### SNIA Solid State Storage Initiative PCIe SSD Task Force Meeting No. 3

Monday 07 MAY 2012





Meeting No. 3 Monday 07MAY2012 4:00 PM - 5:30 PM PST

WELCOME!

#### 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 <u>www.snia.org/forums/sssi/pcie</u>

### **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 pciessd@snia.org
- Send Questions to pciechair@snia.org

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



Topics	09APR12 Mtg No. I	23APRI2 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		×							
Test Platforms		×							
Performance			×						
System Integration			х						
System Arch Form Factors				×					

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	17AUG12 Flash Memory Summit
Review Discussion Q & A	×					
Deployment Strategies Market Development		×				
Topic Review Roadmap Goals			x			
Roadmap & Milestones 2012				x		
SSSI Committee Mtg					×	
SSSI Task Force F2F FMS Round Table						×

#### **Tentative Topics for Meetings**

## AGENDA – 07 MAY 12

I. a. b.	Administrative Roll Call; Call Schedule Announcements; Other	4:00 – 4:05 4:05 - 4:10
II.	Business	
Ι.	Coughlin / Handy Survey Input - How Many IOPS are Enough?	4:10 - 4:20
2.	PCIe Test Methodology	
а.	SNIA SSS Performance Test Specification & PCIe - Easen Ho, Calypso	4:20 - 4:35
b.	Performance Saturation - Chuck Paridon, HP	4:35 - 4:50
3.	PCIe System Integration Issues	
а.	PCIe Hardware / Firmware - Hany Eskander, STEC	4:50 - 5:05
b.	PCIe Device v System IOs- Tony Roug, Virident	5:05 - 5:20
III.	Wrap Up	
а.	Discussion	5:20 – 5:30
b.	Close	

## Attendance

Company	09APR12	23APR12	07MAY12	21 <b>MAY</b> 12	04JUN12	18 <b>JUN</b> 12	02JUL12	16JUL12
Agilent	x	x	x					
Allion	x	x						
AMD	x	x						
Apacer								
Cadence	x							
Calypso	x	x	x					
Cisco		x	x					
CLabs								
Corsair			x					
Coughlin Associates	x	x	x					
Dell	x							
eAsic	x							
EMC	x	x	x					
Enmotus	x							
eTron		x						
Fusion-io	x	x						
Greenliant		x	x					
HDS	x		x					
НР	x	x	x					
HGST	x	x	x				Time: 4:0	0 - 4:05

## Attendance

Company	09APR12	23APR12	07MAY12	21 <b>MAY</b> 12	04JUN12	18JUN12	02JUL12	I6JUL12
Huawei	x	x	x					
Hynix (SK Hynix)		x	x					
HyperIO	x	x	x					
IBM	x		x					
Intel	x	x	x					
Lecroy		x	x					
Lenovo	x		x					
Lotes								
LSI	x	x	x					
Lunastar		x						
Marvell	x	x	x					
Micron	x	x	x					
Molex	x	x	x					
Mushkin								
Objective Analysis	x							
ocz	x							
Oracle	x	x	x					
PLX Technology			x					
Phison	x	x	x					
Renesas	x	x	x					
Samsung	x	x	x				Time: 4:0	0 - 4:05

## Attendance

Company	09APR12	23APR12	07MAY12	21 <b>MAY</b> 12	04JUN12	18 <b>JUN</b> 12	02JUL12	16JUL12
Seagate	x							
Smart Storage		x	x					
SNIA	x	x	x					
STEC	x	x	x					
Taejin	x	x	x					
Tektronix								
тмѕ	x							
Toshiba		x	x					
Tyco Electronics	x	x						
Unigen	x	x	x					
Viking								
Virident	x	x	x					
WDC	x	x	x					
Bitsprings		x						
	39	37						

### **Announcements - Eden Kim**

- 1. Announcements / Other
  - a. Minutes Meeting No. 2
  - b. Task Force Charter / Structure Standing Slides
  - c. Announcements:
    - **1. Meeting Schedule**
    - 2. Topics
    - 3. Reminder: Task Force General Survey
    - 4. SSSI PCIe Round Tables: Flash Memory Summit & Storage Developers Conference

