Technology Paper

Enterprise SSD Interface Comparisons

Introduction

PCI Express (PCIe) is a general purpose bus interface used both in client and enterprise compute applications. Existing mass storage interfaces (SATA, SAS) connect to the host computer through host adapters that in turn connect to the PCIe interface.

The SATA interface was designed as a hard disk drive (HDD) interface, and the SAS interface was designed as both a device interface and a storage subsystem interface/infrastructure. As HDDs and system requirements have evolved, requiring faster interfaces and new features, the SATA and SAS interfaces have gone through several revisions.

Solid state drives (SSDs) have quickly added significant new performance requirements to these interfaces, as the data rates of SSDs have gone from tens of MB/sec, to hundreds, and now thousands of MB/sec. In addition to the increase in data rates, the lack of mechanical movement in SSDs have also increased the number of input and output operations per second (IOPS) that these storage devices can perform.

This development has created a need for improved implementations of the existing standards, as well as enhancements to existing interface standards, to manage the new performance requirements while keeping compatibility with existing system architecture.

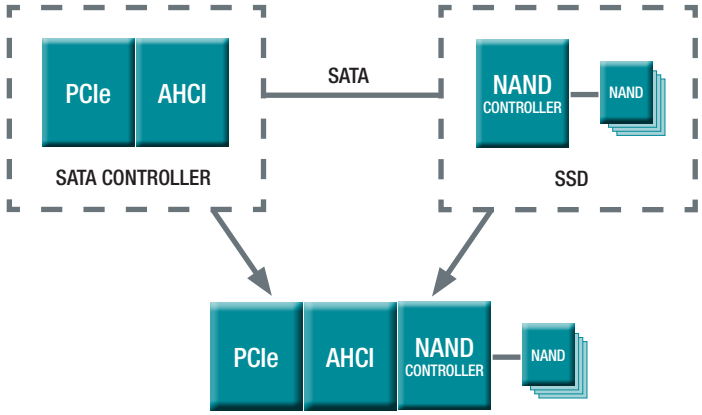
This paper discusses the different interfaces and contrasts the various performance and compatibility trade-offs encountered.

Enterprise SSD Interface Comparisons

SATA Interface

SATA is a low-cost interface designed for point-to-point connection either through a cable or printed circuit board (PCB) trace. The host connection is to an advanced host controller interface (AHCI), which usually resides in the host chipset as the host adapter on the PCIe bus. There are some design issues with this interface that can create a bus overhead of 1µs (or more) for each command. This is not a major issue for HDDs where a 4KB transfer is in the order of 10µs, but SSDs can transfer 4KB of data in 2µs (or less)—thus the overhead becomes significant and the SATA interface less interesting as a high-performance mass storage interface.

SATA is still suitable as a low-cost SSD interface where cost, not performance, is the major decision factor. The SATA architecture can also be consolidated into a host adapter that manages the SATA command-set without actually including the physical SATA interface (PHY) (Figure 1).



Source: Seagate Technology, 2011
Figure 1. Architecture Consolidation

SAS Interface

SAS is also a serial interface, attached to the host through a host adapter, but there are significant differences that make it suitable as an SSD interface:

- Less hardware overhead
- Faster transfer rates
- Wide ports
- Efficient driver-controller interfaces

In addition, SAS includes features not found in SATA that improve reliability and availability of devices connected to the interface:

- Robust serial protocol
- Multiple host support
- End-to-end data integrity
- Dual-port capability
- High degrees of concurrency and aggregation

Less Hardware Overhead

There is not a universal host interface for SAS that would be equivalent to the SATA AHCI controller. Instead, multiple vendors compete in the SAS host adapter market where performance is a major factor—not only to interface individual HDDs but also various RAID systems where the transfer rates of multiple HDD spindles are aggregated for improved transfer speed. Additionally, SAS host adapters are designed to manage higher-performance SSDs and HDDs (such as *short-stroked* 15K-RPM drives). Since the hardware host adapter and the device driver managing that host adapter are designed as a system, new designs optimized for SSDs are starting to become available and further improve not only transfer rates but also the IOPS.

