



NVIDIA LinkX Cables and Transceivers Guide to Key Technologies

Table of Contents

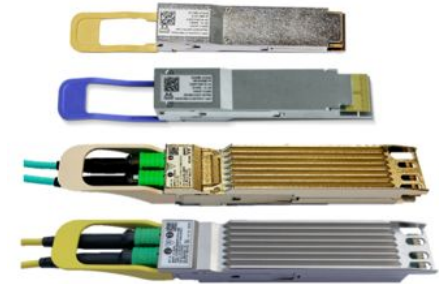
- About This Technology Guide 4**
 - Where to Find More LinkX Documentation 4
- Parts are Described by Use in the Switch 6**
- NVIDIA Networking Overview 7**
 - NVIDIA Spectrum, Quantum, and LinkX Product Lines 7
- Key Technologies 11**
 - Modulation Rates.....11
 - NRZ Modulation 11
 - PAM4 Modulation 12
 - Dual Protocol Capability13
 - Connectors and Cages.....17
 - 100G-PAM4 Series: Twin-port OSFP and OSFP Plugs..... 18
 - QSFP112 19
 - 50G-PAM4 series: QSFP-DD, QSFP56 20
 - 25G-NRZ series: QSFP28, QSFP+ 21
 - Backwards Compatibility 21
 - DACs, ACCs, AOCs, and Transceiver Interconnects.....22
- Optical Connectors 26**
 - MPO-12 Optical Connectors26

LC Duplex	26
Single-mode and Multimode Optical Fibers	28
Single-mode Fiber	28
Multimode Fiber	28
Crossover Fiber Cables	31
Acronyms and Abbreviations	34
Document Revision History	36

About This Technology Guide

This technology guide introduces the basic technologies and terminologies used for NVIDIA cables and transceivers for NVIDIA Quantum InfiniBand and Spectrum Ethernet architectures.

Cables and transceivers are constructed using multiple technologies and a blizzard of buzzwords to describe the parts. This makes discussing cables and transceivers almost a language by itself. This document describes the basic components and terminologies used in constructing and describing NVIDIA® LinkX® cables and transceivers.



Where to Find More LinkX Documentation

This technology guide is to be used in conjunction with other documents located in docs.nvidia.com/networking/ > Interconnect. This site is where the following LinkX cables and transceivers documents are provided.

LinkX Overview Documents:	Review of parts, important notes, and configuration details for linking to NVIDIA switches and adapters <ul style="list-style-type: none">• LinkX Cables and Transceivers Guide to Key Technologies (this document)• LinkX User Guide for 400Gb/s 100G-PAM4 OSFP & QSFP112-based Cables and Transceivers• LinkX User Guide for 400Gb/s and 200Gb/s using 50G-PAM4 and 100Gb/s using 25G-NRZ Modulation Cables and Transceivers
Configuration Maps:	Picture and part number-based PowerPoint® slides for every configuration with NVIDIA switches, network adapters, and DGX GPU systems for 100G-PAM4, 50G-PAM4, 25G-NRZ cables and transceivers <ul style="list-style-type: none">• Configuration Maps

Parts Lists:	<p>Tables summarize by speed, form factor, connector, power, reach, etc. and hyperlinks to individual products specs</p> <ul style="list-style-type: none"> • 400Gb/s (100G-PAM4) Transceivers and Fiber Parts List • 400Gb/s and 200Gb/s (50G-PAM4) and 100Gb/s (25G-NRZ) Cables and Transceivers Parts List using QSFP-DD, QSFP56, QSFP28, SFP28
Product Specifications:	<p>10-to-16-page detailed hardware datasheets with physical, thermal, electrical, and optical specifications for each product</p> <ul style="list-style-type: none"> • docs.nvidia.com/networking/ > Interconnect > <i>select speed and type</i>
Additional Docs:	<ul style="list-style-type: none"> • NVIDIA Cable Management Guidelines and FAQ

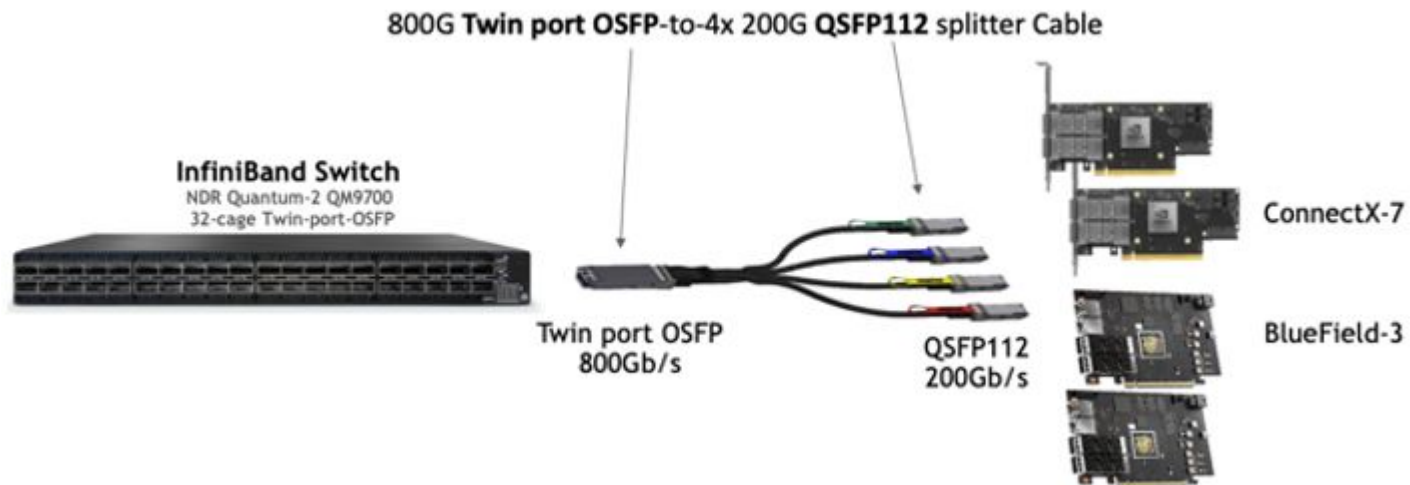
The guides are organized by speeds and modulation rates:

100G-PAM4:	400GbE Ethernet and 400Gb/s NDR InfiniBand using twin-port OSFP, OSFP, and QSFP112
50G-PAM4:	400GbE Ethernet only using QSFP-DD and QSFP56
50G-PAM4:	200GbE Ethernet and 200Gb/s HDR InfiniBand using QSFP56
25G-NRZ:	100GbE Ethernet and 100Gb/s EDR InfiniBand using QSFP28 and SFP28

Parts are Described by Use in the Switch

The DAC, ACC, and AOC cable part number descriptors are based on the connector *used in the switches*, as the devices used in network adapters may consist of multiple connector types such as the single-port OSFP, QSFP112, QSFP56, etc. These plugs or “form-factors” are used to contain optical transceivers and copper cables to form the switching networks linking CPU/GPU compute engines with storage subsystems and to other system clusters in the network. For example, an 800G twin-port OSFP-to-4x 200G QSFP112 splitter copper cable is listed in the parts lists as 800Gb/s twin-port OSFP, and not as 200Gb/s QSFP112.