**STA Tech Showcase Wed PM** 

FMS - email to lance@flashmemorysummit.com

#### Minutes Meeting No. 2 23APRI2



- 1. Attendance:
  - a. (37) Companies present of (58)
  - b. Presentations / Speakers Marvel, Toshiba, Calypso, Virident, STEC, LeCROY
- 2. Administrative:
  - a. Task Force Charter General Survey of PCIe SSD issues; Recommendations for SSSI Committee 2d half 2012
  - b. Task Force Structure -
    - 1. (8) Open Mtgs: 09AP12, 23APR12, 07MAY12, 28MAY12, 04JUN12, 18JUN12, 02JUL12, 16JUL12
    - 2. Meeting Topics:
      - **General** up to (2) major discussions/meeting
      - Presentations Participants may present topics / slides during meetings contact pciechair@snia.org
      - Meeting No. 1 Organizational; Survey Review; Overview of PCIe
      - Meeting No. 2 Standards: A Closer Look; PCIe Test Hardware Refresh
      - Meeting No. 3 PCIe Performance Test Issues; PCIe System Integration Issues
      - Meeting No. 4 System Architecture; Form Factors
      - Meeting No. 5 General Discussion / Q&A Submitted to reflector
      - Meetings no. 6 8 tbd
    - 3. Links: Email Reflector: pciessd@snia.org PCIe Task Force Homepage: www.snia.org/forums/sssi/pts
- 3. PCIe Performance Overview& SNIA SSS PTS Eden Kim, Calypso
- 4. Overview of SSSI RTP Spec Tony Roug, Virident
- 5. PCIe Interposer Boards John Weidermeir, LeCroy
- 6. Next Actions
  - a. Feedback on Meeting Topics post to reflector
  - b. Agenda / Topics Meeting No. 2

Close 5:30PM

#### Announcements 07MAY12



#### 1. Meeting Schedule

- 1. Meetings No. 4 8: 21May; 04June; 18June; 02July; 16July
- 2. 30July Meeting SSSI Committee
- 3. 17AUG Meeting FMS: F2F & SSSI Round Table

#### 2. Topics

1. See Previous slides

#### 3. Reminder - General Survey Period

1. Reflector Emails - many individual posts to chair - don't be bashful

#### 4. FMS & SDC Round Tables & Panels

- 1. FMS (Aug 2012) Closed
- 2. SDC (Sep 2012) Deadline 15MAY12
- 3. Open to PCIe Task Force Members that become SSSI Members
- 4. Contact <u>pciechair@snia.org</u>

#### Charter Standing slide



#### 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

#### Structure Standing Slide



#### 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

### Links Standing slide

#### SSSI

- SSSI homepage
- Understanding SSD Performance Project
- SSS Performance Test Specification (PTS)
- PTS Standard Report Format
- SSSI Bright Talk Webcasts
- SSSI White Papers

#### **PCIe Task Force**

- PCIe SSD Task Force
- PCIe SSD Task Force reflector
- PCle SSD Task Force questions

- www.snia.org/forums/sssi
- www.snia.org/forums/sssi/pts
- www.snia.org/pts
- www.snia.org/forums/sssi/pts
- www.snia.org/forums/sssi/knowledge/education
- www.snia.org/forums/sssi/knowledge/education

www.snia.org/forums/sssi/pcie

pciessd@snia.org

pciechair@snia.org

### PCIe Discussion Topic Tom Coughlin, Coughlin Assoc

- 1. Survey on Performance Needed for Various Applications Mid May Mid June 10 minutes *Tom@tomcoughlin.com*
- Summary Bullets:
- Coughlin Associates and Objective Analysis are going to do a survey on the performance (IOPS, data rate and latency) needed for various common applications that use digital storage
- The survey is meant to determine which applications will get the greatest advantage using flash memory storage and roughly how much flash memory capacity these applications would need
- We plan to start this survey in mid-May and end in mid-June
- Besides contributing to a report on this topic we intend to make a presentation on this at the SDC this Summer (if paper accepted), write articles for SNIA pubs and do a webinar
- We would like you feedback on questions to include in the survey as well as help in finding end user participants to participate

