The final frontier: PCI Express in backplane applications

By Touseef Bhatti

Over the past two years, PCI Express has taken the tech industry by storm. Its high bandwidth, low pin count, and highly scalable architecture has allowed it to establish itself as the interconnect protocol of choice across all market segments. But as recently as two years ago, PCI Express-based systems were virtually nonexistent, with the exception of graphics systems (PCs). Today, all the major chipset vendors are implementing PCI Express technology into their chipsets. It can be found in servers, workstations, storage systems, Redundant Array of Independent Disks (RAID) controllers, Fibre Channel host-bus adapters, routers, switches, and in an array of instrumentation applications.

The lone market segment that PCI Express has yet to conquer is the backplane application space, where Ethernet maintains its stronghold. Until PCI Express came along, no viable contender had stepped forward. Backplane system designers really had no option other than Ethernet. The prominent backplane specifications today are the AdvancedTCA, MicroTCA, and BladeCenter T specifications (released jointly by IBM and Intel). These specifications require a standard serial interface in the backplane. Therefore, PCI, PCI-X, and other parallel interfaces that were compliant with the CompactPCI specification are no longer options. Routing parallel signals at high clock rates over long distances is a challenge and leads to perils such as reflections and bit-skewing.

The AdvancedTCA and BladeCenter T specifications are fairly generic in that all standard serial protocols are supported, including Serial RapidIO, InfiniBand, and the Advanced Switching Interface (ASI). Serial RapidIO, which was targeted at embedded systems, requires an ecosystem that supports it, such as a Serial RapidIO-compliant processor (for example, PowerPC), switches and/or bridges, and endpoints. Most bladed systems today, however, use an x86-based processor from Intel or AMD, which does not support a Serial RapidIO interface. On the other hand, x86-based processors generally connect to a chipset that supports PCI Express. Interestingly, the new PowerPC processors are adding PCI Express support to their feature sets. Furthermore, all the new chipsets from major vendors have native PCI Express built into the chipset. Hence, while the ecosystem prominent in today's bladed systems favors PCI Express, it does not lend itself to Serial RapidIO.

InfiniBand technology is not ideal for backplanes, either, as it's used primarily for interconnection between high-end, high-density cluster systems. It is a more complex technology with a heavy software burden; its software is not compatible with either legacy PCI, and compatibility with Ethernet requires additional software. System designers are much more eager to change the system's hardware than their software. Moreover, InfiniBand technology has always been a niche solution. It never gained much traction in the mainstream marketplace and, therefore, never caught on in the backplane space.

ASI was primed to be the ideal backplane interface. It built upon the PCI Express technology, with added Quality of Service (QoS) and the capability of supporting multiple hosts via its host-agnostic, address-based protocol. However, the technology never gained traction, as the specification kept getting more and more complex. What began as a few upgrades to the PCI Express specification turned into a major overhaul. Ultimately the complexity of the specification kept increasing while PCI Express continued to gain traction in the marketplace. Now, with PCI Express having gained mass adoption as the interconnect of choice, and PCI Express 2.0 around the corner, ASI has missed its time to market and is no longer a viable technology. PCI Express 2.0 will support 5 Gbps signaling and a new PCI Express extension called I/O Virtualization (IOV) will support multihost capabilities, filling a large need that ASI intended to serve. The ASI-SIG has disbanded and has handed the technology over to PICMG.

While all these technologies were proving to be unsuitable for backplanes, designers were left with two choices: go with a proprietary solution or use Ethernet. Proprietary solutions using ASICs quickly proved to be too expensive. Ethernet, on the other hand, was a tried-and-true solution with existing software. So, Ethernet took the lead and has never looked back. However, an argument can be made that PCI Express is poised to unseat Ethernet in backplanes.

Ubiquity

When PCI Express first arrived, designers were not willing to abandon Ethernet just yet. After all, they had spent years building upon the same software. From Ethernet to GbE to 10 GbE, Ethernet had proven itself to be a dependable backplane interface; its ubiquity was its strongest asset.

But with its mass adoption over the past few years, PCI Express has become ubiquitous itself. Most chipsets today support PCI Express, and due to its versatile nature, PCI Express can communicate with protocols such as Ethernet, Fibre Channel, InfiniBand, and legacy PCI and PCI-X systems. Moreover, since PCI Express is based upon legacy PCI software, minimal software investment is required.

