

10 Gbits/sec and beyond: High speed in the data center

The need for higher speed Ethernet, InfiniBand, and Fibre Channel all have their own cable and connectivity requirements.

is growing everywhere in the data center, but at different rates for different applications. Even as adoption of 10GBase-compliant fiber and copper solutions grow, some user groups, such as video-content providers, financial institutions, and consumer-broadband providers, are already clamoring for data rates higher than 10 Gbits/sec.

Recognizing this, the Institute of Electrical and Electronics Engineers' (IEEE; www.ieee.org) High Speed Study Group (HSSG) recently agreed to develop a standard, 802.3ba, that covers both 40- and 100-Gbit/sec Ethernet speeds. The work of the IEEE Task Force starts in earnest this month, and the standard is expected to be approved in 2010. But it is relevant now to ask what types of cabling and connectors will be expected to be used at these speeds, as well as how best to maximize the life-cycle of data center cabling as silicon and transceivers evolve to support speeds beyond 10 Gbits/sec.

Existing Ethernet specifications

The typical life of a data center can reach 10 to 15 years, and with regular maintenance, the facilities infrastructure and cabling plant are both expected to last as long. Active equipment refreshes commonly occur every 3 to 5 years, so cabling systems are expected to support multiple generations of information technology (IT) equipment.

Most data centers are a mixture of 1-Gbit/sec and 10-Gbit/sec structured cabling systems, with 1 Gbit/sec the prevalent data rate for most data link connections and 2 or 4 Gbits/sec for most storage link connections. It is generally predicted, however, that a significant number of data center links will need to carry 10 Gbits/sec or greater within the next decade. To address this need for increased bandwidth, the IEEE created standards to deliver 10-Gigabit Ethernet performance over fiber

(10GBase-SR/LRM/LX4) and copper (10GBase-T).

10GBase structured cabling systems are backward compatible with 1-Gbit/sec active equipment, and can handle rapid levels of growth in network traffic. So, it is important to understand how choice of cabling media,



The modularity of preterminated fiber and copper cabling also enables high densities per unit of rack space, which promotes efficient use of valuable data center real estate and makes them suited for deployment throughout the data center.

connectors, and electronics combine to support present as well as future data rates.

Backbone cabling in the data center consists of longer permanent links that connect the main distribution area (MDA) with the entrance room, the telecommunications room, and to horizontal distribution areas (HDAs) in the data center. The MDA includes the main crossconnect (MC), which is the central patching point for distribution of the structured cabling system.

The reach and bandwidth needed to cable the MDA

ROBERT ELLIOT and **ROBERT REID** are product development managers at Panduit Corp. (www.panduit.com).

Reprinted with revisions to format, from the January 2008 edition of **CABLING INSTALLATION & MAINTENANCE**
Copyright 2008 by PennWell Corporation

typically requires fiber media to carry aggregated signal loads at high data rates (commonly, 10 Gbits/sec). The most economical 10-Gbit/sec fiber network channels are 10GBase-SR channels that deploy laser-optimized 50/125-μm fiber (ISO/IEC designated as OM3) with serial transceiver electronics.

OM3 multimode fiber is the first media option to consider because it is specified by the physical medium dependent (PMD) as being able to support 10 Gbits/sec up to 300 meters (the standard recognizes that other multimode cabling systems may support that rate over varying distances). OM3+ (extended reach OM3) is the other primary fiber choice, as both OM3 and OM3+ media can support very long reach channels throughout the network and can support future data rates beyond 10 Gbits/sec.

The preferred transceiver for 10GBase-SR channels is the short-wavelength (850-nm) VCSEL-based serial type. These low-cost modular electronics are optimized and standardized for use with OM3 fiber up to 300 meters, and are also compatible with OM2 fiber grades. Fiber for these devices is optimized for the 850-nm wavelength window, but can also support transmission in the 1310-nm window. By default, the use of serial transceivers requires the use of LC and SC connectors as the means of attachment to the fiber media.

