Filename: SAS - PCI Express Bandwidth - Infostor v5.doc

Maximizing Server Storage Performance with PCI Express™ and Serial Attached SCSI

Article for InfoStor November 2003
Paul Griffith – Adaptec, Inc.

Server Storage Trends

Tightening IT budgets are increasing pressure on organizations to use system resources with greater efficiency as they look to data centers to serve more clients. More than ever, IT managers must choose technologies that make the most efficient use of their server and storage components as the amount information they manage continues to grow. One system resource that has come under increasing scrutiny is the storage system. The reason: these systems are at the heart of a data center's ability to make digitized data available 24x7, seamlessly scale capacity and increase server storage performance while reducing the cost of managing information.

Data centers have relied on the parallel bus interface as the workhorse disk drive interconnect for more than 20 years. Parallel technologies remain in widespread use and continue to meet the performance requirements of many of today's enterprise solutions. However, significant technical challenges have made it economically impractical to use parallel technologies to meet increasing demand for more robust data integrity, higher system performance, greater storage flexibility and scalability, and smaller system designs.

On a parallel bus, for example, slight length variations and other small differences in individual wires make it difficult to synchronize the arrival of signals at the receiver and to determine when it needs to sample (or clock) the signals, a process called clock skew, one factor that limits system performance. In addition, parallel signal transmissions are susceptible to crosstalk across wide data paths that add line noise and can cause signal errors – a pitfall that has been remedied by slowing the signal, limiting trace length or both. Terminating parallel signals is another difficulty, requiring that individual lines be terminated, usually by the last device, to avoid signal reflection at the end of a cable. Finally, parallel's large cable and connector size make it unsuitable for increasingly dense computing environments.

Serial architectures have emerged to deliver higher performance by allowing more bandwidth per device pathway than their parallel counterparts. Serial architecture connections consist of a single pair of transmission signals that contain an embedded clock for self-clocking, enabling clock speed to be easily scaled. Serial bus architectures also support a network of dedicated point-to-point device connections, versus the multi-drop architectures of parallel buses, to deliver full bandwidth to each device, eliminate the need for bus arbitration, reduce latency, and greatly simplify hot-plug and hot-swap system implementations. This dedicated serial connection also eliminates the single point of failure found in today's parallel environments.

Serial Technologies Deliver New Capabilities for Next-Generation Server Storage

Serial storage technology, specifically Serial ATA, Serial Attached SCSI and PCI Express, address the architectural limitations of their parallel counterparts to deliver highly scalable performance. The technology draws its name from the way it transmits signals – in a single stream, or serially, compared to multiple streams for parallel. The main advantage of serial technology is that while it moves data in a single stream, it wraps data bits into individual packets that are transferred up to 30 times faster than parallel technology data.

In addition, serial technology's point-to-point architecture features dedicated connections that deliver full bandwidth to each device.

Serial ATA extends the parallel ATA technology roadmap by delivering disk interconnects speeds starting at 1.5 Gb/sec (150 MB/sec). Due to its lower cost-per-gigabyte, Serial ATA will continue as the prevalent disk interface technology in desktop PCs, sub-entry level servers and sub-entry level networked storage systems where cost is a primary concern.

Serial Attached SCSI, the successor technology to the parallel SCSI interface, leverages proven SCSI functionality and features while expanding SCSI's proven performance, scalability and reliability for enterprise storage. Serial Attached SCSI offers many features not found in today's mainstream storage solutions such as drive addressability up to 16,000 devices per port, and reliable point-to-point serial connections at first-generation speeds of up to 3 Gb/sec. In addition, Serial Attached SCSI's small connector supports full dual-ported connections on 2.5-inch hard disk drives, a feature previously found only on larger 3.5-inch Fibre Channel disk drives. Dual-ported connections are essential for applications that require redundant drive spindles in a dense server form factor such as blade servers.

The Serial Attached SCSI interface will also be compatible with lower cost-per-gigabit Serial ATA drives, giving system builders the flexibility to integrate either Serial Attached SCSI or Serial ATA devices and substantially reduce procurement, inventory and other costs associated with supporting two separate interfaces.

