

26-port Ethernet Switch Boosts 10-Gbps Density for Data Centers



**T E C H N O L O G Y
B A C K G R O U N D E R**

Introduction

Several trends in network demands are putting particularly strong pressure on data centers to increase port density and performance while reducing cost per port and power consumption — goals that can only be met by using more highly integrated network switch chips. As continuing enhancements to Ethernet standards prove capable of serving the requirements of the converged data center, 10-Gbps Ethernet (10GbE) switches are becoming the ideal technology for meeting the need.

By increasing switch integration to 26 ports in a single chip, Fujitsu Microelectronics is enabling a significant improvement in port density for data center interconnects. This technology background profiles the forces driving data center requirements and shows how the Fujitsu MB86C69RBC switch chip provides the capabilities to meet those requirements — capabilities that reach far beyond high port density.

Bandwidth Demand

Many applications are now pushing the need for high-capacity, high-bandwidth backbones to carry traffic and ultra-large-capacity data centers to store the data (Figure 1). These rapidly expanding applications include social networking, peer-to-peer mobile communication, business-to-business commerce and video services.

Additionally, data services associated with handheld PDA-phones such as the iPhone are putting more pressure on data center capabilities. Telecom analysis firm Light Reading has reported that 85 percent of iPhone users access news and other data services and 35 percent or more use it to access video services.

To satisfy cellular users and others who want a variety of data and media services, data centers must handle the traffic efficiently and with a reasonable quality of service (QoS). To do that, data centers need high-performance, converged fabrics that reliably interface to interconnects while utilizing state-of-the-art congestion management to handle different traffic classes such as voice, data, storage, high-performance computing clusters (using inter-process communication, IPC) and web services.

While these requirements are pushing data centers to increase bandwidth and capacity, green-IT initiatives are legislating a move to lower power consumption. As a result, servers must offer higher throughput per Watt.

All of these factors spur the need for inexpensive, standardized, high-performance, high-bandwidth interconnect. 10GbE stands out as the interconnect technology of choice.

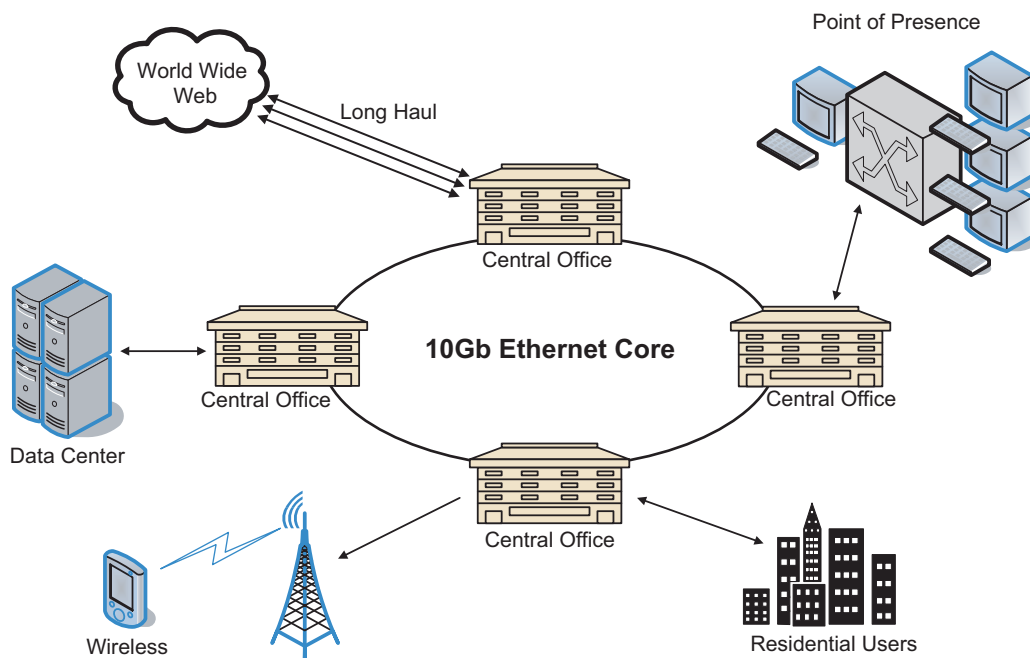


Figure 1 – Applications Pushing Bandwidth Requirements

Evolving Ethernet Market

Ethernet has kept pace with bandwidth demands by providing speed increases well ahead of market requirements. With this proactive approach, Ethernet has captured more than 85 percent of the interconnect market. While 1-Gigabit (1G) Ethernet and Fast Ethernet now dominate, market analysts expect that by late 2008, the availability of low-power, low-cost 10GbaseT PHYs will help ramp 10GbE into a large market share. Ethernet's resiliency, power economics, and economies of scale will continue to deliver the cost/performance characteristics needed by all applications.