Cable Descriptions Denoted by the Switch-side Connector



NVIDIA Networking Overview

The NVIDIA LinkX line of cables and transceivers offers a wide array of products for configuring any network switching and adapter system. The NVIDIA LinkX cable and transceiver product line focuses exclusively on data center lengths of up to 2 kilometers for accelerated artificial intelligence computing systems. To minimize data retransmissions, the cables and transceivers are designed and tested to extremely low bit error ratios (BER) required for low-latency, high-bandwidth artificial intelligence and accelerated computing applications.

These are designed to work seamlessly with NVIDIA's entire networking product line:

- NVIDIA Quantum InfiniBand network switches
- NVIDIA Spectrum™ Ethernet network switches
- NVIDIA ConnectX® PCIe network adapters
- NVIDIA BlueField® PCIe DPUs

NVIDIA designs and builds complete subsystems based on the above switches, adapters, and interconnects. All these subsystems, including the switch, network adapter, DPU, CPU, GPU and transceiver integrated circuits, are designed, manufactured, and/or sourced, all from one supplier NVIDIA, and tested to work optimally in NVIDIA end-to-end complete configurations - specifically for artificial intelligence and accelerated applications.

NVIDIA Spectrum, Quantum, and LinkX Product Lines

NVIDIA's networking offering consists of the Quantum InfiniBand platform and the Spectrum Ethernet platform. Each of these consists of NVIDIA Quantum or Spectrum network switches, ConnectX PCIe-based network adapters, and BlueField PCIe-based Data Processing Units or DPUs. These are all interconnected using the LinkX cable and transceiver product line which consists of:

- Direct attached copper (DAC) cables
- Linear active copper cables (ACC)
- Multimode transceivers and fibers
- Active optical cables (AOCs)
- Single-mode transceivers and fibers.

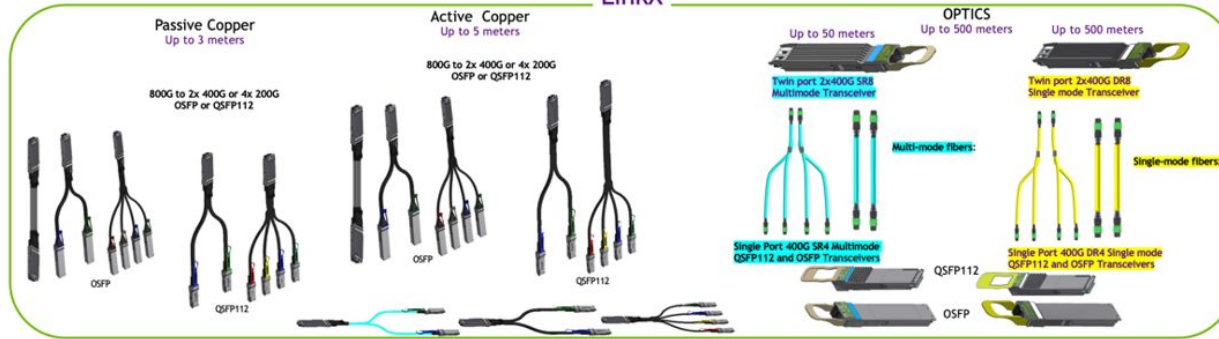
Cable and Transceiver Product Line for 100G-PAM4 and 50G-PAM4 for InfiniBand and Spectrum Ethernet

LINKX 400GBE/NDR PORTFOLIO

Based on 100G-PAM4 Modulation in OSFP and QSFP112



InfiniBand & Ethernet LinkX



InfiniBand & Ethernet

ConnectX-7

ConnectX-7/QSFP112 or OSFP
200G & 400G



BlueField-3

BlueField-3/QSFP112
200G & 400G



LINKX 400GBE ETHERNET PORTFOLIO

Based on 50G-PAM4 Modulation in QSFP-DD + QSFP56

Direct Attach Copper (DAC)



400GbE

400GbE-to-2x200GbE

400GbE-to-4x100GbE

Active Optical Cables (AOC)

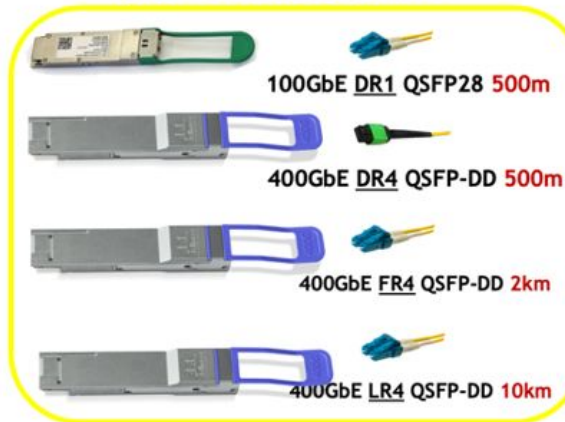


Optical Transceivers - Multi-mode



400GbE SR8 QSFP-DD 100m

Optical Transceivers - Single-mode



100GbE DR1 QSFP28 500m

400GbE DR4 QSFP-DD 500m

400GbE FR4 QSFP-DD 2km

400GbE LR4 QSFP-DD 10km

LINKX 200GBE/HDR PORTFOLIO

Based on 50G-PAM4 Modulation in QSFP56

200G QSFP56 -to- 200G QSFP56

InfiniBand + Ethernet

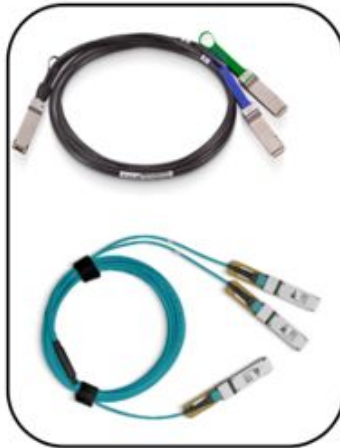


DACs

AOCs

200G QSFP56 -to- 2x 100G QSFP56

InfiniBand + Ethernet



200G QSFP56 -to- 4x50G SFP56

Ethernet only



QSA28
SFP28-to-QSFP28 port adapter

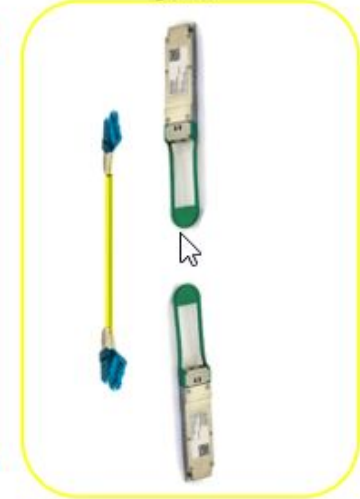
200GbE SR4

850nm Multi-mode 100-meters
QSFP56



200GbE FR4

1310nm Single-mode 2km
QSFP56



Key Technologies

This document examines key technologies used in constructing LinkX cables and transceivers for 100G-PAM4, 50G-PAM4, and 25G-NRZ -modulation based interconnects used to create 800G, 400G, 200G, 100G and 25Gb/s aggregate data rates. The following technologies are used in various combinations to create various cables and transceivers with different modulation rates, copper wires, connector shells, protocols, transceivers, optical connectors, and fibers.

