

Preface

Thank you for downloading this white paper.

The “Evolution of SAS” white paper was authored by PMC-Sierra, Inc. to help educate the storage community about SAS technology.

Suggested Links

Readers of this white paper may find the PMC-Sierra links below relevant.

White Papers

http://www.pmc-sierra.com/go/storage_wp

Topology Discovery Webinar

http://www.pmc-sierra.com/go/discovery_webinar

Storage Products:

<http://www.pmc-sierra.com/go/storage>

Evolution of SAS

White Paper

by Tim Symons

Issue No. 1: April 2007

PMC-Sierra, Inc.

Legal Information

Copyright

Copyright 2007 PMC-Sierra, Inc. All rights reserved.

The information in this document is proprietary and confidential to PMC-Sierra, Inc., and for its customers' internal use. In any event, no part of this document may be reproduced or redistributed in any form without the express written consent of PMC-Sierra, Inc.

PMC-2070555 (01)

Disclaimer

None of the information contained in this document constitutes an express or implied warranty by PMC-Sierra, Inc. as to the sufficiency, fitness or suitability for a particular purpose of any such information or the fitness, or suitability for a particular purpose, merchantability, performance, compatibility with other parts or systems, of any of the products of PMC-Sierra, Inc., or any portion thereof, referred to in this document. PMC-Sierra, Inc. expressly disclaims all representations and warranties of any kind regarding the contents or use of the information, including, but not limited to, express and implied warranties of accuracy, completeness, merchantability, fitness for a particular use, or non-infringement.

In no event will PMC-Sierra, Inc. be liable for any direct, indirect, special, incidental or consequential damages, including, but not limited to, lost profits, lost business or lost data resulting from any use of or reliance upon the information, whether or not PMC-Sierra, Inc. has been advised of the possibility of such damage.

Trademarks

For a complete list of PMC-Sierra's trademarks and registered trademarks, visit:
<http://www.pmc-sierra.com/legal/>.

Patents

The technology discussed in this document may be protected by one or more patent grants.

Abstract

SAS is evolving. SAS, which has replaced cabled parallel SCSI as the interconnect technology of choice, provides high bandwidth and high data throughput and is compatible with existing SATA topologies. This paper provides an overview of SAS and discusses how second generation SAS controllers, expanders and port multiplexers can be optimized to ensure better bandwidth utilization, easier management and network robustness.

About PMC

PMC-Sierra is a leading provider of broadband communications and storage semiconductors for metro, access, fiber to the home, wireless infrastructure, storage, laser printers, and fiber access gateway equipment. PMC-Sierra offers worldwide technical and sales support, including a network of offices throughout North America, Europe, Israel and Asia. The company is publicly traded on the NASDAQ Stock Market under the PMCS symbol and is included in the S&P 500 Index. For more information, visit www.pmc-sierra.com.

About the Author

Tim Symons is a Principal Engineer and Storage architect at PMC-Sierra and represents PMC-Sierra in the storage standard bodies. Prior to joining PMC-Sierra, Mr. Symons was a Storage Systems Architect and Technologist with Adaptec Inc. (formerly Eurologic Systems). Mr. Symons has a Bachelor of Science (Honors) degree in Electrical & Electronic Engineering from the University of Plymouth, UK.

Revision History

Issue No.	Issue Date	Details of Change
1	April 2007	Document created.

Table of Contents

Legal Information.....	2
Abstract	3
About PMC	3
About the Author.....	3
Revision History.....	3
1 The Value of SAS.....	6
1.1 Data Center Topologies.....	6
1.2 High Bandwidth, High Throughput Protocols	7
1.3 SATA Support	8
2 Second Generation SAS.....	9
2.1 6 Gbit/s SAS Link Rates and Multiplexing	9
2.2 Zoning	10
2.3 Self-configuring Expanders	12
2.3.1 First-generation SAS Initialization.....	12
2.3.2 Second-generation SAS Initialization.....	13
2.3.3 Disk Drive Initialization	13
2.4 Diagnostics.....	14
2.5 Multi-Affiliation Support	15
3 Leveraging Software Commonality	17
4 Conclusion	19

List of Figures

Figure 1	Data Center Connectivity	6
Figure 2	Relative Cost SAS / SATA	8
Figure 3	Link Multiplexing.....	10
Figure 4	Zoned Portion of Service Delivery Subsystem (ZPSDS)	11
Figure 5	Discovery Without Self-configuring Expanders.....	12
Figure 6	Discovery With All Self-configuring Expanders.....	13
Figure 7	SAS and SATA Link Initialization Duration	14
Figure 8	Active/Active Port Selector With Multi-Affiliation Support	15
Figure 9	Common Software Interface	17

