

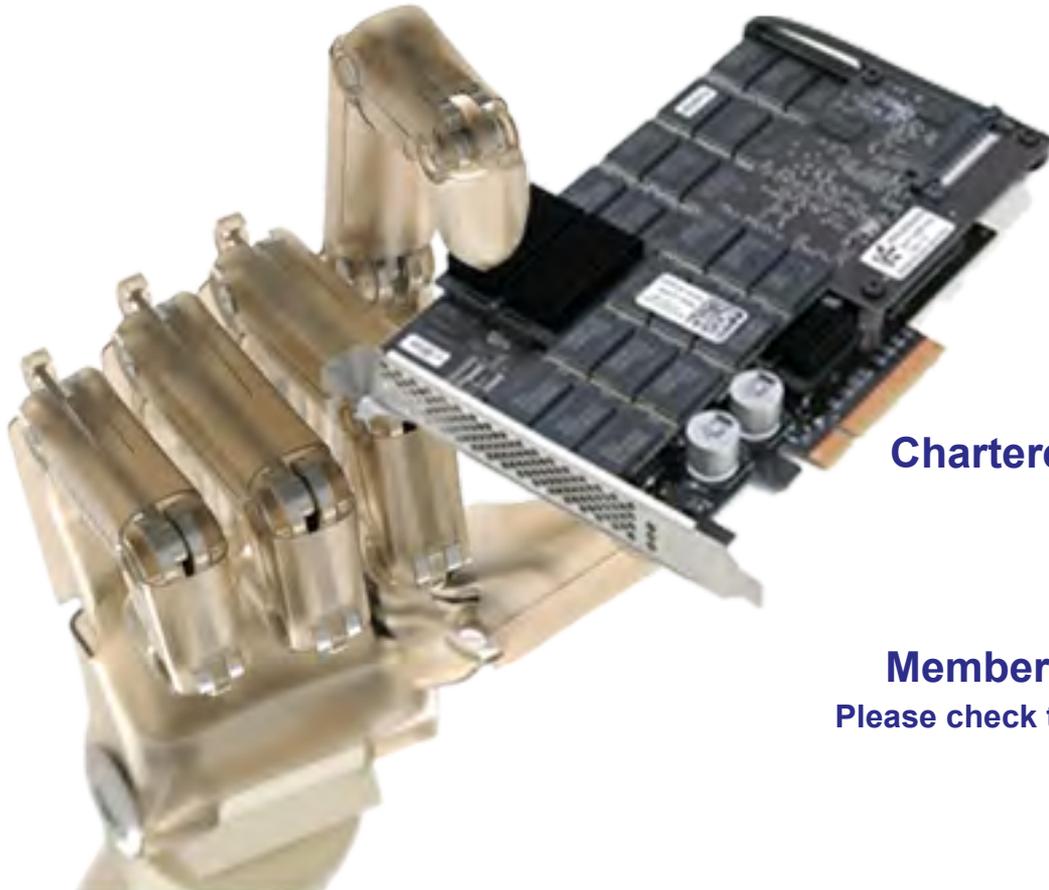
**SNIA Solid State Storage Initiative
PCIe SSD Task Force Meeting No. 5**

Monday 04 June 2012



WELCOME!

Meeting No. 5
Monday 04JUN2012
4:00 PM - 5:30 PM PST



Welcome to the SNIA
Solid State Storage Initiative
PCIe SSD Task Force

This is an Industry Task Force
Chartered to investigate, discuss & educate
All things PCIe SSD

Membership is Complimentary for (90) Days
Please check the homepage at www.snia.org/forums/sssi/pcie

Task Force Participants



Concall Guidelines:

- Use your mute button
- **Sign in webex w/ company name**
john smith (ABC Co.)
- Be on time for roll call
- Un Mute when talking
- Use webex chat to ask questions

- Respond to Feedback Requests
- Email String Topic Discussions
- Email comments to reflector
pciesd@snia.org
- Send Questions to
pciechair@snia.org

(8) OPEN Meetings - Apr - Jul SSSI Committee Aug - Dec 2012

Topics	09APR12 Mtg No. 1	23APR12 Mtg No. 2	07MAY12 Mtg No. 3	21MAY12 Mtg No. 4	04JUN12 Mtg No. 5	18JUN12 Mtg No. 6	02JUL12 Mtg No. 7	16JUL12 Mtg No. 8	30JUL12 SSSI Committee
Kick-Off Mtg Issue Identification	X								
Standards		X							
Test Platforms		X							
Performance			X						
System Integration			X						
System Arch Form Factors				X					

Goals: Issue Identification & Committee 2012 Roadmap

(8) OPEN Meetings - Apr - Jul

SSSI Mtg & FMS Round Table

Topics	04JUN12 Mtg No. 5	18JUN12 Mtg No. 6	02JUL12 Mtg No. 7	16JUL12 Mtg No. 8	30JUL12 SSSI Committee	20AUG12 Flash Memory Summit
Big Picture What's it all Mean	X					
Deployment Strategies Market Development		X				
Where do we go from here?			X Holiday?			
Roadmap & Milestones 2012				X		
SSSI Committee Mtg					X	
SSSI Task Force F2F FMS Round Table						X SNIA / SSSI FMS Reception

Tentative Topics for Meetings

Time: 4:00 – 4:05

AGENDA – 04 JUNE 12

I.	Administrative	
a.	Roll Call; Call Schedule	4:00 – 4:05
b.	Announcements	4:05 - 4:10
II.	Business	
1.	PCIe I/O Performance Considerations - Tom West, HyperIO	4:10 - 4:30
2.	PCIe Standards: How they fit together - Paul Wassenberg, Marvell	4:30 - 4:50
3.	PCIe, SCSI Express, and a world beyond disk drives; What happens when you get new native flash capabilities - Gary Orenstein, Fusion-io	4:50 - 5:20
III.	Wrap Up	
a.	Discussion	5:20 – 5:30
b.	Close	

Attendance

I of 4

Company	No	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Agilent	1	x	x	x	x	x			
Allion	2	x	x		x	x			
AMD	3	x	x						
Apacer	4								
BitSprings	5				x	x			
Cadence	6	x							
Calypso	7	x	x	x	x	x			
Cisco	8		x	x	x				
CLabs	9								
Corsair	10			x					
Coughlin Associates	11	x	x	x					
Dell	12	x							
eAsic	13	x			x	x			
EMC	14	x	x	x	x	x			
Enmotus	15	x			x	x			
eTron	16		x						
Fusion-io	17	x	x		x	x			
Greenliant	18		x	x	x	x			

Attendance

2 of 4



Advancing storage & information technology

Company	No	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
HDS	19	x		x	x				
HP	20	x	x	x	x	x			
HGST	21	x	x	x	x	x			
Huawei	22	x	x	x	x	x			
Hynix (SK Hynix)	23		x	x	x				
HyperIO	24	x	x	x		x			
IBM	25	x		x					
Intel	26	x	x	x	x	x			
Kingston	27								
Lecroy	28		x	x	x	x			
Lenovo	29	x		x		x			
LiteOnIT	30								
Link-A-Media	31								
Lotes	32								
LSI	33				x	x			
Lunastar	34	x	x	x	x				
Marvell	35		x		x	x			
Micron	36	x	x	x	x	x			
Microsoft	37	x	x	x					