Fueled by data center performance demands and declining price per port, the 10GbE switch market is poised for rapid growth in the next few years – 185 percent growth from 2008 to 2009 –and could easily surpass 4.7M ports in 2011 according to market research firm Gartner, Inc. (Figure 2).

During 2006 and 2007, developers of integrated silicon solutions made tremendous progress in refining 10GbE technology and its infrastructure. Improvements to the technology promote wider adoption in large and growing subsections of the computing market, such as blade servers, networked enterprise switches and video servers.

By reducing the cost per port, 10GbE based on SFP+ optical transceivers and 10GBase-T technology (IEEE 802.3an) is

enabling many new applications, using 10GbE NICs and switches for low-cost interconnect. Aimed at applications in the central office, data center and the enterprise, Fujitsu's products offer the option of interconnecting with CX4 or KR (serial Ethernet backplane) directly to the chassis backplane.

Longer term, the demand for 10-Gigabit speeds comes primarily from requirements for converged services that include mobile data services, video delivery, VoIP, storage, system backup, teleconferencing and surveillance. These services demand high bandwidth in switches and server chassis, which is ultimately realized through the use of a high-performance chassis backplane interconnect. For all these applications, 10GbE offers the dramatic improvements in cost-effectiveness that system developers require for deploying viable products into the marketplace.

26-port 10GbE Switch IC

Following the successful launch of the 10GbE switching chip line, Fujitsu has introduced the newest member of this family – a small-footprint, low-power, 26-port 10GbE switch chip capable of running at 520+ Gbps wire speed (Figure 3). In addition to high port density and high throughput, the Fujitsu MB86C69RBC device offers low-latency performance and high QoS. The MB86C69RBC chip was specifically designed to address the data center requirements described earlier.