- Modulation rates
- Protocol support for InfiniBand and Ethernet
- Connector cages and plugs
- Optical connectors
- Optical fibers
- Straight and splitter fiber crossover cables
- Optical patch panels

Note: Aggregate rates can be created by several different combinations of technologies, e.g.,

- 400Gb/s = 4x100G-PAM4 or 8x50G-PAM4
- 200Gb/s = 2x100G-PAM4, 4x50G-PAM4, 8x25G-NRZ
- 100Gb/s = 1x100G-PAM4, 2x50G-PAM4 or 4x25G-NRZ

Modulation Rates

High-speed digital signaling uses several types of voltage modulation. Varying electrical voltages create digital pulses that vary in voltage amplitude or intensity. Modern data centers typically use NRZ for slower speeds and PAM4 for higher speeds.

NRZ Modulation

Early years of digital 1,0 signaling, a digital zero was inserted between every data bit so the receiver clock could synchronize on the data signal. This was called “Return-to-Zero” modulation. Later, as electronics became faster, the inserted zero was eliminated and the pulse synchronized on the edges of the data signals. This became known as “Non-Return-to-Zero” or NRZ. NRZ became the industry standard for 1G, 10G, and 25Gb/s and used for 1G, 10G, 25G,

40G, 100G aggregate data rates. Continuing to 50G became a problem as the electrical wires turned into radio antennae's at high frequencies and caused the signal energy to be lost, which dramatically increased costs to contain the signals on circuit boards and wires.

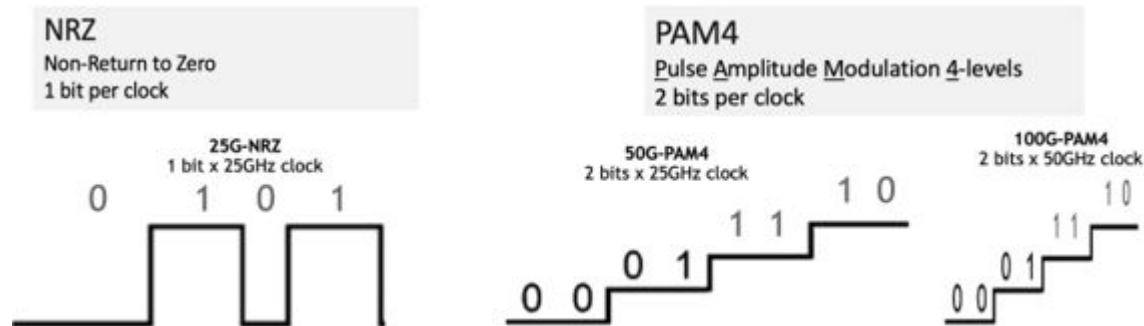
PAM4 Modulation

Industry standards groups created a new modulation scheme that sent two data signals with a single clock pulse by varying the voltage intensity levels to four levels instead of two with NRZ. The four levels created two data bits per clock pulse or {00, 01} and {10, 11} visualized as two 1,0 data bits stacked on top of each other. This became known as Pulse Amplitude Modulation to 4-levels or PAM4.

50G-PAM4 kept the 25GHz slower clock speed of NRZ with two data bits stacked which enabled maintaining low costs. Later, faster electronics enabled using 50GHz clock with two data bits for 100G-PAM4. Soon, the industry will have 100GHz clocks and two data bits for 200G-PAM4.

- 100G-PAM4 modulation is used for 400Gb/s NDR InfiniBand and Spectrum-4 400Gb/s Ethernet systems:
 - 800Gb/s = 8-channels 100G-PAM4
 - 400Gb/s = 4-channels 100G-PAM4
 - 200Gb/s = 2-channels 100G-PAM4
- 50G-PAM4 is used for 400G and 200G Spectrum-2/Spectrum-3 Ethernet and HDR InfiniBand:
 - 400Gb/s = 8x50G-PAM4 used with QSFP-DD devices for Spectrum-3 Ethernet only systems
 - 200Gb/s = 4x50G-PAM4 for 200GbE and HDR InfiniBand
- 25G-NRZ is used for 25G/100G Spectrum/Spectrum-2 Ethernet systems and 100Gb/s EDR InfiniBand:
 - 100Gb/s = 4x25G-NRZ
 - 50Gb/s = 2x25G-NRZ
 - 25Gb/s = 1x25G-NRZ

Four-level signal modulation for 100G-PAM4 and 50G-PAM4 vs. two levels for 25G-NRZ



Dual Protocol Capability

NVIDIA is the only provider of both InfiniBand and Ethernet networking. As the electrical and optical physics are the same for both protocols, NVIDIA combines the protocol support in the firmware coding of its network adapters, DPUs, cables and transceivers. The protocol is automatically enabled by the switch that the cables and transceivers are inserted into as the switches have specific Ethernet or InfiniBand protocols.

This feature is unique to NVIDIA and enables customers to better utilize adapter and interconnect set ups and for migrating systems and combining protocols based on different computing needs. One set of cables and transceivers and adapters can serve two protocols by simply changing the switches they are connected to. For example, DGX-H100 8 GPU systems may use InfiniBand for GPU-to-GPU networking and both InfiniBand and Ethernet for storage networking and other cluster communications links. The dual protocol capability is available with the 100G-PAM4 series, but fragments for the older 50G-PAM4 and 25G-NRZ devices as some devices are InfiniBand- or Ethernet-specific.

The 100G-PAM4 LinkX cables and transceivers, ConnectX-7 adapters, and BlueField-3 DPUs all support **both** InfiniBand and Ethernet protocols **in the same device** and use the **same part numbers**. The protocol of the network adapters and interconnects combination is determined when inserted into a Quantum-2 NDR InfiniBand or Spectrum-4 Ethernet switch.