Faster Transfer Rates

SAS ports currently support up to 6Gb/s data rates. Companies such as LSI and PMC-Sierra are sampling designs currently in development to support 12Gb/s data rates and greater than 2 million IOPS, with the possibility of 24Gb/s in the future.

Wide Ports

Inherent in the SAS architecture is the concept of wide ports—where multiple links can be aggregated to allow multiple, simultaneous paths between one or more hosts and a device. The current SAS drive connector defines two ports for the drive. As a design choice, current HDDs do not support wide port—only dual port, where each port has a different SAS address that prevents configuration as a wide port.

Accepted proposals for *SAS-3* (12Gb/s) allow an increase in the number of ports on the drive to four, all of which could connect to the same domain, or in pairs to different domains. A very limited number of SSDs can support wide port in addition to dual port on a two-port device.

Enterprise SSD Interface Comparisons

Robust Serial Protocol

The SAS serial protocol provides for training of the serial transmitters and receivers. This improves the signal quality on the cable or the backplane by compensating for channel length, impedance mismatch and inter-symbol interference. The SAS serial protocol also manages error detection and retransmission at the hardware level. This allows for faster recovery from intermittent signaling issues.

Multiple Host Support

The SAS interface and switching fabric allow multiple hosts to access the same device. This feature can be used to manage host failures, as well as data path failures for improved data availability.

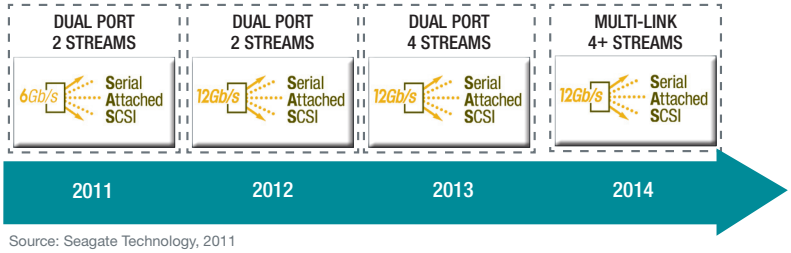
End-to-End Data Integrity

The SAS interface can verify data integrity through cyclic redundancy checks (CRC) of the data from the time it is created in the host data buffer, through the transfer across the PCIe interface and the SAS interface until it is stored on the device and again read and transferred to the host data buffer. This allows for multiple checkpoints along the path from applications through RAID controllers and at devices. This feature is sometimes called protection information (PI).

Dual-Port Capability

SAS target devices support dual-port operation. This provides the ability to create two fault domains and provides increased availability. Even if a failure occurs in one of the paths to a port preventing access along that path, a device is still accessible using the second port.

Historically, Seagate has driven interface adoption in the market. Seagate is working with the SCSI Trade Association (STA) and other industry leaders to leverage the widely deployed, existing SAS infrastructure (Figure 2). Table 1 shows how 12Gb/s SAS and multi-link benefit system builders and end-user organizations.



Source: Seagate Technology, 2011

Figure 2. SAS Interface Evolution

PCI Express Interface

PCI Express (PCIe) is the fundamental interface that connects peripheral devices to the host processor and through a memory controller to the memory architecture in the system. Both the SATA and SAS interfaces discussed earlier connect through a PCIe interface (or host adapter) to the host processor and memory.

Table 1. Benefits of SAS and Proposed Mulit-Link Enchancements SAS

Multiple Links (BW)	X4 (4x600MB/s)
Power Available	25W (2.5-inch)
Total Latency	very low
Multi-Host Protocol	**Yes**
High Availablity	**Yes (Dual Port)**
Scalability	**Excellent**
Robust, Proven Protocol Stack	**Yes**
Hot Swap Serviceable	**Yes**
Compatible With Existing Management SW	**Yes**

