



Education

PCIe SSD Storage

Gary Kotzur, Dell

- ◆ 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.

- Benefits of PCIe SSD Storage
- Tradeoffs versus other options
- Enabling technologies and standards

- Multiple vendors shipping today
 - ◆ FusionIO, Virident, LSI, Micron, OCZ, Smart Modular ...
- Performance versus SAS
 - ◆ BW: PCIe is 4-5X faster
 - ◆ IOPs: PCIe is 3-6X faster
 - ◆ Latency: PCIe is 2-3x lower
- Form Factors
 - ◆ Cards: Low-profile, Full-height/half-length, Full-height/full-length
 - ◆ Appliances: Enclosures, single Racks and multiple Racks
- Media
 - ◆ SLC
 - ◆ MLC

PCIe as a SSD Interface

➤ PCIe is high performance

- ◆ Full duplex, multiple outstanding requests, and out of order processing
- ◆ Scalable port width (x1 to x16)
- ◆ Scalable link speed (2.5 GTps, 5 GTps, 8 GTps)
- ◆ Low latency (no HBA overhead or protocol translation)
- ◆ Low Overhead – encoding is 1.5%

➤ PCIe is low cost

- ◆ High volume commodity interconnect
- ◆ Direct attach to CPU eliminates HBA cost

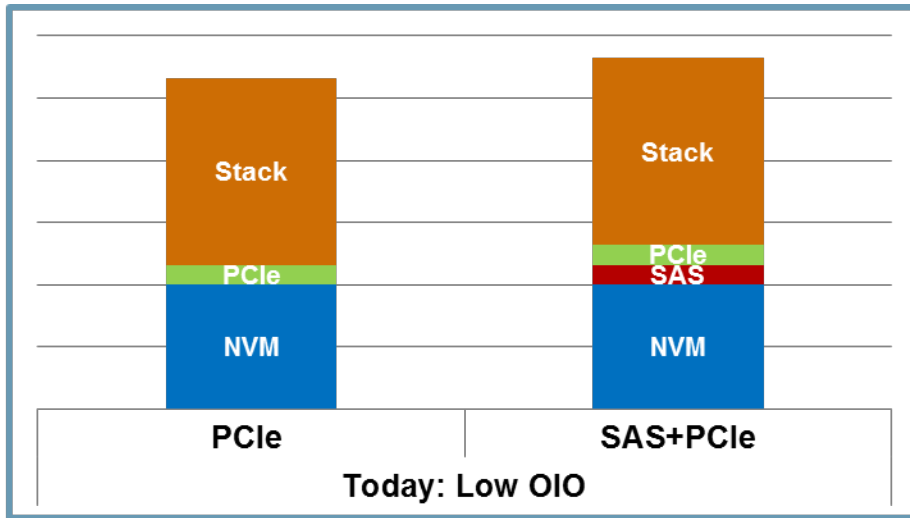
➤ PCIe power management capabilities

- ◆ Direct attach to CPU eliminates HBA power
- ◆ Features include: Link power management, Optimized Buffer Flush/Fill (OBFF), Dynamic Power Allocation, Slow Power Limit, etc

PCIe Storage Strengths

- Current PCIe SSD cards:
 - ◆ Well received by customers, have attained highest performance to date.
- Complement existing storage protocols
 - ◆ Providing highest IOPs and lowest latency for demanding applications
- Obvious advantages: Reduced path components
 - ◆ Lower costs
 - ◆ Less real estate
 - ◆ Less Power
 - ◆ Higher reliability
 - ◆ Lower latency

More on Latency ...



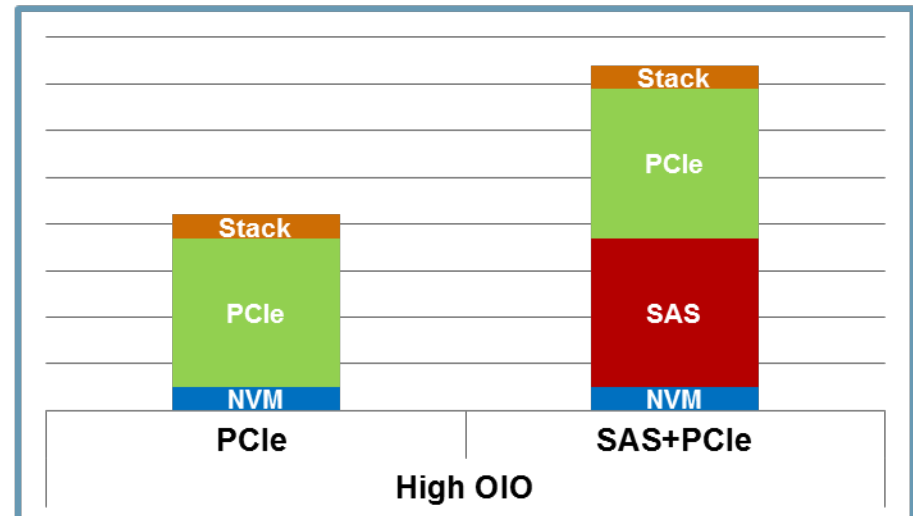
Today: Low OIO (Outstanding IO)

- High latency NVM and legacy stack can diminish interface latency benefits

Today: High OIO

Future: Upcoming advances

- Parallelism reduces NVM and stack aggregate latency, seen today in database work loads
- Future NVM can achieve low latency even at low OIO
- Path latency of a PCIe solution can be much lower than SAS solution, providing significant performance improvement

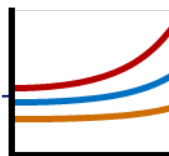


Why PCIe Storage Standards?