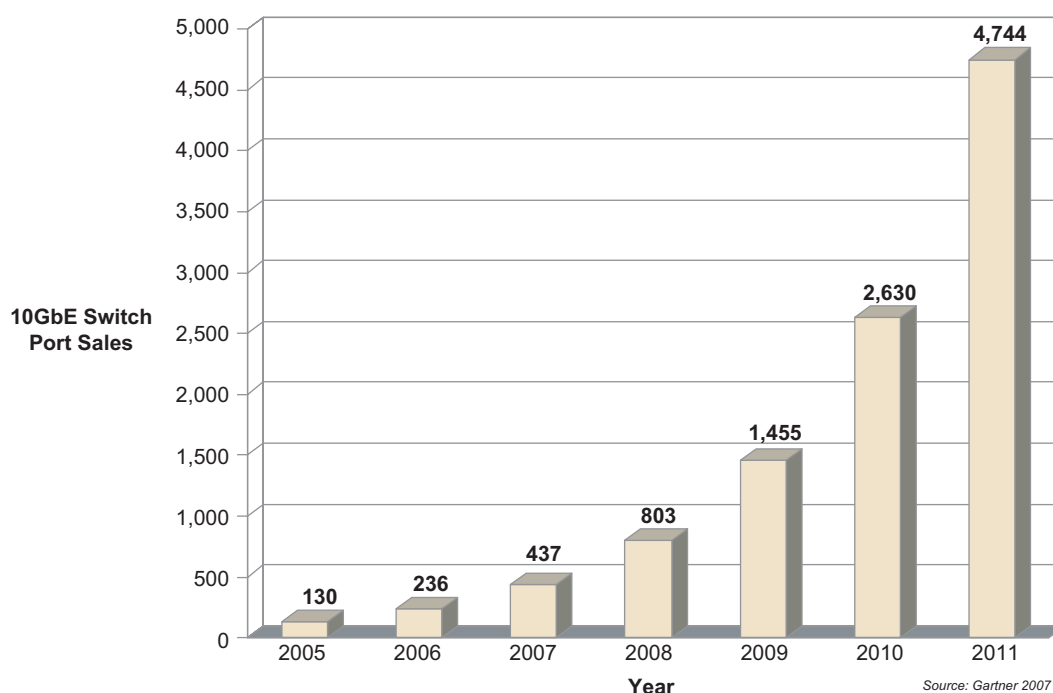


Figure 2 – 10GbE Switch Port Growth

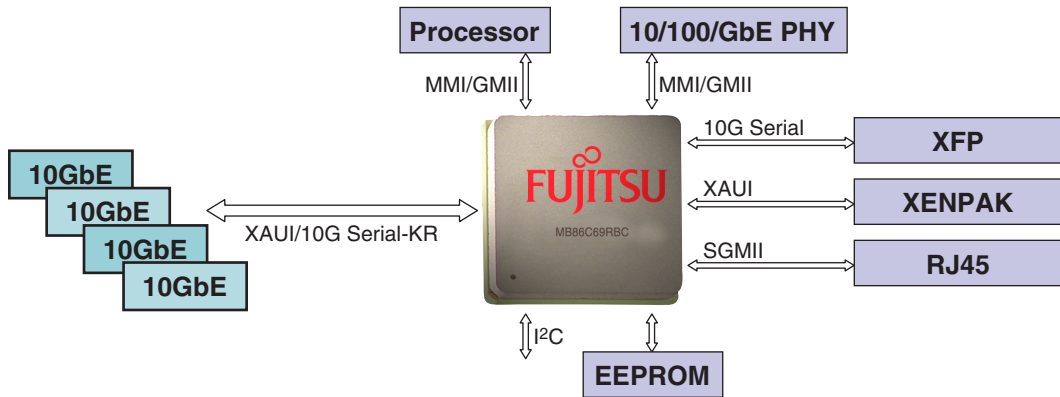


Figure 3 – The Fujitsu MB86C69RBC

This chip incorporates a new multi-rate PHY that supports the recently standardized 10GbE for backplane (802.3ap). The PHY can thus drive 1m of single-channel backplane (also known as KR), XFI (the electrical interface for XFP modules), SFP+, the 1000KX or SGMII, and the now standard CX4 PHYs. The latter provide reliable 25m drive capability (Figure 4) or 1m of FR4 backplane. The device supports KR on all 26 ports, and users can configure any of the ports to run in CX4, 1000 baseX, SGMII or KR mode.

The small footprint and low power (22W under full-load conditions) of the Fujitsu MB86C69RBC make it ideal for small form-factor cards such as Advanced Mezzanine Cards, microTCA controller hubs and blade servers. By substantially reducing the number of chips required to implement an industry-standard switch, the MB86C69RBC cuts the overall cost, improves density, decreases power consumption and helps improve system reliability.

In addition to 520+Gbps non-blocking performance, the MB86C69RBC offers exceptionally low latency — as low as 300 ns. This low latency makes the device ideal for high-performance computing, cluster interconnects and server switches.

The MB86C69RBC also provides Ethernet Layer 2 and Layer 3 (L2 and L3) capabilities well suited for low-cost, high-density,

high-performance switching applications. The full set of L2 features include 802.1s spanning tree, multiple spanning tree and rapid spanning tree protocols; 802.1Q VLAN, Q-in-Q; 801.2p QoS; 802.3ad link aggregation; 802.1D learning and aging; and 802.1D MAC address classification, VLAN classification, MAC-address-based switching/forwarding, and VLAN-based switching/forwarding. Along with essential network-maintenance functions, these features help ensure full wire-speed performance.

Layer 3 capabilities include L3 forwarding as well as L3 access control list (ACL). The chip utilizes the full 16K address table to forward IP addresses to assigned ports. The ACL allows the chip to be programmed to interpret various fields for security, content management and forwarding.

The on-chip micro-engine of the MB86C69RBC executes a variety of macro commands. Using the chip’s high-level application programming interface (API), software developers can create embedded code for applications. The micro-engine enables software developers to reduce the amount of embedded software development required. The chip accepts macro commands through its two Ethernet-based management interfaces and can also be initialized and made fully operational through its EEPROM interface. This feature helps engineers progress with hardware debug while the embedded firmware is in development.