The performance of fiber structured cabling channels depends on the ability of each channel to meet bit error rate (BER) performance requirements at minimum receiver power as specified in 802.3ae, and insertion loss of individual components within channel insertion loss (CIL) requirements. CIL is the sum of all signal attenuation losses incurred along a fiber-cable channel by both media and connectivity. The challenge is to deploy the best combination of fiber media and connectivity to meet both your CIL budget and your cabling plant design.

CIL budgets for 10-Gbit/sec data rates are strict compared to historical, lower-speed Ethernet variants. For example, the CIL budget for a 300-meter 10GBase-SR channel is 2.6 dB, with 1.1 dB allocated for the fiber media itself. Any connectors placed in the fiber path will each consume a maximum of 0.5 to 1.0 dB of the total insertion loss budget, depending on the quality of the connector system.

The good news is that links in a data center environment are typically much shorter than the full extent of the channel specified in the standard (i.e., >90% of links will be less than 50 meters), and the insertion loss conserved over shorter channels and lower power penalties (associated with fiber modal dispersion) can be returned to the overall power budget. This con-

Ethernet 10GBase physical layer specifications					
Type	PMD	Technology	Connector	Media	Reach (m)
Copper	10GBase-CX4	4 lanes	CX4	Twinax	15
	10GBase-T	4 pairs	RJ-45	Category 6 UTP	37
				Category 6 STP, Category 6A UTP/STP	100
Fiber	10GBase-SR/W	850 nm VCSEL, serial	Dual LC and SC	OM1/OM2/OM3 MMF	33/82/300
	10GBase-LRM	1310 nm LD, serial		OM1/OM2/OM3 MMF	220/220/300
	10GBase-LX4	1310 nm LD, WDM		OM1/OM2/OM3 MMF	300
				OS1 and OS2 SMF	10,000
	10GBase-LR/W	1310 nm LD, serial		OS1 and OS2 SMF	10,000
	10GBase-ER/W	1550 nm LD, serial			40,000
	Note: OM1/OM2/OM3 = 200 MHz.km/500 MHz.km/2,000 MHz.km bandwidth.				
OS1/OS2 = 9/125 singlemode/9/125 low water peak singlemode					

served loss can then be used to design additional connectivity into the channel, such as an MDA crossconnect.

Other strategies to minimize total CIL include deploying the lowest loss connectors available, specifying higher grades of fiber that effectively lower power penalties associated with modal dispersion (i.e., OM3+), and using high-quality factory terminated patch cords and cable assemblies (see “InfiniBand network example,” page 35).

10GBase-T and twisted-pair copper

The bulk of data center active equipment is made up of servers to run the applications and processing tasks that keep a business going. To connect active equipment back to switch areas, horizontal cables are typically routed out from the HDA patch field and terminated in the equipment distribution area (EDA). Traditionally, copper solutions have been the preferred structured cabling medium for these links due to cost-effective electronics and easy termination and installation in the field.

The IEEE 10GBase-T standard (802.3an) specifies copper cabling that has a bandwidth specified out to 500 MHz and in lengths up to 100 meters (90 meters across the horizontal and 10 meters in patch cords). 10GBase-T copper structured cabling systems make use of familiar category-grade twisted-pair cables and patch cords terminated with RJ-45 jacks.

Key developments in copper cable and connectors to meet the specifications of this standard include:

- Compensated circuit boards used in the connectors to assure improved higher-frequency responses for many frequency-domain parameters, such as return loss (reflection) and crosstalk;
- Improved dielectric materials to reduce cable insertion loss;
- Improved field-termination techniques to advance installation reliability and performance;

- Tighter twist ratios on cabling to decrease coupling between conductor pairs in the cable, thereby improving near end crosstalk (NEXT) and far end crosstalk (FEXT);
- The use of techniques within the cable to improve alien crosstalk (AXT)—the interference that arises when signals carried along one cable couple to another cable located close to the first.

