Annotation Summary

5 annotations on 1 page by Gary J Little

x1 x4

data throughput. Second generation PCIe supports 5GT/s over a single lane, third generation PCIe increased performance by 60 percent to 8GT/s. A third generation x8 slot can, therefore, support 64GT/s, or in more familiar terms, a theoretical maximum of around 8,000 Megabytes per second (MB/s), as shown in Table 1.

1 PCI Lane

8 PCI Lanes

PCIe	2.x	3.x
Theoretical burst	4000MB/s	7877MB/s
Typical	3200MB/s	6400MB/s

#1

p.2

High performance storage systems normally aggregate multiple SAS lanes to support higher data rates. A third generation SAS x4 wide port, for example, provides a maximum half-duplex aggregate transfer rate of 48Gb/s or 4800MB/s, as shown in Table 2.

Gen 2

Gen 3

SAS x4	6Gb/s	12Gb/s
Maximum	2400MB/s	4800MB/s
Typical	2200MB/s	4400MB/s

Table 2 – Example of some maximum and typical data throughput in 6Gb/s and 12Gb/s SAS in a

#2

p.2

LSI
Storage. Networking. Accelerated.

White Paper

12Gb/s SAS: Busting Through the Storage Performance Bottlenecks

Key Points:

- Storage performance needs to improve to keep pace with the data deluge
- Third generation 12Gb/s SAS helps meet that need, but all disks need to support this new standard for the system to achieve that level of performance
- LSI® DataBolt™ bandwidth aggregation technology is designed to deliver 12Gb/s performance in third generation SAS systems using 3Gb/s and 6Gb/s SAS and SATA drives

The data deluge continues unabated, fueled by demanding applications like business analytics, social media, video streaming and grid computing, among many others. This relentless growth in data presents many challenges for system architects. How to deal with such vast volumes of data cost-effectively? How to extract meaningful information from more and more data? How to protect and backup so much data? And perhaps most importantly: How to maintain or even improve storage performance?

It is difficult to find anyone who is unaffected by the data deluge. Even home users are generating more and higher quality digital videos and photographs that now consume a substantial amount of storage space on PCs and smartphones. This is one reason why social networking sites, such as Facebook, now receive a half Petabyte of new data daily (according to Facebook Engineering Notes from 2012).

As the volume of data increases, so too can the time it takes to access it. This is the rationale behind continual advances in storage technologies like Serial-Attached SCSI as it transitioned from 3 Gigabits per second (Gb/s) to 6Gb/s and now 12Gb/s. SAS provides a scalable, enterprise-class storage solution, and its third generation throughput of 12Gb/s puts SAS performance on a par with far more expensive Fibre Channel solutions that top out at 16Gb/s. This article describes the ways third generation SAS technology is busting through various storage bottlenecks, including one of its own.

A Brief History of Storage Bottlenecks

When designing storage systems for high-performance, it is essential to understand where the bottlenecks can occur. This is especially true given that the bottlenecks change with each new generation of technology along the data storage path. The three most critical elements that affect storage performance are the server's Peripheral Component Interconnect Express (PCIe®) bus, the SAS solution as implemented in host bus adapters (HBAs) and expanders, and the disk drives themselves, which can have either a SAS or a SATA (Serial Advanced Technology Attachment) interface. Each element is considered in turn here.

Facebook, 2012. Facebook Engineering Notes. [Online]
Available at: https://www.facebook.com/notes/facebook-engineering/under-the-hood-scheduling-mapreduce-jobs-more-efficiently-with-corona/10151142560538920

PCIe

PCIe is a serial bus capable of transmitting data over high-speed lanes, and multiple lanes can be aggregated to achieve even higher throughput. Most servers contain several PCIe slots designated as x1, x4, x8 or x16 where the integer represents the number of PCIe lanes. Slow devices can use single lane slots, while high-speed controllers, such as a SAS HBA, require an x8 slot to deliver satisfactory performance.

PCIe throughput is usually quoted in terms of Gigatransfers per second (GT/s) owing to the need to combine data and clock information on the bus, and this encoding overhead has an effect on data throughput. Second generation PCIe supports 5GT/s over a single lane; third generation PCIe increased performance by 60 percent to 8GT/s. A third generation x8 slot can, therefore, support 64GT/s, or in more familiar terms, a theoretical maximum of around 8,000 Megabytes per second (MB/s), as shown in Table 1.

1 PCI Lane 8 PCI Lanes

PCIe	2.x	3.x
Theoretical burst	4000MB/s	7877MB/s
Typical	3200MB/s	6400MB/s

Table 1 – Example of some maximum and typical data throughput in second and third generation PCIe buses

SAS

SAS is now in its third generation, and the performance has doubled with each new generation from the original 3Gb/s to 6Gb/s and now 12Gb/s. Like PCIe, SAS uses lanes, and high-performance storage systems normally aggregate multiple SAS lanes to support higher data rates. A third generation SAS x4 wide port, for example, provides a maximum half-duplex aggregate transfer rate of 48Gb/s or 4800MB/s, as shown in Table 2.