1 The Value of SAS

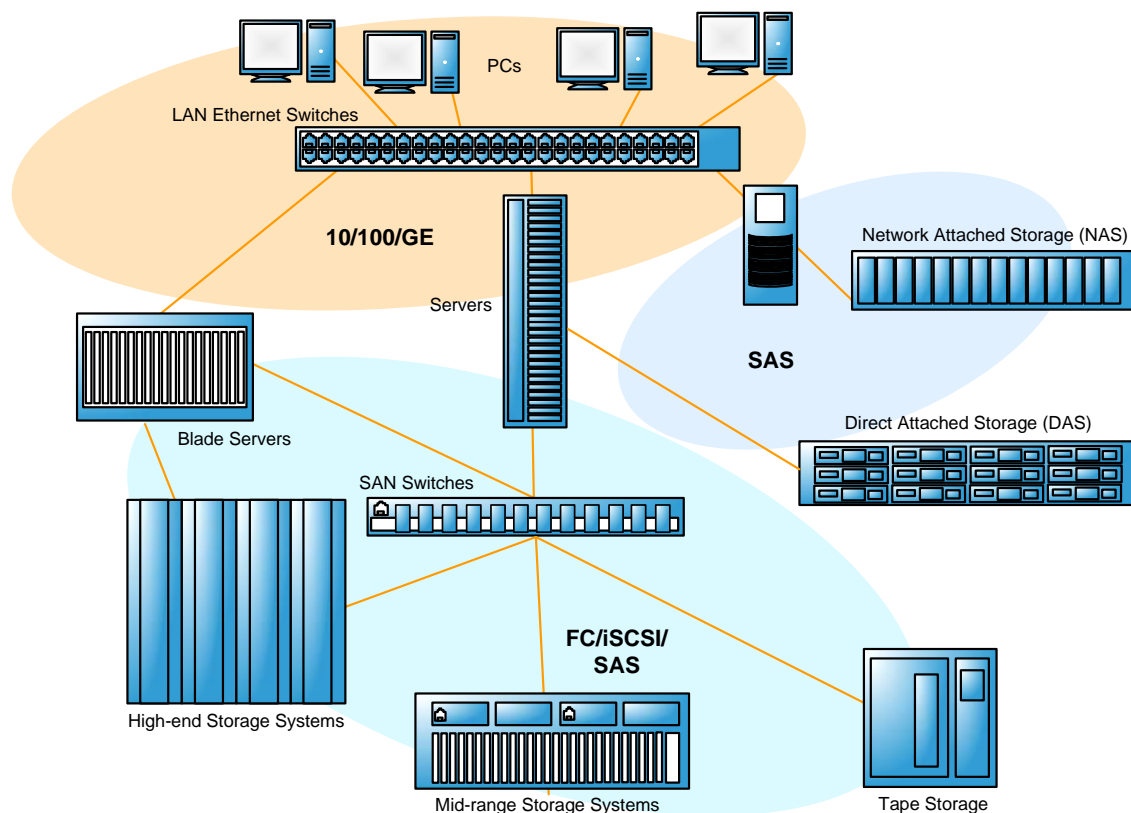
SAS provides a scalable point-to-point topology capable of addressing storage connectivity at many levels. As SAS enters its second generation, the standard is evolving to enable better bandwidth utilization, easier management mechanisms, and network robustness. This paper examines the growing value of SAS.

1.1 Data Center Topologies

Increasingly, data centers are growing more complex, driven by the demand to support rich data content and greater system throughput, and legislation to support data storage, protection recovery, and archiving.

Serial Attached SCSI (SAS) has replaced cabled-parallel SCSI as the interconnect technology of choice. As shown in Figure 1, SAS addresses storage connectivity at many levels.

Figure 1 Data Center Connectivity



The SAS architecture supports multiple hosts and unlimited device attachments, accommodating systems scaling from small direct-attached disk drives to many thousands of network-attached disk drives and controllers.

However, the strength of SAS lies in its ability to:

- Provide high bandwidth and high data throughput
- Compliment Serial ATA (SATA) topologies

1.2 High Bandwidth, High Throughput Protocols

The three protocols that operate within the SAS transport layer are:

- Serial SCSI Protocol (SSP) — Enables end-to-end connections between SAS (SCSI) disk drives, tape drives, etc.
- Serial Management Protocol (SMP) — Enables configuration and management of the SAS domain
- SATA Tunneling Protocol (STP) — Enables compatibility with SATA drives

SSP enables end-to-end connections. A host can establish an open link at one end of a topology and have it connect to a target at the other end of the topology before communication is initiated. Since data transfers are in packet format, the link is not reserved for long periods of time. This ensures that the topology can enable balanced throughput for all devices. Additionally, multiple links (wide ports) may be used independently to provide higher bandwidth and enhance system throughput.