Time: 4:00 – 4:05

Attendance

3 of 4



Advancing storage & information technology

Company	No	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Molex	38	x	x	x	x	x			
Mushkin	39								
NetApp	40				x				
Objective Analysis	41	x			x	x			
OCZ	42	x							
Oracle	43	x	x	x					
PC Perspectives (Allyn M.)	44								
PLX Technology	45			x	x				
Phison	46	x	x	x	x	x			
Renesas	47	x	x	x	x	x			
Samsung	48	x	x	x	x	x			
Sandisk	49	x	x						
Seagate	50	x			x	x			
Semtech Snowbush IP	51								
Smart Storage	52		x	x	x	x			
SNIA	53	x	x	x	x	x			
STEC	54	x	x	x					

Time: 4:00 – 4:05

Attendance

4 of 4 - TOTAL (64) Companies

Company	No	09APRI12	23APRI12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Taejin	55	x	x	x					
Tektronix	56								
TMS	57	x							
Toshiba	58		x	x	x				
Tyco Electronics	59	x	x						
Unigen	60	x	x	x					
Viking	61								
Virident	62	x	x	x	x				
WDC	63	x	x	x	x				
Paul Mitchell (Ind)	64								
		39	37	34	33	27			

Announcements - Eden Kim

1. Announcements / Other

a. Minutes Meeting No. 4 - next slide

b. Task Force Charter / Structure - Standing Slides

c. Announcements:

1. Meeting No. 6 - Deployment Strategies & Market Development

a. Walter Hubis, Fusion-io - PCIe as Persistent Memory

b. Jim Handy, Objective Analysis - Market Issues

c. Andy Rudolf, Intel - Developing an Open Kernel NVM Programming Model

2. Topics - Please email if you are interested to present at Meetings no. 7-8

a. Meeting no. 7 - "Where do we go from Here?" 02JUL12 - Cancel due to 04JUL Holiday?

b. Meeting no. 8 - "Roadmap & Milestones for PCIe Committee"

c. Meeting no. 9 - "SNIA & SSSI - presentations to Task Force"

3. SNIA/SSSI Flash Memory Summit Reception: email to pciechair@snia.org if you plan to attend

What: Reception : SSSI PCIe Committee - Roadmap 2012

Why: New Member Introduction; SSSI PCIe Committee Inauguration

Who: Invitation only : sponsored by SNIA and the SSSI

When: Monday 20AUG12 5:30 - 7:00 pm

Where: Santa Clara Convention Center 2d floor

Minutes Meeting No. 3

21MAY12

1. Attendance:

- a. (34) Companies present of (65)
- b. Presentations / Speakers - Micron, PCI SIG, Agilent

2. Administrative:

- a. **Task Force Charter** - General Survey of PCIe SSD issues; Recommendations for SSSI Committee 2d half 2012
- b. **Task Force Structure** -

- 1. (3) **Open Mtgs left:** 18JUN12, 02JUL12, 16JUL12
- 2. (1) **Special Mtg:** 30JUL12

3. Meeting Topics:

Meeting No. 5 - "What's the Big Picture"

Meeting No. 6 - Marketing Issues & Deployment Strategies

Meeting No. 7 - "Where do we go from here?"

Meeting No. 8 - Recommended Roadmap & Milestones for SSSI PCIe Committee

Meeting No. 9 - Presentation of SNIA / SSSI to Task Force

Meeting No. 10 - SNIA / SSSI FMS Reception 20AUG12 at Santa Clara Convention Center

Links: Email Reflector: pciessd@snia.org PCIe Task Force Homepage: www.snia.org/forums/sssipoints

- 3. **PCIe 2.5" Form Factor** - Janene Ellefson, Micron
- 4. **PCI SIG** - Rick Eads, Agilent
- 5. **PCIe System & Form Factor Issues** - Rob Vezina, Agilent
- 6. **Next Actions** -
 - a. **Feedback on Meeting Topics** - post to reflector
 - b. **Agenda / Topics** - Meeting No. 5

Close 5:30PM

Task Force Charter: GENERAL SURVEY OF PCIE ISSUES

1. Provide Guidance to Marketplace about PCIe SSDs

1. Educational Materials
2. Best Practices Documents
3. Industry Standards Work

2. Coordinate w/ other Industry Organizations

1. Complement other groups
2. Avoid Overlap
3. Fill Voids

3. Open Industry Forum to SSSI Committee

1. (90) Day Free Trial Membership
2. SNIA SSSI Membership Required Aug 2012
3. No IP/NDA - No Confidential Information may be discussed
4. Identify Issues & Define Roadmap for Committee

Task Force Structure:

- 1. Webex Meetings - Every other Monday**
 1. Starting Monday 09APR12 and every two weeks thereafter
 2. 4:00 PM - 5:30 PM PST
 3. (8) Open Calls prior to SNIA/SSSI Membership Requirement

- 2. Email Reflector - pciessd@snia.org**
 1. Agenda, Minutes & Discussion via reflector until 30JUL12
 2. Post Meeting Survey's for feedback and agenda preparation
 3. Email reflector becomes SSSI member only starting 30JUL12

- 3. Target Objectives for (90) Day Public Forum Period**
 1. Table of Standards Groups
 2. Recommendation on PCIe Hardware Test Platform Standard
 3. Identification of PCIe SSD Performance Issues
 4. Hosting of PCIe Round Table Panel
 5. Other Objectives defined by Task Force
 6. Identity Issues & Recommend SSSI PCIe Committee Roadmap for 2012

SSSI

- SSSI homepage www.snia.org/forums/sssi
- Understanding SSD Performance Project www.snia.org/forums/sssi/pts
- SSS Performance Test Specification (PTS) www.snia.org/pts
- PTS Standard Report Format www.snia.org/forums/sssi/pts
- SSSI Bright Talk Webcasts www.snia.org/forums/sssi/knowledge/education
- SSSI White Papers www.snia.org/forums/sssi/knowledge/education

PCIe Task Force

- PCIe SSD Task Force www.snia.org/forums/sssi/pcie
- PCIe SSD Task Force reflector pciessd@snia.org
- PCIe SSD Task Force questions pciechair@snia.org

PCIe I/O Performance Considerations

Tom West, HyperIO



Advancing storage & information technology

How IOs Traverse the I/O Software/Hardware Stack

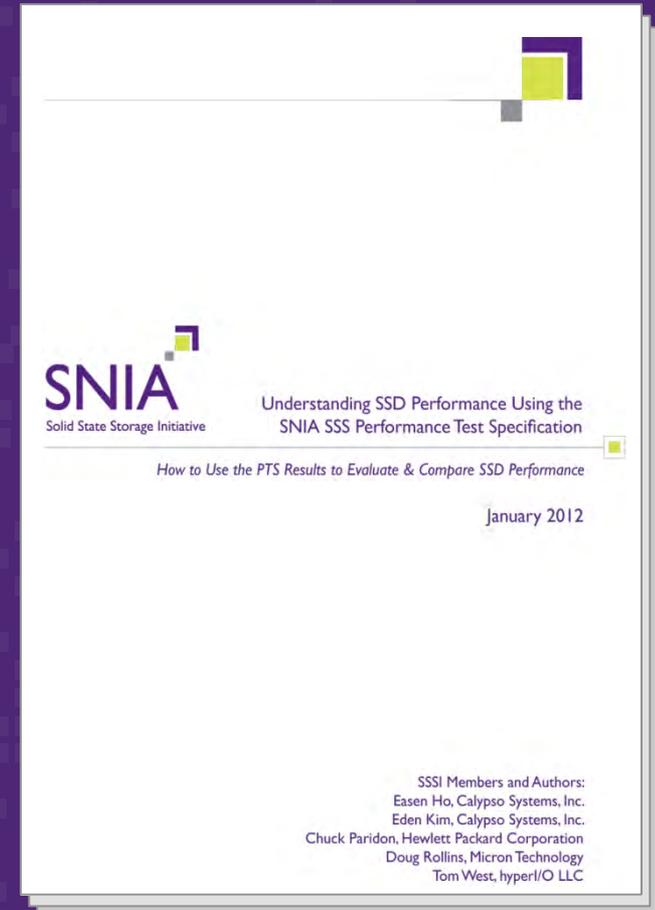
And Its Impact on Performance and Testing