#### Survey on How much perf needed Suggested Questions - how many IOPS needed



- 1. Roughly how many IOPS do you believe is the most your storage systems can take advantage of before some other bottlenecks in your normal processes gets in the way:
  - 10
  - 100
  - 1,000
  - 10,000
  - 100,000
  - 1,000,000
- 2. What is the source of the number in (1):
  - Measured
  - Estimated
- 3. What is the latency required for your dominate applications?
  - 1 second
  - .1 second
  - .01 second
  - .001 second
  - .0001 second
  - .00001 second
- 4. What is your typical application that you use storage systems for? (& storage capacity needed as a separate Q)
  - On-line transaction processing
  - Mail server and storage
  - Databases
  - Video creation or distribution
  - Scientific or Military
  - Education
  - Other\_

We plan to let folks know how the survey turns out if they give us an email address at the end of the survey No. users / app; R/W Mix and Block Size; IOPS Burstiness; How sophisticated is the target audience? What is the computing environment (server based / client)? Server limits? Total bytes written over Time period (SLC/MLC & endurance cycles)?

Tom wants suggested questions for the survey

## PCIe Test Methodology Easen Ho, Calypso

- 1. PTS Test Methodology
- 2. PCIe Test Issues
- 3. Recommended PCIe Performance Tests
- 4. Other

## **Two specifications released**

SNIA Solid State Storage Performance Test Specification (PTS)					
TS-E	PTS Enterprise ver 1.0	PTS-C	PTS Client ver 1.0		
	SNIA		SNIA		
Advancing storage & Information technology		Advancing storage & Information technology			
Solid State Storage (SSS) Performance Test Specification (PTS) Enterprise		Solid State Storage (SSS) Performance Test Specification (PTS) Client			
	Version 1.0	Version 1.0			
The document has been remained and approved by the SNIA. The SNIA between that the data, manuallegare and technologue described in this document accurately represent the SNA goals and are appropriate for whiteward definitution. Suggestion for revealer should be directed to http://www.ania.org/faceback/		This occurrent has been revealed and approved by the SNA. The SNA beneves alters, methodologues and increasing its described in the document accurrent repr SNA goals and are accelerable for wedespread distribution. Suggestion for reveal directed to http://www.ama.org/faceback/			
			SNIA Technical Position		
			August 6, 2011		
	SNIA Technical Position				
	April 26, 2011				

### **Basic Test Flow**



## **Tests Contained In PTS-E I.0 SPEC**

- Enterprise Performance Test Specification (PTS-E) VI.0 encompasses:
  - A suite of basic SSS performance tests
  - Preconditioning and Steady State requirements
  - Standard test procedures and reporting requirements



## Tests Contained In PTS-C 1.0 SPEC

Advancing storage & information technology

### Client IOPS

- Random Access
- R/W:
- 100/0, 95/5, 65/35, 50/50, 35/65, 5/95, 0/100
- BS:
  - 1024KiB, 128KiB, 64KiB, 32KiB, 16KiB, 8KiB, 4KiB, 0.5KiB

#### • Range Restriction:

- 100% & 75% LBA
- 2048 Segments

#### • Active Footprint Restriction:

• 8 & 16 GiB

### Client TP

- Sequential Access
- R/W:
- 100/0, 0/100
- BS:
- 1024KiB
- Range Restriction:
  - 100% & 75% LBA
  - 2048 Segments
- Active Footprint Restriction:
  - 8 & 16 GiB

### **Client Latency**

- Random Access
- R/W:
- 100/0, 65/35, 0/100
- BS:
  - 8KiB, 4KiB, 0.5KiB
- Range Restriction:
  - 100% & 75% LBA
- 2048 Segments
- Active Footprint Restriction:
  - 8 & 16 GiB

