

Guide to Myri-10G PHYs

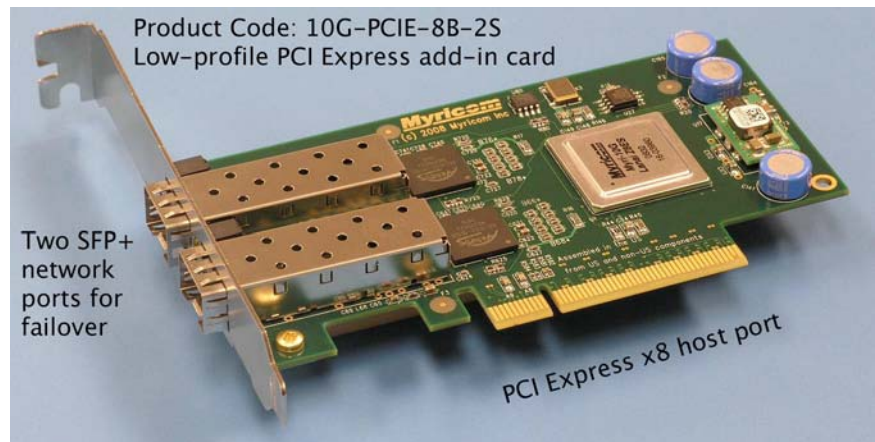
Myricom, Inc.
7 October 2008

Myri-10G is standards-based 10-Gigabit Ethernet networking, and more. At the Physical level (PHY, layer 1), the ports of Myri-10G components conform to IEEE 802.3 10Gb Ethernet signaling specifications (see the Appendix). At the Data Link level (layer 2), the network protocols may be either 10Gb Ethernet or 10Gb Myrinet.

Some Myri-10G ports use connectors and cables specified in the 10Gb Ethernet standards, but it is common practice with 10Gb Ethernet for network ports to employ other connectors or sockets for pluggable modules. Thus, Myri-10G product codes include the symbols shown below for the port connectors or sockets rather than for the PHYs¹:

Product-code Symbol	Port Connector	Example PHY(s)
QP	QSFP transceiver socket	XAUI over ribbon fiber
I	IBM BladeCenter H	XAUI
C	CX4	10GBase-CX4
CJ	Jackscrew CX4	10GBase-CX4
M	DensiShield (mini-CX4)	10GBase-CX4
R	XFP transceiver socket	10GBase-SR, 10GBase-LR
S	SFP+ transceiver socket	10GBase-SR, 10GBase-LR

For example, a 10G-PCIE-8B-2S Network Interface Card (NIC) has a two sockets for SFP+ optical-fiber transceivers. The sockets support the 10Gb Ethernet 10.3125 GBaud 64b/66b-encoded “SFI” signals, electrical power, and the standard management interface for the SFP+ transceiver. Depending upon the type of SFP+ transceivers used, the 10Gb Ethernet PHY for the port may be 10GBase-SR (850nm wavelength, short range), 10GBase-LR (1310nm wavelength, long range), or others. In



addition, these SFP+ sockets can be used with SFP+-terminated twinax copper cables up to 5m, available from multiple sources, or with SFP+-terminated Electrical-Optical-Electrical (EOE) cables such as Finisar Laserwire cables.

¹ Please see the Appendix for definitions of the acronyms.

Depending on the software and firmware used with a Myri-10G NIC, its network port may communicate using either Ethernet or Myrinet protocols at the Data Link level. However, the layer-2 protocol for a Myri-10G switch port is fixed, either Ethernet or Myrinet. For example, a 10G-SW32LC-16ES line card has 16 SFP+ sockets, indicated by the “S,” where the “E” in this part of the product code indicates that these ports communicate using Ethernet protocols.

Preferred Choices of PHYs

10GBase-SR and 10GBase-LR: Myri-10G “S” ports with SFP+ transceivers are the best choice today for 10GBase-SR and 10GBase-LR 10Gb Ethernet applications found commonly in data centers, campus networks, storage, telecom, and many other uses. SFP+ transceivers, whose production has been ramping up recently, are lower cost and lower power than the XFP transceivers used in Myri-10G “R” ports. Myricom will continue to supply XFP transceivers and components with “R” ports, but we expect these XFP components to be displaced over time by the more modern SFP+ components.

Myri-10G components with “R” or “S” ports for serial fiber have slightly higher latency, power, and cost than components with “QP,” “I,” or CX4 ports, so are rarely used in HPC applications.

HPC Clusters: Until recently, CX4 ports were preferred for Myrinet-protocol communication in HPC clusters due to their low cost and low latency. However, Myri-10G “QP” ports and QSFP-terminated Electrical-Optical-Electrical (EOE) cables have now become the best choice of PHY and cabling for HPC clusters and for switch-to-switch links in large switch networks. CX4 cables are bulky, and limited in length to 15m. Standard lengths of QSFP-terminated EOE cables are up to 25m, with lengths up to 200m available on special order. The QSFP transceivers in the cable ends exhibit almost zero latency, so that the only latency is in the fiber itself, ~5ns/m.



Although these EOE cables have a higher list price than CX4 cables at equal cable lengths, the QP ports of Myri-10G components are priced correspondingly less than 10GBase-CX4 “C,” “CJ,” or “M” ports on Myri-10G NICs and switches. Thus, one can today enjoy the benefits of fiber interconnect – thin, lightweight, flexible cables – at the same or lower cost than that of bulky CX4 cables.