Time: 4:10 - 4:30

How IOs Traverse the I/O Software/Hardware Stack

And Its Impact on Performance and Testing

SSSI White Papers - Expanded from “*Understanding
SSD Performance Using the SNIA SSS PTS*”
www.snia.org/forums/sssi/knowledge/education

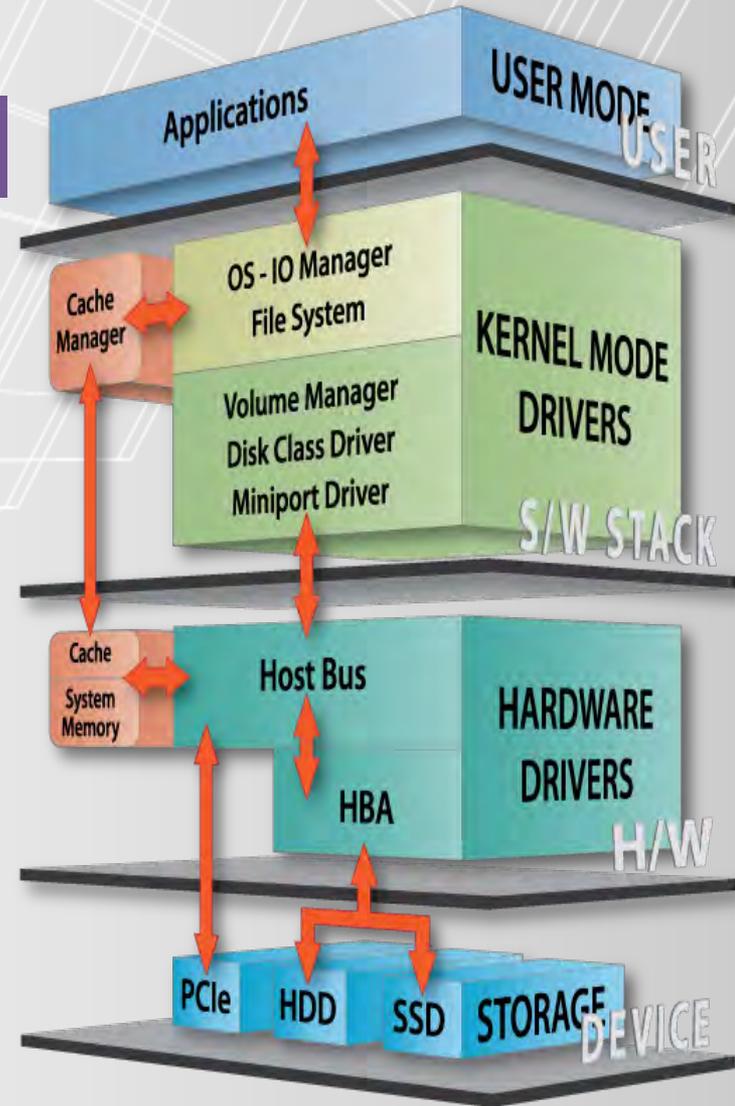


Download the deck at: www.snia.org/forums/ssi/knowledge/education

1. Brief Background Information about the PTS
2. The IO Software/Hardware Stack
3. Examples of How IOs Traverse the IO SW/HW Stack
4. Impact of Traversing the IO SW/HW Stack
5. Conclusion

What is the IO Software/Hardware Stack?

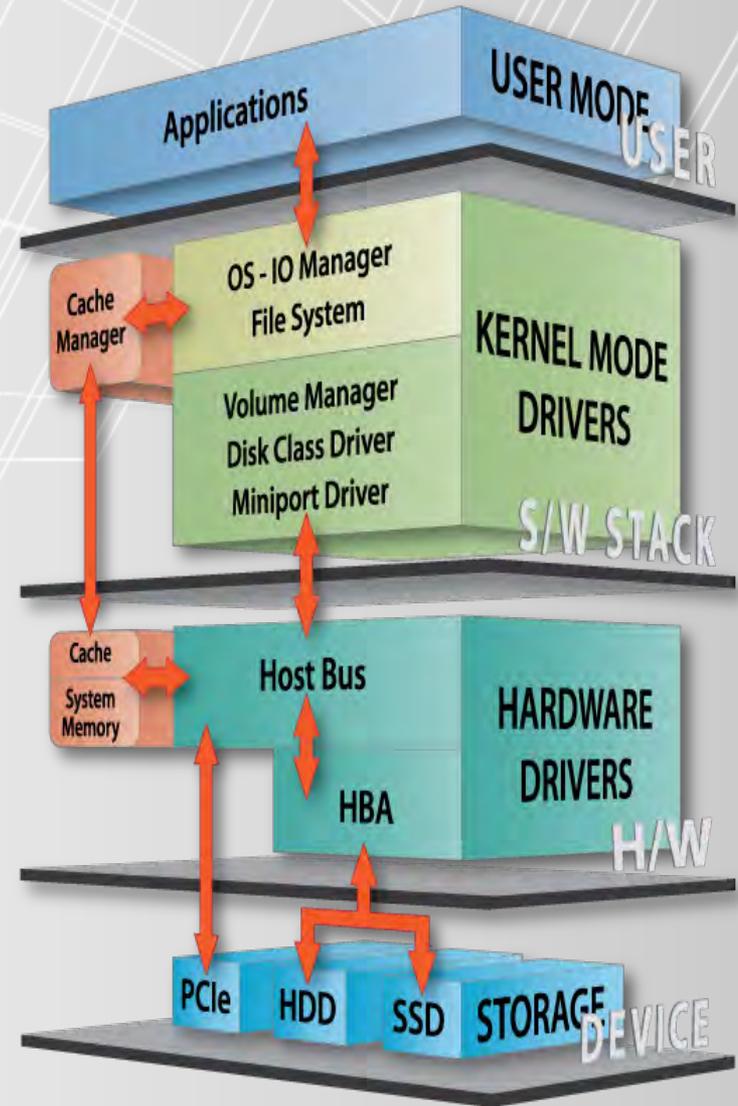
- Disk I/O operations are handled by a particular stack of I/O-related components (the IO SW/HW stack)
 - ✓ Operating-system (OS) software
 - ✓ Hardware components
- IO operations issued by applications traverse through the various components within the IO SW/HW stack
- A variety of factors and variables can impact:
 - ✓ The use of these components
 - ✓ The interactions amongst these components



IO SW/HW Stack: OS Software and Hardware Components

Why Is Understanding the IO Stack Important?