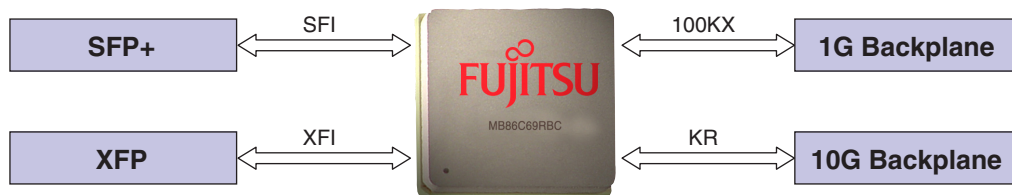


Figure 4 – Multi-rate PHY

In applications that require more than 26 ports, the MB86C69RBC can easily be scaled by using multichip enabling features. Specifically, the chip supports source port routing (SRP) and outbound-tag routing (OTR) for multichip architectures. These techniques combine with link aggregation and Backward Congestion Notification (BCN) to allow designers to fix the paths and balance loads in multichip/multi-path scenarios.

Quality of Service (QoS)

The Fujitsu MB86C69RBC switch IC provides QoS control through IEEE 802.1p as well as with IPV4 and IPV6 Layer 3 DiffServ Service classification. These features enable the switch to manage traffic from latency-sensitive voice and video applications, which are becoming increasingly common in today’s converged networks.

By providing QoS capabilities for multiple protocols, the chip enables system developers to meet a number of critical performance requirements. For example, switching products based on the chip satisfy the requirements for service-aware

systems as well as latency-, delay- and jitter-sensitive applications.

The MB86C69RBC also supports improved scheduling algorithms at the egress as well as early drop capability at both the ingress buffer and the egress. These techniques mitigate head-of-line blocking and improve TCP/IP performance.

To support Data-Center Ethernet (DCE) or Data Center Bridging (DCB) the MB86C69RBC supports state-of-the-art congestion management techniques like BCN and per-priority pause, which is an extension of the traditional 802.1x PAUSE (Figure 5). In heavy overload-traffic situations, the ability to handle congestion instantaneously is critical, and per priority PAUSE (PPP) frames can help manage these situations. PPP allows the device to stop inflow of low-priority traffic. The traffic stopped by PPP is programmable to any traffic class of choice coming from the upstream link. The PPP frame is the modified PAUSE frame.

