SAS: The Fabric for Storage Solutions

Marty Czekalski
President, SCSI Trade Association - Emerging Interface and Architecture Program Manager, Seagate Technology

Greg McSorley
Vice President, SCSI Trade Association
Technical Business Development Manager, Amphenol
The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:
- Any slide or slides used must be reproduced in their entirety without modification
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Abstract

- SAS is the backbone of nearly every enterprise storage deployment, rapidly evolving, adding new features, enhanced capabilities and offering “no compromise” system performance. SAS not only excels as a device level interface, its versatility, reliability and scalability have made it the connectivity standard of choice for creating new Enterprise storage architectures.

- This presentation covers the advantages of using SAS as a device interface, and how its capabilities as a connectivity solution, are changing the way data centers are being deployed. Advantaging 12 Gb/s SAS transfer rates, bandwidth aggregation, SAS Fabrics (including switches) active connections, and multi-function connectors (connectors that support SAS as well as PCIe Attached Storage devices) allows data center architects to create sustainable storage solutions that scale well into the future.
SAS Market Overview & Technology

 Basics of SAS Architecture & Deployment
SAS Market Evolution: Preserving the Past, Creating the Future

Preserve Legacy SCSI
- 25 years of SCSI middleware

Future Architected
- Protocol extends to new technologies
  - Serial, switchable
  - SFF connectors

Customer Choice
- 3.5” and 2.5” form factors
- Plug compatible
- Multi-protocol

Usability
- Dual-ported
- Point-to-point
- Cost equal to SCSI

Performance
- Wide ports
- Low overheard

Scalable
- 1000s of connections
SAS/SATA Compatibility

Disk Drive Connectors

SAS Connector Flip Side

Port B

SAS Backplane Connector

Accommodates both SAS & SATA Drives

SAS: The Fabric for Storage Solutions
© 2015 Storage Networking Industry Association. All Rights Reserved.
SAS & SATA Span the Storage Spectrum

Direct Attach Storage
- Controllers/ROCs/HBAs*
- Expanders
- SAS/SATA HDDs
- SAS/SATA SSDs
- Storage blades

SAS Fabrics
- NAS/SAN heads
- Native SAS connect
- Controllers/ROCs/HBAs*
- Expanders
- SAS/SATA HDDs
- SAS/SATA SSDs
- SAS/SATA tape

External Storage
- SAS/SATA tape
- Drive carriers
- Drive controllers

HDD/SSD
- SAS drives
- SATA drives

* ROC = RAID on a Chip
HBA = Host Bus Adapter

SAS is the Predominate Enterprise Drive Interface
Enterprise Market Revenue

Source: TRENDFOCUS, March 2015

SAS: The Fabric for Storage Solutions
© 2015 Storage Networking Industry Association. All Rights Reserved.
SAS Continues to Evolve
- Performance Gains without Protocol Changes

Expected Improvements w/12Gb/s SAS
- Protocol execution
- Application hints
- OS improvements
- Controller caching

Performance (4K Sequential IOPS)

Note: 12Gb/s SAS shipped at >1M IOPS in 2013!

6-10X performance gains in ~3 years
SAS: Bandwidth Aggregation

- Each SAS Link (Rx and Tx)
  - 12Gb/s → 24Gb/s (full-duplex)

- Wide Ports
  - Combine SAS links (12Gb/s SAS)
    - 2 ports > 48Gb/s (full duplex)
    - 4 ports > 96Gb/s (full-duplex)

- Efficient aggregation with Rate Matching and Frame Buffering

- Concurrency Brings Higher Performance Demands
  - Multiple concurrent I/Os
    (lots of drives operating concurrently)

Bandwidth **DOUBLES** with 12Gb/s SAS!
Connector Types

Mini-SAS
SAS 4X Connector

Preferred external connection scheme for 6Gb/s SAS
16 links in F.H. PCIe slot

Mini-SAS HD
SAS 4X Connector

Preferred external connection scheme for 12Gb/s SAS
16 links in H.H. PCIe slot

Preferred external connection scheme for 6Gb/s SAS
16 links in F.H. PCIe slot
Scalable Storage Ecosystem