Expanders and controllers use SMP to configure and manage the SAS domain. The protocol operates in-band through the SAS links. SMP provides notifications of changes in a domain, such as the removal or addition of a disk drive or another type of device. It also provides reporting functionality for status logging. Expander and controller devices are SMP initiators.

Note that SAS 2.0, which is currently being drafted, provides additional status and reporting information to facilitate diagnostic functions. This ensures that optimal system operation can be maintained. In the event of a fault, this status data can be used to identify, isolate, and analyze fault and error conditions.

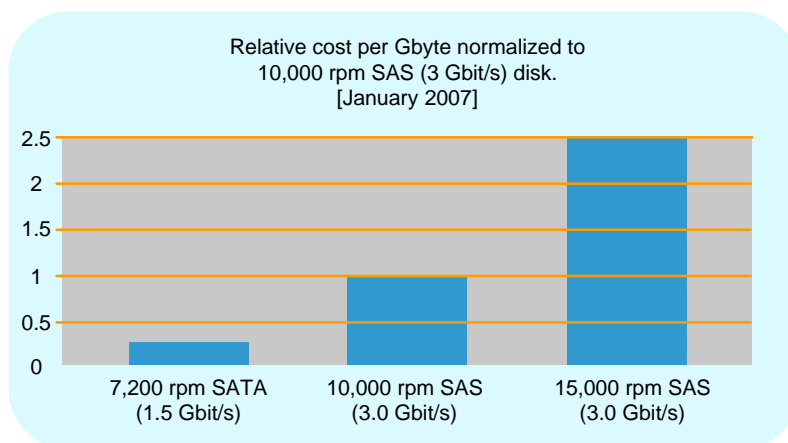
STP provides the link between SAS and SATA. A SAS controller uses STP to wrap the SATA protocol and data packets into SAS packets. SAS expander devices are able identify when a SATA device is attached to their ports, and will use STP to remove the SAS wrapper and present the SATA disk drive with the native SATA protocol.

There are several advantages to being able to use SATA devices in a SAS fabric, as are discussed in the next section.

1.3 SATA Support

SAS is designed to be compatible with high-capacity SATA drives. This is valuable as SATA provides the highest capacity at the lowest cost-per-gigabyte of any storage media. Refer to Figure 2. Additionally, the use of a SATA Active/Active port selector to “dual-port” a SATA hard disk drive (HDD) enables the design of fully redundant path storage-system architectures that provide greater system fault tolerance.

Figure 2 Relative Cost SAS / SATA



While SAS provides higher performance and reliability than SATA, it provides lower capacity. At the time of writing, the largest SATA drive is a 1000 Gbytes, while the largest SAS drive is 300 Gbytes.

The tradeoff for SATA is reduced performance and lower reliability than enterprise-class SAS and Fibre Channel (FC) drives. However, data storage centers typically incorporate backup and RAID to enable recovery of data in the event of a device failure, which can offset some of the reliability issues. Where SATA drives are used for infrequently accessed data, near-line storage or for backup RAID is commonly used to resolve the reliability risks of SATA storage.

The synergy that exists between SAS and SATA enables further network compatibility and scalability.

2 Second Generation SAS

First generation controllers, expanders, and port multiplexers have allowed systems to successfully evolve from bus-based parallel SCSI to link-based Serial SCSI. These links run at 3.0 Gbit/s. Multiple links (wide ports) can be used to increase the available bandwidth between devices. Expander devices can be added to scale the SAS topology up to 16,000 devices in a single domain. Successful interoperability sessions have been held to confirm compatibility between different vendor devices and systems.

SAS is now entering its second generation. SAS 2.0 is currently being drafted. It defines a higher link rate, improved bandwidth utilization, and many features to improve the robustness and manageability SAS topologies.

A few of the major changes include:

- 6 Gbit/s SAS — Doubles the link rate and bandwidth
- Multiplexing — Optimizes bandwidth by enabling two 3 Gbit/s links to share a 6 Gbit/s port
- Zoning — Enables partitioning of a domain into smaller sets of accessible devices
- Self-discovering expander devices — Accelerates topology initialization and detection of changes
- Diagnostics and robustness — Improves status reporting and error notification
- Affiliation support — Enables a SATA disk drive to respond to more than one host