First-generation 10GBase-T active equipment emerged in 2007, at a relatively high cost and low port density. Vendor roadmaps predict that 10GBase-T switches and network interface cards (NIC) will be available early this year, and that costs will decrease as the market matures. The promise of 10x the throughput in the same rack space is too great for the market to ignore for long. Vendor roadmaps indicate that 24-port 1U densities using 10GBase-T products will be achieved sooner rather than later.

Most data center interconnects between servers, and from the server farm to the switch, are accomplished by Ethernet structured cabling solutions. Although the Ethernet protocol supports many areas of the network, the data center contains areas with specialized active equipment architectures that require specialized cabling solutions.

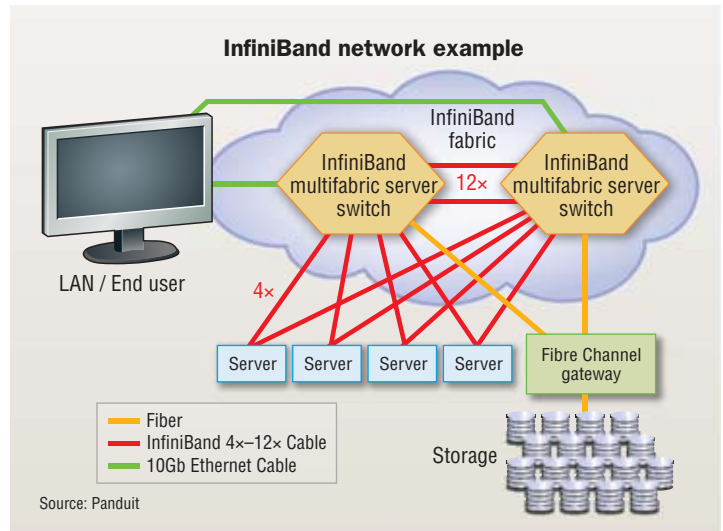
One example is high-performance computing (HPC), where copper-based InfiniBand solutions are finding increased use. Another is in the data center storage area, where the high bandwidth and small-form-factor of fiber media and connectivity have made ANSI Fibre Channel the incumbent solution.

Server resource coordinator

InfiniBand is a scalable, high-speed interconnect architecture that has extremely low latency. Its strongest niche is HPC clusters, and it is increasingly used alongside virtualization technologies to coordinate server resources and maximize both processing capability and server/ storage efficiency.

Standard InfiniBand 4X copper cables comprise eight concurrent runs (or four lanes) of 100-Ω impedance twinaxial cable (i.e., twinax) that transmit at 2.5 Gbits/sec per lane to allow an overall speed of 10 Gbits/sec. Today, 4X copper twinax is used to support data rates up to 20 Gbits/sec. Higher data rates are obtained by using more lanes, and 12X twinax supports up to 60 Gbits/sec. Within two years, use will be made of so-called Quad Data Rate (QDR) technology to support 40 Gbits/sec over 4X cables and 120 Gbits/sec over 12X.

Twinax differs from twisted-pair copper used in Ethernet applications in that pairs of insulated conductors are laid side-by-side and enclosed in a foil shield. The assembly operates in dual simplex (simultaneous bidirectional) mode, where one send run and one receive run in each lane support data independently. Unlike the RJ-45 connector, the InfiniBand con-



nectors are based on a precision screened and balanced edge card design. InfiniBand channels are commonly deployed in point-to-point configurations to help maximize the link loss budget by minimizing the number of connectors per channel.

Finally, the InfiniBand specification specifies a maximum insertion loss over the channel rather than a maximum channel length. In general, signal attenuation increases as cable length increases, and InfiniBand cable manufacturers often use a lower gauge wire (i.e., thicker conductors) to partially compensate for the higher attenuation, which results in a higher overall cable thickness and bend radius compared to twisted-pairs. This has resulted in standard 4X InfiniBand copper cables having a reach of approximately 15 meters (50 feet) for 10-Gbit/sec data rates.

