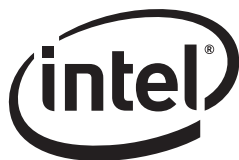# Deploying 10 Gigabit Ethernet into HPC Clusters, Server-to-Server/Server-to-Storage Infrastructure and Workstations over CX4 Copper Interconnects

The advent of 10 Gigabit Ethernet (10GbE) over CX4 copper cabling offers a cost-effective means of extending high-bandwidth, low-latency connectivity to many network areas requiring the added performance advantages of 10GbE. In particular, 10GbE CX4 connectivity can now be extended over a great number of servers and clusters, and it offers an affordable, standards-based solution in areas previously dominated by expensive proprietary fabrics, such as server-to-server and server-to-storage infrastructure. The lower cost of 10GbE CX4 also makes it attractive for deployment into special-use workstations in highly collaborative, data-intensive applications that can benefit from the increased bandwidth and lower latency of 10GbE.

FUJITSU

(intel)

Table of Contents

Introduction

Since its introduction in 2002, 10 Gigabit Ethernet (10GbE) over fiber media has proved itself as a high-bandwidth, low-latency solution in Local Area Networks (LANs), Metropol-itan Area Networks (MANs), and Wide Area Networks (WANs). Until now, these applications of 10GbE have been primarily as a network backbone where the longer reach of fiber (up to 10 kilometers over single-mode fiber) allows the deployment costs to be spread out and the price of fiber-optic modules is justifiable.

Recent technological advances and new standards defini-tions are now encouraging further 10GbE deployment by making the low-cost advantages of copper interconnects available. This allows IT Managers and System Integrators to justify further investments in 10GbE, allowing growth into other network infrastructure areas where the use of 10GbE used to be prohibitively expensive. One of the key elements behind this is the 2004 ratification of the 802.3ak*
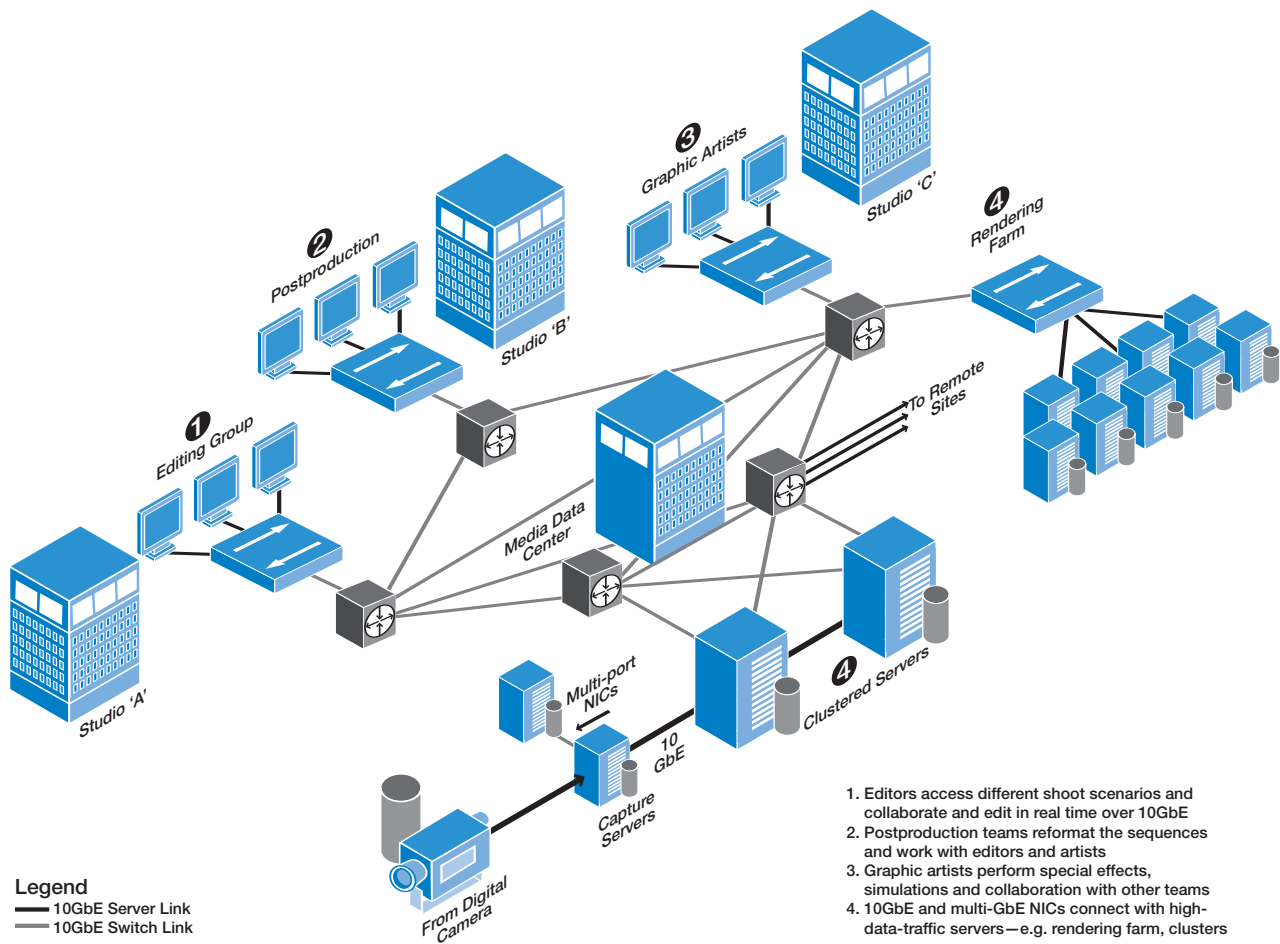
10GBase-CX4 standard that defines 10GbE operation over copper cabling. The other key element is availability of 10GbE CX4 network interconnect components such as the Intel® PRO/10GbE CX4 server adapter and the Fujitsu XG700 10GbE CX4 switch, which offers low-latency, non-blocking switching of CX4 copper links. These networking components are available at a fraction of fiber-based connectivity costs owing to their usage of copper-based technologies.