Serial Attached SCSI improves drive addressability and connectivity using an expander that enables one or more Serial Attached SCSI host controllers to connect up to 128 ports, which may include other host connections, other Serial Attached SCSI expanders for even greater scalability, or hard disk drives. Connecting multiple expanders together will achieve connectivity to over 16,000 devices. This highly scalable connection scheme enables enterprise-level topologies that easily support multi-node clustering for automatic failover availability or load balancing.

PCI Express, a new serial host interconnect architecture, is designed to address a wide range of current and future system interconnect requirements by delivering the flexibility, scalability, and performance bandwidth needed to support upcoming technologies like 10 Gbit Ethernet and Serial Attached SCSI. A point-to-point architecture with hot-plug and hot-swap support, PCI Express is software-compatible with PCI and PCI-X to simplify the design of next-generation serial computing systems.

PCI Express uses a dual simplex serial data stream with an embedded clock to overcome many of the performance limitations of parallel bus architectures. A PCI Express link consists of two low-voltage, differentially driven pairs of signals: a transmit pair and a receive pair. And like Serial Attached SCSI, a data clock is embedded using the 8b/10b encoding scheme to achieve very high data rates. Each point-to-point interconnect may have 1, 2, 4, 8, 12, 16, or 32 dual simplex 2.5 Gb/sec lanes (2.0 Gb/sec effective rate), providing scalable bandwidth up to 128 Gb/sec (16 GB/sec) between nodes. In comparison, a typical 64-bit, 133 MHz PCI-X 1.0 device provides approximately 1 GB/sec bandwidth.

Wide SAS Ports

First-generation Serial Attached SCSI will deliver throughput of 3Gb/sec per link and succeeding generations up to 12 Gb/sec to keep pace with technology and application advances. In addition, Serial Attached SCSI's full-duplex, point-to-point architecture supports simultaneously active connections among multiple initiators and high-performance Serial Attached SCSI targets. Devices can transfer data in both directions at once to effectively double the useable bandwidth of the link rate. These multiple links, in turn, can be combined into wide ports, allowing system designers to aggregate the performance of Serial Attached SCSI initiators and Serial Attached SCSI expanders to increase

total available bandwidth. Grouping four or eight links, which is typical, can produce bandwidth of 12 Gb/sec or 24 Gb/sec, respectively.