- How IOs traverse the IO SW/HW stack has a direct impact upon performance (also data integrity, cost, etc.)
- Better understanding of the IO stack can be beneficial when:
 - ✓ Resolving performance problems
 - ✓ Managing storage Quality-of-Service (QoS)
 - ✓ Configuring overall system options and usage
 - ✓ Evaluating storage device options, etc.
 - ✓ **Considering IO performance testing approaches and results**

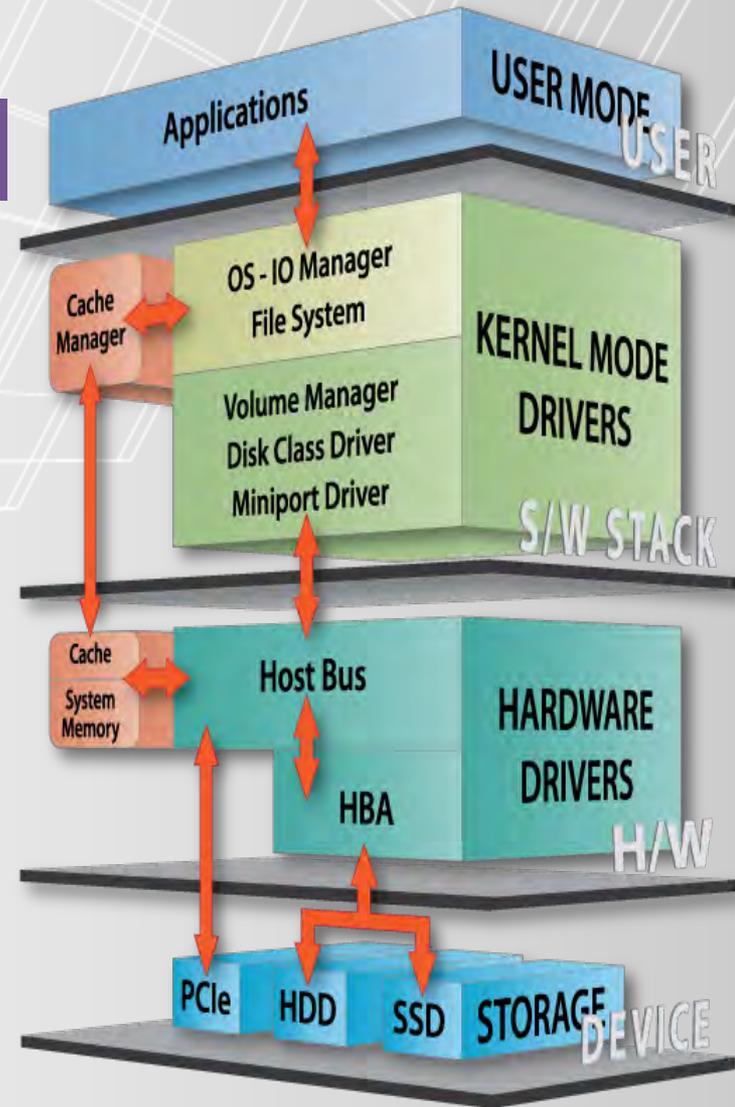


IO Stack – Directly Impacts IO Performance and Testing

IOs are Impacted by Traversing the IO Stack

The various paths that an IO operation might take through the IO SW/HW stack can have an impact upon (as examples):

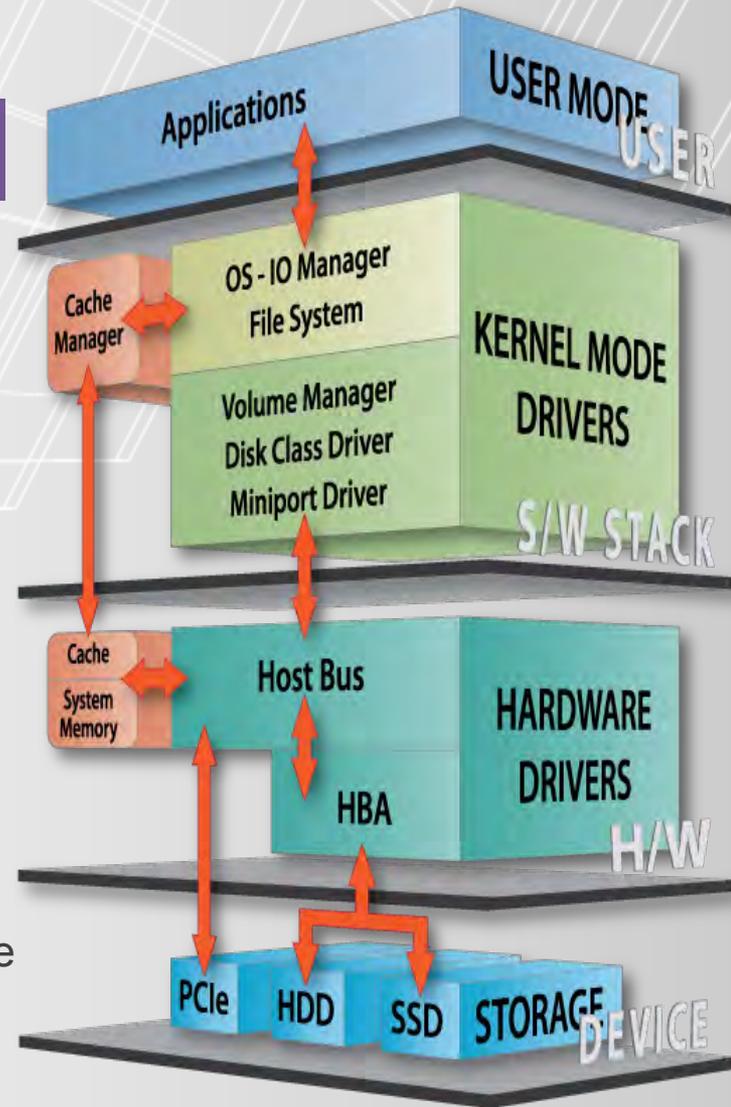
- ✓ IO operation response times
- ✓ Storage device usage and utilization
- ✓ CPU and overall system utilization



Traversing the IO Stack can have a **significant** impact upon various performance factors

Several Key Impact Considerations

- Application IO activity does not necessarily equate to storage device IO activity
 - ✓ IO operation performance metrics can be different
- Consider various viewpoints of IO operation performance:
 - ✓ An individual IO SW/HW stack component perspective
 - ✓ An overall system perspective
 - ✓ The particular application/workload perspective



Perspective: Observed Performance depends upon where you are standing in the stack

PCIe I/O Performance Considerations

Tom West, hyperI/O LLC



Advancing storage & information technology

- **I/O Software Path Length/Time**

- Within the I/O Software/Hardware Stack (SST)
- A larger proportional component of overall latency
- Where microseconds now matter more



- **“Antithetical” Approaches?**

- “Communications of the ACM” (January 2012, Volume 55, Number 1)
 - ✓ “Revamping Storage Performance”, pp 20 – 22 (access NV storage from app)
 - ✓ “I/O Virtualization”, pp 66 – 72

- **The “IOPS Arms Race”**

- More IOPS is “better” (and inherently so – the Iometer benchmark wars)
- How many IOPS can your applications issue – and **actually do** issue?
 - ✓ Get empirical metrics that reflect actual application/system usage

PCIe I/O Performance Considerations

Tom West, HyperIO



Advancing storage & information technology

- **Discussion / Questions & Answers**

Time: 4:10 - 4:30

PCIe SSD Standard Building Blocks

Paul Wassenberg, SSSI

PCIe Standards

How They Fit Together

PCIe SSD Standard Building Blocks

Paul Wassenberg, SSSI



Advancing storage & information technology

Building Block	Building Block Owners	Client PCIe SSD	Enterprise PCIe SSD
Physical Interface	PCI-SIG (PCI Express) SATA-IO (SATA Express) SCSI Trade Assn (SCSI Express)	Up to 2 lanes	Up to 4 lanes
Driver Interface	Intel (AHCI) NVM Express Group (NVMe) T10 (SOP/PQI w/support for NVMe)	AHCI or NVMe or SOP	SOP/PQI or SOP/NVMe or NVMe
Host Connector	SATA-IO (SATA Express) SSD FF WG (multifunction connector) Small FF Committee (SFF-8639)	SATA Express or SFF-8639	SFF-8639
Device Connector	SATA-IO (SATA Express) SSD FF WG (multifunction connector) Small FF Committee (SFF-8639)	SATA Express	SFF-8639

- SATA-IO defined the # of lanes for SATA Express. Driver interface was not defined.
- SCSI Trade Association defined # of lanes and the SOP/PQI driver interface for SCSI Express.
- SSD Form Factor Working Group defined a “universal” connector that was standardized by the SFF committee as the SFF-8639.