The far lower cost of ownership for 10GbE CX4 networking components opens the way for 10GbE use in a much wider range of applications. Inexpensive 10GbE CX4 connectivity can now be extended to distribution and workgroup server clusters and load-balanced web, database and hosting servers. In particular, 10GbE CX4 offers an economical, standards-based solution in areas previously dominated by expensive proprietary fabrics, such as server-to-server and server-to-storage infrastructure. Additionally, the lower cost of 10GbE CX4 makes it an ideal performance enhancement for high-end workstations in workgroups dealing with data-intensive, highly collaborative applications. Such applica-tions include computer-generated imaging (CGI), animated films, simulation and modeling, medical imaging and health sciences, and computer-aided design and manufacturing (CAD/CAM). This white paper discusses 10GbE CX4 technology and its use in providing high-bandwidth, low-latency connectivity between servers, between servers and storage, and for workgroup workstations.

Data-Intensive Environments Demand 10GbE Performance

The state of the art in many computer-based applications is reaching the point of being data bound. In other words, the art cannot progress further without the capacity to handle more data faster and with less response delay or latency. A highly visual example of this is digital filmmaking, particu-larly in the areas of animation and special effects. Compare

Figure 1. Example of 10 Gigabit Ethernet used in filmmaking to provide high-bandwidth, low-latency links between various studios or workgroups.

Legend
— 10GbE Server Link
— 10GbE Switch Link

1. Editors access different shoot scenarios and collaborate and edit in real time over 10GbE
2. Postproduction teams reformat the sequences and work with editors and artists
3. Graphic artists perform special effects, simulations and collaboration with other teams
4. 10GbE and multi-GbE NICs connect with high-data-traffic servers—e.g. rendering farm, clusters

the quality of today's special effects and animated features to those of a decade or even a few years ago. The smooth transitions, fine detail of motion, texturing, shadowing, and tonal qualities achieved today require massive amounts of CGI data and highly interactive and responsive editing that can only be achieved through high-bandwidth, low-latency systems. An example of such a system is shown in Figure 1, where 10GbE is used extensively.