Some manufacturers provide cable offerings in a variety of gauges. For example, InfiniBand copper cables with very thin gauge conductors (e.g., 30 to 32 AWG) feature a smaller bend radius for improved cable management but a very short reach of about 3 to 4 meters. Also, the use of active circuitry located within the connector can improve overall link margin, which effectively increases the length of a copper cable assembly by two to three times.

Additionally, as data rates increase, the maximum allowable length of copper cable is reduced, which can limit the number of nodes available for other network components and cabling. For example, an InfiniBand-based cluster might be configured with a few thousand nodes using 15-meter 4X copper assemblies and a maximum speed of Single Data Rate (SDR). If the processing speed of this cluster is upgraded to require InfiniBand copper cables supporting Double Data Rate (DDR), the copper assembly length would be reduced to 7 to 8 meters and the reduced distance available between servers and switches would permit only several hundred nodes in the

InfiniBand networks, such as the one illustrated here, use 4X and 12X cables engineered specifically for the protocol.

cluster. In this case, unconventional rack and cabinet layouts (e.g., a “U” shape) can increase the number of nodes interconnected using shorter cables.

Fiber InfiniBand cables are deployed in situations in which extended-reach solutions are required. The standard InfiniBand 4X SDR fiber cable is a 12-fiber multimode 50/125-μm ribbon matrix having a reach up to 200 meters, and terminated with a multi-fiber array connector. Fiber cable assemblies have a smaller diameter and bend radius than copper twinax, which can offer routing and management benefits in tight data center spaces. Since most switches come

with copper CX4 connectors, fiber-to-copper transceivers (media converters) must be attached to CX4 ports to yield these long reach solutions.

In practice, a balance of short-reach applications and cost considerations have helped copper twinax evolve into the preferred medium for InfiniBand cabling deployments.

The distance between switches and servers in cluster applications rarely exceeds 15 meters, so the reach benefits of fiber are not typically required. The costs for copper and fiber InfiniBand cabling assemblies are about equal, and the cost of copper electronics and adapter cards traditionally has been less than that for fiber electronics. In the future, as modular transceivers and active cable assemblies become more available, manufacturers will increasingly integrate fiber electronics into switch gear.

Fibre Channel fabrics

The other commonly used data center interconnect is Fibre Channel for storage area network (SAN) applications. Although copper may be used in the storage environment over short runs, deployment of Fibre Channel fabrics are typically based on fiber optics due to high bandwidth, small-form-factor connectivity and media, and wide availability.

OM3 is the first fiber grade to consider for storage areas, based on its recommendation in TIA-942, its ability to carry error-free 10-Gbit/sec traffic, and to scale to data rates beyond 10 Gbits/sec.

InfiniBand physical layer specifications					
Type	PMD	Technology	Connector*	Media	Reach (m)
Copper **	4X	4 lanes, dual simplex	CX4 (MicroGiGaCN)	Twinax	15 SDR; 10 DDR; 5 QDR
	12X	12 lanes, dual simplex	Pluggable (X-Pak, XFP,X2, XENPAK)	Twinax	15 SDR; 10 DDR; 5 QDR
Fiber	1X-SX	850 nm VCSEL, serial duplex	Dual LC	OM1/OM2/OM3 MMF, two fibers	125/250/500 SDR
					65/125/200 DDR
					33/82/300 QDR
	4X-SX	850 nm VCSEL, 4 lanes parallel	Single multi-fiber array	OM1/OM2/OM3 MMF, one 12-fiber ribbon	75/125/200 SDR
					50/75/150 DDR
					N/A*** QDR
	8X-SX	850 nm VCSEL, 8 lanes parallel	Dual multi-fiber array	OM1/OM2/OM3 MMF, two 12-fiber ribbons	75/125/200 SDR
					50/75/150 DDR
					N/A*** QDR
	12X-SX	850 nm VCSEL, 12 lanes parallel			75/125/200 SDR N/A*** DDR N/A*** QDR
	1X-LX	1310 nm LD, serial duplex	Dual LC	OS1 and OS2 SMF	10,000 SDR, DDR, QDR
	4X-LX	1310nm LD, serial duplex	Dual LC	OS1 and OS2 SMF	10,000 SDR
					N/A*** DDR
					N/A*** QDR

* For active cables, the QFSP specifications defines the form factor. ** Active InfiniBand copper assemblies can extend reach.
*** N/A = Assembly currently not available on the market.

