

October 2004

SAS: The Path to Maximum SATA Scalability

A JOINT WHITEPAPER BY:
LSI Logic and Seagate Technology



www.lsillogic.com

www.seagate.com

Introduction

The arrival of Serial Attached SCSI (SAS) marks a new era in storage scalability, wherein both the *type* and quantity of drives can easily be optimized. SAS compatibility with Serial ATA (SATA) enables seamless deployment of desktop-class SATA drives and enterprise-class SAS drives *in the same SAS domain*, giving IT managers unprecedented flexibility to specify the most appropriate drive for both online (transactional, high availability) and nearline (archival, low availability) duties. Moreover, employing a common SAS infrastructure minimizes hardware redundancy (for example, a single enclosure can house both types of drives), further enhancing efficiency.

But there is another, less immediately obvious benefit to SAS/SATA compatibility: It boosts SATA scalability far beyond the limits imposed by SATA-based infrastructures. Deploying hundreds, even *thousands*, of SATA drives in a single SAS domain is a straightforward affair, requiring only standard SAS host bus adapters (HBAs) and expanders.

Simply put, enterprise-class SAS infrastructure enables SATA disc drives to transcend the scalability constraints inherent in their desktop DNA.

Rooted in the Desktop

As an evolutionary development of parallel ATA, it's no surprise that SATA shares its predecessor's emphasis on desktop storage. Of course, SATA's modern serial architecture brings a host of improvements (faster throughput, improved scalability, no master/slave and termination issues, compact cabling and connectors). But the fundamental objective remains the same—delivering desktop-class storage at the lowest cost/GB.

To achieve this goal, the authors of the SATA 1.0 specification wisely focused on those capabilities applicable to desktop environments—incorporating additional features of marginal relevance would only serve to needlessly drive up costs. Scalability, of course, is not a primary consideration among desktop users, who seldom have reason to install more than one additional drive. As such, SATA's limited scalability should not be viewed as an architectural flaw, but rather a logical by-product of diligent efforts to maximize SATA's cost-effectiveness.

SATA II and Port Multipliers

But as enterprises increasingly turn to SATA disc drives for nearline, backup/restore and other low-availability storage duties, the need for greater scalability has steadily grown. The new SATA II specification addresses this rising demand with the advent of devices known as Port Multipliers (PMs). While SATA 1.0 allows only one drive per host controller port (requiring additional ports to accommodate extra drives), SATA II's Port Multiplier functions like a hub, enabling each PM-equipped port on the host controller to connect up to 15 drives. (Note that PMs with four- or eight-drive connectivity will likely prove more common).

To be sure, Port Multipliers clearly enhance SATA's scalability in workstation and high-performance PC environments. Video editing, audio editing, graphic design and video games are just a few of the applications that can benefit from the improved performance and capacity of PC RAID configurations utilizing Port Multipliers. Furthermore, such RAID configurations can significantly boost the effectiveness of low-end servers. Nevertheless, Port Multipliers fall short of meeting the enterprise's needs in a data center environment. Specifically, SATA infrastructure, including its Port Multipliers, presents challenges in the following areas:

- **Compatibility**

Port Multipliers require SATA II host controllers that are Port Multiplier-aware; legacy SATA 1.0 controllers will need to be replaced.

- **Expandability**

Unlike conventional hubs, Port Multipliers cannot be linked together, severely limiting their flexibility and expandability in a network environment. PM's can quickly become prohibitively complex as drive quantities reach enterprise-class levels. For example, based on the theoretical limit of 15 drives per PM, deployment of 100 SATA drives would require *seven* PMs. Connecting 100 drives via the more common eight-port PM configuration would require *13* Port Multipliers.

- **Performance**

Switches are the *de facto* enterprise standard because they enable simultaneous communication between multiple initiators and targets. But Port Multipliers support only one active host connection, significantly slowing effective throughput. Furthermore, each time communication is initiated with a drive, a time-consuming drive reset must occur.

- **Data Integrity**

Enterprise storage infrastructures typically employ a multitude of disc drives; addressing protects data integrity by assigning a unique address to each drive, thus ensuring data is consistently directed to the correct drive. Desktop environments (entailing minimal drive counts) don't require such addressing capabilities, thus SATA does not support them. Furthermore, Port Multipliers don't allow persistent drive connections; the host may only address one drive at a time, the PM dynamically closing the connection to one drive and opening a new connection to another. With each closed connection drive history (e.g., data source, destination drive, data and command context) is lost, thus with each opened connection the chance of misidentification and sending data to the wrong drive is increased.