Not only are EOE cables much more economical than QSFP fiber transceivers in QP ports with quad-ribbon-fiber cables plugged into the transceivers, there is no possible problem with dust particles in the optical-fiber cable ends or with laser safety. Thus, Myricom no longer offers QSFP transceivers for sale.

CX4 ports and cables: Although we expect the use of CX4 cables to be displaced with newer technology such as EOE cables, there are still many applications for 10GBase-CX ports and cables, such as to connect to existing 10Gb Ethernet switches that have 10GBase-CX4 ports.

Industry-standard components: Some 10Gb Ethernet vendors (you may know who they are) disable transceiver ports in which the transceiver fails to present the vendor's own ID, thus requiring the use of transceivers that they sell. Myri-10G components do **not** require the use of transceivers, EOE cables, or other industry-standard components purchased from Myricom. Of course, Myricom recommends and warrants the optical-fiber transceivers, EOE cables, fiber cables, and copper cables that we sell, inasmuch as we have qualified these products in rigorous testing and have long and favorable reliability experience with our current suppliers. Myricom can offer these products at attractive prices because of the volume in which we purchase them and because our profit margins on these "pass-through" items are intentionally small. You are free to purchase any of these industry-standard components from the vendor of your choice, but if you encounter interoperability problems because those third-party components do not conform to 10Gb Ethernet specs, Myricom cannot be responsible.

Appendix: Myri-10G PHYs

Here is a technical summary of the standards, characteristics, and applications of the 10Gb Ethernet PHYs supported in Myri-10G components.

XAUI (Product-code symbols: QP, I)

Specification: IEEE 802.3ae

Signaling: 4 lanes of 3.125 GBaud signals, 8b/10b encoded, resulting in a data rate of 2.5 Gb/s per lane and an aggregate data rate of 10Gb/s. XAUI includes built-in mechanisms for physical-level link synchronization, large lane-to-lane skew tolerance, and clock-skew tolerance.

Comments: XAUI, the "X Attachment Unit Interface," is the *de facto* standard for 10Gb/s communication on circuit boards and backplanes, and is the common-denominator PHY for 10Gb Ethernet. There are many merchant-silicon components that translate between XAUI and other 10GbE PHYs.

Implementation: XAUI is a native PHY of Myricom's custom-VLSI Myri-10G chips. Myricom's implementation includes ~3dB pre-emphasis to improve operating margins on circuit boards, on switch backplanes, and in blades.

Applications: Both 10GbE and HPC solutions. On circuit boards, backplanes, and through Quad Small Form factor Pluggable (QSFP) transceivers or QSFP-terminated Electrical-Optical-Electrical (EOE) cables, XAUI exhibits minimal cost and latency, and is thus the preferred PHY for 10Gb Myrinet HPC applications as well as 10Gb Ethernet and 10Gb Myrinet in blade and switch backplanes and midplanes.

Length limitations: ~2m on circuit boards; up to 200m over ribbon fiber or EOE cables.

10GBase-CX4 (Product-code symbols: C, CJ, M)

Specification: IEEE 802.3ak

Implementation: These ports include a low-latency XAUI \leftrightarrow 10GBase-CX4 conversion chip that performs dynamic equalization on the RX signals, as required by the 10GBase-CX4 specifications².

Applications: Both 10Gb Ethernet and 10Gb Myrinet HPC solutions.

Length limitations: The cable length limit per IEEE 802.3ak is 15m.

10GBase-SR, 10GBase-LR (Product-code symbols: R, S)

Specification: IEEE 802.3ae

Implementation: XAUI \leftrightarrow XFI or SFI conversion to an XFP or SFP+ pluggable fiber transceiver. XFI = X (10) Fiber Interface, and SFI is an electrical variant of XFI with signal limiting for SFP+ transceivers. The signaling at the transceiver is a 10.3125 GBaud 64b/66b-encoded serial stream. The conversion adds ~430ns latency TX+RX. Thus, a cluster using serial fiber between hosts and switches would see ~860ns larger latency than an equivalent cluster that uses QSFP-terminated EOE cables.

Applications: Primarily 10GbE solutions due to cost, power, and latency.

Length limitations: XFP and SFP+ transceivers sold separately.

- 10GBase-SR, 850nm wavelength, multimode fiber to 26–300m depending on the transceiver and fiber.
- 10GBase-LR, 1310nm wavelength, single-mode fiber to 10km.
- 10GBase-ER, 1550nm wavelength, single-mode fiber to 40km (these transceivers are not supplied by Myricom).

Comments: SFP+ has cost and power advantages over the older XFP transceivers.

10GBase-T (not currently offered by Myricom)

Specification: IEEE 802.3an

Implementation: Merchant-silicon chips to convert between XAUI and 10GBase-T are starting to appear, but operate at relatively high power levels, and do not yet appear to have any advantages over the PHYs listed above. However, Myricom is able to produce components with 10GBase-T ports when and if sufficient demand appears.

Applications: Primarily 10GbE solutions due to cost, power, and latency.

Length limitations: 30-100m.

² Note that 10GBase-CX4 cables are not the same as InfiniBand cables. 802.3ak specifies dynamic equalization for RX signals at the port, whereas InfiniBand uses a passive equalization network inside the cable end.