However, this dual protocol capability is fragmented for 400GbE, 200GbE, HDR and 100GbE, EDR.

- 400GbE QSFP-DD switches based on 8x50G-PAM4 are Ethernet only, as InfiniBand does not use the QSFP-DD form-factor.
- 200Gb/s cables and transceivers based on 4x50G-PAM4 are mostly dual protocol with a few specific parts unique to Ethernet or InfiniBand, as InfiniBand requires a much lower bit error rating than Ethernet which increases testing costs.

- 100Gb/s cables and transceivers based on 4x25G-NRZ are a mix of combination and specific protocol parts. InfiniBand EDR is only 100Gb/s. SFP28 is not used for InfiniBand.

One dual-protocol LinkX cable, transceiver, ConnectX-7, and/or BlueField-3 DPU adapter assembly for various switch protocols

InfiniBand
NDR Quantum-2 QM9700
32-cage Twin-port-OSFP



Ethernet
Spectrum-4 400GbE SN5600
64-cage Twin-port-OSFP



Creates InfiniBand Networks

Creates Ethernet Networks

InfiniBand & Ethernet

LinkX



ConnectX-7

ConnectX-7/QSFP112 or OSFP
200G & 400G



BlueField-3

BlueField-3/QSFP112
200G & 400G



Connectors and Cages

Electronics, optics, and copper wires are housed in metal shell plugs called form-factor plugs. Plugs have a cage counterpart that is in the network switch front panels and on top of network adapters. The metal plugs have many code name extensions based on the single-channel; small-form-factor plug (SFP).

SFP can be preceded by Q for quad or 4-channels (QSFP) and for 8-channels, quad-double density (QSFP-DD), and octal (OSFP). NVIDIA created an 8-channel transceiver called the twin-port OSFP that has 8 electrical channels and two optical 4-channel ports. Numbers at the end indicate the maximum Gb/s speed rating of the connector e.g., 28, 56, 112 e.g., QSFP56, QSFP112 to contain the signal EMI noise. Also note that InfiniBand and Ethernet use slightly lower speed ratings than the connector maximum speed e.g. QSFP56 supports 50G rates, and QSFP112 uses 100G rates.

The LinkX 100G-PAM4 line uses three connectors and switch and adapter cage types:			
• Twin-port OSFP	800G	8-channels	Switches only - Quantum-2 InfiniBand and Spectrum-4 Ethernet
• Single-port OSFP	400G	4-channels	ConnectX-7/OSFP adapters only - InfiniBand and Ethernet
• Single-port QSFP112	400G	4-channels	ConnectX-7/QSFP112 and BlueField-3 DPUs - InfiniBand and Ethernet

The LinkX 50G-PAM4 line uses two connectors and switch and adapter cage types:			
• QSFP-DD	400G	8-channels	Spectrum-3 Ethernet switches only
• QSFP56	200G	4-channels	InfiniBand HDR, 200Gb/s Ethernet Spectrum-2 switches, ConnectX-6, BlueField-2 DPUs

The LinkX 25G-NRZ line uses two connectors and switch and adapter cage types:			
• QSFP28	100G	4-channels	Spectrum Ethernet and InfiniBand EDR, ConnectX-5
• SFP28	25G	1-channel	Spectrum Ethernet, ConnectX-5 (InfiniBand does not use SFP28)

100G-PAM4 Series: Twin-port OSFP and OSFP Plugs

The octal small form-factor plug, or OSFP, has become the preferred form-factor for high-speed applications such as artificial intelligence and HPC networking as it offers future expansion with more channels, more space for components, and higher power dissipation capabilities. The twin-port OSFP 800G plug has 8-channels of electrical signaling for the switch and two 400Gb/s engines inside the transceiver that exit to two 400G optical or copper ports. Extra cooling fins are used on top to support 17-Watt transceivers- hence the name “2x400G twin-port OSFP finned-top”.

❗ Twin-port OSFP “finned-top” cables and transceivers are **only** used in NVIDIA Quantum-2 NDR InfiniBand and Spectrum-4 SN5600 400GbE Ethernet systems.

The 800G twin-port OSFP is also offered in a flat top version and a 400Gb/s single-port, flat-top OSFP is offered. These are all the same size, but the twin-port OSFP finned top version is taller due to the heat sink.

The three OSFP versions are:

- 800G **finned-top, twin-port**, 8-channel, 2x400G OSFP for Quantum-2 and Spectrum-4 SN5600 Ethernet air-cooled switches.
- 800G **flat-top, twin-port**, 8-channel, 2x400G OSFP for linking DGX H100 Cedar7 GPU links which use internal cage riding, air-cooled, heat sinks for liquid-cooled systems. This has the same internals as the finned top version.
- 400G flat-top, single-port, 4-channel, OSFP for ConnectX-7/OSFP network adapters using cage riding heat sinks.

❗ Single-port, 400G OSFP or QSFP112 devices **cannot** be used in twin-port OSFP switch cages -- **only adapters and DPUs**.

800G twin-port OSFP finned-top connectors are also used to construct passive copper cables, active copper cables, and active optical cables using twin-port OSFP, OSFP, QSFP112, and QSFP56 connector ends. These parts are available in various combinations including 800G twin-port OSFP with flat tops instead of fins.

Lastly, a long reach single mode, 2km twin-port OSFP called 2xFR4 transceiver will be available at the end of 2023 that uses a finned-top but with a lid on top of the fins creating a closed channel for additional cooling.

Twin-port OSFP Open Finned Top



Closed Finned Top



QSFP112

QSFP112 form-factor is a single-port, 4-channel, 400G device specifically for ConnectX-7/QSFP112 adapters and BlueField-3/QSFP112 DPUs.