Without 10GbE, a system, such as shown in Figure 1, could only be achieved through expensive proprietary fabrics or a far more complex system of teamed Gigabit Ethernet (GbE) links. Moreover, for the switch links shown in Figure 1, the Intel® PRO/10GbE LR server adapter and Fujitsu XG1200 switch over single-mode fiber provide a range to 10 kilometers (or 300 meters over multi-mode fiber with the Intel® PRO/10GbE SR). This extended transmission range to 10

Intel® PRO/10GbE Server Adapter
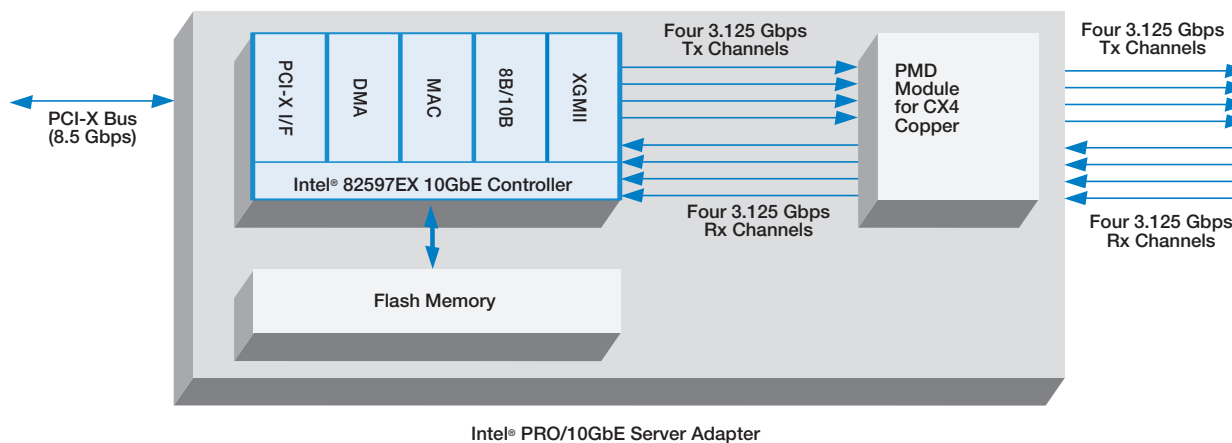
kilometers in single-mode fiber is important for supporting widely separated facilities with high-bandwidth, low-latency connectivity.

The introduction of 10GbE CX4 provides the crucial final link to workgroup workstations. Because of its much lower price, 10GbE CX4 connectivity is a very cost-effective means of giving graphic artists, editing and postproduction groups the fast-response capabilities needed to create today's eye-popping special effects.

It is also important to note that 10GbE CX4 as provided by the Intel PRO/10GbE CX4 server adapter and the Fujitsu XG700 10GbE CX4 switch can be used with either Intel® architecture-based systems or Apple* G5 systems. Interoperability with Apple G5 is particularly important for the graphic arts community, and Small Tree Communications has already packaged a low-cost, high-performance solution using the Apple G5, Intel PRO/10GbE CX4 server adapter and Fujitsu XG700 switch combination[1].