- **Reliability**

Port Multipliers offer only passive failover capability vs. the added security of active failover. Should a PM's primary host controller fail, any backup controller must be manually configured to restore PM function. Active-failover devices employ dual ports to ensure uninterrupted service; should one controller fail, the device *automatically* switches ports to access the remaining controller.

- **Cabling**

Lack of flexibility in PM/drive deployment is further exacerbated by the short cable lengths (one meter) permitted in the SATA specification.

These issues preclude Port Multipliers from efficiently meeting the enterprise's demand for greater SATA drive scalability. But there is an elegant alternative, one that delivers both extensive scalability and unmatched flexibility. Not surprisingly, it comes from an interface purpose-built for the rigors of enterprise use (high availability, 24/7 duty cycles): Serial Attached SCSI.

SAS: Superset of SATA

Leveraging their common serial, point-to-point architecture, SAS encompasses all of SATA's virtues and then surpasses them with a comprehensive range of enterprise-class capabilities far beyond those of its desktop-centric sibling. Specifically designed as a superset of SATA, SAS is able to synergistically interoperate with SATA, significantly enhancing the value of both technologies.

SAS, of course, is optimized for online, high-availability applications in the most demanding enterprise environments. To that end, it incorporates an impressive array of strengths (full-duplex, dual-port operation for maximum transfer rates and failover capability, rock-solid reliability, rich and mature SCSI command set, advanced command queuing, sophisticated verification/error correction) to deliver the throughput and dependability mission-critical environments demand.

But the very strengths that make SAS disc drives an ideal performance solution render them a relatively over-engineered (and costly) choice for nearline, bulk storage chores. Conversely, SATA disc drives (preferably enterprise-optimized) are ideally suited to such duties, where maximum capacity per dollar supersedes such factors as reliability and high availability.

SAS for Performance, SATA for Capacity				
Device	Application	Duty Cycle	MTBF (Typical)	Seek Time (Typical)
SAS (15K RPM)	Online, high availability, random reads	24 hrs/day, 7 days/week	1,200,000 hours @ enterprise workloads	3.6 msec
SATA (7200 RPM)	Nearline, low availability, sequential reads	8 hrs/day, 5 days/week	600,000 hours (1,000,000 hours*) @ desktop workloads	9.5 msec

*MTBF rating for Seagate NL35 Series enterprise-optimized SATA disc drives

Figure 1

The Serial Attached SCSI standards committee well understood the complementary nature of these two storage technologies, and the synergies (both fiscal and physical) that would result if SAS and SATA drives could share a common storage infrastructure. To effectively serve both interfaces, such an infrastructure must seamlessly blend scalability, flexibility and affordability. SAS infrastructures neatly satisfy all of these criteria via devices known as *expanders*.

Expanders: Key to SATA Scalability

SAS expanders are high-speed switches that enable a single SAS domain to contain over 16,000 drives (SAS and/or SATA). There are two types of expanders: *edge expanders*, capable of connecting up to 128 drives; and *fan-out expanders*, one of which can aggregate up to 128 edge expanders in a single SAS domain.

Unlike Port Multipliers, SAS edge expanders can be linked—either to a fan-out expander (which employs table routing) or to another edge expander (which employs subtractive routing, and optionally, table routing). Increasingly, SAS vendors are recognizing the value of incorporating table routing into their edge expanders, thus eliminating the need for a fan-out expander and allowing multiple edge expanders to be cascaded together in daisy-chain fashion.

Overview: Expanders vs. Port Multipliers			
Device	Maximum Number of Connections	Drives Supported	Notes
SATA II Port Multiplier	15 drives	SATA only	Requires SATA II, Port Multiplier-aware host controller
SAS Edge Expander	128 drives; one edge expander, one fan-out expander	SAS and SATA	Multiple edge expanders can be linked if table routing incorporated
SAS Fan-out Expander	128 edge expanders and/or drives	SAS and SATA	Maximum: one fan-out expander, 16,384 SAS devices in a single domain

Figure 2

In addition to expanders, there is another key component of SAS/SATA scalability—the Serial ATA Tunneling Protocol (STP). STP enables SAS HBAs to identify and communicate with Serial ATA devices. When data is directed to a SATA drive that's connected to a SAS backplane with an edge expander, an STP connection is immediately opened to enable SATA frames to pass through the connection to the drive. STP operates transparently in the background, with virtually no impact on system throughput.