SAS x4	6Gb/s Gen 2	12Gb/s Gen 3
Maximum	2400MB/s	4800MB/s
Typical	2200MB/s	4400MB/s

Table 2 – Example of some maximum and typical data throughput in 6Gb/s and 12Gb/s SAS in a 4-lane port

Disk Drives

Both hard disk drives (HDDs) and solid state disks (SSDs) can support either SAS or SATA storage interfaces. SAS is normally used in servers, while SATA is preferred for PCs. One of the advantages of SAS is that it also supports the use of SATA drives, and these are often preferred in slower tiers of storage owing to their lower cost.

HDDs vary in performance according to the interface type and the rotational speed of the media. Fast-spinning SAS drives with a 6Gb/s interface can deliver a data rate of up to 230MB/s (or 200MB/s with 10K RPM drives), whereas SATA drives with a 6Gb/s interface normally deliver from 100 to 170MB/s. SSDs are capable of delivering data rates closer to the peak throughput of the 6Gb/s SAS interface, or around 550MB/s. Such peak performance should also be possible for high-performance SSDs with a third generation 12Gb/s SAS interface. There is no industry roadmap to 12Gb/s for SATA drives.

Bottlenecks and Application Behavior

Because the way applications access data has an impact on disk I/O and throughput performance, the nature of the applications involved is also a consideration when designing a storage system. Broadly speaking, applications can be divided into two types: those needing frequent access to small amounts of data in a random fashion; and those accessing large amounts of data sequentially.

For random access with HDDs, the biggest bottleneck often occurs in the time it takes to position the heads over the correct track (seek time latency) and waiting for the desired sector to rotate under the heads (rotational latency). This results in fewer I/O operations per second (IOPs) and, therefore, a lower overall throughput. No such latencies exist for SSDs when accessing data randomly, which is why SSDs have both a higher IOPs and a higher throughput in general.

For sequential access with either HDDs or SSDs, the majority of time is normally spent actually transferring data. Because sequential data is ideally stored in contiguous blocks, HDD head movement is kept to a minimum and rotational latency becomes less relevant. The storage bottleneck for sequential applications is, therefore, normally the data transfer rate of the interface.

Most data centers host a mix of applications requiring both sequential and random access. With a storage area network (SAN) or network-attached storage (NAS), there may be little opportunity to optimize for sequential or random access. With direct-attached storage (DAS), by contrast, such optimization is possible. For example, servers can be configured with HDDs to support sequential applications more cost-effectively, while random access applications that require high performance can be run on servers configured with SSDs for either primary storage or caching.

Storage System Performance Guidelines

When designing a storage system for high performance, it is necessary to understand the throughput limitations of each element. The typical throughputs of different types and generations of the three most critical elements discussed in this paper are summarized in Table 3.

HDD (6Gb/s SAS)	230MB/s
SSD (6Gb/s SAS)	550MB/s
6Gb/s SAS x4	2200MB/s
12Gb/s SAS x4	4400MB/s
PCIe 2.x	3200MB/s
PCIe 3.0	6400MB/s

Table 3 – Example of typical throughputs for different disk types, and generations of SAS and PCIe

Table 4 provides a summary of some sample configurations showing where the bottleneck exists when configured with a "full complement" of disks (the slowest element in the system). As shown, the need to support more disks (for capacity) requires the use of later generations of SAS and/or PCIe, and/or more SAS lanes. Looking at it another way, in systems with a small number of disks, their relatively low aggregate throughput becomes the bottleneck, so there is no need to "over-design" the configuration with later generation technologies and/or more SAS lanes. The disks referenced in the Table 4 example all have a 6Gb/s interface with a throughput of 230MB/s and 550MB/s for the 15K RPM HDDs and SSDs, respectively.

Configuration	Bottleneck (MB/s)	# of HDDs	# of SSDs
6Gb/s SAS x4 / PCIe 2.x	SAS (2200)	9	4
6Gb/s SAS x8 / PCIe 2.x	PCIe (3200)	14	6
12Gb/s SAS x4 / PCIe 2.x	PCIe (3200)	14	6
12Gb/s SAS x4 / PCIe 3.0	SAS (4400)	19	8
12Gb/s SAS x8 / PCIe 3.0	PCIe (6400)	28	12

Table 4 – Sample storage configurations showing each one's bottleneck and the number of drives supported at their peak throughput

Note that Table 4 assumes all drives are operating at their maximum throughput simultaneously, and this does not always occur. It is also important to note that IOPs is often more critical than throughput in many applications today, depending on the circumstances. For these reasons, each configuration is normally able to support many more disks than indicated.

Migrating to 12Gb/s SAS

The primary issue in the migration to third generation SAS is a familiar one: investment protection. Most organizations have made a significant investment in SAS disks, and want to preserve that investment when migrating to 12Gb/s SAS technology. The problem is: The third generation SAS standard maintains backwards compatibility by throttling down to the slowest SAS data rate in the system.

In small-scale point-to-point configurations, this is not always an issue because the migration would require upgrading both an Initiator and its Target. But in most organizations, such point-to-point configurations are rare. The system-level "slowest data rate" performance limitation, therefore, means that in organizations without point-to-point configurations would not be able achieve the 12Gb/s performance boost until all disks support this new standard.