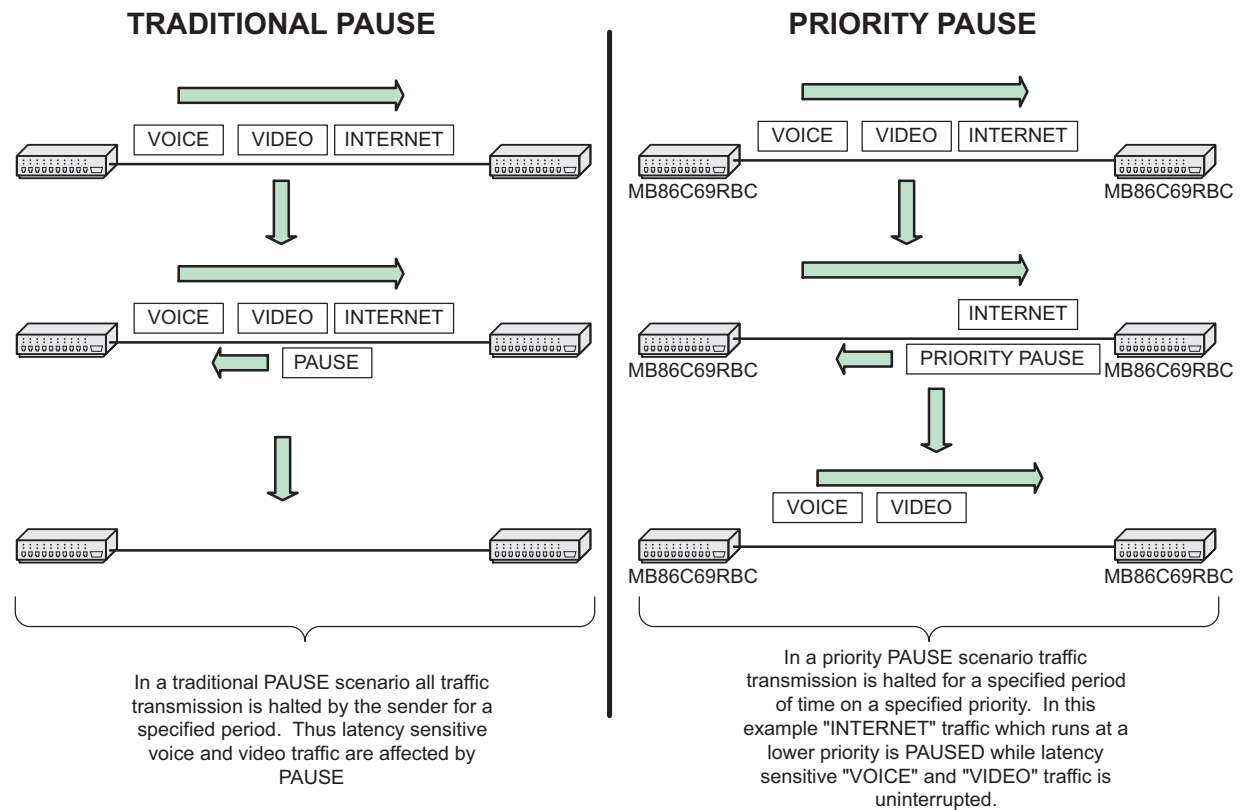


Figure 5 – Traditional PAUSE vs. Priority PAUSE

Like its predecessor, the MB86C69RBC has deep ingress buffers that allow it to deliver lossless performance when an overflow occurs in the shared buffer due to output-based congestion. A congested switch can use Backward Congestion Notification (BCN) to notify sending nodes (through return packets) that their transmissions are causing congestion. The BCN frames contain information about which packet caused the congestion and the state of the queue. A node causing congestion should accordingly rate-limit its transmission. This rate control can be achieved through the application of 802.1Qau rate limiting. The MB86C69RBC supports both pre-standard formats of BCN (802.1Qau based) and PPP (802.1Qbb).

For more information about the congestion-management features and other QoS capabilities of the MB86C69RBC, please consult the Fujitsu Microelectronics technology

backgrounder, “Essential Ethernet Switch Features for Data Centers,” available at: <http://www.fujitsu.com/us/services/edevice/microelectronics/networkingassps/whitepaper/10gdatactr.html>

System Solutions

To support development of a variety of the MB86C69RBC applications, Fujitsu Microelectronics provides a suite of software, evaluation board (Figure 6) and a switch reference design. The evaluation board and its design collateral provide reference for board development.

Fujitsu offers a full 24-port 10GbE Switch reference to enable customers a faster Time to Market. The reference design comes with the necessary software, schematics, layout files and mechanical design.