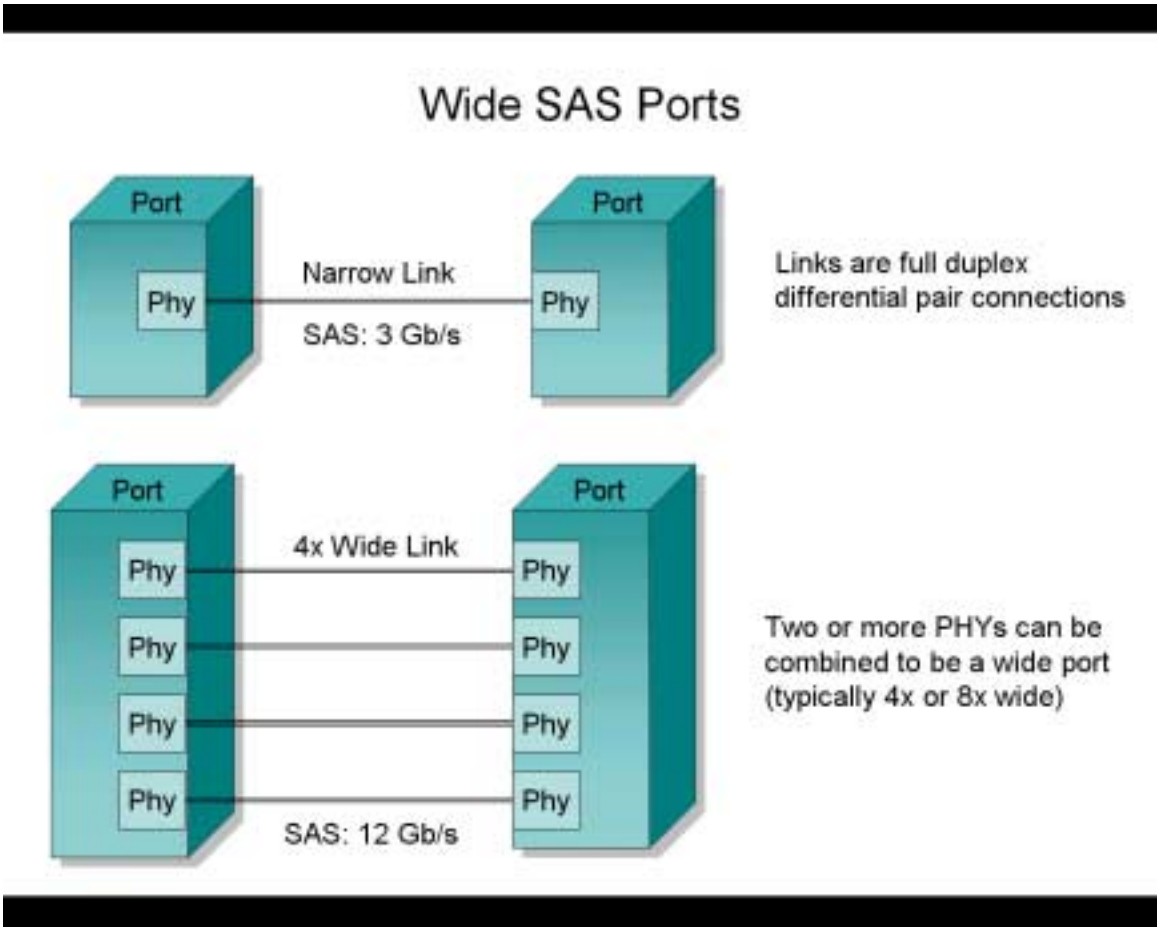

Figure 1. Multiple PHYs can be combined to form wide ports that support the significant bandwidth requirements of large Serial Attached SCSI topologies.

Serial Attached SCSI Expanders

The scalability of parallel buses is limited because they share connection paths, and adding more buses with multiple initiators does little to extend this limited sharing ability. Serial Attached SCSI uses expander hardware as a switch to simplify configuration of large external storage systems that can be easily scaled with minimal latency while preserving bandwidth for increased workloads. This expander hardware enables highly flexible storage topologies of up to 16,256 mixed Serial Attached SCSI and Serial ATA drives.

One type of expander, a fan out, for example, can connect up to 128 devices per each PHY, including initiators, Serial Attached SCSI and Serial ATA drives, and edge expanders with either narrow or wide links. These additional initiators and edge expanders can in turn be linked to other hosts and drives, providing additional connection nodes. The SCSI Management Protocol (SMP) within Serial Attached SCSI manages the point-to-point connections in the topology.