Fortunately there is a way to overcome this limitation, and that requires understanding a little about how SAS expanders function. A SAS expander makes it possible for a single (or multiple) Initiator(s) to communicate with multiple Targets concurrently, as shown in Figure 1. Expanders help make SAS remarkably scalable, and because each is capable of supporting multiple disks, expanders also makes it possible to aggregate the throughput of those disks.

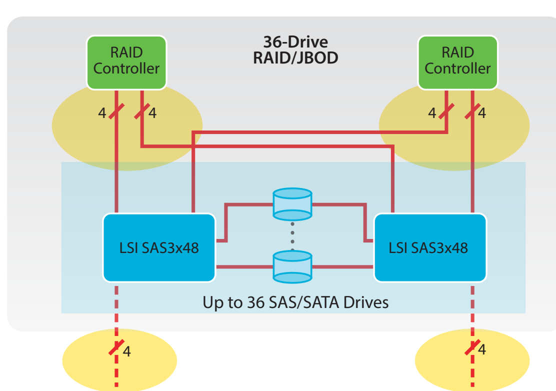

Figure 1 – The two 48-port SAS expanders shown here in this example are each connected to 12 Initiators and 36 Targets

Figure 1 shows redundant 48-port expanders connected to two RAID controllers and a shared disk enclosure. Additional disk enclosures can be connected in a daisy-chained configuration, and some SAS expanders are capable of supporting upwards of 2000 disks in such configurations.

According to the third generation SAS standard, if any of the SAS or SATA disks operate at 6Gb/s, the interface to the RAID controllers will also operate at 6Gb/s. But what if the bandwidth from two 6Gb/s disks could be aggregated into a single 12Gb/s channel, much like multiple lanes can be aggregated with PCIe?

Such Target (disk) bandwidth aggregation is the rationale behind what LSI calls DataBolt technology. Aggregation enables 12Gb/s SAS to be introduced into an existing 6Gb/s storage array in a way that provides an immediate doubling of overall system performance. In fact, with bandwidth aggregation, there is no need to use any 12Gb/s SAS disks to achieve 12Gb/s SAS performance at the system level.

Bandwidth aggregation works by using a 12Gb/s buffer on each port to enable what appears to be 12Gb/s communication between SAS Initiators and any speed (3Gb/s or 6Gb/s) SAS or SATA Target device. While the aggregation technology itself is proprietary, it operates completely internal to the expander, which enables all of the interfaces to be fully standardized. So a 12Gb/s SAS port connected to a 6Gb/s SAS disk would operate as a backwards-compatible second generation SAS port at 6Gb/s, but the system would operate at the third generation 12Gb/s data rate.

Conclusion

Storage bottlenecks migrate among the successive generations of the various technologies involved end-to-end. With the advent of third generation PCIe, for example, second generation SAS became the new storage bottleneck. Third generation SAS is now able to take full advantage of third generation PCIe's performance, making PCIe the new bottleneck in systems using 12Gb/s SAS.

DataBolt bandwidth aggregation technology used in third generation SAS solutions from LSI affords an immediate boost in IOPs and throughput performance, while protecting the investment in first and second generation SAS storage. SAS technology from LSI also affords substantial scalability with the LSI SAS Expanders able to support up to 2,000 devices and the LSI SAS 3008 I/O controllers used in the HBAs able to support up to 1,000 devices. In addition, Mini-SAS HD connectors on LSI SAS HBAs support longer distances, rack-to-rack configurations and other expanded topologies using optical and active cabling. To learn more about how your organization can benefit from third generation SAS solutions with DataBolt bandwidth aggregation technology, visit www.lsi.com.

For more information and sales office locations, please visit the LSI website at: www.lsi.com

North American Headquarters
San Jose, CA
T: +1.866.574.5741 (within U.S.)
T: +1.408.954.3108 (outside U.S.)

LSI Europe Ltd.
European Headquarters
United Kingdom
T: [+44] 1344.413200

LSI KK Headquarters
Tokyo, Japan
T: [+81] 3.5463.7165

LSI

Storage. Networking. Accelerated.

LSI, the LSI & Design logo, Storage.Nework.Accelerated., SandForce, DuraWrite, WarpDrive, MegaRAID and Nytro are the trademarks or registered trademarks of LSI Corporation in the United States and/or other countries. All other brand or product names may be trademarks or registered trademarks of their respective companies.

LSI Corporation makes no representations, warranties or guarantees regarding the compatibility or performance of specific hardware or software products. The information herein is provided "AS IS" without warranty of any kind express or implied for use at your sole risk. Without limiting the forgoing, the warranties of merchantability and fitness for a particular purpose are expressly disclaimed. This document is not warranted to be error-free. LSI Corporation has no responsibility to update the information herein and reserves the right to make changes to this document and to any of its products and services at any time without notice. LSI Corporation does not assume any responsibility or liability arising out of any reliance on or use of this information.

Copyright ©2013 by LSI Corporation. All rights reserved. > 1213