PCIe SSD Standard Building Blocks

Paul Wassenberg, SSSI



- **Discussion / Questions & Answers**

PCIe, SCSI Express, and a world beyond disk drives

Gary Orenstein, Fusion-io



Advancing storage & information technology

PCIe, SCSI Express, and a world beyond disk drives;
What happens when you get new native flash capabilities



4 June 2012

PCIE, SCSI EXPRESS, AND A WORLD BEYOND DISK DRIVES

What happens when you get native flash capabilities

A photograph of a server room with multiple racks of servers. The servers are dark-colored with perforated front panels. Many of the server units have glowing blue lights, indicating they are powered on. The racks are arranged in rows, and the perspective is from a slightly low angle, looking down the length of the server aisle.

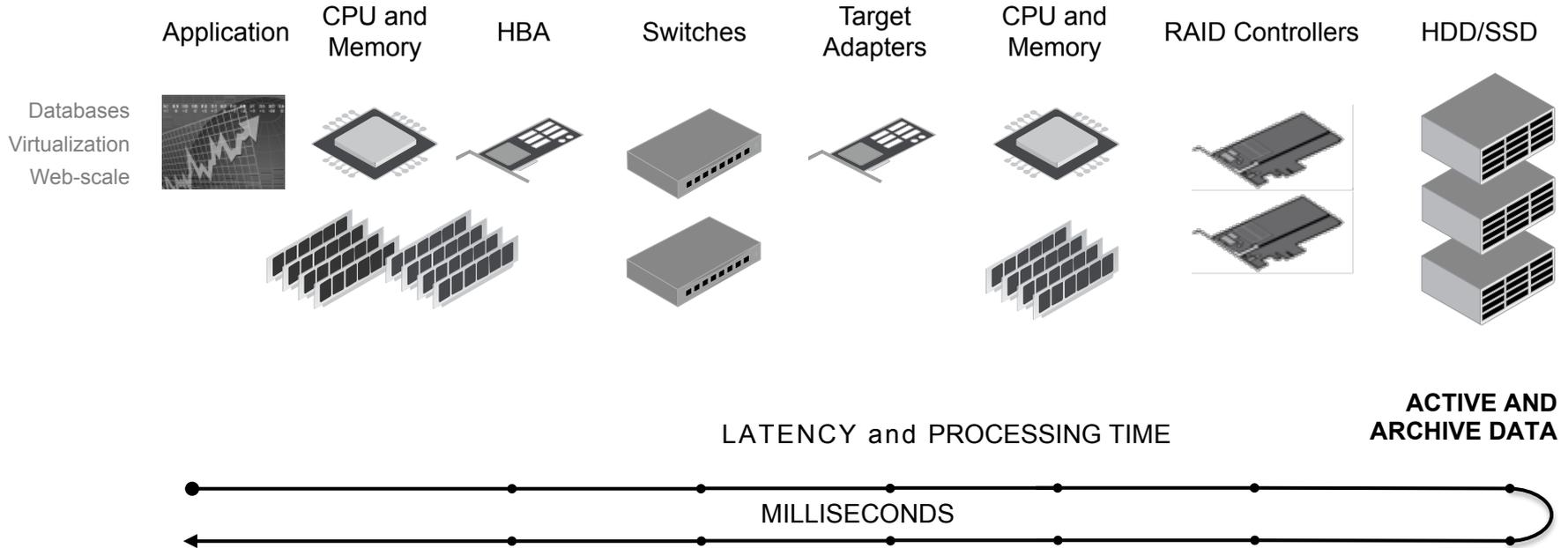
Lessons from the front lines

TRADITIONAL ARCHITECTURE

SERVERS

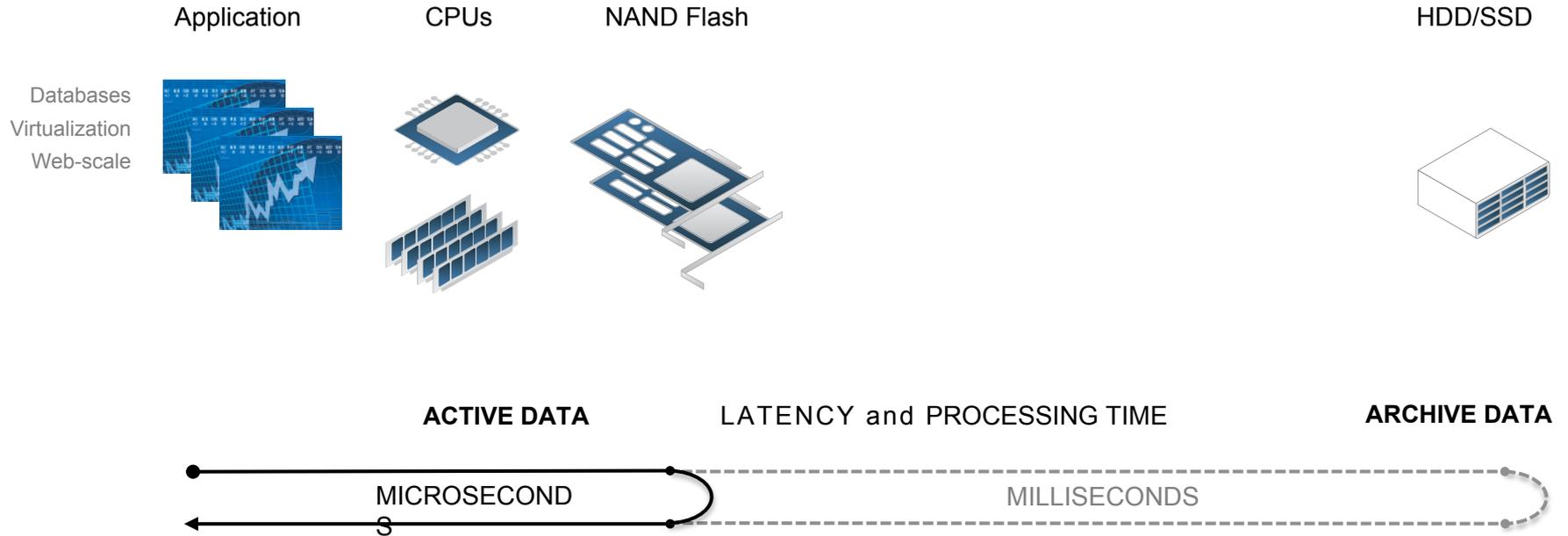
NETWORK

STORAGE (performance optimized)