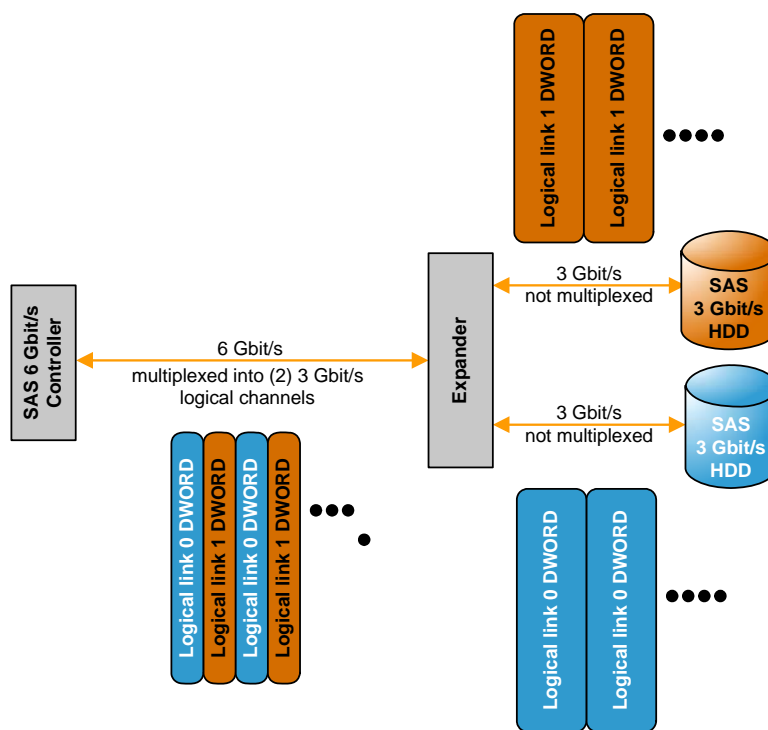
Second generation controllers are optimized to take full advantage of PCIe 2.0 (5 Gbit/s per link) connectivity, aligned to SAS 2.0 6 Gbit/s links. Sustained throughput is higher than any previous generation devices. These SAS 2.0 improvements ensure that SAS systems are faster, provide better bandwidth utilization, are easier to manage, and have improved system robustness. SAS is evolving.

2.1 6 Gbit/s SAS Link Rates and Multiplexing

Second generation SAS expander devices provide the connectivity fabric to complement higher performance controllers. 6 Gbit/s links provide double bandwidth availability to support more disk drives and better throughput.

6 Gbit/s links can be multiplexed to provide two interleaved 3 Gbit/s links sharing a single port. This enables very high-throughput controllers to concurrently access more disk drives with fewer connections. In a system where both 3 Gbit/s and 6 Gbit/s devices are used, some ports may be configured as multiplexed ports while others run at native 6 Gbit/s rates to optimize the system performance and retain backwards compatibility. See Figure 3 for details.

Figure 3 Link Multiplexing



2.2 Zoning

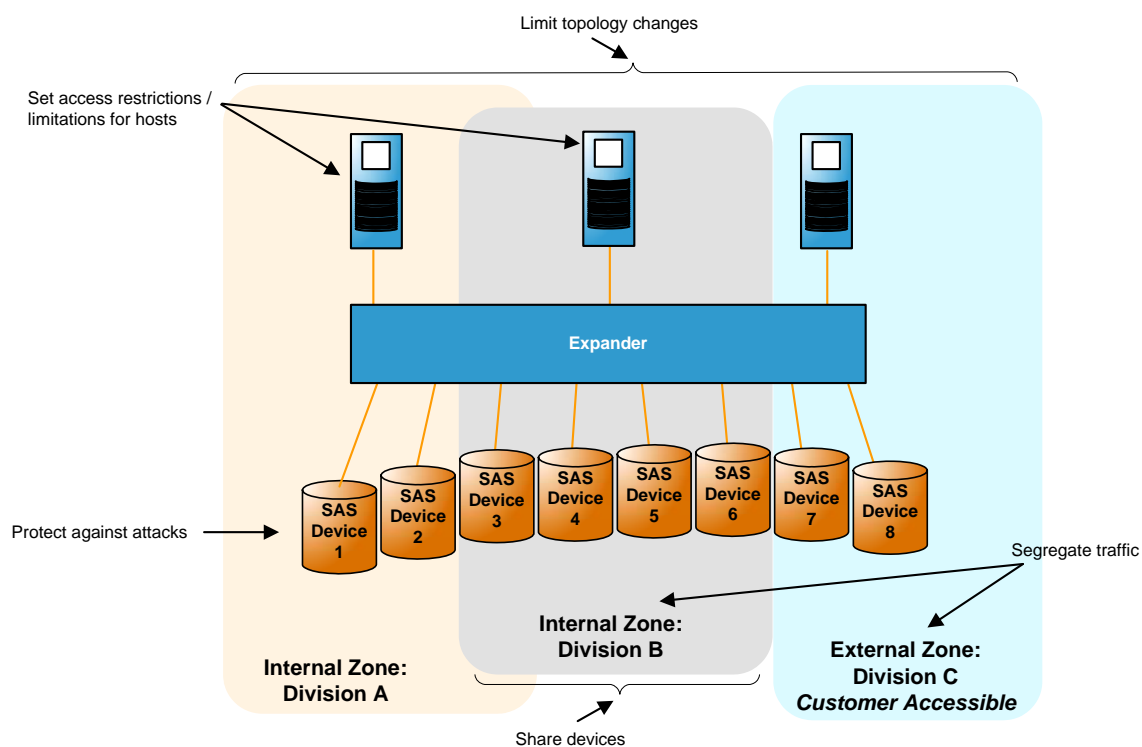
Zoning provides a mechanism to allocate sets of devices to specific controllers within a single SAS subsystem. Zoning allows groups of devices in the SAS subsystem to be linked together or isolated from each other to simplify the management of data services.