With the emergence of PCI Express, most blades connecting to the backplane are running PCI Express (see Figure 1). As mentioned earlier, both x86 and newer PowerPC processors support PCI Express, and all new chipsets support PCI Express as well. Designs requiring a mid-plane, or connectivity to the backplane, are also using PCI Express to fan out to (or aggregate from) connectivity modules (see Figure 2). The only piece of traffic not running on PCI Express is the backplane traffic. If the backplanes can also be designed around PCI Express, the entire system can run off of one protocol.

PCI Express is a suitable *in the box* solution, serving as the best chip-to-chip

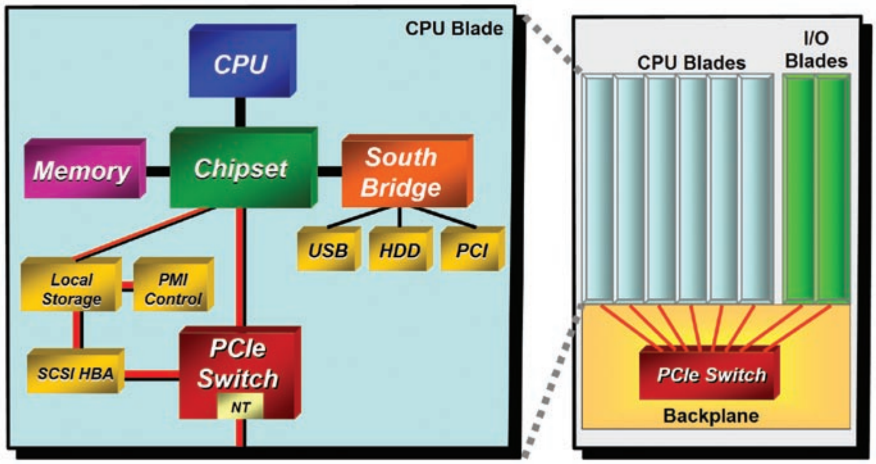

Figure 1

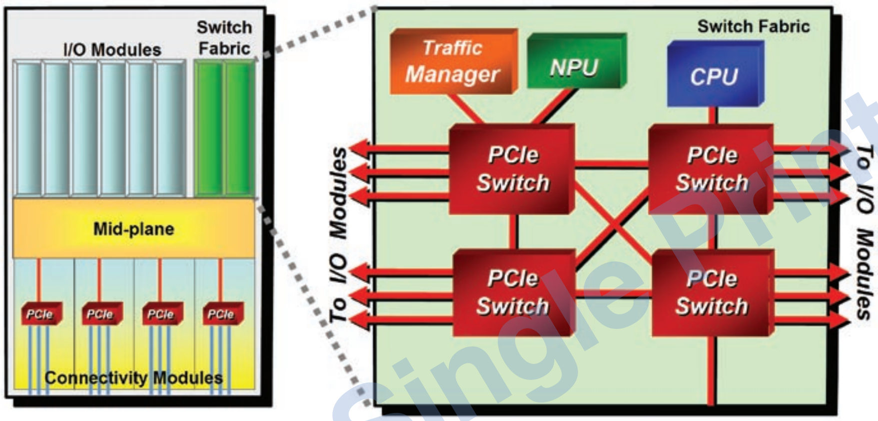

Figure 2

interconnect, while Ethernet provides a viable *out of the box* solution. With the recent introduction of the PCI Express Cable Specification, PCI Express capabilities have been extended by a few meters. Nonetheless, PCI Express is not in any position to replace Ethernet in the LAN/MAN/WAN spaces. The backplane, however, is still very much *in the box*, where PCI Express strengths lie.

Scalability

The AdvancedTCA and BladeCenter T backplane standards support up to four lanes per switch fabric interface. This means a x4 PCI Express link can be supported per blade by backplanes and yield 10 Gbps (2.5 Gbps x 4) theoretical bandwidth in each direction. But the true aggregate throughput is 80 percent of that (due to 8b/10b encoding), which yields 8 Gbps in each direction. This falls short of the 10 Gbps (3.125 Gbps x 4 x 80 percent) supported by 10 GbE. However, very few applications today require an aggregate data rate of 10 Gbps. For most applications, especially backplane applications, even 8 Gbps is more than enough.