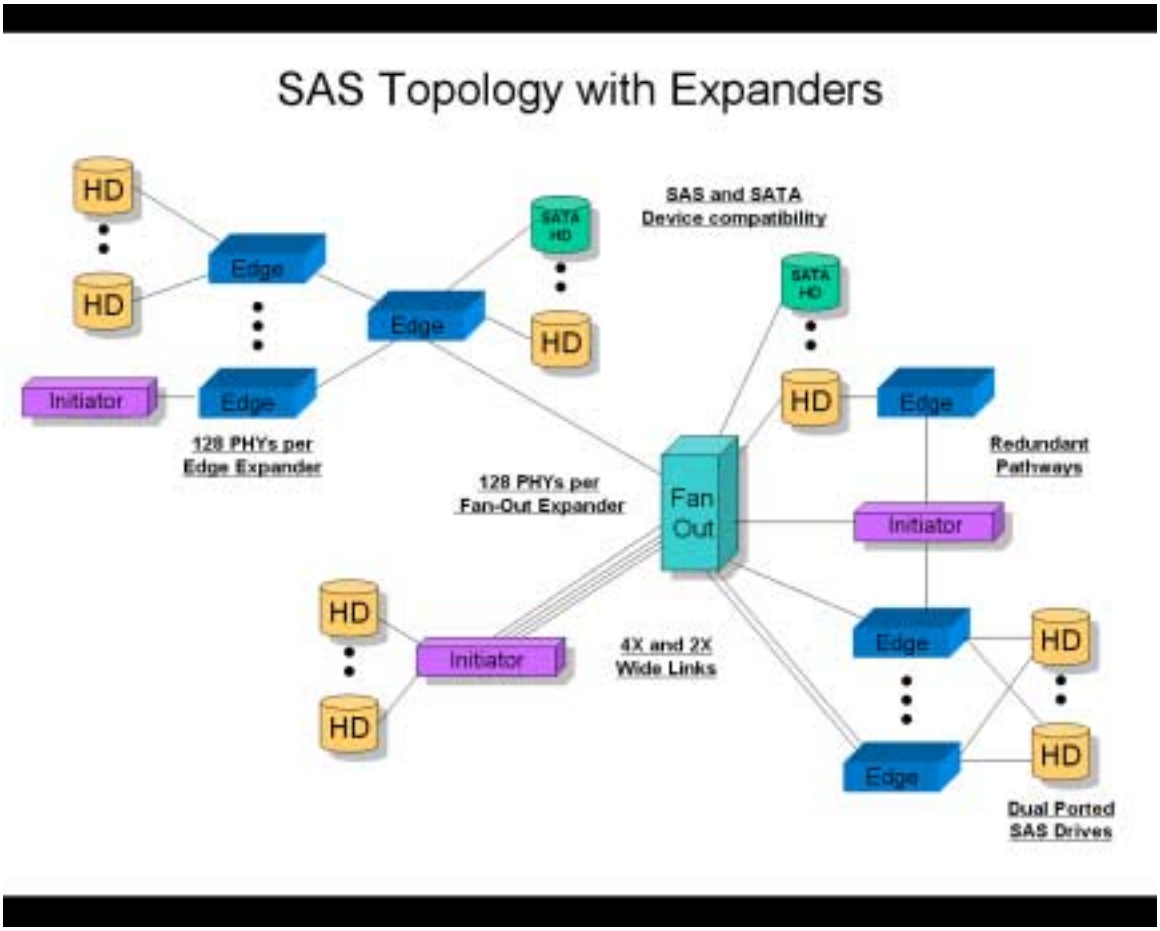

Figure 2. Expanders enable the design of very large storage topologies. Each fan out expander can be connected to up to 128 physical devices per PHY, including multiple initiators and other edge expanders.

SAS Bandwidth

First-generation Serial Attached SCSI link rate is 3 Gb/second (300 MB/sec) and supports full duplex data transfers for up to 600 MB/sec bandwidth. And because SCSI protocols are not restricted to half duplex operation, Serial Attached SCSI will also support full duplex transfers, allowing data to be transferred simultaneously in both directions to maximize bandwidth. For example, a device can simultaneously transfer data from a previously queued read operation while receiving data for a write operation. Although full duplex will not be used during all transfers, this feature can double the useable bandwidth of the link rate. In contrast, Ultra320 SCSI's shared bus architecture is restricted to 320 MB/sec for all attached devices. Serial Attached SCSI's support for wide ports further improves throughput by enabling several disks to communicate with a single port address simultaneously.

For example, a Serial Attached SCSI controller with four 3 Gb/second links configured as a wide port will support data transfer rates of 1200 Gb/sec or 1.2 GB/sec at half duplex. A Serial Attached SCSI controller configured with eight links supports data rates of 2.4 GB/sec. or 2400 MB/sec at half duplex.

Currently, a 15,000-RPM disk drive will sustain data rates up to 75 MB/sec. At these sustained data rates, two disk drives will saturate a Serial ATA 1.5 GB/sec bus. The shared Ultra320 SCSI bus supports a total of 320 MB/sec or the sustained data rates of four to five disk drives. By contrast, a 4-wide Serial Attached SCSI port supports as many as 16 hard drives before becoming saturated.

Serial Attached SCSI's ability to aggregate bandwidth through the use of wide ports will support the performance scalability required by next-generation servers and storage systems. However, while Serial Attached SCSI can supply the bandwidth for next-generation storage I/O, it requires a proficient host interconnect to optimize total system performance.