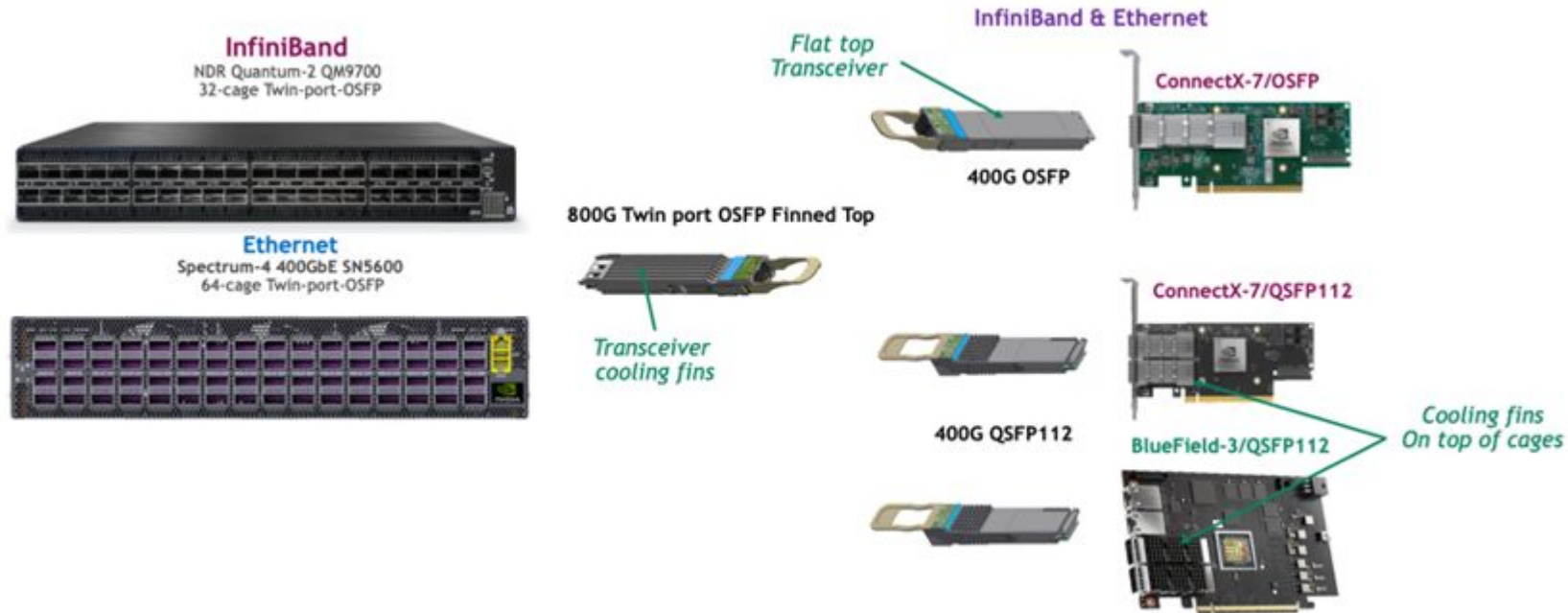
- QSFP112 has a flat top without cooling fins on top and uses the cooling fins located on the adapter and DPU connector cages.
- QSFP112 cannot be used in twin-port OSFP switch cages or single-port OSFP ConnectX-7 adapters.

⚠ ConnectX-7 is offered in **both** OSFP and QSFP112 versions. BlueField-3 DPUs **only** use the QSFP112.

Twin-port OSFP, OSFP, and QSFP112 connector plugs

			
<i>Twin-port 2x400G finned-top OSFP</i>	<i>Twin-port 2x400G flat-top OSFP</i>	<i>Single-port OSFP</i>	<i>Single-port QSFP112</i>

Twin-port OSFP for Switches, OSFP, and QSFP112 ConnectX-7, and BlueField-3 DPUs



50G-PAM4 series: QSFP-DD, QSFP56

50G-PAM4 modulation enables 400GbE with 8-channels and 200GbE and 200G HDR InfiniBand with 4-channels.

- **400G QSFP-DD** has two rows of 4-channel electrical lanes, hence the name Double Density. Based on 8-channels of 50G-PAM4, QSFP-DD is used only in Spectrum-3 SN4000 400G Ethernet and Spectrum-4 SN5400 switches.
- **200G QSFP56** is used for 4-channels of 50G-PAM4 for HDR InfiniBand and 200GbE Spectrum-2 Ethernet networking.

25G-NRZ series: QSFP28, QSFP+

The NVIDIA Spectrum SN2000 series switches and ConnectX-5 adapters use the QSFP28 plug and some models SFP28. These parts date back to 2015 but are still in widespread use.



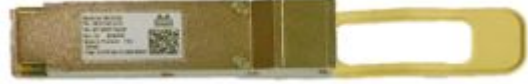


- **100G QSFP28** is used for 4-channels of 25G-NRZ for EDR InfiniBand and Ethernet 100GbE networking.
- **40G QSFP+** is used for 4-channels of 10G-NRZ for QDR InfiniBand and Ethernet 40GbE networking.
- **25G SFP28** is used for 1-channel of 25G-NRZ for Ethernet (not used for InfiniBand).
- **10G SFP+** is used for 1-channel of 10G-NRZ for Ethernet (not used for InfiniBand).

Backwards Compatibility

The QSFP evolved over many years and generations and offers backwards compatibility of newer cages with older devices. QSFP-DD plug is the same height and width as the 4-channel QSFP56 and QSFP28 used for 200G and 100G systems, but longer to support two rows of 4-channel electrical pins creating 8-channels. This enables the QSFP-DD cages in switches to support backwards compatibility to 4-channel QSFP56, QSFP28, and QSFP+ devices, which engage only the first row of 4 pins in the 8-channel QSFP-DD cage supporting 4-channels.

- QSFP-DD is not used for InfiniBand systems as of 2023 and not for 100G-PAM4 Ethernet systems.
- QSFP-DD cages in switches can accept QSFP112, QSFP56, and QSFP28 devices using 50G-PAM4 or 25G-NRZ modulation.
- QSFP112 cages in adapters and DPUs can accept QSFP56, QSFP28, and QSFP+ cables and transceivers.
- QSFP56 cages devices in switches and adapters accept QSFP28, QSFP+ cables and transceivers.
- ConnectX-6/7 and BlueField-2/3 are not offered using QSFP-DD cages -- only for switches.
- NVIDIA does not offer a QSA port adapter to convert QSFP112 devices to be inserted into twin-port OSFP cages.
- NVIDIA does offer a QSA+ and QSA28 port adapter enabling SFP devices to be inserted into QSFP cages and passing through one channel. QSA56 is not offered.

QSFP-DD, QSFPxx, and SFPxx connector plugs

QSFP-DD	
QSFP112, QSFP56	
QSFP28, QSFP+	
SFP28, SFP+	
QSA28, QSA+	

DACs, ACCs, AOCs, and Transceiver Interconnects

There are two main ways to link switches and adapters by using either copper wires or optics. Copper has a length or reach limitation of less than 5 meters and two different optical technologies enable using different technologies for the least cost to fit the application. Multimode short reach optics typically have 50m-100m maximum reaches. Single-mode, mid to long reach optics are typically 100m, 500m 2km, 10km, and 40km maximum reaches.

There are many different technology combinations of optical connector, plugs, optical connectors, electronics, and optics. This document concentrates on high-volume products offered by specifically NVIDIA for accelerated AI data center-oriented cables and transceivers.

Direct Attach Copper cables (DACs) consist of a connector plug (QSFP or OSFP) and basically copper wires and shielding. DAC wires tend to radiate high-speed electrical signals, like radio antennas, hence are limited to 2, 3, 5-meter lengths depending on the speed. DAC cables are very popular due to their low cost, almost no power consumption and latency delay. DAC cables are complete assemblies and cannot separate into plugs and wires.