Expanders = Efficient Scalability

As shown below, SAS infrastructure addresses many of the SATA infrastructure challenges previously discussed, thus ensuring optimal scalability for both SATA and SAS drives:

- **Compatibility**

Offering an unprecedented degree of compatibility and efficiency, SAS HBAs and expanders enable deployment of both high-performance (SAS) drives and high-capacity (SATA) drives in the same infrastructure. By eliminating the need for separate and redundant infrastructures, SAS reduces both hardware and IT management costs.

- **Expandability**

Each SAS edge expander is capable of connecting up to 128 devices (SAS HBAs, SAS and/or SATA drives and other SAS expanders). Thus a single SAS HBA port connected to a SAS edge expander is theoretically capable of addressing **over eight times** as many SATA drives as a single SATA host controller port connected to a 15-port Port Multiplier. Should additional drive ports be necessary, another edge expander can simply be cascaded off the first expander, maximizing the value of each HBA port.

SAS delivers exceptional expandability and flexibility by enabling direct cascading of multiple edge expanders (each expander connecting up to 128 devices). For still greater scalability, a single fan-out expander can aggregate up to 128 edge expanders, yielding a theoretical total of over 16,000 devices in a single SAS domain (see Figure 3).

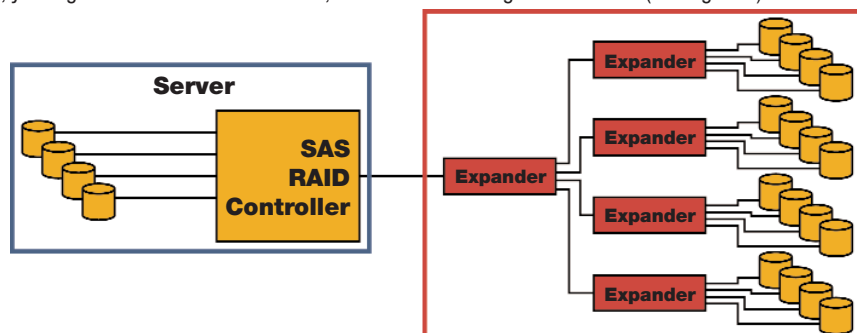


Figure 3

The dual-port and point-to-point architecture of SAS also facilitates expandability with the improved performance of bandwidth aggregation. Constructing wide links from multiple ports enables transmissions from several drives to be aggregated over a single, large pipe to the host and between expanders, eliminating the performance bottleneck that can arise when too many drives are connected to a single host port (see Figure 4).

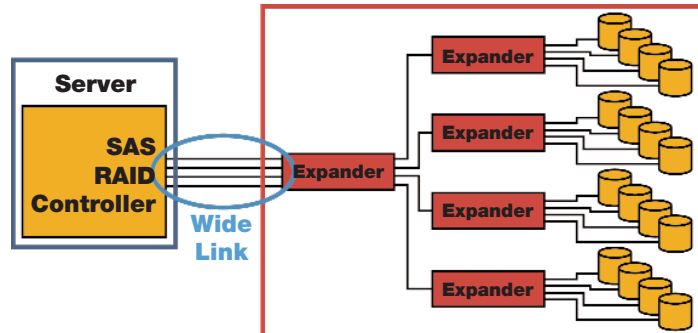


Figure 4

- **Performance**

SAS edge expanders and fan-out expanders are true enterprise-class, high-speed switches that enable simultaneous communication between multiple initiators and targets. In addition, expanders benefit from SAS's full-duplex architecture. Full-duplex transmission means a drive can transmit data to the host at the same time the host is sending additional commands to the drive. Separate interface sessions aren't needed to send commands, freeing up more bandwidth for data transmission (see Figure 5).