Filmmaking is not the only data-intensive endeavor that can benefit from 10GbE and 10GbE CX4 connectivity. Many modeling, simulation, forecasting, design, and manufacturing efforts can benefit from, and advance to higher levels of capability through, 10GbE connectivity. Weather modeling and forecasting, geophysical analyses, financial forecasting, medical imaging and diagnostics, or complex CAD/CAM typical of aerospace—all are data-intensive, fast-response applications that can benefit. In fact, the filmmaking system diagram in Figure 1 is widely applicable to many of these applications. Just change the digital camera to an array of data-collection devices—such as strain gauge and accelerometer arrays—and change the Editing Group, Postproduction and Graphic Artist labels to other workgroup names—such as airframe design, flight dynamics analysis, and so forth. Regardless of the labels, the end result is the same for any data-intensive application—cost-effective, high-bandwidth, low-latency 10GbE connectivity down to the workstation for greater workgroup capability and productivity.
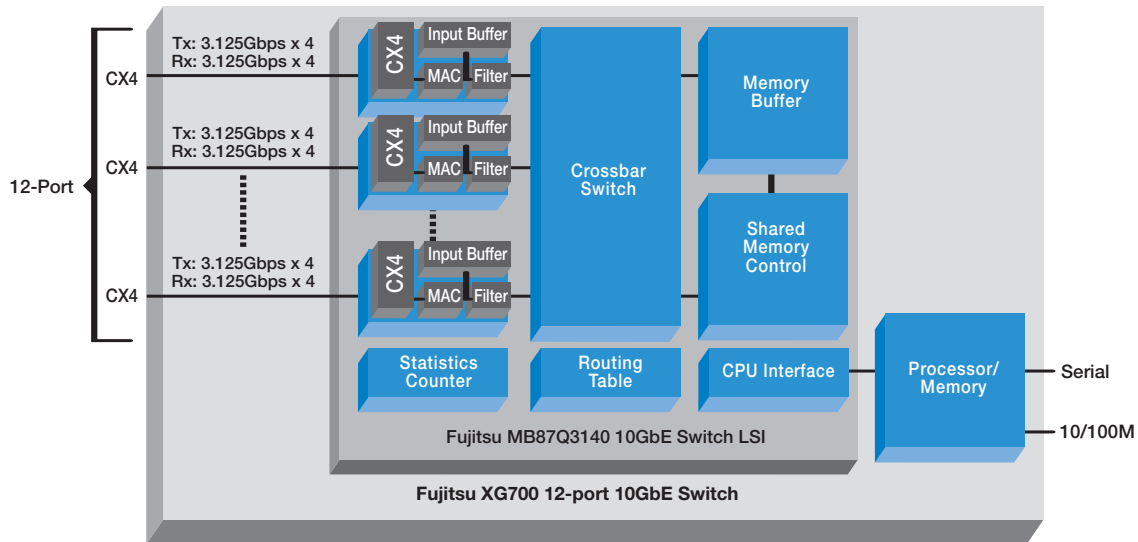
Overview of 10GbE Technology

The 10 Gigabit Ethernet Standard is an extension of the basic IEEE 802.3* standard protocols to a wire speed of 10 Gbps. As an extension, 10GbE is still fully Ethernet compatible and retains the key Ethernet architecture, including the Media Access Control (MAC) protocol, the Ethernet frame format, and the minimum and maximum frame size. Just as Gigabit Ethernet followed the standard Ethernet model, 10GbE continues the evolution in speed while using virtually the same architecture used in other Ethernet specifications. This allows implementation of 10GbE while maintaining compatibility with the rest of the network architecture, and retains the existing principles of network operation and management.

There are, however, some key differences in 10GbE. These differences, which are performance related and do not affect Ethernet compatibility, are illustrated in Figure 2 and involve primarily the MAC interface to the Physical Layer (PHY) of the Open Systems Interconnection (OSI) model

and the Physical Medium Dependent (PMD) connection to the network transmission medium, CX4 copper in this case.

As shown in Figure 2, the Intel® 82597EX 10GbE Controller connects to the adapter's PMD by means of four transmit (Tx) channels and four receive (Rx) channels. Each channel has a 3.125 Gbps bandwidth so that the aggregate bandwidth in the Tx and Rx directions exceeds 10Gbps each way. The separate Tx and Rx channels also allow the server adapter and 10GbE traffic to operate in full-duplex mode (simultaneous Tx and Rx). By contrast, Ethernet standards prior to 10GbE operated in simplex mode and required use of a Carrier-Sensing Multiple-Access with Collision Detection (CSMA/CD) protocol to avoid contention between network devices seeking access to the server. By using full-duplex operation and eliminating CSMA/CD processing, 10GbE can provide faster, lower-latency responses to network transaction requests.