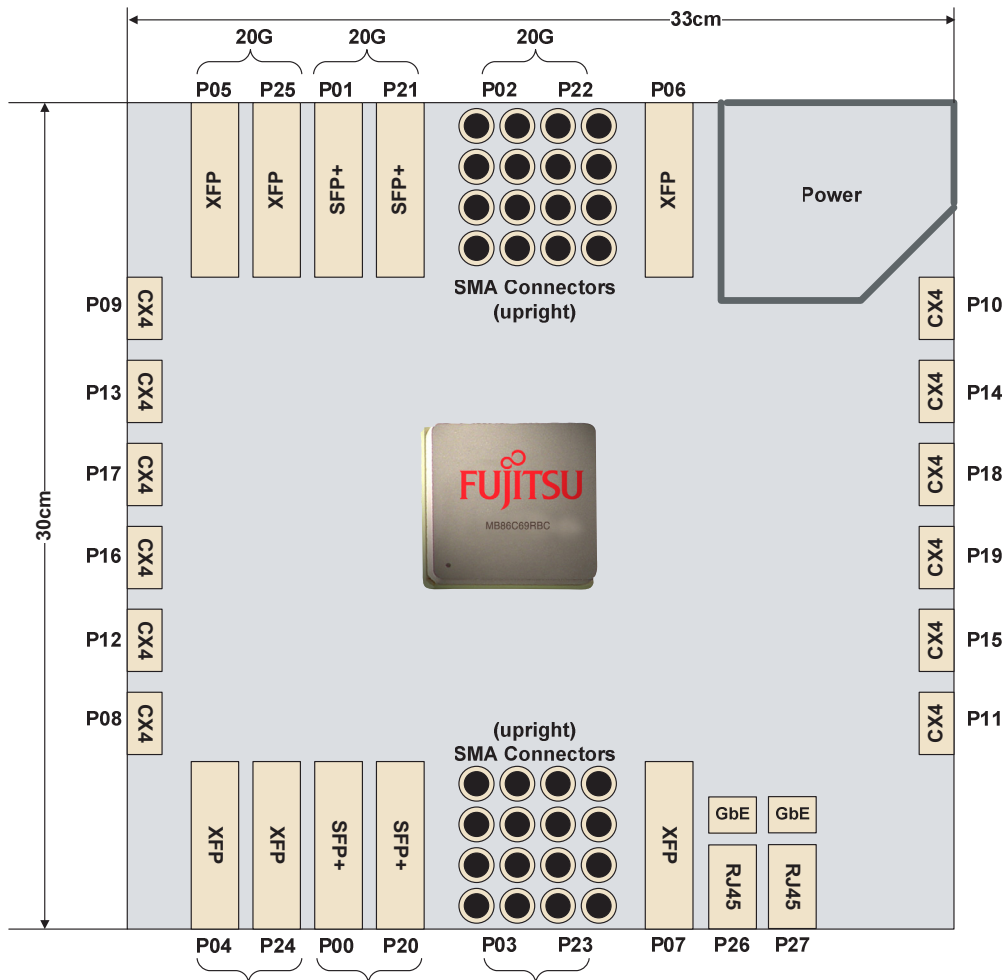


Figure 6 – The Fujitsu MB86C69RBC Evaluation Board

Profiles of applications that require 10GbE switching show the various ways in which the MB86C69RBC suits system requirements:

Switch Blade for Bladed Servers

- Chassis usually has 16 or 18 slots, of which two are switch blades
- 16 MB86C69RBC 10GbE ports into the 10GbE backplane
- Eight 10GbE MB86C69RBC ports for uplinks
- Two spare 10GbE ports

Standalone Switch

- 16 MB86C69RBC 10GbE ports of SFP+
- Eight MB86C69RBC 10GbE ports of CX4
- Two 10GbE MB86C69RBC ports for a network processor for enhanced L3 or metro support

Switch Fabric for ATCA Chassis

- 16 MB86C69RBC 10GbE ports for XAUI or 10GbE serial backplane
- Six MB86C69RBC 10GbE ports for three AMCs
- Two MB86C69RBC 10GbE ports for RTM
- Two 10GbE CX ports for the front panel

Switch for L2 High-Performance Computing (HPC)

- 24 MB86C69RBC 10GbE ports
- All ports configured for SFP+ or CX4

Summary

The Fujitsu MB86C69RBC IC specifically targets the requirements of a virtualized data center and offers many advantages that are unavailable in competing fabric chips. Features that clearly differentiate the MB86C69RBC from other fabric chips include:

- Direct 802.3ap (KR) and SFP+ support
- Low power
- Small footprint
- Large buffer memory (largest currently available)
- Low latency (lowest available)
- State-of-the-art congestion management features
- Programming ease utilizing built-in micro-engine

For More information

For more information on the Fujitsu MB86C69RBC chip and supporting products, please visit the company web site at <http://us.fujitsu.com/micro/10gethernet> or send e-mail to inquiry@fma.fujitsu.com

FUJITSU MICROELECTRONICS AMERICA, INC.

Corporate Headquarters
1250 East Arques Avenue, M/S 333, Sunnyvale, California 94085-5401
Tel: (800) 866-8608 Fax: (408) 737-5999
E-mail: inquiry@fma.fujitsu.com Web Site: <http://us.fujitsu.com/micro>

©2008 Fujitsu Microelectronics America, Inc. All rights reserved.
All company and product names are trademarks or registered
trademarks of their respective owners.
Printed in U.S.A. 10GE-TB-21301-4/2008