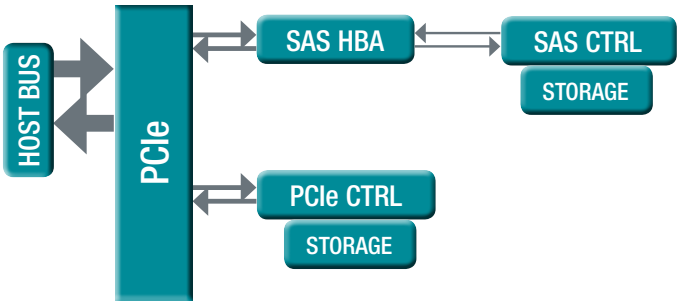
Source: Seagate Technology, 2011

The PCIe interface is a serial implementation of the original PCI interface that provided a parallel address/data connection between peripherals and host processor/memory. The PCIe interface communicates over one or more *lanes* that consist of one transmit and one receive serial interface for each lane. Up to 32 lanes can be used to connect a host to a device. The serial data rate on each lane depends on the version of the PCIe standard implemented, the current version is 3.0 and the data rate is approximately 1GB/s.

For a 1U server, the PCIe interface is designed to utilize a single connector on a (client) motherboard, or a two-connector, right-angle adapter on a (server) motherboard. A cabling system is also available (though seldom used). A 2U, 4U or 7U server has many more PCIe slots, similar to client implementations. The PCIe specification also uses transmitter (and receiver) training to adapt to the impedance variations of a configuration, but is targeted as shorter length transmission channels than SAS.

PCIe switches can accommodate singleroot I/O virtualization (SR-IOV) and multi root I/O virtualization (MR-IOV)—methods used to improve controller performance in virtual (hypervisor) systems with a single or multiple hosts. SR-IOV is just starting to become generally available in adapters; however, VMware may not yet take advantage of it. MR-IOV is typically not supported on adapters.

Enterprise SSD Interface Comparisons



Source: Seagate Technology, 2011
Figure 3. SAS Interface Evolution

Storage devices that connect using the PCIe interface do so either through a direct register connection, or through a host adapter that then connects to the device through additional cabling or a backplane-type interface.

Currently there are a number of different implementations of both architectures. SATA uses a host bus adapter implementation in the system chipset *(southbridge)*—either the Intel or AMD AHCI—requiring different AHCI drivers but mapping to compatible IDE legacy implementations. These interfaces also implement various RAID management features.

SAS has multiple vendors of HBAs, with additional expanders and RAID controllers available, all using proprietary device drivers and BIOS to satisfy various needs for performance and configurability.

The PCIe driver-controller interface is implemented in the NVM Express specification and in the proposed SCSI over PCIe (SOP) specification.

The SATA consolidated architecture described above is another example of the PCIe direct register connection.

PCIe SSDs Today

There are two primary types of PCIe SSDs in the market today: the native and the aggregator. The native controller attaches to the host PCIe bus and then directly controls multiple flash memory buses. These typically use a software interface that is proprietary to the manufacturer and used only for the specific device. Some of these implementations place the burden of address translation and other functions on the host CPU and memory. This in turn causes reduced system resources for applications when the devices are used under heavy workloads. Additionally, being relatively new to the marketplace, these unique drives and hardware combinations are sometimes prone to instabilities, as their ecosystems are still evolving.

The aggregator model takes a different approach to design. This approach utilizes an existing SAS or SATA RAID controller, to which are attached multiple SAS or SATA SSDs. These are packaged together on a single PCIe card. The RAID controller aggregates the performance of multiple devices to offer high levels of performance. Being based on existing proven enterprise-class hardware and software interfaces, these designs are very stable and mature. Additionally, these designs use intelligent controllers that perform address translations and other functions, allowing full use of system CPU cycles and memory by applications, even under heavy I/O workloads.

The Future of PCIe SSDs

Both SOP and NVMe approaches are architecturally similar. However, NVMe is being developed in an industry working group, whereas SOP is being developed in a recognized open standards forum. NVMe is targeted only at use for nonvolatile memory devices, while SOP is also being targeted at use for host bus adapters and RAID controllers with features for bridging between various SOP devices. Additionally, SOP heavily leverages existing industry architectures and features, while NVMe uses a new, very limited instruction set and queuing interface.