The same Ethernet principles apply for the Fujitsu XG700 and XG1200 Ultra-low Latency 12-port switches, making

them interoperable with the Intel PRO/10GbE CX4 server adapters. Figure 3 shows a functional block diagram of the XG700 switch, which features non-blocking Layer 2 switching for 12 10GbE CX4 ports with a 450-nanosecond fall-through latency.[2] The XG700 switch is also unique in that it uses the Fujitsu MB87Q3140 switch-on-a-chip, which significantly reduces the 12-port CX4 switch footprint and switch cost over previous switch fabric implementations. The reduced switch size is particularly important in size-sensitive applications such as HPC and grid computing applications, and the low cost is particularly important for deploying 10GbE capability to workgroup workstations.

10GbE CX4 Cost/Benefit Advantages

Some cost/benefit advantages of 10GbE CX4 over 10GbE fiber have already been mentioned. Specifically, CX4 components and connectivity cost far less than fiber components. With 10GbE CX4, there is no need for expensive optical modules and subcomponents, assembly is easier, and CX4 has a longer life in a field environment. This gives 10GbE CX4 a distinct advantage as an econom-ical solution for providing 10GbE connectivity to platforms within the CX4 15-meter range, such as workgroup workstations or close-proximity servers in a server cluster or farm. For longer distance requirements, such as LAN backbones or switch links like those illustrated in Figure 1, fiber connectivity via the Intel PRO/10GbE LR or SR server adapters and the Fujitsu XG1200 switch can provide ranges to 10 kilometers, depending on the fiber medium used.

CX4 Versus Other Technologies. In terms of relative costs, the Intel PRO/10GbE CX4 server adapter costs a fraction of that of the Intel PRO/10GbE LR and SR server adapters for fiber. In terms of Gigabit Ethernet for Category 5 twisted pair, the CX4 adapter is about on par price-wise with two
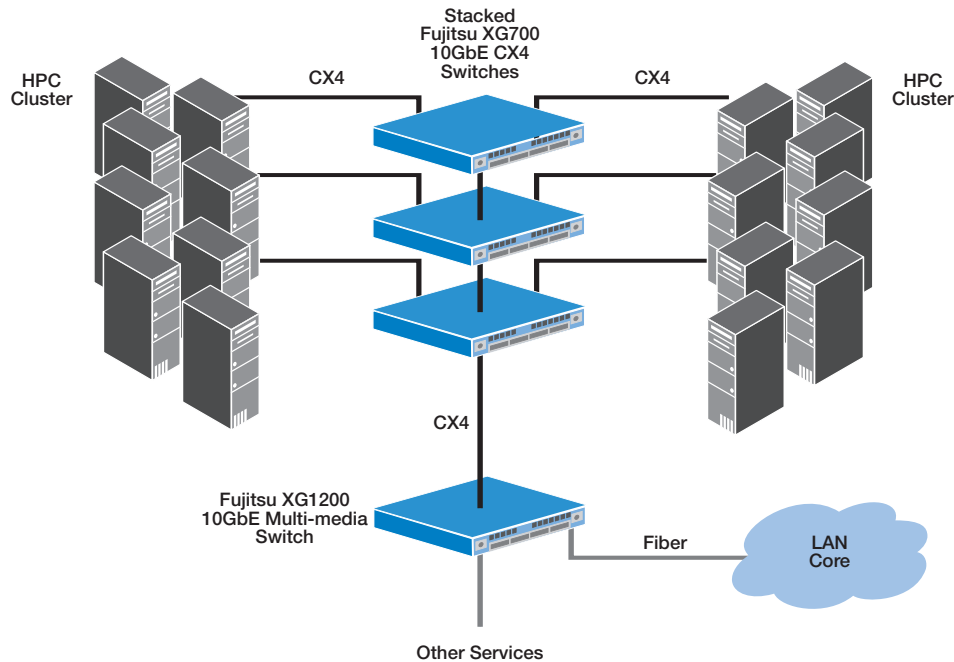
teamed Intel® PRO/1000 MT Quad Port server adapters. However, while eight teamed Gigabit Ethernet (GbE) ports can provide nearly the bandwidth of 10GbE, eight teamed ports consume at least two PCI slots on the server and eight switch ports as compared to the higher bandwidth of 10GbE from a single server slot and a single switch port. On the other hand, teamed GbE adapters do have the advantage of a 100-meter range in Category 5 twisted pair compared to the 15-meter range of 10GbE CX4, and may be a better choice in some network installations for this reason.