Host Attach
SAN or NAS or SAS

SAS Switch (packaged view)

Various JBODs

Embedded Controllers
SAS Connected

SAS Switch

SAS Connected

SATA HDDs
Nearline SAS HDDs
SAS HDDs and SSDs

SAS: The Fabric for Storage Solutions
© 2015 Storage Networking Industry Association. All Rights Reserved.
SAS in Servers – Features & Benefits:

- Internal storage scalable via expanders
- Robust 12Gb/s SAS performance
- Scalable I/O performance
- High reliability & redundancy
- Point-to-point serial - simplified cabling
- Compatible with SATA & SAS HDDs & SSDs
Scaling outside the Box

**SAS External Drive Enclosures**

- 1U to 8U Form Factor
- Supports SFF 2.5-inch & 3.5-inch SAS/SATA/SSD Drives
- Higher Capacity with LFF SAS & SATA
  - Higher Port Count/Rack
  - Lower Power/Drive
- Scalable Expander or Low Cost Non-expander base Chassis Solution
- Hot-swappable Drive Carriers, Cooling Fans & Power Supplies
- Single & Dual Expander Options Available
- Large configurations for archive applications
High Availability-Dual Path to External Storage

- Dual paths from host to external storage enclosures as well as dual paths to individual (dual-ported) SAS drives

Cascade/Daisy Chain
Dual Domain/Redundant Path: 4 enclosures behind 2 SAS ports on a single server

Server/Cluster in a box

Dual path to each enclosure, and to individual dual ported drives
Switch Benefits and Management:

- Central management
- Multiple servers to one or more storage JBODs
- Efficient scale out
- OS independent
- Direct Attached Storage (DAS) or Shared Storage
- Port based zoning
- Drive bay zoning
- High Availability
SAS Connects the Tiers

- **Managed** - More scale, more cables, greater need for management
- **Distance** - Active copper (20m), optical (100m)
- **Performance** - 6Gb/s & 12Gb/s SAS
- **Density** - More “beachfront,” More ports
- **Consistency** - Standard method to scale distance & management
- **Converged** - One solution for external active & passive

![Tiers Diagram]

- **Tier 1** – SAS/SATA SSDs (few) – data frequently accessed
- **Tier 2** – SAS HDDs (dozens) (10k, 15K, Hybrid)
- **Tier 3** – SAS/SATA HDDs (100’s) (7.2K)
- **Cold Storage** – SATA HDDs (>100’s or tape) data infrequently accessed
12Gb/s SAS, 24Gb/s SAS, Advanced Connectivity
SAS Roadmap

First Plugfest (leading edge)

First End-User Products (approximately 12–18 months later)


3Gb/s SAS

12Gb/s SAS

24Gb/s SAS

*SAS Roadmap –SCSI Trade Association –March 2014*
12Gb/s SAS

- Standards Completed (SPL and SAS-3)
- Fourth 12Gb/s SAS Plugfest held Sept 2014
- Doubles the throughput while maintaining the same distance use cases
  - Copper, Active and Optical cable options
- Expect Controller capabilities >1 million IOPs
- Volume ramp for end users now
SAS Advanced Connectivity Objectives

- Drive market consistency
- Simplify cable & connector options
- 2X density improvement
- Provide converged high-density connectivity
- Provide managed connectivity standards
- Provide active copper solution to 20m
- Provide optical solution to 100m
- Support 6Gb/s & 12Gb/s SAS deployments

Supply power here for active cabling

Cable provides active component for optical or copper

SAS-3 standardizes OOB for active cables

Internal similar to External

Passive, Active Copper, or Optical use same connector
SAS Advanced Connectivity Roadmap
Managed Cable System

- New to SAS
- Managed Cables simplify configuration and ease of use
- OoB (Out of Band) method of controlling the interface
- Every pluggable device has an EEPROM or microprocessor that communicates with the system via a low-speed, two-wire interface.
- Allows each port to support short passive copper cables to 100m active optical cables
Connection Distances Comparison