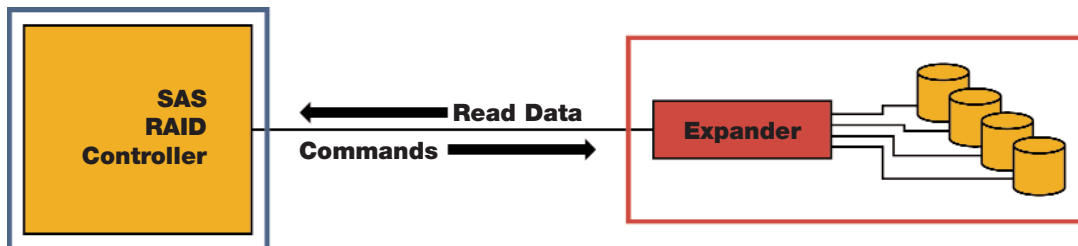


Figure 5

- **Data Integrity**

SAS expander route tables contain addresses for all attached SAS and SATA drives, ensuring they can be readily located and sent data regardless of their location in the SAS domain. This greatly reduces the risk of losing data by sending it to the wrong drive, a key consideration as drive counts continue to climb. SAS further ensures accurate data transmission with its more advanced header information, which includes the source, destination and context of each command that accompanies data transmission. This effectively logs drive activity in a SAS domain, vastly improving connection reliability. For SATA drives in SAS infrastructures, this enhanced header information passes between SAS HBAs and expanders: for SAS drives this header info travels directly between HBAs and SAS drives.

- **Reliability**

SAS's dual-port architecture also enables seamless, active failover capability to ensure reliability under intense enterprise traffic. Dual ports enable SAS devices to be connected to multiple hosts; should one controller fail, the SAS device will automatically switch to another available controller (see Figure 6, next page).

- **Cabling**

Maximum cable length per discrete connection between two SAS devices is eight meters (total SAS domain cabling distance can run into thousands of feet). This offers exceptional flexibility to locate servers and storage arrays in the most cost-effective, space-efficient configurations possible

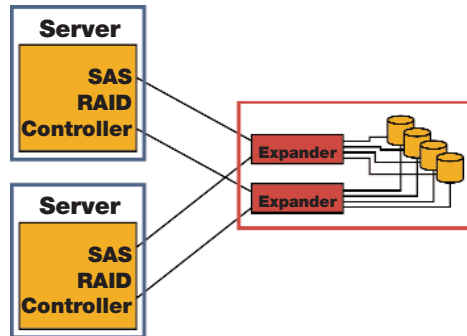


Figure 6

Furthermore, SAS infrastructure components (HBAs, expanders) are built to withstand the performance requirements of enterprise use. In contrast, SATA components are primarily designed for less demanding, cost-conscious desktop environments.

With the above points in mind, it's tempting to assume that SAS infrastructure will be substantially more costly than its SATA counterpart. To be sure, the initial expenditure on SAS infrastructure will surpass that of a comparable SATA deployment (preliminary industry estimates put SAS HBAs and RAID backplanes in rough price parity with equivalent parallel SCSI pieces). But it is the Total Cost of Ownership (TCO) story that will ultimately prove more telling.

It cannot be overemphasized that the true cost-effectiveness of any storage infrastructure goes well beyond the initial expenditure it entails. What is the cost in manpower and downtime/lost productivity when a drive fails and a RAID volume must be rebuilt? How many additional drives must be purchased to ensure an adequate supply of spares? How well does the infrastructure adapt and scale as needs inevitably change? Is it labor-intensive to deploy and administer? How many SKUs must be qualified, purchased and inventoried? The answers to such questions play a key role in determining the long-term efficiency and value of any storage infrastructure. Consider the following two common scenarios for SATA deployment in the enterprise:

Scenario One: Raw Capacity

To achieve maximum storage capacity at minimum cost, an obvious approach might be to connect a SATA host server to an external SATA enclosure. This enclosure would be equipped with SATA RAID backplanes, carrying multiple Port Multipliers. As more SATA drives were needed, additional backplanes (and thus PMs) could be added to the enclosure.

But because Port Multipliers cannot be cascaded, each additional Port Multiplier requires an additional host controller port. As disc drive (and thus backplane/PM) quantities escalate, the server's host controller must be upgraded to a more costly host controller with higher port count. Before long the maximum port count for host controllers (up to 32 in theory, 8 or 16 in practice) is reached.

By contrast, a SAS infrastructure is easily able to handle a multitude of SATA drives, all with a single SAS HBA port. As more SATA drives and SAS RAID backplanes are added to the enclosure, the additional SAS expanders (incorporating table routing) on those backplanes can be seamlessly cascaded together.