Areas to Address

Performance Trends

Processor vs. Storage
Gap Increasing



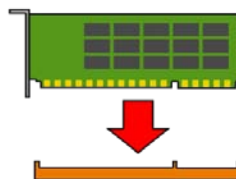
Serviceability

Internal Access
Cold-Plug



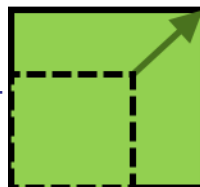
Interoperability

Card Form Factor
Varying Card Sizes



Scalability

Performance
& Capacity



PCIe SSD Benefits

Minimize Gap

Improved Latency
Improved IOPs

Remove Constraints

External Access
Hot-Pluggable

Common Form Factor

Drive Form Factor
Multi-protocol

Increased Slots

External Slots
“Live” Scaling

*other brands and names may be claimed as the property of others

- Increased Performance of PCIe
- High Availability and Serviceability
- Compatibility:
 - ◆ SAS/SATA/PCIe
 - ◆ Standard driver for each OS
- Improved Power Efficiency
- Reduced TCO

*other brands and names may be claimed as the property of others

- SSD FF WG was developed by industry consortium of 49+ members and is directed by a 5 company Promoter Group
- Charter
 - ◆ Promote Enterprise Storage usage of PCIe SSDs, by enabling serviceability, high-availability, ease of integration, interoperability and scalability of Solid-State Storage
- Elements
 - ◆ Form Factor
 - ◆ Connector
 - ◆ Hot-Plug

Working Group Key Elements

Form Factor



**Benefit from current 2.5”
HDD form factor**

Expand power envelope

Connector



Multiple protocols:

PCIe 3.0, SAS 3.0, SATA 3.0

Management Bus

Dual port (PCIe)

Multi-lane capability (PCIe/SAS)

Power pins

**SAS Drive Backward
Compatibility**

Hot-Plug



Hot-Plug Connector


**Identify desired drive
behavior**

**Define required system
behavior**

- 2.5” Form Factor Specification is released
 - ◆ Rev. 0.7 in Feb’11 released internally to working group members
- Drive Connector Mechanical and Pinout Specification released
 - ◆ Rev. 1.0 released in March’11
 - ◆ Rev. 1.1 released in May’11
 - ◆ Released to SFF
 - › Content in SFF-8639
 - › Actively working towards industry alignment for “one connector” for PCIe and SAS
- More information
 - ◆ Website: www.ssdformfactor.org
 - ◆ Email: info@ssdformfactor.org

- NVM Express is a scalable host controller interface designed for Enterprise and Client systems that use PCI Express* SSDs
 - ◆ Includes optimized register interface and command set
- NVMe was developed by industry consortium of 80+ members and is directed by a 10 company Promoter Group
- NVMe 1.0 published on March 1st, available at *nvmexpress.org*

NVMe: Efficient SSD Performance

	AHCI ¹	
Uncacheable Register Reads Each consumes 2000 CPU cycles	4 per command 8000 cycles, ~ 2.5 μs	0 per command
MSI-X and Interrupt Steering Ensures one core not IOPs bottleneck	No	Yes
Parallelism & Multiple Threads Ensures one core not IOPs bottleneck	Requires synchronization lock to issue command	No locking, doorbell register per Queue
Maximum Queue Depth Ensures one core not IOPs bottleneck	32	64K Queues ¹ 64K Commands
Efficiency for 4KB Commands 4KB critical in Client and Enterprise	Command parameters require two serialized host DRAM fetches	Command parameters in one 64B fetch

- NVMe drives broad adoption of PCI Express* SSDs by:
 - ◆ Enabling standard drivers across a wide range of OSES
 - ◆ Driving a consistent feature set across SSDs
 - ◆ OEM does not need to qualify separate driver for each SSD
- Drivers are coming online for major OSES
 - ◆ Linux* driver available at nvmexpress.org
 - ◆ IDT*, Intel, and SandForce* actively developing Windows* driver that will be released open source in Q1 '12
- The University of New Hampshire IOL is creating an interoperability test suite and integrator's list
 - ◆ LeCroy PCIe Protocol Analyzer includes NVMe command decode

➤ NV Memories

- ◆ Flash
- ◆ Phase Change Memory
- ◆ MRAM
- ◆ Memristor
- ◆ More ...

➤ SSD controller enhancements

- ◆ Higher parallelism
- ◆ Improved performance consistency
- ◆ Additional features: Power, Reliability, Endurance

➤ PCIe Switches

- ◆ Increased lanes
- ◆ Version 3.0

Enabling Technologies (con't)

- PCIe storage support
 - ◆ Hot-plug
 - ◆ Error-reporting
- Stack optimizations
 - ◆ Lower latency
- Applications
 - ◆ Adaptations to SSD accesses
- OS optimizations
 - ◆ Trim
 - ◆ Align to SSD behaviors

Customer Benefits Summary

- **Increased Performance of PCIe**
 - ◆ High Throughput
 - ◆ Low latency
- **High Availability and Serviceability**
 - ◆ Extended RAS capability in a common form factor
 - ◆ Known drive replacement behavior
- **Compatibility**
 - ◆ Standardization reduces issues
 - ◆ Single connector for SAS/SATA/PCIe
 - ◆ Standard driver(s) for multiple OSes
- **Improved Power Efficiency**
 - ◆ Higher performance from same media
 - ◆ Improved IOPs/Watt
- **Reduced TCO**
 - ◆ Reduce component complexity
 - ◆ Improved \$/IOPs

- Please send any questions or comments on this presentation to SNIA: tracktutorials@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**

- SNIA Education Committee

**Jim Pappas, Intel
Jason Leone, EMC
Berndt WinkleStraeter, FTS
Adam Roberts, IBM
Amber Huffman, Intel**