Building an All-Flash Array with SAS, NVMe or SATA

Cameron T Brett
Board of Directors, SCSI Trade Association
Director, SSD and Storage Solutions, Toshiba Memory America
All-Flash Array Type Forecast (2013-2022)

- 2017 AFA $6.2B
- IDC sees AFAs growing to 51% of all arrays by 2022
- Steep decline of HDD-only arrays
- Slow decline of hybrid arrays

Source: IDC, September 2018
SAS Is the Main Enterprise Storage Interface

WW Enterprise Drive Unit Forecast (2017-2022)

WW Enterprise Drive Capacity Forecast (2017-2022)

SAS Infrastructure Enables >70% of Enterprise Storage Drives and >85% of Enterprise Storage Capacity thru 2022

Source: IDC, May 2018
All-Flash Arrays
Built on SAS
SAS Overview

- Developed in early 2000’s. Based on SCSI protocol (1986)
- Data rates\(^1\): Up to 12Gb/s, with 24G SAS coming soon
- Cable lengths: Up to 100m optical; up to 10m copper
- Typical deployment: server; enterprise storage; high-performance applications; large scale requiring up to 1000s of devices
- Supports both SAS and SATA HDDs and SSDs
- SCSI protocol is supported with other storage technologies, such as FC, USB, 1394/FireWire, iSCSI
- Top SAS SSD suppliers – Toshiba, HGST/WD, Seagate, Samsung
- 33.07M SAS drives sold in 2017\(^2\) (3.0M SSDs and 30.07M HDDs)

\(^1\) Single-lane performance  \(^2\) Source: TRENDFOCUS, 2018
SAS AFA Enterprise Data Center Use

**Storage Area Network**

- Servers
- Ethernet or FC switches
- SAS AFAs

**Direct Attached Storage**

- Servers
- SAS AFAs
SAS AFA – Closer Look Inside

Network

Fibre Channel or Ethernet to a SAN

SAS HBA if DAS to a server

Backplane could be active or passive

CPU

SAS IOC

SAS Expander

SAS IOC

SAS Expander

NIC

Backplane

SAS SSDs
The Value Pillars of SAS

- **Reliable**
- **Secure**
- **Trusted**
- **Data Center Ready**

**Dependability**
- SAS and SATA
- SSDs and HDDs
- Backward Compatibility

**Flexibility**
- 1000s of End Devices
- Capacity Storage
- Enables “Fabric” Solutions

**Scalability**
- $ / Performance
- $ / GB
- Highest SSD capacities
- Huge Established Ecosystem
- Development

**Economics**
What’s Next?  24G SAS Highlights

**Physical Layer Enhancements**
- Double the effective single-lane bandwidth of 12Gb/s SAS
- Higher throughput and IOPs performance
- Enhanced 20-bit Forward Error Correction (FEC)
- More robust data reliability and connectivity
- SAS-4 enhanced transmitter training algorithm
- Continuous optimal signal tuning

**Protocol & Block Level Enhancements**
- Performance consistency across large and mixed protocol topologies
- Improves SSD efficiency, latency and QoS
- Prioritizes management-class communications for complex, deep topologies
- Fairness enhancements
- Storage intelligence and persistent connections
- SMP priorities
All-Flash Arrays Built on NVMe
NVMe Overview

- Introduced in 2011 with large industry backing; designed for SSDs
- Works over PCIe. Ethernet, Fibre Channel, IB for NVMe-oF.
- Robust roadmap plans for protocol, NVMe-oF and management
- Data rates\(^1\): Up to 16Gb/s
- Connection lengths: 1m internal, external not defined for PCIe
- Typical deployment: server, HFA/AFA, cache tier, high-performance, lowest-latency applications
- Many, many companies developing NVMe storage products
- 3.93M NVMe drives sold in 2017\(^2\)

\(^1\) Single-lane performance  \(^2\) Source: TRENDFOCUS, 2018
NVMe AFA Enterprise Data Center Use

Storage Area Network

- Servers
- Ethernet or FC switches
- NVMe AFAs

NVMe-oF Network (Disaggregated)

