Switched-Fabric Interconnects

William Carson
VMETRO Inc. Field Application Engineer
1880 Dairy Ashford, Suite 400
Houston, TX 77077
281-584-0728
wcarson@vmetro.com

Tom Bohman
VMETRO Inc. VP Business Development
1880 Dairy Ashford, Suite 400
Houston, TX 77077
281-584-0728
tbohman@vmetro.com

BACKGROUND

In this paper, we discuss the various switched-fabric interconnect architectures being developed to increase I/O throughput. Several technologies are competing to address the I/O bottleneck of current shared bus architectures.

Most system architects would agree that I/O bottlenecks are the key aspect holding back increased computer performance. To address this need, several standards bodies are racing to develop switched-fabric interconnects, resulting in a confusing multitude of proposed solutions. Each of these standards has its own champions debating the relative merits of their solution versus the rest. One sure bet is that at least one of these technologies will be the next architecture that enables computer performance increases.

The five leading technologies vying to be the next interconnect architecture are the product of their respective trade groups:

- INFINIBAND Trade Assoc.,www.infinibandta.org
- RapidIO Trade Assoc., www.rapidio.org
- HyperTransport Technology, www.hypertransport.org
- PCI Express, Spec, PCI SIG, www.pcisig.com
- StarFabric Trade Assoc., www.starfabric.com

There are others, but they lack the critical mass to compete in this ever-tightening competition. More than at any other time, the 'who-you-know' is as important as what you know. The winner of this war will bring the right combination of marketing savvy along with technical competency. Complex teaming arrangements are as much a part of this competition as are the technicians in the lab.

Categories and Subject Descriptors

High-Speed Interconnect Technologies

General Terms

Design, Standardization, Networking, Communications, Theory.

Keywords

InfiniBand, RapidIO, HyperTransport, PCI Express, StarFabric, networking, communications, protocols, switched fabrics, embedded systems, VXS, Serial RapidIO, StarFabric.

1. INTRODUCTION

This paper compares and contrasts the five competing technologies, looking for relative strengths and weaknesses. From a hardware perspective, we look at the physical topologies such as, signaling technologies, link layers, along with such key attributes as latencies, and sustained throughput. The many differences here can have a significant impact on performance.

From a software and firmware standpoint, we will examine software overhead required for framing and routing and the transition path from bus-based systems.

2. CAPABILITIES

2.1 Common Capabilities

From the five (5) competing technologies come six (6) competing interconnects, because RapidIO has both parallel and serial incarnations. The trade groups behind these proposed solutions are battling to be your next computer interconnect architecture.

To clarify the evaluation, we look at the parallel interconnects separately from the serial ones. The parallel interconnects, HyperTransport and Parallel RapidIO, are best suited for the chip-to-chip connection (i.e., the local bus on the board or board-to-board inside-the-box of a microcomputer). The two parallel standards provide little in high level functionality, because this is not needed when dealing with local links.

The four (4) serial interconnects, InfiniBand, Serial RapidIO, PCI Express, and StarFabric, are targeted to applications board-to-board and box-to-box. Of these four, the PCI Industrial Computers Manufacturers Group (PICMG®) has designated three as next-generation telecommunications standards, AdvancedTCA™ - Advanced Telecom Computing Architecture. InfiniBand has been designated PICMG 3.2, StarFabric is PIMG 3.3, and PCI Express was recently added as PICMG 3.4.

The horizontal axis in Figure 1 shows the system levels that each of these interconnects targets. Inside the box (left side), we see a tight grouping of standards vying for the internal interconnect. As we venture further out of the box (toward the right), we reach a point where only InfiniBand remains, reaching out toward the existing Internet.

The vertical axis in Figure 1 depicts the product timeline to convey the relative maturity of each product. The higher vertical location of a 'bubble' implies a further advanced product.

2.1.1 Serial Interconnects

For all of the serial contenders, several issues should be retired from this discussion, because they are common to all. First is scalability, because all of these point-to-point switched fabric interconnects offer excellent scalability.

Secondly, we may ignore the signaling technology at the physical layer. InfiniBand, Serial RapidIO, PCI Express, and StarFabric use the standard SERDES (serializer deserializer) that incorporate 8B/10B encoding for line balancing and clock extraction. This well proven hardware comes from Fibre Channel.

The common link data rate is based upon 2.5 Gigabits per second, except for StarFabric at one-fourth that speed (622 Mb/s).

Throughput is scaled up by providing multiple paths or “lanes”. All of these interconnects use LVDS signaling and have bidirectional capability.