<table>
<thead>
<tr>
<th>Meters</th>
<th>10</th>
<th>100</th>
<th>1000</th>
<th>10,000</th>
<th>Unlimited</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMB Remote site/Campus</td>
<td>Large Campus Data Centers</td>
<td>Enterprise Multi-floor infrastructure Many-to-many</td>
<td>Global Enterprise Disaster Recovery</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SAS**
- Passive
- Active

Cascadable links have been demonstrated several hundred meters

**FC**

**FCoE**
- Not IP Routable

**iSCSI**
- IP Routable

• As distance increases, performance can be negatively impacted.
• This slide represents a generalization. In certain cases, distances can be extended with multiple switches, tunnels, buffering/repeaters, etc.
24Gb/s SAS Benefits

- Preserves existing SAS use models
- Continues 6Gb/s SATA compatibility
- Maintains and supports SAS backward compatibility
  - Must be backward compatible 2 generations:
    - 12Gb/s SAS and 6Gb/s SAS
- Maximizes link utilization when using devices operating at less than 24Gb/s
- Improves storage system RAS attributes
- Doubles effective data transfer rate
- Maintains backward intermateability with existing connector options
24Gb/s SAS Status

Investigation results

- Connector studies indicate backplane, drive, and cable connectors (Mini-SAS HD) can be extended to work at 24Gb/s
- Signal loss simulations indicate channel lengths (cable and backplane) consistent with existing use models are achievable with encoding changes and forward error correction
- Leverage work done on other standards operating in the same frequency ranges

Conclusion

- Line of sight to 24Gb/s solutions
- Change from 8b/10b to 128b/130b encoding
- Add Forward Error Correction
- Better board materials can help
- SSDs will be a driving factor
SSDs, MultiLink SAS™, and Express Bay

Increasing SAS Bandwidth
MultiLink SAS™ Roadmap: Backplane Slot Location
Wide Port SAS for Increased Throughput

Server

SAS Controller

SAS SSD

SAS SSD

SAS SSD

SAS SSD

2.4 GB/s full-duplex per SSD
### Bandwidth per Device Connected

<table>
<thead>
<tr>
<th></th>
<th>SATA</th>
<th>SAS</th>
<th>Wide, Dual Link SAS</th>
<th>Wide, 4 Link SAS</th>
<th>PCIe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of Links / Lanes</strong></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Transfer Rate per Link/Lane</strong></td>
<td>Half-duplex 6 Gb/s</td>
<td>Full-duplex 12 Gb/s</td>
<td>Full-duplex 12 Gb/s</td>
<td>Full-duplex 12 Gb/s</td>
<td>Full-duplex 8 Gb/s</td>
</tr>
<tr>
<td><strong>Max Bandwidth</strong></td>
<td>0.6 GB/s</td>
<td>2.4 GB/s</td>
<td>4.8 GB/s</td>
<td>9.6 GB/s</td>
<td>8 GB/s</td>
</tr>
<tr>
<td><strong>Interface 4KB Random I/O Potential</strong></td>
<td>100K</td>
<td>450K</td>
<td>900K</td>
<td>1800K</td>
<td>1500K</td>
</tr>
</tbody>
</table>
Express Bay: PCIe, SAS, SATA and SATAe

Express Bay

- Up to 25 Watts
  - For both SAS (std and MultiLink) and PCIe
- SFF-8639 connector
- PCI-SIG electrical specification

Objectives

- Preserve the enterprise storage experience for PCI Express storage
- Meet SSD performance demands
- Serviceable, hot-pluggable Express Bay opens up new possibilities …

Note: Specific configuration, protocol and power support will be OEM specific
SAS Connector Compatibility