Zoning is also particularly useful in blade server applications where multiple controllers on different processor blades are connected to a common storage blade. Zoning allows the HDDs on a storage blade to be partitioned among the various processor blades so that each controller is only aware of a subset of the HDDs on the storage blade.

In a zoned portion of the service delivery subsystem (ZPSDS), a host is allocated to a set of devices to access a subset of devices from the entire topology. Some devices may be shared between multiple hosts and others may be in a hot-spare group, ready for allocation to any groups than needs additional capacity or a backup drive.

A zoned subsystem is shown in Figure 4. Zoning may be managed remotely or directly through a device in the system. An application management client uses SMP functions to configure all devices in the domain.

Figure 4 Zoned Portion of Service Delivery Subsystem (ZPSDS)



Zoning not only segregates the data services, but also the management and diagnostic notifications that are retained within each zoned subsystem. This ensures that a failed link or disruption in one zone, does not affect other zones and ensures that maximum performance is maintained in each portion of the ZPSDS at all times.

In a SAS system, there may be a mix of device types:

- Multiple Host controllers
- High-performance SAS drives
- High-capacity SATA disk drives
- SAS tape drives
- SAS DVD drives

Furthermore, applications accessing a SAS storage subsystem may have mixed requirements from the storage system:

- High availability drives
- Near-line access drives
- Remote boot drives

- Backup drives
- RAID protected drives
- Encryption protected drives
- Hot spare drives

In addition to application criteria, legislation requires that different storage strategies should be employed for personal and financial records as well as E-mail archiving. Meeting these requirements in a data centre is greatly simplified by the ability to use zoning.

2.3 Self-configuring Expanders

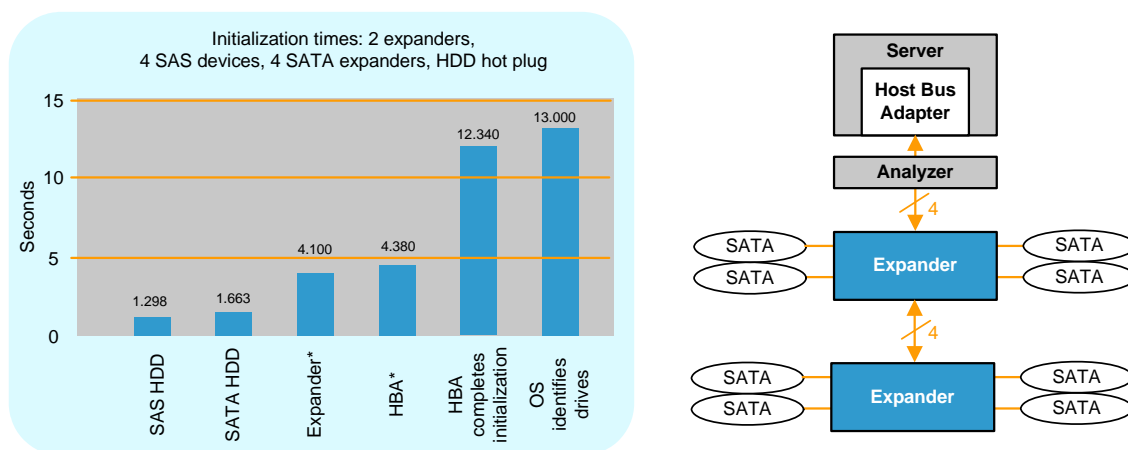
First generation SAS expanders do not assist with the SAS topology discovery process. SAS controllers process all of the device identification and mapping of route tables within each expander device.