PROCESS CRITICAL DATA CLOSE TO THE CPU

SERVERS



APPLICATION ACCELERATION EXAMPLES

AVERAGE DATABASE THROUGHPUT

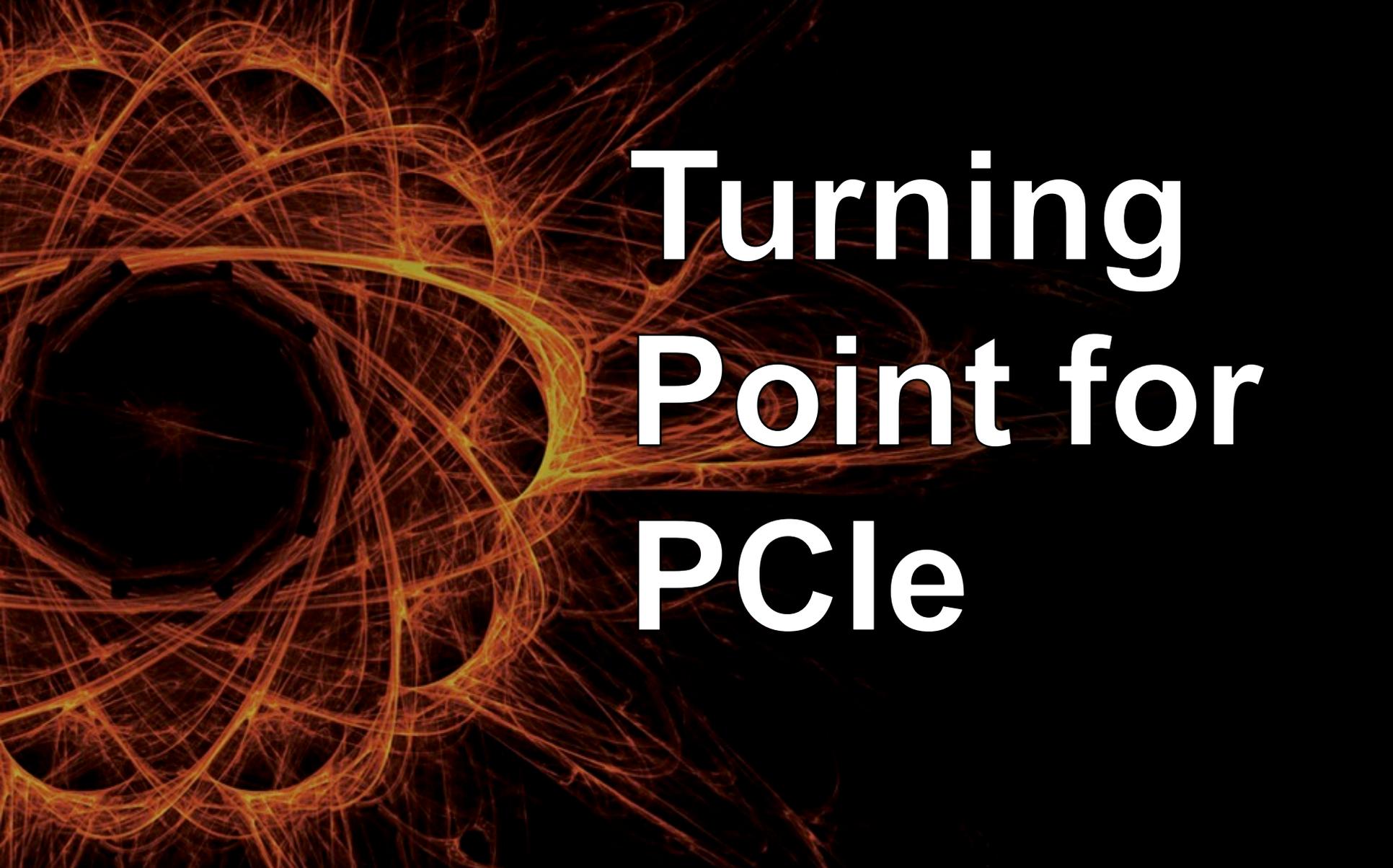


PERFORMANCE PER RACK UNIT



QUERY TIME



The background of the slide features a complex, abstract pattern of thin, glowing orange and yellow lines that resemble fibers or a dense network of connections. These lines are set against a solid black background, creating a high-contrast, energetic visual. The lines are most concentrated on the left side of the image, where they form a dense, circular-like structure, and then radiate outwards towards the right.

Turning Point for PCIe

THE WAY WE WERE

Consumer

- USB
- SATA
- IDE

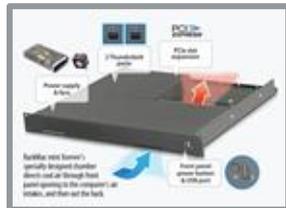
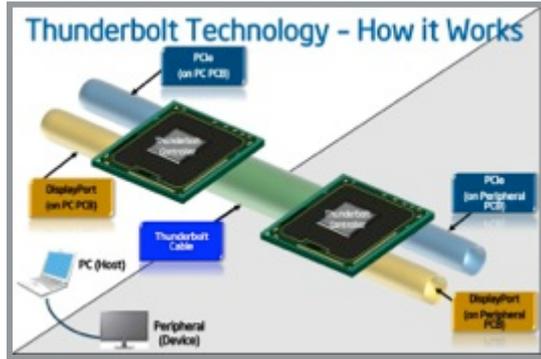
SMB/SME

- SATA
- SAS
- FC

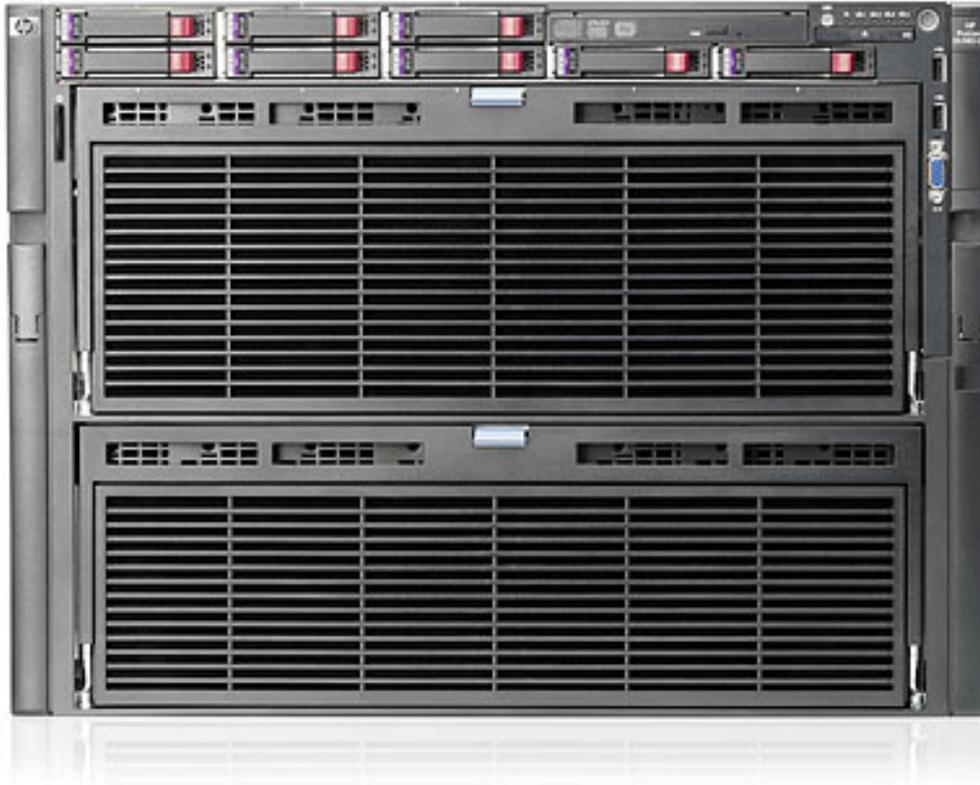
Enterprise

- SAS
- FC
- IB

THUNDERBOLT PCIE DEVICES



HP DL 980



Up to 11 full
height/full length
slots supported

2410 GB x 11

26.5TB per
server



x11

WHERE WE ARE HEADED

Consumer

SMB/SME

Enterprise

PCIe

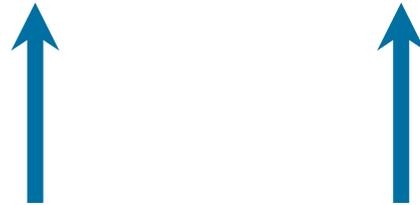
Architecture



INITIAL STAGE OF NAND FLASH DEPLOYMENT

Applications

Block I/O



NAND FLASH

ARCHITECTURAL DEPLOYMENT DIFFERENCES

Applications

Block I/O

Architected like
a disk

Architected like
memory

BLOCK AND MEMORY ACCESS

Applications

Block I/O

SSD
SAS
SATA

Block
I/O

Beyond
Block I/O

Memory

Native Flash

SCSI Express



SCSI EXPRESS

A set of industry initiatives delivering a PCIe Express based enterprise storage solution