Compared to InfiniBand*, Fibre Channel*, and proprietary fabric technologies, 10GbE CX4 offers distinct cost-of-ownership benefits while providing comparable or better performance. Fibre Channel, while it does offer connectivity in fiber, coax, or twisted pair, is performance-limited to 4 Gigabits per second (4 Gbps), although there are plans for a 10-Gbps version.

InfiniBand and 10GbE CX4 are the most comparable in terms of performance. Both use CX4 copper cabling and have comparable data rates (10 Gbps). However, InfiniBand, like Fibre Channel, is more of a niche technology targeting server and storage fabrics. In contrast, 10GbE CX4 adheres to the ubiquitous Ethernet standard and provides greater network compatibility and a broader range of applications. As a result, 10GbE CX4 does not require special switch ports or a specialized skill set for installation and mainte-nance, which makes 10GbE CX4 more cost-effective to apply and use in the long term.

Benefits of Ethernet. The same cost-effectiveness for 10GbE CX4 also holds true for comparisons to other proprietary fabrics. Proprietary fabrics, because they are proprietary and low-production-volume items, are generally far more expen-sive than 10GbE CX4. Moreover, proprietary fabrics require highly specialized skill sets for installation, operation, and

maintenance, which significantly adds to their cost of ownership. 10GbE CX4 offers a much lower initial equipment cost and can be installed, operated, and maintained with the same Ethernet resources and skill set used across the network infrastructure. This lower initial cost and lower cost of ownership make 10GbE CX4 extremely effective as a high-performance, low-latency solution for 10-Gbps wire-speed connectivity to workstations as well as a replacement for proprietary fabrics in server-to-server and server-to-storage infrastructures.

Deploying 10GbE CX4 into HPC Clusters

Because of its high performance and comparatively low cost, 10GbE CX4 is a highly effective fabric for connectivity within high-performance computing (HPC) clusters. As the term "cluster" implies, HPC platforms are typically located in close proximity to each other, generally well within the 15-meter range of 10GbE CX4 connectivity. This being the case, 10GbE CX4 provides an excellent connectivity solution for HPC clusters and at far less cost than fiber connectivity or proprietary fabrics.

Figure 4 illustrates the concept of 10GbE CX4 connectivity in HPC clusters. Intel PRO/10GbE CX4 server adapters are used in each server for server-to-server connectivity through the stacked Fujitsu XG700 switches.

In addition to their low unit cost, the Fujitsu XG700 switch has a low profile and small footprint, making it ideal for stacking to provide high-density 10GbE CX4 switching. This low profile is a result of full integration of the 12-port switch onto a single LSI chip. The high level of integration also results in lower power consumption, which is also a critical consideration in high-density HPC cluster environments.

Figure 5. Example of using 10GbE CX4 connectivity in the data center.
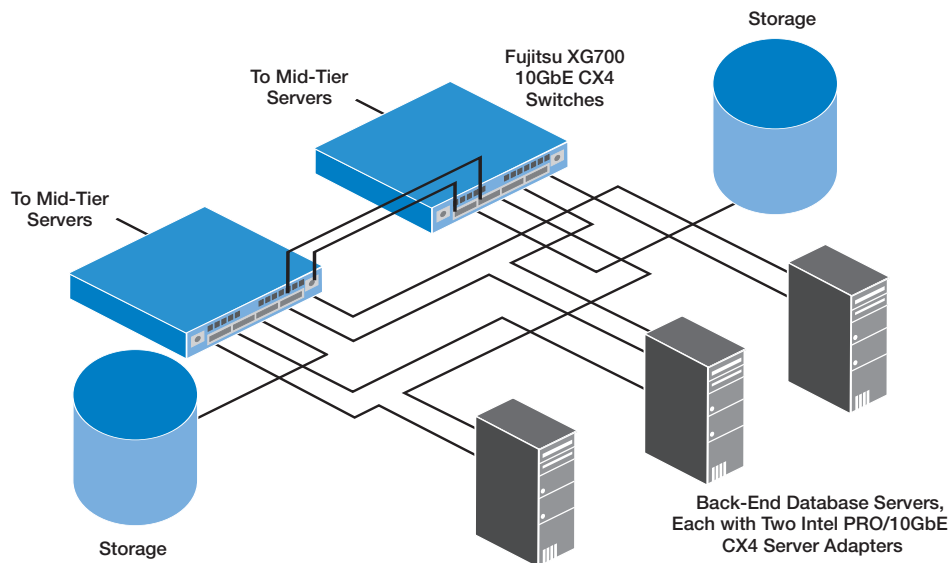
Deploying 10GbE CX4 into Server-to-Server and Server-to-Storage Network Infrastructure