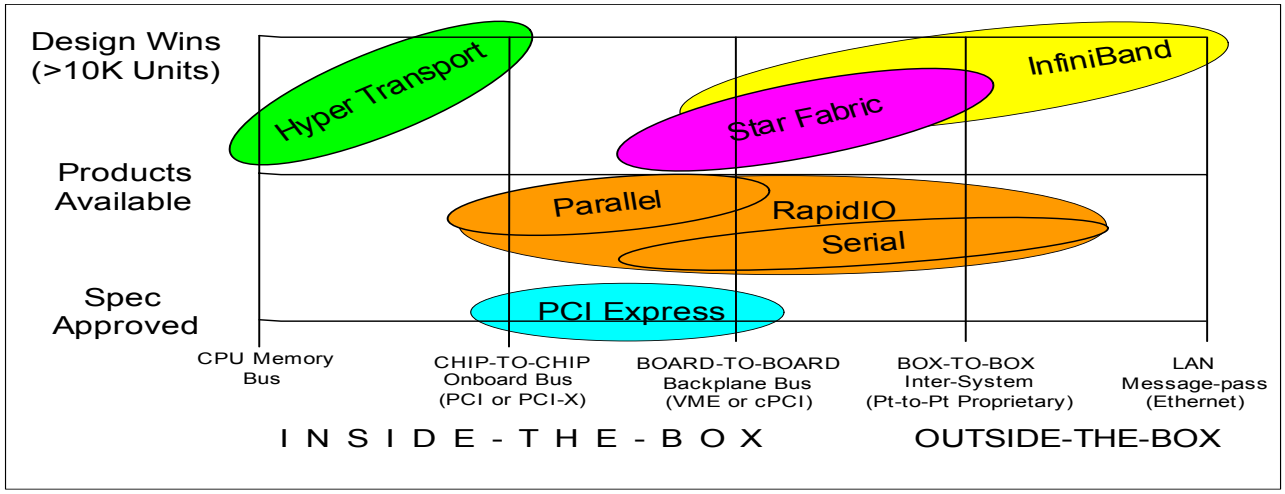

Figure 1: Interconnect Battlefield

2.1.2 Parallel Interconnects

As we see in figure 1, the parallel interconnects (HyperTransport and Parallel RapidIO) provide support for chip-to-chip interconnects. Parallel architectures are limited in deployable distance by clock skew; the disparity in signal time of arrival in seemingly identical parallel circuits. Short runs, as in onboard chip-to-chip, have minimal skew. Conversely, the serial are not as appropriate for the chip level. The extra resources required for serialization and de-serialization are wasted in short links.

2.2 Unique Capabilities

2.2.1 InfiniBand (Sys I/O, NGIO, Future I/O)

InfiniBand is emerging as the leader in High-Performance Embedded Computing (HPEC) systems and computer cluster intercommunication. With industry backing, InfiniBand is appearing in data centers from academia to industry. InfiniBand is said to be able to co-exist with Fibre Channel (FC) and Ethernet. The capability of InfiniBand to reach between the Server and the Internet is unique within this group.

2.2.2 Parallel RapidIO

RapidIO is a rapidly maturing technology, although it is still in the design phase. Parallel RapidIO is targeted to in-the-box applications to connect chip-to-chip and board-to-board. Thanks to its parallel architecture and use of classical load/store methodology, parallel RapidIO is in the comfort zone of bus users. In the area of parallel local bus interconnects, HyperTransport is the key competitor of Parallel RapidIO.

2.2.3 Serial Rapid I/O

The latest entrant into the interconnect war is Serial RapidIO. The RapidIO Trade Association issued Serial RapidIO Specification to enable a larger reach for RapidIO links. To prove its quick acceptance, Serial RapidIO has already been defined in the VMEbus Switched Serial (VXS) by the VMEbus Standards Organization under VITA 41. This Serial variant of the original RapidIO parallel architecture fills critical gaps in the architecture, while the common Transport layer unifies the two physical specs.

2.2.4 HyperTransport (formerly LDT)

HyperTransport and Parallel RapidIO are the two (2) parallel standards we are examining. HyperTransport is primarily a product of AMD and was started as a processor interconnect for AMD and MIPS-based CPUs.

HyperTransport has been described as host-centric daisy-chained model. This means that data flows between the host CPU and devices directly attached such as RAM and local peripherals. More recently, Hyper Transport has moved towards a switch-based paradigm. As with the other interconnects shown on the left side of figure 1, this operates as a load/store model. Any interconnect design targeting the local bus must use this model.

2.2.5 PCI Express (formerly 3GIO)

PCI Express (formerly Intel’s 3GIO) has a long history. PCI-SIG has endorsed this Intel entry into the fabric war. A major asset for PCI Express is its compatibility with the current PCI software environment. Having backward compatibility makes PCI Express appear to be the low risk approach, however, this attribute also symbolizes excess complexity versus a completely new interconnect.

PCI Express is a serial, load/store, packet based architecture. Each link is capable of up to 8 gigabytes per second. Multiple lane widths available are 1, 2, 4, 8, 12, 16, and 32 bits. Embedded clocking scheme enables extensive frequency scalability.

2.2.6 StarFabric

StarFabric was named by StarGen, a fabless semiconductor company. Marketed as an evolutionary step from PCI, StarFabric plans to use PCI Express AS (Advanced Switching) as its roadmap to higher speeds. StarFabric has been called InfiniBand Light. The primary difference between StarFabric and the other standards is the choice of low-cost, serial, physical-layer using 622Mbps LVDS links versus the 2.5 Gbps physical links. These

links are capable of bi-directional 1.25Gbps total bandwidth. Each StarFabric link is made up of four such pairs, creating a 2.5 Gbps full duplex link ‘port’.