Active Copper Cables (ACCs) are DAC copper cables but include a signal booster IC in the end to extend the length to 3, 4, and 5-meters depending on the speed. ACC are very popular due to their lower cost than optics, very low power consumption, and low latency delay. ACC cables are complete assemblies and cannot separate into plugs and wires. NVIDIA’s 100G-PAM4 ACCs, aka linear ACCs, use a pre-emphasis IC that offers very low latency, only 1.5-Watts, and can reach lengths of 5-meters.

Multimode Transceivers use a large, 50-um light carrying core in the optical fiber. A digital pulse consists of many individual photons that travel down the fiber in different paths or “modes” bouncing off the fiber walls. As the fiber length increases, the different photon paths arrive at the receiver at different time and distort the signal pulse at the receiver receiver limiting the maximum reach.

- Multimode maximum length is 100-meters for 50G-PAM4 and 25G-NRZ but reduces to 50-meters for 100G-PAM4. Multimode transceivers and AOCs use 850nm vertical cavity surface emitting lasers (VCSELs) and due to the ease of aligning fibers with lasers and detectors are significantly less expensive than single-mode transceivers.
- Multimode transceivers are described as Short Reach or SR, SR4, SR8, etc.
- Parallel optical connectors use 8 or 16-fiber Multiple-Push-On (MPO) connector or 2-fiber Lucent Connector (LC). UPC is an Ultra flat polish and APC is an Angled polish.

Name	Description	Reach	Optical Connector	Speeds	Modulation
SR	Short Reach 1-channel	100m	LC	1, 10, 25Gb/s	25G-NRZ
SR4	Short Reach 4-channel	50-100m	MPO-12/UPC	100, 200Gb/s	25G-NRZ, 50G-PAM4
SR8	Short Reach 8-channel	100m	MPO-16/APC	400Gb/s	50G-PAM4
2xSR4	Short Reach 2x 8-channel	50m	2x MPO-12/APC	2x 400Gb/s	100G-PAM4

Active Optical Cables (AOCs) consist of two multimode optical transceivers with the optical fibers bonded inside and not removable. AOCs offer lower costs than two transceivers and separate fibers as only electrical testing is required in manufacturing. AOCs are offered up to 100 meters and are typically used in configurations with easy cabling access. AOC cables are complete assemblies and cannot separate into plugs and fibers. AOCs are very popular at 100G, 200G, and 400GbE speeds.

- AOCs based on 100G-PAM4 are not offered with 800G twin-port OSFPs, due to the large OSFP connector size and possibility of breaking the fiber during installation. A 4x 200G AOC splitter would have five, large OSFP connectors and ne very difficult to install in crowded infrastructures.
- OSFP AOCs are offered for backwards compatibly links of 2x200G-to-2x200GbE and 2xHDR and 2xHDR100.

Single-mode Transceivers use a tiny 9-um light carrying core in the optical fiber, which is difficult to align lasers and detectors in manufacturing, hence are more expensive than multimode transceivers. An individual 1310nm light pulse travels down the fiber in a single path or single mode and can travel great distances without distorting the pulse. Single-mode optics used in data centers is offered up to 40km.

For 100G using 4x25G-NRZ single mode transceivers, the optics was defined by several individual industry groups and the IEEE.

Name	Description	Max. Reach	Optical Connector	Speeds
PSM4	Parallel Single-Mode 4-channel	500m	MPO-12/APC	100Gb/s
CWDM	Coarse Wavelength Division Multiplexed 4-channel	2km	LC	100Gb/s
LR4	Long Reach 4-channel, multiplexed	10km	LC	100Gb/s
LR	Long Reach 1-channel, multiplexed	10km	LC	25Gb/s

For 50G-PAM4 and 100G-PAM4, the IEEE group standardized on different naming terminologies that everyone agreed to:

- PSM4 became Datacenter Reach 4-channel (DR4) for 500m
- CWDM4 became Far Reach 4-channel (FR4) for 2km

Name	Description	Reach	Optical Connector	Speeds	Modulation
DR1	Data center Reach, 1-channel	500m	LC	100Gb/s	25G-NRZ, 100G-PAM4
DR4	Data center Reach, 4-channel	500m	MPO-12/APC	200G, 400Gb/s	50G-PAM4, 100G-PAM4
2xDR4	Data center Reach, 2x 4-channel	500m	2x MPO-12/APC	800Gb/s	100G-PAM4
FR4	Far Reach, 4-channel	2km	LC	200G, 400Gb/s	50G-PAM4
2xFR4	Far Reach, 2x 4-channel	2km	2x LC	800Gb/s	100G-PAM4
LR4	Long Reach 4-channel	10km	LC	200, 400Gb/s	50G-PAM4
ER	Extended Reach (WDM4)	40km	LC	100Gb/s	25G-NRZ

As a result, the naming of an optical transceiver consists of many descriptors such as:
“400G DR8 single-mode 500m, 8-channel electrical, 4-channel multiplexed optical, MPO-12/APC optical connector”.

Optical Connectors

Several optical connector types are used to link optical fibers together which are inserted into transceiver ends. A fiber cable consists of multiple optical fibers and two optical connectors for a straight cable and up to 5 optical connectors for 4x splitter cables.

Optical connector types typical in high-speed data centers are:

<ul style="list-style-type: none">• MPO-12/APC• LC duplex• MPO-12/UPC• MPO-16/APC	8-fiber 2-fiber 8-fiber 16-fiber	50G-PAM4, 100G-PAM4 25G-NRZ, 50G-PAM4 25G-NRZ, 50G-PAM4 50G-PAM4
----------------------------------------------------------------------------------------------------------------------------	-------------------------------------------	---------------------------------------------------------------------------

MPO-12 Optical Connectors

The Multiple-Push-On (MPO) optical connector is a ceramic block with holes that contain the ends of multiple optical fibers epoxied in as either single-mode or multimode types. The ceramic blocks are made with different numbers of holes 8, 12, 16, 24, etc. but data centers typically use the 8-fiber but labeled MPO-12 with 4 unused, or 16-fiber MPO-16 versions.

Some of the light sent into a fiber reflects backwards from the fiber end face. Slower speed electronics and optics are **less sensitive** to back reflection created inside the optical fiber. Hence, 25G-NRZ (4x25G-NRZ SR4) and some 50G-PAM4 transceivers (4x50G-PAM4 SR4) use the MPO-12/UPC or **Ultra-flat Polished Connector**.

8x 50G-PAM4 400G and all 100G-PAM4 transceivers are **more sensitive** to back reflections and use the MPO-12/APC or **Angled Polished Connector**. This has an 8-degree polish on the end that causes the back reflections to be diverted into the fiber side cladding and away from the transmitter.