Fiber links to storage areas typically are less than 50 meters, so it is possible to deploy OM2/OM2+ fiber in SAN areas, and/or repurposed existing runs of OM2 fiber as less-expensive solutions than OM3; however, it is worth noting that OM2+ extended-reach fiber may not support next-generation high-speed services.

Cable management in the SAN equipment area is a challenge. In traditional installations, large numbers of duplex zipcord fiber assemblies are routed to high-port-count chassis-based SAN directors, and it can be difficult to locate and alter cables in such high-density SAN fabrics. It is also important to keep equipment light-emitting diodes (LEDs) visible and to prevent having to disconnect the cables when removing neighboring blades. Cabling must also not block airflow or access to the power-supply chassis.

Small-form-factor cabling solutions that deploy LC connectors and array-based connectors with ribbonized cabling elements serve to minimize congestion in the SAN. These types of cabling systems combine multi-fiber array patch cords and panels with quick deployment cassettes or hydra assemblies to efficiently cable and manage the SAN. Hydra assemblies on the SAN director side connect to multi-fiber array trunking assemblies through patch panels in the director rack. The trunking assemblies are distributed to SAN equipment racks, where racking space is typically at a premium. Zero-U multi-fiber array brackets in these SAN equipment racks con-

nect hydra assemblies to the trunk assemblies, forming Fibre Channel links between hosts and the SAN director.

ANSI/INCITS (American National Standards Institute/InterNational Committee for Information Technology Standards) have developed Fibre Channel standards along a Base2 migration path for signaling at the SAN fabric edge. The standard defines signaling requirements over fiber and copper media at speeds of 1, 2, and 4 Gbits/sec, with 8 Gbits/sec under review and 16 Gbits/sec projected.

A Base10 path has also been approved to support the aggregation of edge traffic over the SAN core/interswitch link (ISL) connections, and is designed to migrate upward as edge data rates increase (i.e., 4 Gbits/sec edge speed with 10 Gbits/sec ISL migrates to 8 Gbits/sec edge and 20 Gbits/sec ISL).

To improve Fibre Channel's competitiveness, ANSI is sponsoring two approaches based on speeds available to Ethernet through Base-T standards. The first is FC-BaseT, which was designed to define the use of twisted-pair Ethernet channels for the transport of Fibre Channel information. It defines Fibre Channel operation at 1, 2, and 4 Gbits/sec over Category 5e or better cabling at a reach of up to 100 meters.

ANSI's T11 committee is also developing a standard called Fibre Channel over Ethernet (FCoE), which proposes wrapping Fibre Channel information within Ethernet frames and running it over standard Ethernet links. The FCoE protocol represents a step toward converged infrastructures in the data center, leveraging existing 10GBase fiber and copper physical-layer solutions to move Fibre Channel data across the network.

40- and 100-Gbit/sec applications

In pursuit of still higher data rates, the IEEE 802.3 committee set up a study group to assess the viability of developing an Ethernet standard with 10x the highest data rate (i.e., 100 Gbits/sec). After much internal discussion, the road forward is set to commence this month with development of a combined 40-Gbits/sec and 100-Gbits/sec data transmission standard. This standard will be written for copper and fiber cabling solutions.