The same 10GbE CX4 benefits of low-cost, high-performance, standards-based connectivity discussed in the previous sections of this paper also apply to server-to-server and server-to-storage network infrastructure. For example, network attached storage (NAS) is a case where 10GbE CX4 can provide the necessary bandwidth and low-latency performance at substantial savings over proprietary fabrics.

Server consolidation and multi-tiered data centers are other candidates for the performance-enhancing, cost-reducing features of 10GbE CX4. Figure 5 provides an example of 10GbE CX4 implementation in the Back-End Tier of a three-tiered data center. In this example, each database

server has two Intel PRO/10GbE CX4 server adapters to provide link redundancy. This is supported with cross-connection between two Fujitsu XG700 switches to provide switch fault tolerance. Additionally, as indicated in Figure 5, 10GbE CX4 can also provide connectivity for NAS.

In a manner similar to Figure 5, the data-center front-end and mid-tier can also be migrated to 10GbE CX4. The possible exception would be tier-to-tier connections beyond the 15-meter CX4 limit, which would require fiber connectivity for the additional range. However, within tiers, connectivity with the 15-meter limit typically is not a problem.

Prior to the advent of 10GbE CX4, deploying 10GbE throughout the data center had to be done using fiber. This was difficult to cost justify for the number of optical links needed to provide switch fault tolerance as shown in Figure 5. Because of its much lower cost, 10GbE CX4 now makes

Figure 6. Deploying 10GbE to workstations using Intel PRO/10GbE Server Adapters and a Fujitsu XG700 Switch for 10GbE CX4 switching.
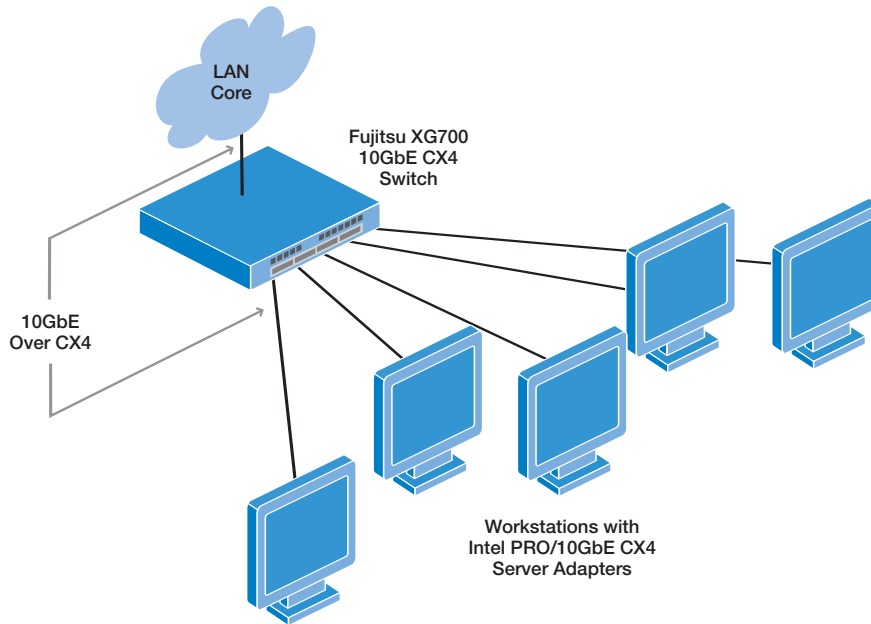
it possible to deploy 10GbE throughout the data center for high-performance support of numerous data-intensive applications, including deployment of 10GbE down to the workgroup and workstation level.

