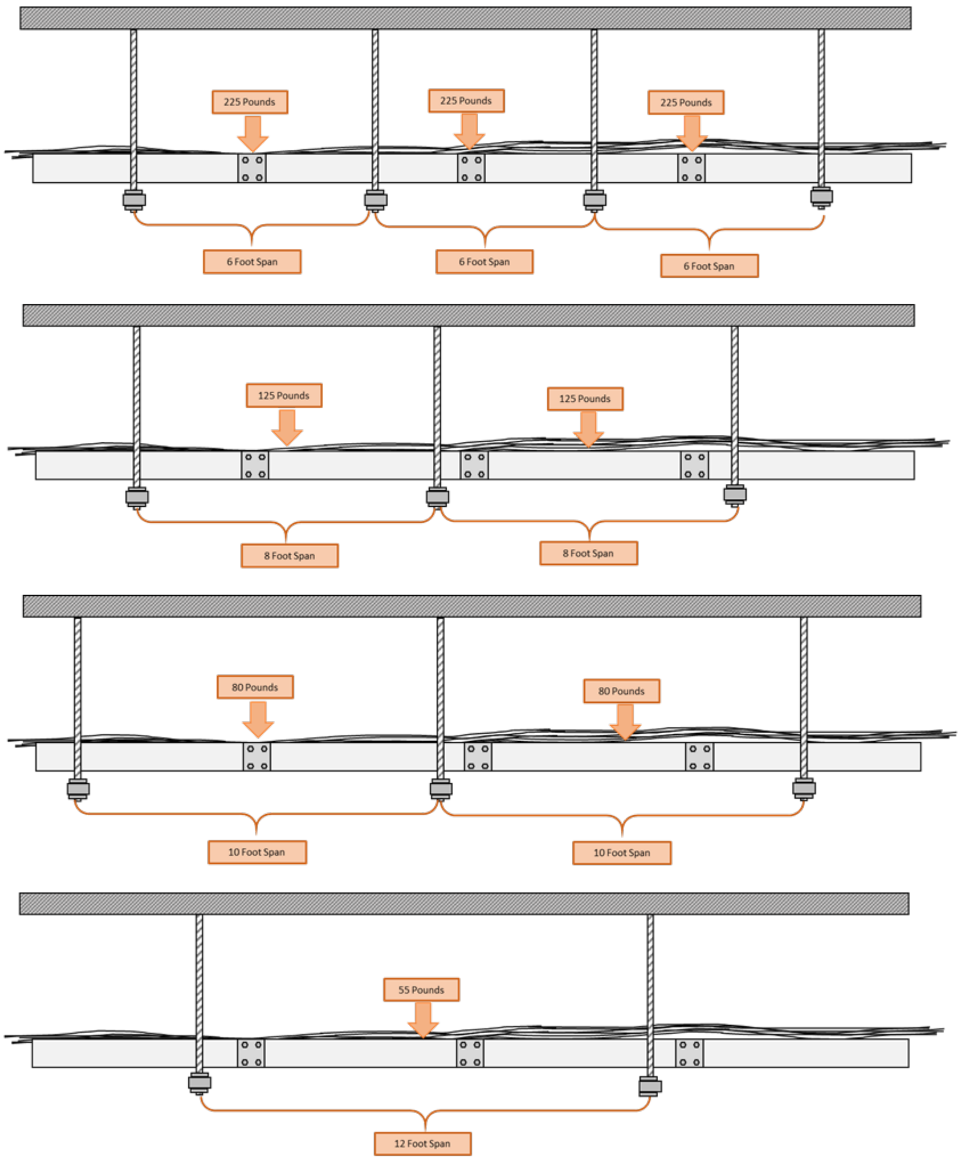
Figure 40. Distributing weight across the cable support system



7.4 Span Load Examples

As shown in Figure 41, the maximum tray load decreases as the tray span width increases.

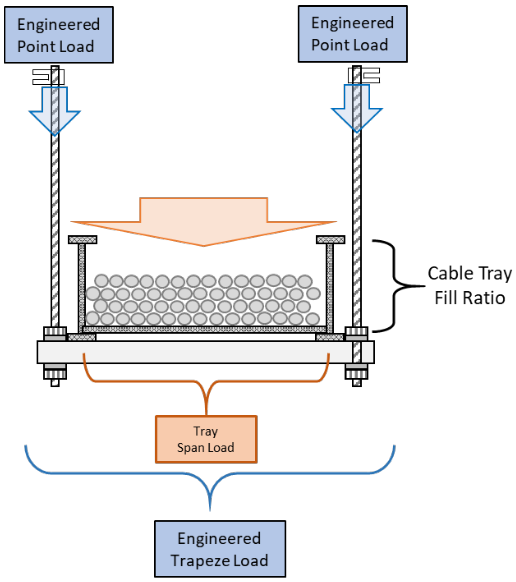
Figure 41. Tray load decreases a tray span increases



7.5 Trapeze Single Layer

Figure 42 illustrates a side view of a cable tray system.