Data rates of up to 100 Gbits/sec are envisaged to support aggregation of the data paths carrying 10 Gbits/sec, so these

Commonly used Fibre Channel physical layer specifications					
Type	PMD	Technology	Connector	Media	Reach (m)
Copper	4GFCBase-T	4 pairs	RJ-45	Category 6/6A	40/100
	2GFCBase-T	4 pairs	RJ-45	Category 5e/6/6A	60/70/100
	1GFCBase-T	4 pairs	RJ-45	Category 5e/6/6A	100
Fiber	8GFC 800-SM-LC-L*	1310 nm LD, serial	Dual LC and SC	OS1 and OS2 SMF	10,000
	8GFC 800-M5E-SA-I*	850 nm VCSEL, serial		OM3 MMF	300
	8GFC 800-M5-SA-I*			OM2 MMF	100
	8GFC 800-M6-SA-S*			OM1 MMF	40
	4GFC 400-SM-LC-L	1310 nm LD, serial	Dual LC and SC	OS1 and OS2 SMF	10,000
	4GFC 400-M5E-SN-I*	850 nm VCSEL, serial		OM3 MMF	380
	4GFC 400-M5-SN-I			OM2 MMF	150
	4GFC 400-M6-SN-I			OM1 MMF	70
	2GFC 200-SM-LC-L	1310 nm LD, serial	Dual LC and SC	OS1 and OS2 SMF	10,000
	2GFC 200-M5-SN-I	850 nm VCSEL, serial		OM2 MMF	300
	2GFC 200-M6-SN-I			OM1 MMF	150
1GFC 100-SM-LC-L	1310 nm LD, serial	Dual LC and SC	OS1 and OS2 SMF	10,000	
1GFC 100-M5-SN-I	850 nm VCSEL, serial		OM2 MMF	500	
1GFC 100-M6-SN-I			OM1 MMF	300	
* Standards currently in draft.					

rates are expected to find applications in core and aggregation layer switching. The 40-Gbit/sec data rates will be used mainly in access layer applications, and could also be used within HPC clusters. In this case, high-speed Ethernet is likely to become a close competitor to alternative techniques such as InfiniBand.

The copper part of the standard will be based on cabling solutions for short links at both 40 and 100 Gbits/sec. Although the approach for the copper cable is not known at this stage, it is possible that IEEE will model the standard on proven technologies from familiar, more-specialized applications, such as InfiniBand. Here, cable technology has been adequately demonstrated to show that data rates with different coding approaches can be supported by eight lanes operating at 4x the basic data rate (i.e., QDR).

Data rates at 120 Gbits/sec can be supported by another commonly available cable, the so-called 12X, which consists of 24 unidirectional shielded twin-axial conductor pairs. Other schemes include four lanes of 25 Gbits/sec, favored within the fiber-optic community.

Although it is predicted that copper solutions at very high speeds will be substantially different from familiar twisted-pair cabling, fiber solutions should remain similar in terms of media and connectivity. The strategy is to advance transceiver technologies to use the full bandwidth of higher-grade fibers and keep structured fiber cabling in place. Recently, several fi-

ber electronics manufacturers have begun to move beyond current IEEE 10-Gbit/sec standards and work on defining coarse wave-division multiplexing (CWDM) transceivers to carry signals at speeds of 40 and potentially 100 Gbits/sec.

Historically, the market has embraced serial transceiver technology on the basis of cost, so one key design goal is to develop a 40-Gbit/sec CWDM serial multimode PMD for use with OM3 fiber at short reach to leverage existing investments in fiber cabling media. (For an example of some industry efforts underway, see X40msa.org.) OM3+ may be required for use with a short-reach 100 Gbits/sec serial multimode PMD.

Manufacturers have this same incentive to develop serial solutions at 100-Gbit/sec speeds; however, a transceiver may only be available in a parallel-optics footprint. If the IEEE

specifies only parallel solutions for this data rate, additional media would need to be added.

Prepare for the future

In the future, advances in fiber- and copper-media technology will enable all data center links to be capable of achieving error-free 10-Gbit/sec data rates. High-speed applications, such as core-to-core switching, HPC clusters, and server/SAN virtualization will continue to drive data rates to 40 and even 100 Gbits/sec.

However fast your enterprise moves toward these speeds, this information will help you maximize the life of your structured cabling system. 