However, second generation expanders implement self-configuration features. Each expander device discovers the devices attached to it and completes its own route table. Since all expanders are initializing at the same time, the overall system topology is resolved quickly. This is most apparent with large topologies or where there are multiple hosts in a domain.

2.3.1 First-generation SAS Initialization

Figure 5 shows the measured results of an actual SAS system using a first-generation SAS controller and expander devices without the self-configuring option.

Figure 5 Discovery Without Self-configuring Expanders



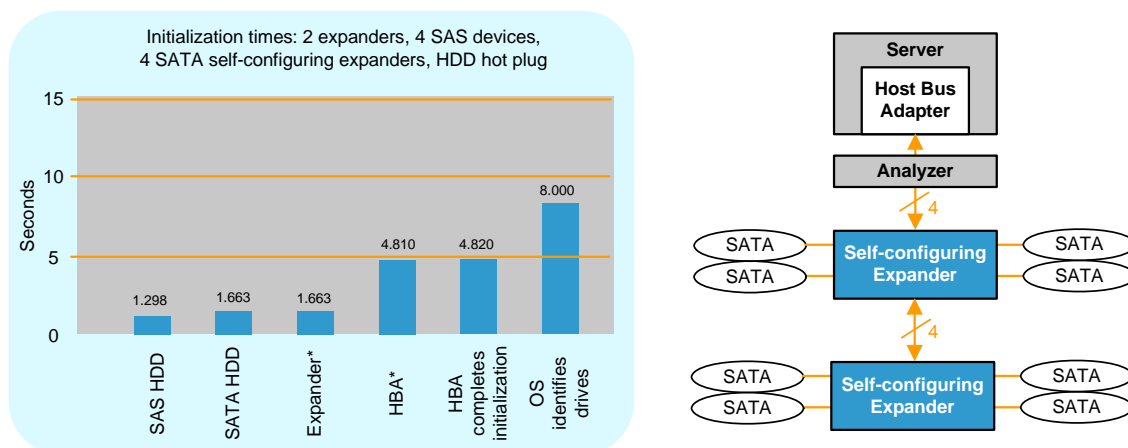
The initialization time of the route tables in the expander devices is around 4 seconds. The HBA completes its discovery and interaction with the OS drivers after 12 seconds. The OS notifies the user about the drives after 13 seconds.

In this test, the disk drives were powered up and spinning prior to the host initialization.

2.3.2 Second-generation SAS Initialization

Figure 6 shows the measured results of an actual SAS system using a first-generation SAS controller and second-generation SAS expander devices that have self-configuring capability. Note the 33% improvement in initialization time.

Figure 6 Discovery With All Self-configuring Expanders



The initialization time of the route tables in the expander devices is around 4.8 seconds. The HBA completes its discovery and interaction with the OS drivers after 4.8 seconds. The OS notifies the user about the drives after 8 seconds.

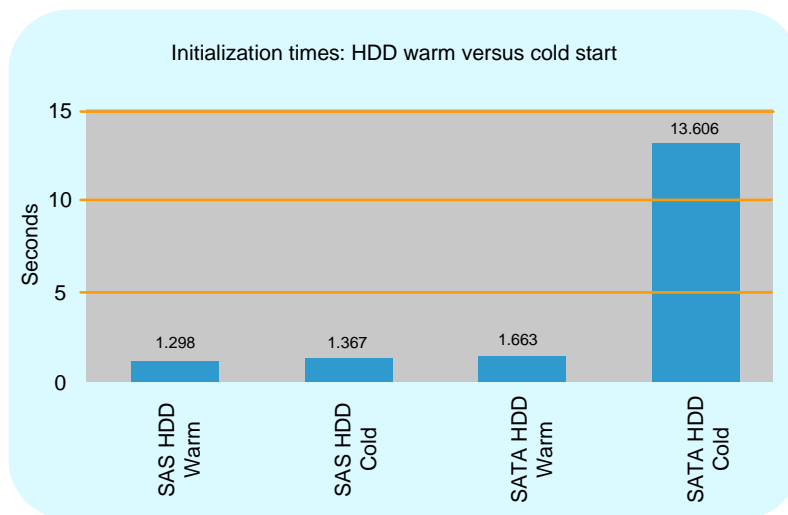
In this test, the disk drives were powered up and spinning prior to the host initialization.

It should be expected that a second-generation host controller would further improve the initialization time because it would not attempt to write the expander device route tables.

2.3.3 Disk Drive Initialization

In the previous measurement tests, initialization was recorded from the power-off (cold) state. It was noted that all SAS devices initiate their links immediately, but most SATA devices do not negotiate the link connection until after the spin-up sequence. This is typically measured at 13.6 seconds before negotiating the link. See Figure 7. The additional link initialization delay is an extra 13 seconds to the overall boot time.