The benefit of this SAS-based approach to SATA storage is significant: Requiring only a minimal number of ports on the HBA and with more ports available on the expanders (theoretical maximum of 128 vs. 15 maximum on Port Multipliers), a SAS infrastructure becomes increasingly cost-effective as the number of drives grows. And as savvy IT managers know, mushrooming drive counts is a given.

Scenario Two: Performance and Capacity

For many enterprises, the mix of performance/high-availability storage and capacity/low-availability storage is not fixed but dynamic, changing as business needs continually evolve. SAS is particularly appropriate for such environments, offering the compatibility and flexibility needed to bridge these two distinct storage applications.

As illustrated in the first scenario, investing in SATA infrastructure to support only bulk storage needs becomes increasingly inefficient as capacity and drive counts grow over time. But when both performance and capacity requirements must be met, the value proposition of a SATA infrastructure *immediately* plummets. As noted earlier, SATA is simply not designed for online, high-availability storage duty in the enterprise. When greater need for such storage arises, any initial investment in SATA infrastructure must be augmented by subsequent investment in SAS infrastructure, an unnecessary and costly redundancy.

Using the same server/enclosure model as Scenario One, this infrastructure redundancy not only entails added expenditures on SAS HBAs and SAS backplanes, it also requires the purchase of additional enclosures. In effect, infrastructure costs are almost doubled when SATA is initially installed and then followed by SAS deployment. Furthermore, it increases the workload on IT departments who must administer two separate infrastructures.

Selecting a SAS infrastructure from the outset to meet existing capacity needs conveys the benefits noted in Scenario One, while ensuring the flexibility to deploy SAS drives by merely plugging them into the existing infrastructure. Thus a single enclosure equipped with a SAS RAID backplane can address both capacity (nearline) and performance (online) applications. Growing firms can purchase SAS infrastructure for initial use with SATA drives only, secure in the knowledge that such equipment will not become obsolete/unusable when their storage needs expand and enterprise-class drives are needed.

Beyond the obvious efficiency of specifying the optimal disc drive for a given application, standardizing on the SAS platform will significantly reduce the cost and complexity of the data center by minimizing the number of individual components that must be qualified, purchased, inventoried and maintained. Such component rationalization also results in a smaller data center footprint and places fewer demands on management resources and support staff.

Confluence of Performance, Capacity and Density

The flexibility of SAS infrastructure extends to disc drive form factors as well. SAS drives will be available in the industry-standard 3.5-inch form factor for storage systems utilizing a common backplane. A single storage subsystem will thus be able to house a low-cost 7200-RPM SATA drive in the same enclosure as the preferred online enterprise solution, mainstream 15K-RPM SCSI drives. SAS drives will also be offered in the 2.5-inch small form factor. For denser computing environments in which raw capacity takes a back seat to higher throughput (IOPS/U), these compact SAS drives will play an increasingly prominent role.

Note that choosing 2.5-inch SAS drives for transactional applications (for example, database storage for ERP and CRM software) doesn't preclude use of 3.5-inch SATA drives for periodic backup and restore. For example, servers and storage subsystems of varying sizes (1U, 2U, 4U, and so forth) stacked on top of one another in a cabinet can transfer data interchangeably between SAS and SATA drives. Depending on the cabinet's configuration, Serial ATA Tunneling Protocol (STP) could be accomplished at an HBA or expander, thus eliminating the need for SATA and SAS drives to share the same backplane.

Conclusion

Greater storage efficiency continues to be an urgent priority for every enterprise, and SAS is uniquely positioned to facilitate this key goal. SAS was proactively engineered as a superset of SATA, and this innovative architecture pays multiple dividends. Not only does SAS infrastructure (via its compatibility with SATA disc drives) enable IT managers to select the most cost-effective storage solution for any task, it also vastly improves SATA's scalability. In terms of maximum possible drive quantities and long-term deployment flexibility, SAS infrastructure simply outperforms its SATA counterpart. Furthermore, SAS eliminates the redundancies and inefficiencies of purchasing and maintaining separate infrastructures for high-performance (SAS) and high-capacity (SATA) storage applications. Clearly, SAS establishes a new paradigm for efficient enterprise storage.