Industry Initiative	Focus
SCSI Over PCIe (SOP)	Streamline SCSI command set optimized for solid state
PCIe Queuing Interface (PQI)	Flexible and extensible transport layer
Universal drive connector	Supporting current and emerging devices
PCIe physical layer	Drive error handling and asynchronous hot add/remove
Native OS support	Standard driver-enablement to support range of devices

SCSI EXPRESS BENEFITS

- ▶ Embrace PCIe
- ▶ Fill gap between DRAM and HDD
- ▶ Embrace SCSI
- ▶ Work together on standards
- ▶ Ensure a quality ecosystem

**Don't
forget
software**



CONVENTIONAL ARCHITECTURE

Applications and File Systems

Storage Stack

Physical Device Operations

STUCK ON DRIVES AND BLOCK I/O

Applications and File Systems

Block I/O

read()

write()

Physical Device Operations

Entirety of software and physical stacks optimized for rotating disks

FLASH ONLY AS A FAST DISK

Applications and File Systems

Storage Stack

Physical Device Operations

Flash

Flash

Flash

Flash

Flash

Disk-
centric
approach

Legacy
stacks
remain

BUT FLASH IS DIFFERENT

- ▶ Asymmetric read/write latencies
- ▶ Write-impact on durability
- ▶ Unique erase characteristics

FLASH AS A NEW ARCHITECTURE

Applications and File Systems

Flash Translation Layer

Flash

Flash

Flash

Flash

Flash

Flash-centric approach

Retain backwards compatibility with conventional block I/O

FLASH TRANSLATION LAYER 101

Input

Logical Block Address (LBA)

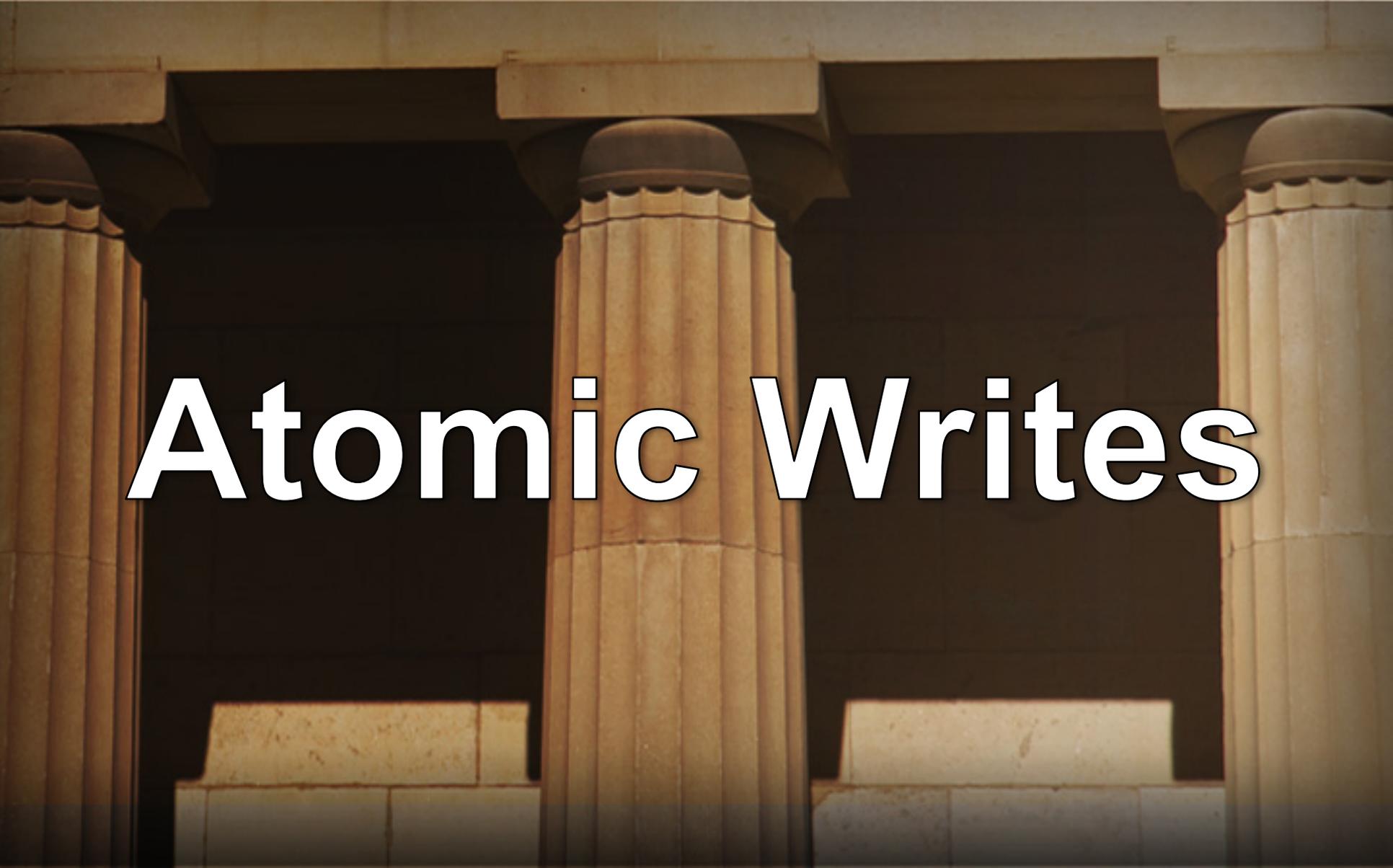
Flash Translation Layer

Output

Commands to Physical NAND flash

OPTIMIZING SOFTWARE STACKS FOR FLASH

- ▶ Virtualize the storage layer
- ▶ Retain compatibility with conventional block I/O
- ▶ Deliver new flash-native capabilities

A photograph of three classical columns supporting a structure, with the text "Atomic Writes" overlaid in the center. The columns are fluted and have papyrus capitals. The background is dark, and the lighting is warm, highlighting the texture of the stone.