- Top of Rack Switch
- Servers
- NVMe AFAs
NVMe AFA – Closer Look Inside

Fibre Channel, Ethernet or IB connectivity

48 PCIe lanes from the CPU

Backplane may be active or passive

NVMe-oF or SAN

NVMe enclosure management not as mature as SAS

PCIe re-timers may be required
NVMe Value Proposition

- Large industry backing; multiple sources
- Optimized for flash storage and next generation NVM
- Lowest latency and highest performance for typical deployments
- Lower AFA BOM cost for smaller topologies (no IOC required)
  - Large and HA topologies require additional switches
- Best for high performance storage tiers (i.e. Tier-0)
- Form factor and performance variations for client, data center and enterprise applications
All-Flash Arrays
Built on SATA
SATA Overview

- Developed in early 2000s
- Roadmap ends with 6Gb/s; no further physical layer development
- Data rates\(^1\): 6Gb/s, 3Gb/s, 1.5Gb/s
- Cable lengths: Up to 1m
- Typical deployment: desktop, server and storage; near-line and cold storage; low-cost, high-capacity applications,
- In an AFA or enterprise server, SATA devices commonly deployed behind SAS infrastructure
- 50.61M SATA drives shipped in 2017 (15.78M SSDs and 34.83M HDDs)\(^2\)

\(^1\) Single-lane performance
\(^2\) Source: TRENDFOCUS, 2018
SATA AFA Enterprise Data Center Use

**Storage Area Network**
- Servers
- Ethernet or FC switches
- SATA AFAs

**Direct Attached Storage**
- Servers
- SATA AFAs
SATA AFA – Closer Look Inside

Network

Fibre Channel or Ethernet connectivity

CPU

NIC

SAS IOC

SAS Expander

PCI bus

Backplane

SATA SSDs

Network or Ethernet connectivity

SSDs

CPU

NIC

SAS Expander

Backplane

SATA SSDs
SATA Value Proposition

- It’s cheap - leverages economies of scale to lower cost
- Lowest power consumption
- Leverages SAS infrastructure
- Both SSDs and HDDs supported
- Will have a lengthy tail and will be around for a long time…..
## Summary Comparison of Typical Deployment

<table>
<thead>
<tr>
<th></th>
<th>SAS</th>
<th>NVMe</th>
<th>SATA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance (IOPS, GB/s)</strong></td>
<td>Better (x1 lane)</td>
<td>Best (x4 lanes)</td>
<td>Good</td>
</tr>
<tr>
<td><em><em>Performance (Read Latency</em>)</em>*</td>
<td>Good</td>
<td>Best</td>
<td>Good</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Best</td>
<td>Good</td>
<td>Better (SAS infrastructure)</td>
</tr>
<tr>
<td><strong>Power (per drive)</strong></td>
<td>Better (9-12W)</td>
<td>Good (up to 25W)</td>
<td>Best (~6W)</td>
</tr>
<tr>
<td><strong>Flexibility</strong></td>
<td>Best (SAS, SATA, HDDs, SSDs)</td>
<td>Good (SSDs)</td>
<td>Better (SSDs, HDDs)</td>
</tr>
<tr>
<td><strong>Manageability</strong></td>
<td>Best (most mature)</td>
<td>Good (recent spec)</td>
<td>Better (SAS infra)</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>Best</td>
<td>Better</td>
<td>Good</td>
</tr>
<tr>
<td><strong>System cost</strong></td>
<td>Higher</td>
<td>Highest (performance premium)</td>
<td>Lowest</td>
</tr>
<tr>
<td><strong>Roadmap future</strong></td>
<td>Long-term</td>
<td>Long-term</td>
<td>Limited</td>
</tr>
</tbody>
</table>

*Latency includes OS, driver, HBA (if required) and flight time, media access times not included
Summary

- SAS is a modern, and mature interconnect
  - Actively being developed with long roadmap plans
  - New features to enhance SSD functionality and performance
  - Highly scalable and flexible
  - Value SAS may cannibalize SATA
  - 3Gb/s, 6Gb/s, 12Gb/s, 24G (sampling today)

- NVMe is still new, but will mature over time
  - Highest performance with lowest latency
  - Very active development and robust, long-term roadmap
  - Development of specifications for networked NVMe storage and management

- SATA has no future roadmap
  - Unit shipments of SATA devices still strong; will have a long tail