Interface Benefits and Issues

Each of the storage architectures described have benefits as well as issues. Depending on the overall system design, the benefits of using a specific architecture may outweigh the issues associated with that architecture, and a careful analysis is required to make the appropriate decision. That decision must also include consideration of compatibility with an existing system design.

For example, updating a laptop computer system that has an existing 2.5-inch SATA HDD with an SSD would only work with an SSD of the same physical size and with the same (or newer) SATA interface. There will be a limit on how fast the SSD can be in this case; exceeding the existing host SATA interface speed will not add to the performance of the system.

In a similar situation, an enterprise server that is using a short-stroked, 15K-RPM SAS HDD to store a database index can be upgraded using a SAS SSD, which will increase overall system performance, but only to the degree that some other system factor becomes the new bottleneck (CPU, memory, network, adapters, etc.).

In a new system architecture, the addition of solid state storage can significantly increase system performance, but only to the extent that the rest of the system architecture can accommodate the increased data rate and data bandwidth. Faster data rates in SSDs also require more power supplied to the device and more heat dissipation required wherever the SSD is mounted.

Enterprise SSD Interface Comparisons

Table 2. Native and Aggregator PCIe SSD Comparison		
	Native	**Aggregator**
Commands/Transport	Proprietary (FTL[1] in host/main memory) 	SCSI or SATA (Multiple SSDs,controller on card) 
Committee	None	None
Standards-Based	No	**Yes**
Performance With Flash	High	High
CPU Overhead	High	**Low**
Latency With Short Queue	Very Low	Low
Latency With Deep Queue	Moderate	**Low**
Use Case Extensibility	No	**Yes** (RAID, HBA, etc.)
Maturity	Evolving	**Based on Proven Industry Architectures**
Enterprise Feature Set (PI, Security, Mgmt., etc.)	No	**Depends on implementation**

1 FTL: Flash Translation Layer
Source: Seagate Technology, 2011

Table 3. SOP and NVMe PCIe SSD Comparison		
	SOP[1]	**NVMe[2]**
Commands/Transport	SOP/PQI[3] (FTL in controller) 	NVMe/NVMe (FTL in controller) 
Committee	**T10/INCITS[4]**	Industry Working Group
Standards-Based	**Yes (ANSI/ISO)**	No
Performance With Flash	Very High	Very High
CPU Overhead	Low	Low
Latency With Short Queue	Very Low	Very Low
Latency With Deep Queue	Low	Low
Use Case Extensibility	**Yes** (RAID, HBA, etc.)	No (NVM only)
Maturity	**Based on Proven Industry Architectures**	TBD
Enterprise Feature Set (PI, Security, Mgmt., etc.)	**Full Support**	Limited

1 SOP: SCSI over PCI Express
2 NVMe: Nonvolatile Memory Express
3 PCIe Queuing interface
4 INCITS: International Committee for Information Technology Standards
Source: Seagate Technology, 2011

www.seagate.com

Another factor is the timing for availability of the operating system device drivers and BIOS support for these new SSD interfaces, as well as the initial reliability of the software.

Interfaces and Flash SSD Latency Facts

There are many misconceptions about what factors add latency and how much they actually affect application performance. When looking at this aspect, it is important to focus on the overall picture, not just one part of it.

The overwhelming contributors to latency in SSDs are the flash parts themselves. SLC access times are 25µs+; MLC access times are 50µs+, both assuming no access contention. As queue depths increase, the contention for access to the flash parts can add substantially to latency.

Once a flash part starts its access, other requests to the same part must wait. As many as eight flash die share a common bus, which cause die to wait their turn using the bus. Housekeeping activities add additional latency (address translation, garbage collection, wear leveling, etc.).

Another aspect is the operating system, which adds latency regardless of the access protocol and interconnect. These include the file system, volume manager, class drivers and context switching overheads.

Differences in protocols and interconnects have negligible effects on latency as seen by an application (fractions of a microsecond).

AMERICAS Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
ASIA/PACIFIC Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, 65-6485-3888
EUROPE, MIDDLE EAST AND AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00