Atomic Writes

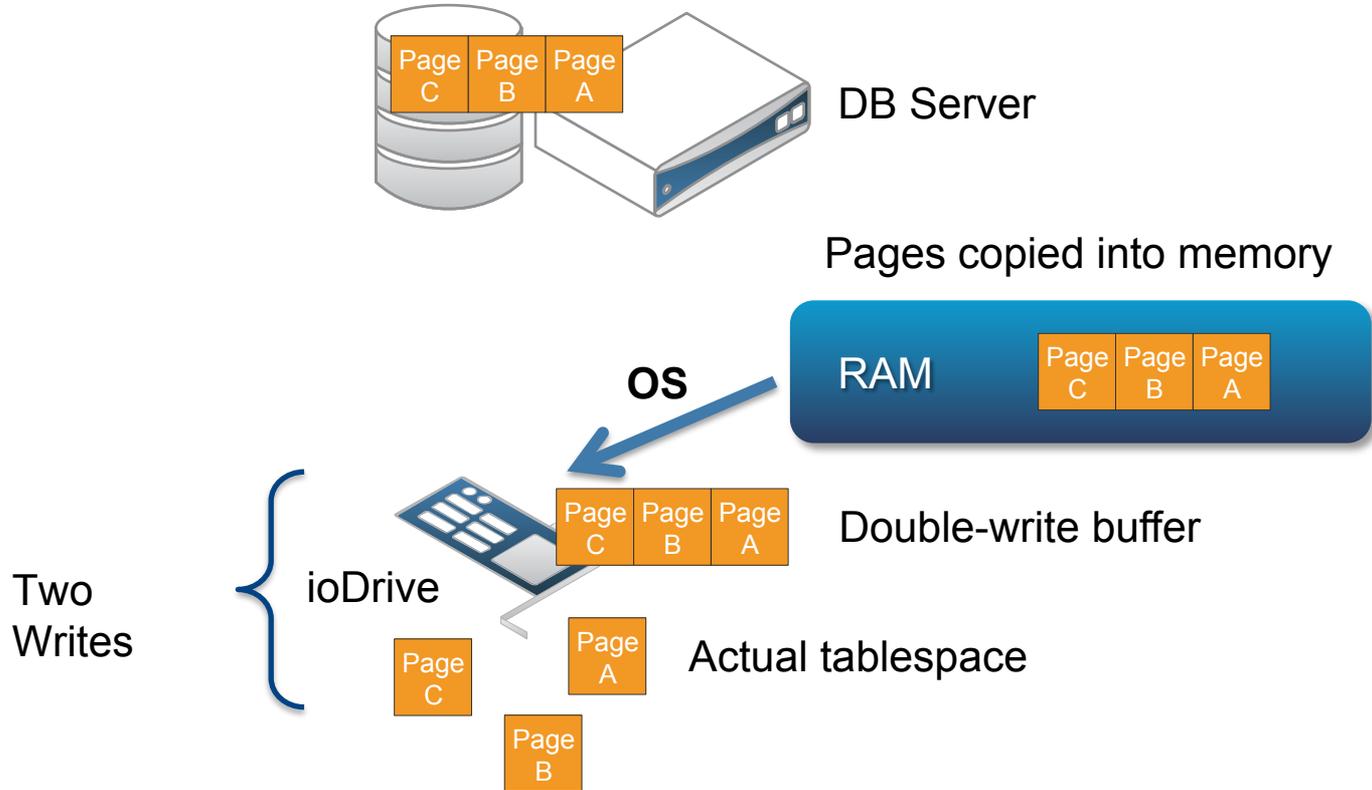
IT IS ABOUT TRANSACTIONS

- ▶ Building block of applications and databases
- ▶ Transactional Semantics
 - Data Integrity
 - Concurrency
 - Crash Recovery
- ▶ Applications
- ▶ File Systems
- ▶ Databases
- ▶ Web Services
- ▶ Search Engines
- ▶ Mission Critical Computing

ATOMIC WRITES

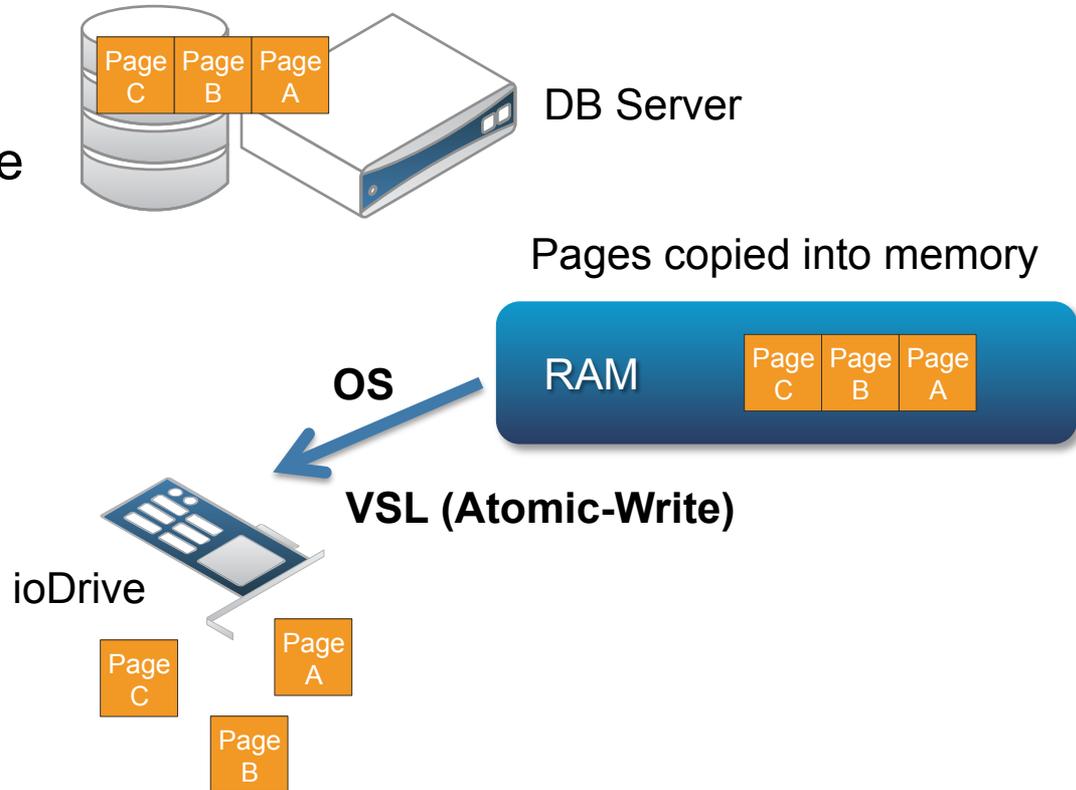
- ▶ Batch multiple I/O operations to multiple independent ioMemory virtual addresses
- ▶ Entire batch written with transactional (ACID) semantics
 - Persisted as a whole or rolled back upon failure
- ▶ Application benefits
 - Simplify
 - Less writes

INNODB (MYSQL) DOUBLE-WRITE



INNODB (MYSQL) ATOMIC-WRITE

- ▶ Less Writes
- ▶ Better Performance
- ▶ ACID Compliance



OPPORTUNITIES FOR ADVANCEMENT

MySQL extension for Atomic-Write

35%

PERFORMANCE

2^x

ENDURANCE

1/2

WRITE WORKLOAD





THANK YOU

PCIe, SCSI Express, and a world beyond disk drives

Gary Orenstein, Fusion-io



Advancing storage & information technology

- **Discussion / Questions & Answers**

Open Discussion

Next Actions

Open Discussion:

Next Actions:

- Email response to reflector: ***Would you attend a private reception at FMS if invited?***
- Survey: Feedback on Meetings; Questions & Comments for Mtgs. No. 7-8
- Agenda & Topics
 - Meeting No. 6 - Big Picture; Deployment Strategies & Marketing Issues
 - a. Walter Hubis, Fusion-io - PCIe as Persistent Memory
 - b. Jim Handy, Objective Analysis - Market Issues - “How Many IOPS Survey”
www.thesdguy.com or <http://www.SurveyMonkey.com/s/KGKVR6X>
 - c. Andy Rudolf, Intel - Developing an Open Kernel NVM Programming Model
 - Meeting No. 7 - Where do we go from here?
 1. ***OPEN - contact chair if you are interested to present***

Supplemental Slides