



# Applying PCI Express switches to network adapter cards

By John Gudmundson

The introduction of many types of products based on PCI Express (PCIe) technology is moving forward at an accelerating pace. Solutions from root complex chipsets to switches, bridges, and endpoints for PCI, PCI-X, InfiniBand, Fibre Channel, Ethernet, and other standards are now available. Processors, both network and fully programmable, are being developed with native PCIe interfaces. Within the realm of switches, high port and lane count solutions are available today. These devices are well suited for applications within storage, server, and communications for backplane switching fabric designs. New generation switches coming to market will expand this portfolio and include lower lane and port count designs targeted at many mass market high-volume applications. These markets include laptop docking stations, printer and other peripheral device controllers, XMC and AMC mezzanine cards, and network adapter cards. Such PCIe switching solutions will need a wide variety of features to fill such diverse applications. One market in particular that can utilize PCIe technology is that of intelligent network adapters. Given the nature of these card architectures, lower port and lane count switches are highly beneficial for I/O aggregation and processor isolation.

## Migration to intelligent adapters

Intelligent network adapter devices are rapidly moving to expand offload engine capability. With a local processor, these adapters can:

- Initiate and terminate TCP/IP connections
- Encode and decode SSL encrypted sessions
- Provide iSCSI protocol processing
- Manage RAID controller functionality and many other functions

It is imperative as the line rates reach and exceed Gbps speeds that these functions be offloaded from the host processor. Innovations within server and PC operating systems are providing compatibility with adapter card software drivers to allow these adapters to be practical. Intelligent adapters will need to be low cost, yet provide high throughput. In order to accomplish a low cost per bit of data transferred and provide high bandwidth, adapter cards are moving to incorporate several ports per card. Multiple ports can enhance bandwidth or add redundancy and to do so will require a method to aggregate multiple flows through a single upstream port to the host.

## Adapter I/O data aggregation

One method to fan in data from multiple ports is through the use of an aggregation switch. Using switches based on PCIe technol-

ogy provides low cost, high bandwidth, Quality of Service (QoS), low latency, and other crucial benefits (see Figure 1). The PCIe switch combines two I/O endpoints and allows aggregation to the PCIe four-lane upstream port. Switches allow peer-to-peer support for data transfer to the local CPU for offload-engine processing before being transferred to the x4 port. This prevents the bottleneck of having data sent upstream to the host and back to the local CPU. I/Os could be based on Gigabit Ethernet, Fibre Channel, InfiniBand, or other outside the box type communications standards. The PCIe devices shown in Figure 1 are bridges from one protocol to PCIe technology. In this design, the one lane of PCIe to the I/O is more than sufficient as this allows a net bidirectional flow of 2 Gbps per direction. The PCIe links are aggregated through the switch to the upstream PCIe host. The local CPU or a network processor handles the offload engine computations. The x2 link to the CPU provides the bandwidth to allow wire-speed operation from two ports. In addition to the bandwidth, these switches offer low latency. Switches will be available with latency from ingress to egress ports under 150 ns, even for packets with larger data payloads. This low latency is not possible with many competing standards but is absolutely required in communications systems.

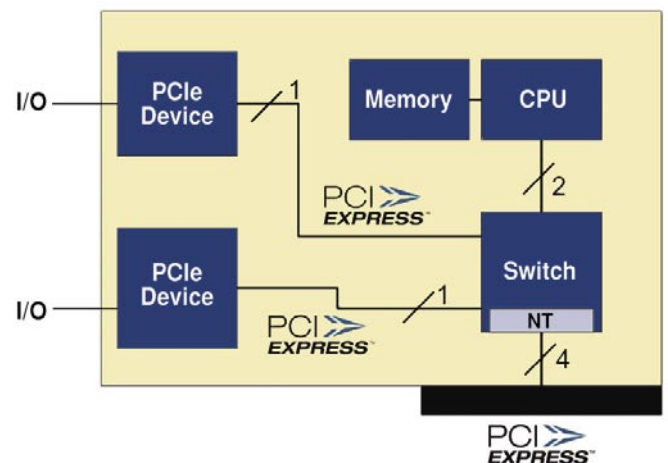



Figure 1

## Processor isolation

Other critical switch features include QoS, end-to-end data integrity, and processor domain isolation. PCIe switches coming to market will provide for multiple Virtual Channels (VCs) along with the mandatory eight types of traffic class. Each class of data is mapped to the VCs available on a link by link basis. The switch

offers a number of selectable VC arbitration techniques to provide the VC prioritization. PCIe technology offers numerous data integrity techniques. One important optional method is based on end-to-end cyclic-redundancy check packet protection. Switches for such adapter applications will provide this capability to ensure reliable transport through an extended series of PCIe links from the adapter through the host side to the root complex. Perhaps the most significant need in these switches is the ability to provide processor isolation. With local intelligence and a system host, system discovery and configuration can result in memory mapping conflicts with these multiprocessor systems. Switches will be enabled with selectable Non-Transparent Bridging (NTB) ports to provide processor isolation. The switch's non-transparent port will prevent the system processor from enumerating and configuring the downstream endpoint devices by making the switch appear to be an endpoint using type "0" configuration status registers. Also provided for NTB is address translation for both address and requestor ID routed packets between these processor domains. Switching through NTB ports does not result in any significant added latency. Next generation systems, soon to be on the market, will take advantage of these advanced PCIe switching solutions. 

***John Gudmundson** is a senior marketing manager for PLX Technology, Inc. Prior to PLX, he held marketing positions at Sunrise Telecom, Nortel Networks, Integrated Telecom, and Ascend Communications. John is an active member of the Advanced Switching Interconnect Special Interest Group (ASI SIG) and is a member of the IEEE and American Marketing Association. He holds a BSEE from the University of California, Los Angeles, a BSBA from the University of California, Berkeley, and an MBA from University of Southern California.*

For further information, contact John at:

**PLX Technology, Inc.**

870 Maude Ave.

Sunnyvale, CA 94085

Tel: 408-328-3500

Fax: 408-774-2169

E-mail: [jgudmundson@plxtech.com](mailto:jgudmundson@plxtech.com)

Website: [www.plxtech.com](http://www.plxtech.com)