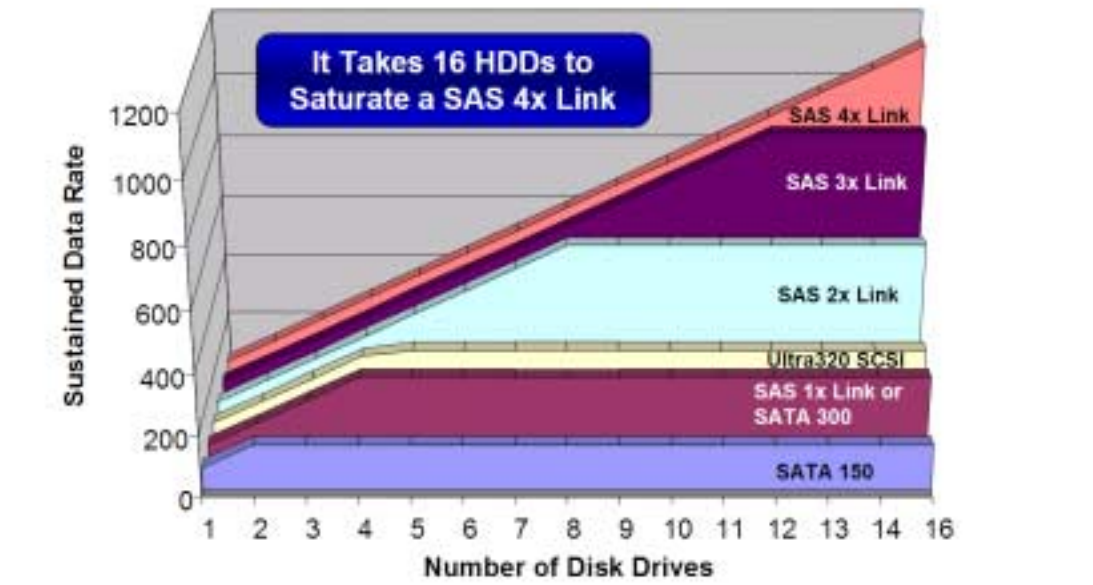

Figure 3. By combining links, Serial Attached SCSI supplies the bandwidth needed to support the sustained data rate of many attached disk drives. Assuming each drive can sustain a data rate of 75MB/sec, a Serial Attached SCSI topology with a 4-wide port could support up to 16 disk drives before throughput saturation.

Serial Attached SCSI and PCI Express: Maximizing Server Storage Performance

System architects typically optimize performance by eliminating bandwidth bottlenecks – a goal typically met by matching interleaving technologies with complementary efficiency and availability levels.

Like Serial Attached SCSI, PCI Express delivers scalable performance by combining multiple data links to create wide data paths. This common capability is the key to optimizing performance between Serial Attached SCSI and PCI Express.

By combining Serial Attached SCSI with PCI Express, a system designer can easily generate bandwidth for at least 16 hard disk drives, with neither technology bottlenecking the performance of the other. With today's drives generating sustained data rates of about 75 MB/sec, 16 drives will require about 1200 MB/sec bandwidth. A Serial Attached SCSI port configured as 4-wide supports 1200 MB/sec while a PCI Express slot configured for 8-wide supports 2048 MB/sec.