PCI Express is also far more scalable than Ethernet. With PCI Express, designers can choose to utilize a 1-, 2-, or 4-lane link. These links would yield a data rate of 2 Gbps, 4 Gbps, or 8 Gbps. Ethernet provides a data rate of only 10 Mbps, Fast Ethernet provides 100 Mbps, GbE provides 1 Gbps, and 10 GbE provides 10 Gbps. The scalability is not as efficient. With increasing throughput requirements, GbE has already been ruled out of the discussion. Even one lane of PCI Express will yield a data rate of 2 Gbps, double that of GbE. The only option after GbE is 10 GbE, a jump from 1 Gbps to 10 Gbps. Ethernet is not as scalable as PCI Express, where data rates of 2 Gbps (x1), 4 Gbps (x2), or 8 Gbps (x4) are achievable.

Latency

In addition to its scalability assets, PCI Express has lower latency and, therefore, much less overhead than does 10 GbE. PCI Express has three D-words of overhead (packet header) for 32-bit processors and four D-words of overhead for 64-bit processors. Ethernet, on the other hand, has 16 D-words of overhead. In terms of bytes, for every four bytes of data (or payload), PCI Express requires a 20-byte packet for four bytes of data, meaning only 16 bytes of overhead. Ethernet, on the other hand, requires a 64-byte packet for every four bytes of payload, meaning 60 bytes of overhead. This means that, with Ethernet, more CPU cycles are being utilized on processing the overhead rather than the actual data transfer. Because of its larger overhead, Ethernet's latency is higher than that of PCI Express, so in latency-sensitive systems, PCI Express will prove to be more efficient than Ethernet.

The primary reason for this higher latency is that Ethernet requires processing of the TCP/IP stack, while PCI Express does not. The impact of the header depends on the granularity of the data. If large data packets are being transferred, the impact of the header is less. However, if smaller packets of 256 bytes are being transferred, the impact of the header is magnified quite a bit. This is the reason PCI Express is often used for control plane applications where the packet sizes are small and require processing by the CPU.

Quality of Service

Last but not least, PCI Express provides better QoS than Ethernet does. PCI Express offers guaranteed error-free packets and delivery, while Ethernet does not. Per the PCI Express specification, PCI Express does not allow for dropped packets and checks acknowledgements, or ACKs, at each hop to ensure that the integrity of the data is maintained and that no packets are dropped. In Ethernet, however, dropped packets can cause large hits to bandwidth. With no hardware error recovery in Ethernet, the system does not know that a packet has been dropped until the data reaches its destination. At that time the system will request for the packet to be resent all over again from its origin. This leads to oversubscription, which in turn erodes the 10 Gbps of available bandwidth.

Going forward

Due to the throughput requirements, those insisting on sticking with Ethernet will be forced to make the jump from GbE to 10 GbE. Since 10 GbE is still relatively new, it can be pricy to implement. Although the investment in Ethernet software is preserved, 10 GbE requires new hardware that is still quite pricy. PCI Express has been in the market for enough time to establish itself technically, and its pricing is competitive as more suppliers are entering the market. While PCI Express is already providing multihost support via the Non-Transparent Bridging de facto standard, with the PCI Express Base Specification 2.0 recently finalized, PCI Express will now also offer multihost support through the new IOV standard. Now, PCI Express can be used to connect to multiple CPU blades in the backplane without needing extra devices for host isolation.

Conclusion

Although Ethernet has been the tried-and-true backplane interface, PCI Express boasts greater scalability, lower overhead, lower latency, higher security, and better cost efficiency than 10 GbE. Ethernet's ubiquity kept it ahead of the pack, but the mass adoption of PCI Express is rapidly gaining steam with no signs of slowing down. Designers are now starting to realize that for applications utilizing smaller payloads and/or packets requiring heavy processing, PCI Express proves to be more efficient than Ethernet. Also, vendors such as PLX Technology are now offering high port-count switches targeted at backplane and control-plane applications. As the PCI Express technology and market matures, vendors and users alike will continue to enhance their understanding of PCI Express and how to better implement it into the backplane.

Touseef Bhatti *is senior product marketing engineer at PLX Technology. He has a BSCE from the University of Michigan.*

For more information, contact Touseef at:

PLX Technology
870 Maude Avenue
Sunnyvale, CA 94085
800-759-3735
tbhatti@plxtech.com
www.plxtech.com