1 Max two links operational
2 Four links operational
3 Two or four links operational depending on host provisioning

SATA 22 Pins

SATA

SATA 22 Pins

SATA

SAS 29 Pins

MultiLink SAS™ 3

MultiLink SAS™ 2

MultiLink SAS™ 1

SCSI Express

SATA 32 Pins

SATA 43 Pins

MultiLink SAS™ 2

MultiLink SAS™ 3

SCSI Express

SATA 68 Pins

SATA

SAS

MultiLink SAS™ 3

SCSI Express
SSD Performance Scales with Power

- \( P = P_{\text{base}} + P_{\text{I/O}} \times \text{IOPs} \)

- **Power Limit Control**
  - Allows system and device to negotiate allowable power usage
  - Both SAS and PCIe have this capability
  - For more bandwidth, additional links are needed
Power Limit Control

What power levels are supported?

e.g. 9W, 15W, 25W

Set Power Level
New Features Nearing Completion

- Zoned Block Commands
- Storage Intelligence
Why Shingled Magnetic Recording?

Areal Density Accelerator

- Much higher track density
- Overall AD growth 25%/Y vs. 15%/Y
- 65% greater maximum capacity by 2020
Conventional vs. SMR Writing

Conventional Writing

Reader

Writer

Track N

Track N+1

Track N+2

Track N+3

Shingled Writing

Reader

Writer

Track N

Track N+1

Track N+2

Track N+3

Track N+4
Zoned Block Device Drive Types

- **Drive Managed**
  - Drive autonomously hides all SMR issues
  - Backward compatible

- **Host Aware**
  - Superset of Drive Managed and Host Managed
  - Backward compatible
  - Extensions to ATA and SCSI command sets

- **Host Managed**
  - Extensions to ATA and SCSI command sets
  - Error conditions for some reads and writes
  - Not backward compatible
  - New device type
Overview of Bands and Zones

- **SMR Bands**
  - Physical construct
  - Boundaries are not known outside the drive

- **Zones**
  - Logical space is divided into zones
    1. Conventional zones
    2. Write pointer zones
      - Each has a write pointer
      - Automatically advances
      - Host issues reset before re-write
    1. Sequential Write Preferred zones
      Random writes are supported
    2. Sequential Write Required zones
      Random writes are not allowed
Write Pointer Zones

- Writes at the write pointer have conventional performance
  - Write pointer automatically advances

- Issue Reset Write Pointer before re-writing

![Write Pointer Zones Diagram]

- **Empty**
  - Write pointer is at start of zone

- **Open**
  - Write pointer is mid-zone

- **Full**
  - No write pointer value
ZBC Summary

- Standards are in final stages of review
  - ZBC - completed letter ballot it T10
  - ZAC - Q2CY2015 letter ballot target

- SMR enables decreasing storage costs and increased capacity
  - Some applications may need to be tuned to optimize performance

- Interest in SSD community
  - Applications could potentially improve performance and endurance of SSDs
    - Reduced degradation in performance as devices age
Streams
- Provides hints to SSD about data sets that have similar expected lifetimes
- Reduces intermixing of data from different applications, thus reducing fragmentation during garbage collection
- Improves performance
- Reduces write amplification and improves endurance

Background Activity Control
- Provides hints to SSD to optimize timing of background activities (e.g., garbage collection)
- Provides more consistent performance during peak activity times

Standards proposals are in final stages of acceptance for inclusion in SBC4
Summary

- Logical SCSI lives across the Storage Spectrum
- Proven Enterprise attributes & growing footprint
- Flexible architecture = Platform for Innovation
  - Enterprise reliability, improved connectivity
  - Low latency device performance, capacity scaling
  - Multitude of other storage dimensions
- A Multi-generational Investment
  - Enterprise-hardened middleware
  - Preserves backward compatibility
  - Roadmaps continue to preserve legacy investments
Attribution & Feedback

The SNIA Education Committee thanks the following individuals for their contributions to this Tutorial.

**Authorship History**
Marty Czekalski/Aug 2012,2013
Updates:
Marty Czekalski/March 2014

**Additional Contributors**
Harry Mason
Greg McSorley
Dennis Martin
Rick Kutcipal
Jay Neer
STA Members

*Please send any questions or comments regarding this SNIA Tutorial to tracktutorials@snia.org*