Deploying 10GbE CX4 into Workstations

Workgroups involved in data-intensive design, simulation and digital imaging processes can gain an immediate performance boost by deploying 10GbE CX4 connectivity to their workstations. Figure 6 illustrates such a deployment, where an Intel PRO/10GbE CX4 server adapter is installed in each workstation and a Fujitsu XG700 or XG1200 switch provides Level 2, non-blocking switching. As mentioned previously, the workstations can be either Intel architecture-based or Apple G5 architecture-based.

Figure 6 shows 10GbE CX4 deployment using a Fujitsu XG700 switch as the workgroup switch. The assumption here is that the LAN Core is within the distance limit for CX4 connectivity. When the connection distance to the core exceeds 15 meters, switching and media conversion from CX4 to fiber can be done through a Fujitsu XG1200 10GbE Layer 2 switch. While the XG700 switch does offer lower-cost switching because of its fixed 12-port CX4 configuration, the XG1200 switch offers the advantage of greater media interfacing flexibility through a selection of XENPAK 10GbE interfaces for CX4 copper, and SR, LR, ER and LW fiber operation. Such media conversion capability makes the Fujitsu XG1200 switch the better choice where media conversion is necessary for 10GbE deployment.

Conclusion

When 10 Gigabit Ethernet was initially introduced for fiber media, network designers were able to deploy higher performance into network backbones to better support emerging data-intensive applications across a broad range of industries. Now, with the emergence of 10GbE in CX4 copper cabling, the performance advantages of 10 Gigabit Ethernet can be cost-effectively deployed across the network, from server-to-server and server-to-storage connections in the data center, through the core, and down to workgroups and individual workstations. This extension of 10 Gigabit Ethernet capacity down to the workstation level provides the added capacity and performance needed to make numerous data-intensive applications even more capable and responsive in a highly cost-effective manner.

Intel and Fujitsu provide a variety of 10 Gigabit Ethernet products with proven interoperability to support the continuing migration of networks to 10GbE capability in fiber and CX4 copper media. These products include:

From Intel:

- Intel® PRO/10GbE CX4 Server Adapter for low-cost network connectivity in CX4 copper to 15 meters in workgroups, server clusters, and network attached storage

- Intel® PRO/10GbE SR Server Adapter for network connectivity to 300 meters in multi-mode fiber

- Intel® PRO/10GbE LR Server Adapter for network connectivity to 10 kilometers in single-mode fiber

From Fujitsu:

- Fujitsu XG700 10Gb Ethernet Ultra-low Latency 12-port Switch for non-blocking, Layer 2 switching of 10GbE CX4 copper links

- Fujitsu XG1200 10Gb Ethernet Ultra-low Latency 12-port Switch for non-blocking, Layer 2 switching through various plug-in media interfaces (copper and fiber XENPAK—CX4, SR, LR, ER, LW)

To find out more about Intel products for 10 Gigabit Ethernet, visit www.intel.com/go/10GbE

**To find out more about Fujitsu products for 10 Gigabit Ethernet, visit www.fcpa.fujitsu.com/products/new-technologies/xg700-xg1200/**

About Fujitsu Computer Products of America, Inc.

Fujitsu Computer Products of America, Inc. is a subsidiary of Fujitsu Limited, a leading provider of customer-focused IT and communications solutions for the global marketplace. FCPA provides innovative solutions for the U.S. marketplace. Current product and service offerings include high performance hard disk drives, magneto-optical drives, scanners, and scanner maintenance. Additional information about Fujitsu is available at www.fcpa.fujitsu.com

About Intel Corporation

Intel, the world's largest chip maker, is also a leading manufacturer of computer, networking and communications products. Additional information about Intel is available at www.intel.com/pressroom.

intel®

FUJITSU
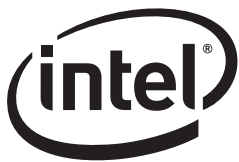
1. Small Tree Communications Announces Superior 10Gb Ethernet Solution for Xserve G5, www.small-tree.com/nov_15_05.htm

2. Fall-through latency is the time taken for a unit of data to pass through the switch without contention.