Figure 7 SAS and SATA Link Initialization Duration



2.4 Diagnostics

The cost of maintaining and servicing any system increases with scale and complexity. A major objective of the SAS 2.0 standard is to improve status reporting throughout a SAS topology.

Controllers, expanders, and SATA port selectors, such as those provided by PMC-Sierra, provide link test, loopback, and self-test functions to exercise and analyze any link at any time. These features coupled with an intelligent diagnostic application provide ongoing status monitoring and testing capabilities for every link in a SAS system.

Error counters, status change counters, and event notification counters can be used to log events on each PHY in a SAS system. Threshold counters can be used to determine and log error rates before and after an actual link failure occurs to improve failure analysis and avoid unnecessary replacement of good devices.

In redundant systems, remote monitoring and system policies can automate a failover recovery process or even bring hot-spare backup drives online to replace a device exhibiting a higher than acceptable error rate. If a device is exhibiting an abnormally high error rate, then it can be replaced before it fails. It may be possible to backup all the data from a potentially failing drive before a failure condition and data is loss occurs.

In a RAID system, high retry rates on any single device will reduce the performance of the entire RAID set. By monitoring retry conditions it is possible to ensure optimal RAID operation, and replace any devices that are operating sub-optimally.

SMP application clients are integrated into controllers and expander devices, which individually monitor PHY status. A diagnostic application uses the data generated by the PHY counters to implement policies for system error conditions and can trigger self test operations to generate historical logs of system events. To keep status information activity to a minimal level, notification events are only triggered when a threshold limit is surpassed.

System monitoring can be managed remotely. This all adds up to improved reliability, serviceability and system robustness.

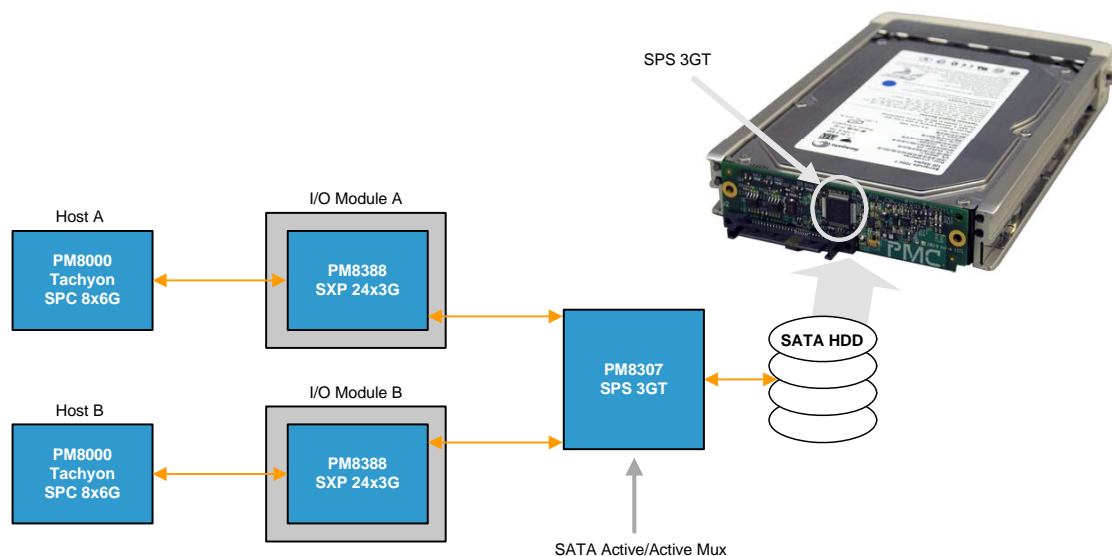
2.5 Multi-Affiliation Support

Enterprise-class storage systems typically implement levels of redundancy. A disk drive is required to have two ports in order to support a fail-over redundancy. To resolve this issue, SATA port selectors are used to multiplex a single SATA interface onto two ports.

PMC-Sierra provides an Active/Active SATA port selector device (PM8307 SPS 3GT), which enables two controllers, one attached to each of the two ports to operate simultaneously. The Active/Active port selector enables system redundancy as well as load balancing, where both controllers share access to the disk drive.

Figure 8 shows the SPS 3GT multiplexer on a dongle board attached to a SATA disk drive. The block diagram indicates connectivity to two PMC-Sierra SPC 8x6G host controller device domains through SXP 24x3G expander devices.

Figure 8 Active/Active Port Selector With Multi-Affiliation Support



The SATA drives implement Native Command Queuing (NCQ). The SATA port selectors provide additional intelligence to process and map the response to the requestor.