MPO/UPC (flat) cannot be mated together with MPO/APC (angled) connectors.

LC Duplex

The 2-fiber Lucent Connector or LC duplex is typically used for a single-channel links using a transmit fiber and a receive fiber. This applies to multimode and single mode optics.

Running parallel fibers over long reaches becomes expensive so transceivers use multiple lasers with different wavelengths (colors) which are combined or “multiplexed” into a single fiber for transmission and filtered and separated back out at the receiver. One fiber for transmission and one for receiving but carrying 4 or 8 channels.

The LC duplex optical connector is used for single-channel multimode SR transceivers and using single-mode optics, the 1-channel, LR transceiver, and multiplexed transceivers such as 100G DR1, 200G FR4, 400G FR4, 2xFR4, LR4 and ER transceivers. These are used to link InfiniBand and Spectrum Ethernet switches together and linking clusters across campuses by using only two fibers in each optical connector to save on fiber costs over long runs

The 800G twin-port OSFP 2xDR4 transceivers use two MPO-12/APC optical connectors. The 2xFR4 transceiver uses two LC optical connectors.

MPO-12/APC, MPO-12/UPC and duplex LC optical connectors

<i>MPO-12/APC</i>	<i>MPO-12/UPC</i>	<i>MPO-12/UPC</i>	<i>Duplex LC</i>
			

Single-mode and Multimode Optical Fibers

NVIDIA sells optical fiber cables based on 4-channels of 100G-PAM4 *only*. These fibers are specific to the 100G-PAM4 NVIDIA offering for NDR InfiniBand and Spectrum Ethernet. The fibers are crossover Type-B fibers that enable directly linking two transceivers together. NVIDIA offers these and 1:2 splitter fiber cables from 1m to 100m in straight single mode and 1m to 50m for multimode and single mode splitters.

- Not offered by NVIDIA are cables using 2-fiber LC, MPO-12/UPC, MPO-16/APC and other fiber cables and splitter cables. NVIDIA recommends sourcing these from numerous third-party suppliers as they are a commodity.

Optical fiber consists of a strand of glass with the outside coating glass a higher density than the inside. This confines the light to travel down the length of the fiber and bending when encountering the higher density glass coating.

There are two types of fibers: single-mode and multimode.

Single-mode Fiber

Single-mode fiber has a tiny 9-um light carrying core. This is small enough to make the light bends inside at a very shallow angle which keeps the data pulse light photon packets together as a group or “single mode” which can travel over great distances.

- Typically used for long reaches 50-meter to 40km but can be used for shorter 1m lengths.
- Single-mode fiber for data centers is optimized to be most transparent at 1310nm wavelength.
- The fiber cable jacket is usually colored yellow, and the transceiver pull tabs yellow.

Multimode Fiber

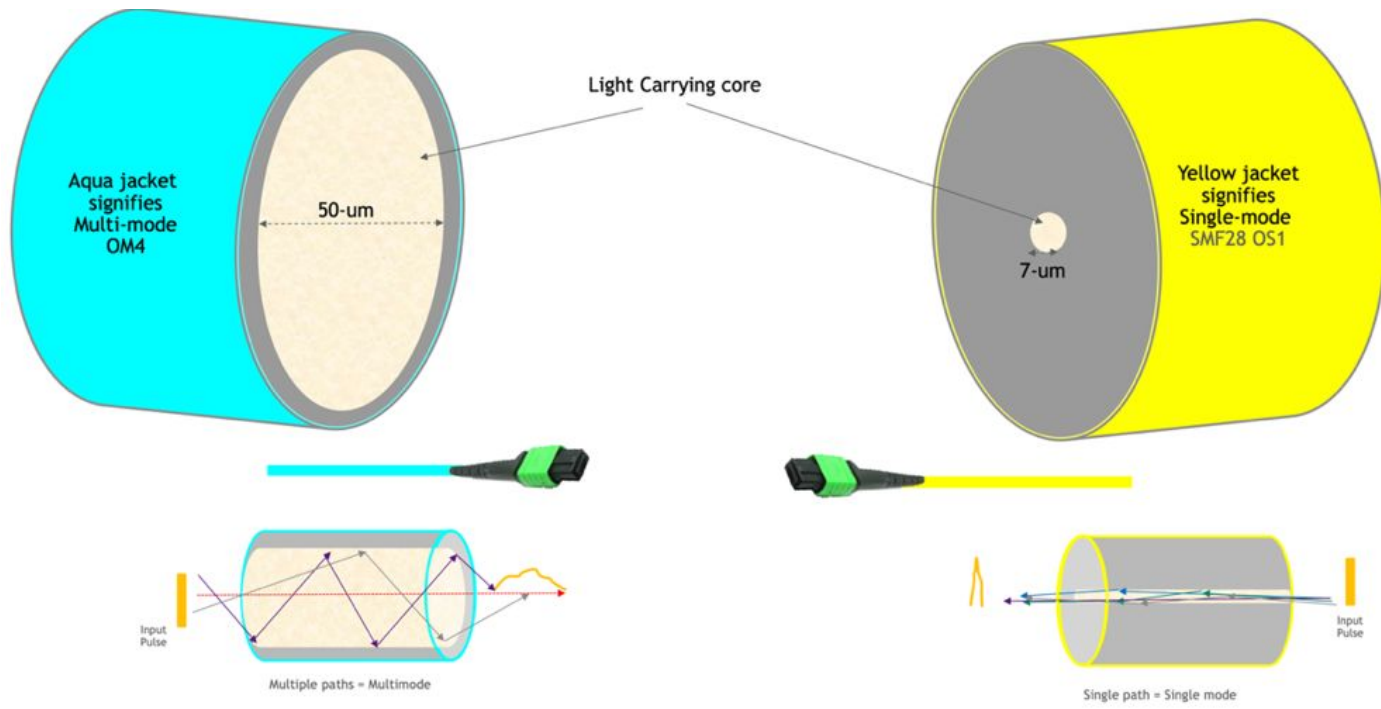
Multimode fiber has a larger 50-um light carrying core and some of the photons in a single data pulse take steeper angles traveling down the fiber. This causes the light pulse to scatter inside the fiber into multiple paths or “multi(ple)modes” with some parts of the pulse taking longer to arrive at the end at the photodetector. This results in the pulse intensity weakening and spreading out in time enough to collide with the next following data pulse which limits the maximum reach.

- Multimode fiber is optimized to be most transparent at 850-nm wavelength.
- Multimode fibers and transceivers based on 850nm cannot operate with 1310nm single-mode fibers and transceivers.
- The fiber cable jacket is usually colored aqua, and transceiver pull tabs tan.