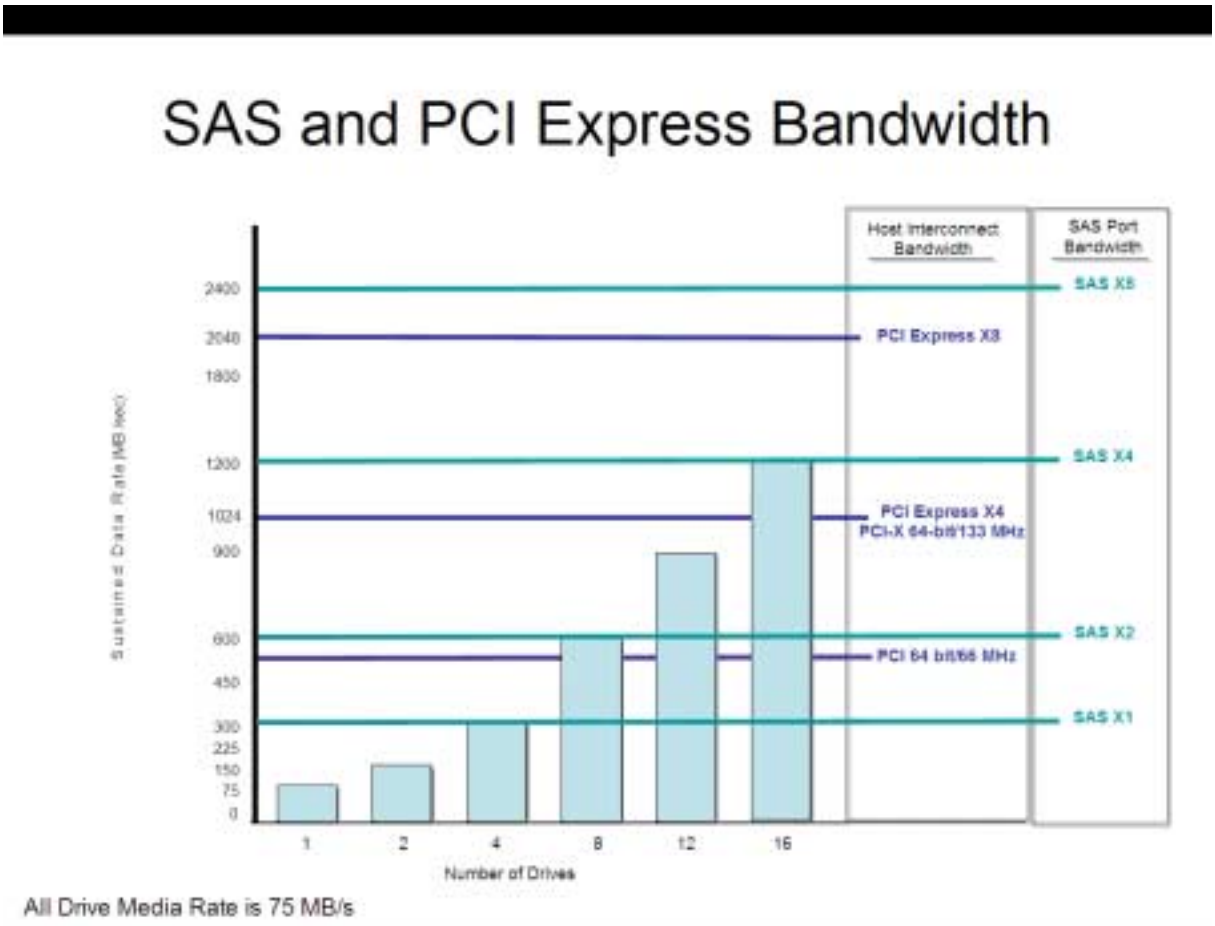

Figure 4. A 4-wide Serial Attached SCSI port supports bandwidth of 1200 MB/sec, enough for up to 16 attached drives. A Serial Attached SCSI host attached to a PCI Express 8-wide lane, which can accommodate 2048 MB/sec, delivers the system bandwidth to support this streaming data rate.

Summary

As data centers are called on to serve more clients, IT managers must choose technologies that optimize the capabilities of their server and storage components. Deploying complementary components within the system is a vital step in that direction.

New serial technologies are emerging to overcome the bandwidth limitations of today's parallel architectures and deliver highly scalable performance for next-generation systems. Serial Attached SCSI delivers a 3.0 Gb/sec per link data transfer rate, the unprecedented flexibility to deploy either low cost-per-gigabit Serial ATA drives or high-performance Serial Attached SCSI drives in the same system, point-to-point connections for high reliability, and highly scalable connectivity to more than 16,000 devices in a single domain. PCI Express offers similar benefits, with each PCI Express lane supporting 2.5 Gb/sec performance with scalability up to a 32-wide lane configuration.

By combining Serial Attached SCSI's wide ports with PCI Express's wide lanes, a system designer can maximize total storage system performance. For example, a 4-wide Serial Attached SCSI port will support bandwidth of up to 1200 Gb/sec for up to 16 hard disk drives. And with the Serial Attached SCSI initiator attached to a PCI Express 8-wide lane, the entire data path through the system will be optimized with bandwidth to spare.

About the author

Paul Griffith is a Strategic Marketing Manager at Adaptec Inc. For more information, visit: http://www.scsita.org.