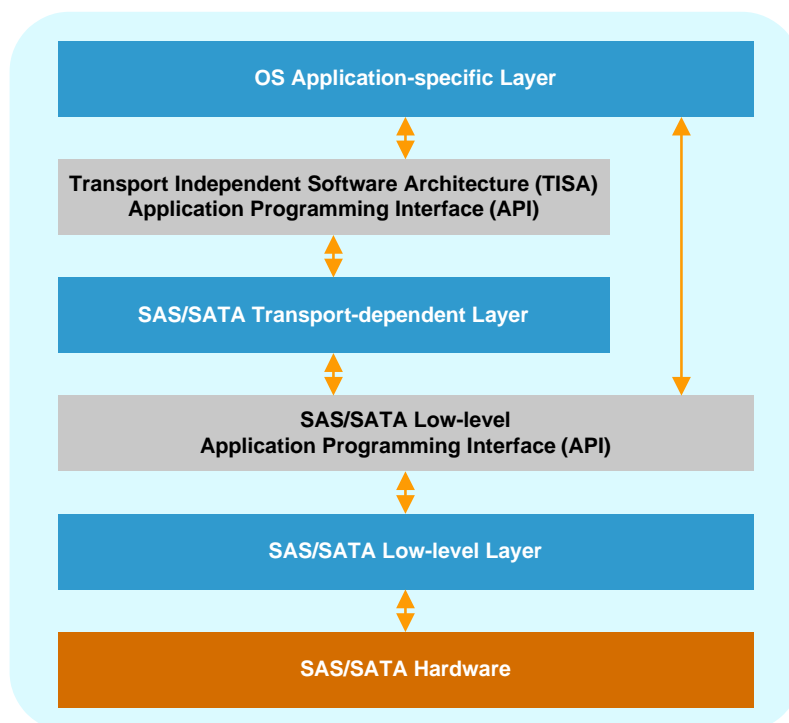
The SATA protocol is designed for desktop environments where there is only a single host controller. The protocol includes tags to link response data to a specific request, but does not include source and destination address information. Where an Active/Active port selector is used, the request and response tags are mapped (or affiliated) to the appropriate host controller device to support more than one host device.

3 Leveraging Software Commonality

Vendors provide the software interface between the SAS controllers and the application. A proven, user-friendly, scalable, and portable interface for device configuration is vital to delivering a high performance, high-quality product to the market place.

PMC-Sierra’s Transport Independent Software Architecture (TISA) provides a common SCSI-like software interface that ensures that SAS implementers can leverage their expertise from previous generations of Tachyon controller devices to deliver the latest generation SAS solutions. See Figure 9.

Figure 9 Common Software Interface



This model is applied to all PMC-Sierra storage devices, from controllers to expanders to SATA port selectors.

The common software interface provides a transport-specific configuration and management interface. It features:

- A platform-independent software interface for all operating systems, enabling faster time-to-market
- Reduced risk by providing a reused common programming environment and tools to improve productivity and reliability

- Reduced servicing and support costs
- Programmability at all layers, even down to the register level

A storage application that is developed using the TISA API can support different SAS and SATA transport protocols (as well as iSCSI and FC) by replacing the transport-specific layers underneath this API. Included with the TISA package is the:

- TISA API
- Low-Level APIs
- Hardware interface

4 Conclusion

SAS is evolving. The second-generation controllers provide more than just higher throughput and bandwidth and SATA interoperability, they interact with the expander and SATA port selector infrastructure devices to improve system diagnostics, status, management, configuration and initialization. SAS 2.0 brings features that enhance system robustness, and enable large topologies with a variety of devices to interoperate successfully.

PMC-Sierra is a leader in both SAS/SATA and Fibre Channel storage devices. Users of PMC-Sierra's Tachyon SAS Protocol Controller (SPC), maxSAS SXP expander, and SPS port selector devices benefit from a complete end-to-end SAS storage solution. Tied together with the user-friendly, common software platform, TISA, PMC-Sierra storage products offer proven technology and robustness extended from the industry-leading Tachyon family.

Further reading on many of the topics covered in this paper may be found at:

- www.sas6g.org: Details about SAS 6G protocol
- www.t10.org: SAS specifications and drafts
- www.pmc-sierra.com/storage: Product features and technology details

Contacting PMC-Sierra

PMC-Sierra
100-2700 Production Way
Burnaby, BC
Canada V5A 4X1

Tel: +1 (604) 415-6000
Fax: +1 (604) 415-6200

Document Information: document@pmc-sierra.com
Corporate Information: info@pmc-sierra.com
Technical Support: apps@pmc-sierra.com
Web Site: <http://www.pmc-sierra.com>