# PTS Follow-On Work (PTS-C/ E I.I...)

Host Idle Test	<ul> <li>See how the drive responds to host idle time amidst continuous access</li> </ul>
Cross Stimulus Recovery Test	<ul> <li>See how drive handles switching between sustained access patterns</li> </ul>
Demand Intensity / Response Time Histogram Test	<ul> <li>See how drive responds to increasing host demands along with detailed response time statistics at key operating points</li> </ul>
Power Efficiency Test	• Measures power efficiency of the device, i.e., IOPS/W, etc.
Enterprise Composite Synthetic Workload	<ul> <li>Synthetic composite workload for Enterprise environments similar to JEDEC workload for endurance testing</li> </ul>

## Example PCIe Test Results RND/4K Write Saturation





## Example PCIe Test Results PTS-E I.0 IOPS



# Example PCIe Test Results PTS-E I.0 IOPS

Advancing storage & information technology

■ 0/100 ■ 5/95 ■ 65/35 ■ 50/50 ■ 35/65 ■ 95/5 ■ 100/0



## Example PCIe Test Results PTS-E I.0 TP



# Example PCIe Test Results PTS-E I.0 Latency



## Applying PTS: PCIe Test Issues

#### RTP HW Update Needed:

- PCle Gen 3, new processors
- additional interface
- solution for power control and measurement
- solution for detailed timing information

### RTP SW

• Needs to be able keep up with DUT

#### PTS Tests Need Additional Reporting Related To:

- Host utilization statistics (CPU/memory usage/?)
- other CPU configuration information (# CPU/# cores used/HT/?)
- other hardware settings (power states/?)

## Example: Effect of CPU Configuration SNIA Advancing storage & information technology

**CPU\_USER** Statistics



# **Example: Effect of CPU Configuration**

Advancing storage & information technology



Time: 4:20 - 4:35

# **Example: Effect of CPU Configuration**



# Additional PTS Tests Specific to PCIe?

- Additional reporting needed for PCIe tests
- All PTS test should be applicable to PCIe as well
- Possible PCIe-specific tests:
  - Tests under different power profile
  - Tests that shows performance dependency on host parameters
  - Tests looking at detailed stack-up of latencies

# PCIe Test Methodology Easen Ho, Calypso

Advancing storage & information technology

• Discussion / Q & A

## PCIe Performance Saturation Chuck Paridon, HP

Advancing storage & information technology

**See Next Slides** 



### Solid State Device Saturation Curve Test

Hewlett Packard

#### PCI SSD Storage Device tested at HP-Roseville, Feb. '10





Advancing storage & information technology

Large Block Saturation Curve



Time: 4:35 - 4:50

Advancing storage & information technology





OIO: 1, 2, 4, 8, 16, 32

#### Configuration Details

- Server Details
  - > Dual Quad-core 2.13 GHz Xeon Processors
  - > 8GB of ram
  - > Windows Server 2003 R2 SP2
- Storage Device Details
  - > PCI Plugged 36 GB SSD.
  - > Server Ram Serves a Cache
- Workload Details
  - > Small Block 100% Random Read 4 KB transfer size
  - > Large Block 100% Sequential (Isolated LBA range) Read 1 MB transfer size
  - > 1/2 hr Warmup, 5 minute Measurement Period

#### Conclusions

- Server Resources and IO Generation Method Have a Profound Affect on High Speed SSD Performance under a small block read workload
- These affects disappear under a large block workload due to longer device latency
- "No free lunch" applies again as high IOP rates -> High CPU Utilizatin

## PCIe System Integration Issues Hany Eskander, STEC

Advancing storage & information technology

1. PCIe Hardware / Firmware Issues



- PCle Gen3 system integration challenge
  - PCIe Gen3 is the next revolution of general purpose PCI Express IO standard
  - PCIe 3.0 (Gen 3) is 2x the bandwidth of Gen2
    - This bit rate represents the most optimum tradeoff between manufacturability, cost, power and compatibility
  - Uses a combination of Protocol and Encoding Changes
    - Use 128b/130b encoding on individual lanes
    - Use PHY layer packetization to identify packet boundaries
      - Remove the "K" codes
    - Scrambling only (no 8b/10b) to provide edge density
  - backwards compatible with previous PCIe generation