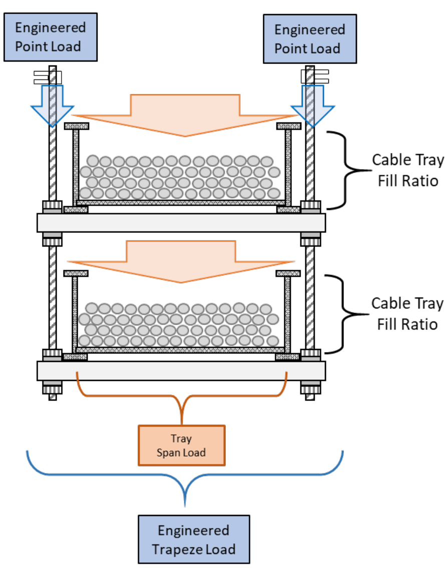
Figure 42. Side view of cable tray system



7.5.1 Trapeze Dual Layer

Figure 43 illustrates multiple cable trays using a trapeze support system.

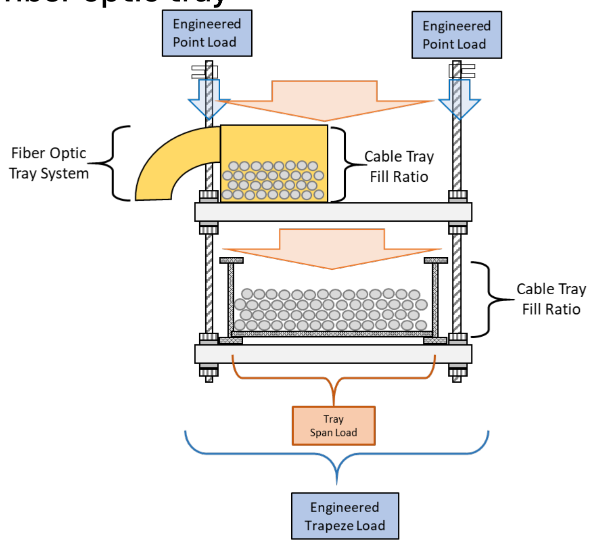
Figure 43. Multiple trays on a common trapeze support



7.5.2 Trapeze Mixed Use

Figure 44 shows a separate fiber optic tray system.

Figure 44. Separate fiber optic tray



7.6 Cable Management Guidance

Here is some guidance for cable management:

- Avoid sharp edges in the cable pathway.
- Keep cable bend radius with supported bend radiuses.
- Segregate cables for power, copper data, and fiber data.
- Minimize cable travel distances.
- Ensure that cables or associated apparatus do not obstruct airflow to equipment.
- Use proper cable strain support.
- Label and document cables.

The three common cable support systems include:

- Fiber optic trays.
- Basket trays or wire mesh.
- Cable ladders.

Choosing the correct cable support system for an application requires careful consideration for a deployment.

The structural load of that cabling is 380 cables * 0.009 Kg per cable per linear meter, or 3.42 Kg per linear meter. The calculations illustrated here are general guidelines, and calculations based on the size and weight specifications of the specific quantities of the actual cables to be used should be performed.

Appendix A. Terms and Definitions

Table 6 contains some terms and definitions used in this document.

Table 6. Terms and definitions

Term	**Definition**
Active Optical Cable (AOC)	An optical fiber cable with an optical transceiver with the fibers bonded inside and not removable. The optical transceiver converts the host electrical signals into light pulses and back
InfiniBand	InfiniBand is a computer-communications standard used in HPC that features very high throughput and very low latency. InfiniBand is commonly used in HPC and hyperscale data centers. InfiniBand is promoted by the InfiniBand Trade Association (IBTA).
Quad Small Form-factor Pluggable (QSFP)	A bidirectional transceiver or cable with four lanes in each direction.
Small Form-factor Pluggable (SFP)	A transceiver or cable with a single lane (channel) in each direction. All cables and transceivers commonly used in data centers are bidirectional.
Transceiver	A converter with an electrical connector in one end and an optical connector in the other end. It can have one or more parallel lanes in each direction (transmit and receive).

Appendix B. References

B.1 Building Cable Standards

Building cable standards include the following:

- ANSI/TIA-568.0-D, Generic Telecommunications Cabling for Customer Premises.
- ANSI/TIA-568.1-D, Commercial Building Telecommunications Infrastructure Standard.
- ANSI/TIA-568-C.2, Balanced Twisted-Pair Telecommunication Cabling and Components Standard, published.
- ANSI/TIA-568-C.3, Optical Fiber Cabling Components Standard
- TIA-569 Commercial Building Standard for Telecommunications Pathways and Spaces.

B.2 Cable Tray Standards

Cable tray standards include the following:

- NEC: The National Electrical Code.
- CSA: Canadian Standards Association.
- NEIS: National Electrical Installation Standards equivalent to NEMA VE1.
- NEMA VE1: National Electrical Manufacturers Association (partnered with CSA) Standard for Metal Cable Tray Systems.
- NEMA VE2: National Electrical Manufacturers Association Standard for Cable Tray Installation Guidelines.
- IEC 61537: International Electrotechnical Contractors Standard for Cable Tray Systems and Cable Ladder Systems for Cable Management.
- IEC 60204: International Electrotechnical Contractors Standard for Safety of Machinery/Electrical Equipment with Machinery.
- NFPA 79: National Fire Protection Association's Standard; equivalent to IEC 60204.
- NFPA 70: National Fire Protection Association's Standard; equivalent to NEC.
- NEMA Standards Publication VE 2 Cable Tray Installation Guidelines.
- NEC ARTICLE 392—CABLE TRAY.

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051

http://www.nvidia.com