- Fiber type is optimized for specific modulation speeds:
 - For 25G-NRZ and 50G-PAM4: OM4 fiber type can reach up to 100-meters and OM3 for up to 70-meters.
 - For 100G-PAM4: OM4 fiber type can reach up to 50-meters and OM3 for up to 30-meters.
 - NVIDIA supplies fiber cables for the 100G-PAM4 line only up to 50-meters multimode and 100-meters single mode for straight fibers, and up to 50-meters for both single-mode and multimode 1:2 splitter fiber cables.
 - NVIDIA supplied fiber cables are MPO-12/APC only and use a green connector plastic shell for both single-mode and multimode optics.

The large diameter of multimode fiber is easy and inexpensive to connect lasers and photodetectors to and less expensive in building transceivers, but it is limited at 100G-PAM4 speeds to 50 meters. Single-mode fiber is more expensive to interface to but can reach beyond 2 kilometers - spanning entire data centers. Multimode optics are the most used optics as most data center interconnects are less than 50 meters.

Multimode and Single-mode fibers



Multimode fiber

Single mode fiber

Crossover Fiber Cables

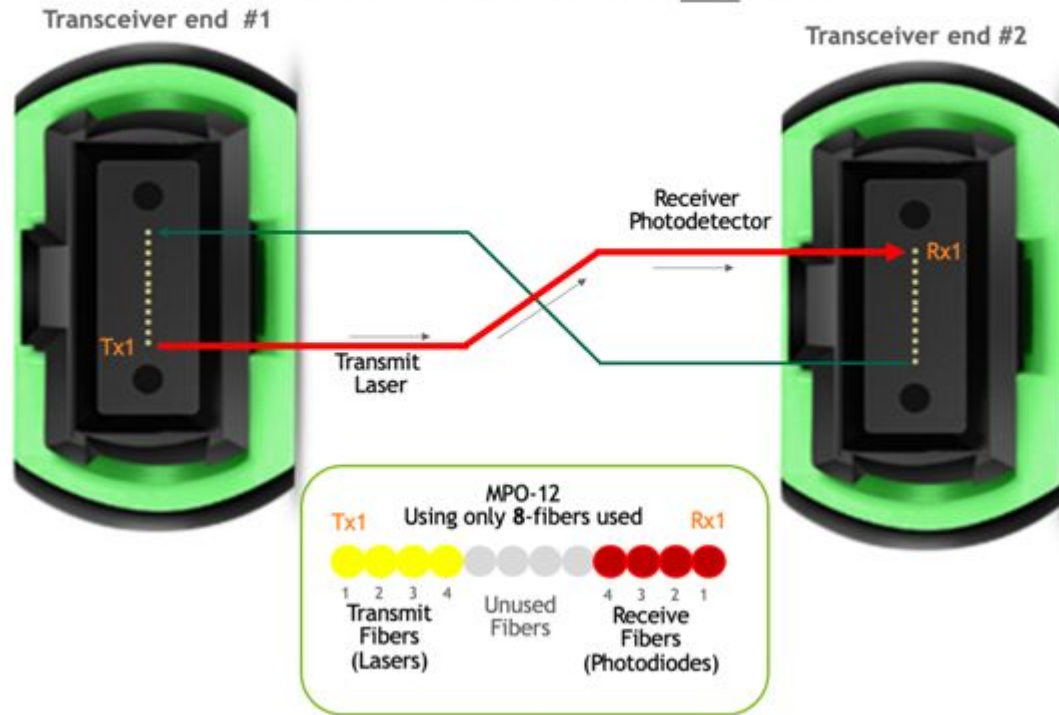
Crossover, Type-B single-mode and multimode fiber cables are offered by NVIDIA for the 100G-PAM4 series only. This enables directly connecting transceivers together and aligning transmit lasers with receiver photodetectors by crossing over the fibers' pin arrangement inside the cable with both optical connectors. E.g., pin 1 in transmit laser side is ***crossed over*** to pin 12 in the receive transceiver's photodetector side optical connector. Type-A fiber cables have parallel fibers and are used for trunk cables between racks and optical patch panels.

- Not offered: non-crossover Type-A parallel straight or splitter fiber cables, trunk cables, and crossover fibers longer than 100 meters.
- Not offered: other fibers such as MPO-12/UPC or MPO-16/APC or LC duplex.

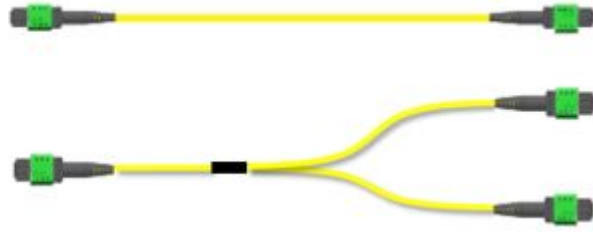
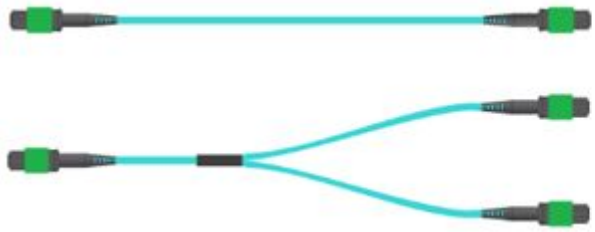
Crossover fiber cable connector ends

Type-B "Crossover" cable

Transmitter lasers 1,2,3,4 have to align with receiver photodetectors 1,2,3,4
Transmit and receive fibers are crossed inside the cable



Multimode (aqua) and Single-mode (yellow) fiber straight and 1:2 fiber splitter cables



Acronyms and Abbreviations

The following terms, abbreviations, and acronyms are used in this document.

Term	Description
ACC	active copper cables
AOC	active optical cables
APC	Angled Polished Connector
BER	bit error ratio
ConnectX®	NVIDIA network adapter product family for InfiniBand and Ethernet
DAC	direct attached copper cable
DPU	data processing unit, e.g. NVIDIA BlueField® products
finned top	extra cooling fins on top of the form-factor plug
flat top	as opposed to finned top
LinkX	NVIDIA's cable and transceivers product line
MPO	Multiple-Push-On
NRZ	Non-Return-to-Zero
NVIDIA Quantum	NVIDIA InfiniBand switch product line
NVIDIA Spectrum™	NVIDIA Ethernet switch product line

Term	Description
OSFP	octal small form-factor plug
PAM4	Pulse Amplitude Modulation to 4-levels
UPC	Ultra-flat Polished Connector

Document Revision History

Version	Date	Description
1.0	August 2023	Initial release

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. Neither NVIDIA Corporation nor any of its direct or indirect subsidiaries and affiliates (collectively: "NVIDIA") make any 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 products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

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

Copyright

© 2023 NVIDIA Corporation & affiliates. All Rights Reserved.