#### PCIe Gen3, system integration HW/FW Hany Eskandar, STEC Advancing storage & information technology

#### PCle Gen3 Bandwidth

• The bandwidth differences between previous Gen and Gen3

PCI Architecture	Raw Bit Rate	Interconnect Bandwidth	Bandwidth Per Lane Per Direction	Total Bandwidth for x16 link
PCIe GenI.x	2.5GT/s	2Gb/s	~250MB/s	~8GB/s
PCIe Gen2.x	5.0GT/s	4Gb/s	~500MB/s	~I6GB/s
PCIe Gen3.0	8GT/s	8Gb/s	~IGB/s	~32GB/s

- Derivation of these numbers in Gen2 and Gen 3
  - Gen2

```
5GT/s * 1 lane = 5 Gb/s
```

- =(5 Gb/S) x (1Byte/10bits) = 500 MB/s per diff pair
- Gen3

8GT/s \* 1 lane = 8 Gb/s

= (8Gb/S) / 8bits= 1GB/s per diff pair

PCle 3.0 is 2x the bandwidth of Gen2, @8GT/s, the data rate only provides 60% boost in bandwidth



PCI Express Hardware topology





#### **PCI Express Hardware topology**



Desktop chipset PCIe support



Server chipset PCIe support

- Advancing storage & information technology
- PCIe 3.0 area of changes (Physical Layer-New Electrical Requirement)





• PCIe 3.0 area of changes (Physical Layer-32-bit Support)



- Advancing storage & information technology
- PCIe 3.0 area of changes (Digital Controller will have to handle 2x data)



## PCIe System Integration Issues Hany Eskander, STEC

Advancing storage & information technology

• Discussion / Q & A

## PCIe System Integration Issues Tony Roug, Virident

- Tony Roug
- Virident
- Solution Architecture

## PCIe system level integration' Tony Roug, Virident

Advancing storage & information technology

#### System Level Issues are different based on architecture of PCIe Device

	SAS/SATA Aggregator	Direct Block, non-standard	Standard		
Architecture	SATA/SAS: HBA/RAID + SATA/SAS SSDs	Block Driver + PCIe board	Standard driver Standard PCIe registers Standard Form Factor		
Examples	LSI WarpDrive, Intel 910	FusionIO, Virident, Marvell	NVM Express SCSI Express		
Form Factor	PCIe board	PCIe board	PCIe board + SATA Express		
Driver	existing PnP	custom	Custom evolving to plug-n-play		
Boot	Mostly in BIOS	Custom BIOS	Goal		
Management	SMART/Custom	Custom	TBD		
Power	25 Watts/Board	25 Watts/board	25 Watts/board or device		
APIs	n/a	Kernel and User direct emerging	TBD		
Other	Non standard capacities, kernel tuning, scaling architectures, caching architectures				

## Performance Testing Tony Roug, Virident



- Single Device versus System Performance
  - Device vendors and computer OEMs typically compare at device level
  - End-users (IT, Web, HPC, etc) interested in solution performance

#### Device testing versus System Testing

	Device Testing	System Testing	
Device	IOPS, Throughput, Variance	Linear performance scale	
Driver	Function	Operates with multiple	
Kernel	Tuning for PCIe	Including scale (e.g. interrupts)	
Memory	Memory per board	Efficient memory scale	
Processor	CPU overhead per board	Efficient CPU scale	
Power	Board within 25 Watts	System can cool multiple boards	
Scale	One board	Units in the system	
Endurance	TB written, Compression, Dedup		
Application	Direct modes	RAID, Caching	

#### SNIA PCIe task force focus priority Tony Roug, Virident Advancing storage & information technology

### 2012 concept proposal

- Evolution of SNIA performance spec for existing PCIe devices and emerging standard architectures
- Define system level canonical architecture based on emerging standard architectures

# PCIe System Integration Issues Tony Roug, Virident

Advancing storage & information technology

• Discussion / Q & A

## Open Discussion Next Actions

Advancing storage & information technology

**Open Discussion:** 

**Next Actions:** 

- Feedback on Meetings
- Agenda & Topics for Meeting No. 4



## **Supplemental Slides**