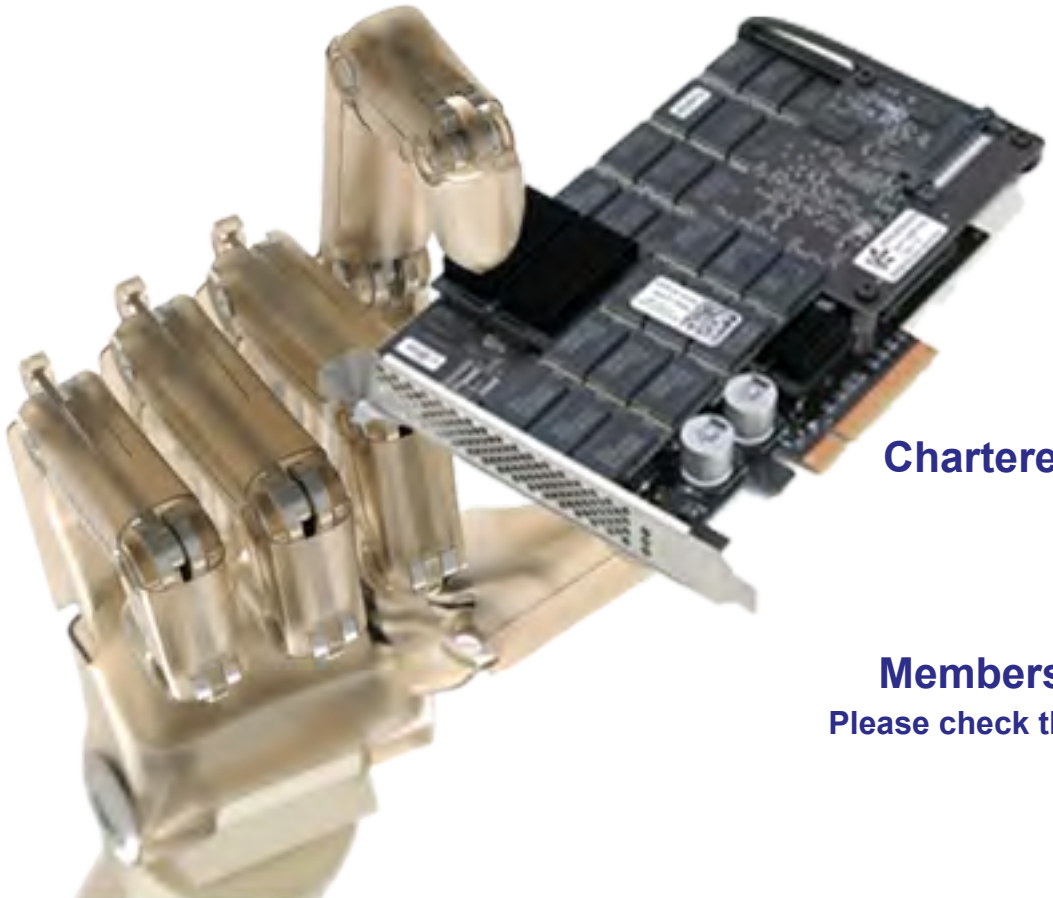
SNIA Solid State Storage Initiative
PCIe SSD Task Force Meeting No. 2

Monday 23 APR 2012

WELCOME!

SNIA
Advancing storage & information technology

Meeting No. 2
Monday 23APR2012
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

SNIA
Advancing storage & information technology

Concall Guidelines:

- Use your mute button
- Sign in webex w/ company name
- 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

SNIA
Advancing storage & information technology

Topics	09APR12 OPEN	23APR12 OPEN	07MAY12 OPEN	21MAY12 OPEN	04JUN12 OPEN	18JUN12 OPEN	02JUL12 OPEN	16JUL12 OPEN	30JUL12 SSSI Committee
Kick-Off Mtg Issue Identification	X								
Standards		X							
Test Platforms		X							
Performance			X						
System Integration			X						
Other tbd									

Goals: Issue Identification & Committee 2012 Roadmap

Time: 4:00 – 4:05

AGENDA – 09 APR 12

SNIA
Advancing storage & information technology

I.	**Administrative**	
a.	Roll Call; Call Schedule	4:00 – 4:05
b.	Announcements; Other	4:05 - 4:10
II.	**Business**	
1.	**Standards - A Closer Look:**	
a.	SATA Express; NVMe/AHCI - Paul Wassenberg, Marvell	4:10 – 4:25
b.	SFF 8639 Connector - Mike Fitzpatrick, Toshiba	4:25 - 4:40
c.	Standards Questions	4:40 - 4:45
2.	**PCIe Test Hardware RTP Refresh**	
a.	SNIA SSS Performance Test Specification - Eden Kim, Calypso	4:45 - 4:55
b.	PCIe Hardware RTP - Motherboards, Components - Tony Roug, Virident	4:55 - 5:05
c.	PCIe Interposer Boards - Min-jei Chong / Rob Vizena, Agilent	5:05 - 5:15
d.	PCIe Interposer Boards - John Weidermeier, LeCroy	5:15 - 5:25
III.	**Wrap Up**	
a.	Discussion	5:25 – 5:30
b.	Close	

Time: 4:00 – 4:05

Attendance

SNIA
Advancing storage & information technology

Company	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Agilent	x	x						
Allion	x	x						
AMD	x	x						
Apacer								
Cadence	x							
Calypso	x	x						
Cisco		x						
CLabs								
Corsair								
Coughlin Associates	x	x						
Dell	x							
eAsic	x							
EMC	x	x						
Enmotus	x							
eTron		x						
Fusion-io	x	x						
Greenliance		x						
HDS	x							
HP	x	x						
HGST	x	x					Time: 4:00	– 4:05

Attendance

SNIA
Advancing storage & information technology

Company	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Huawei	x	x						
Hynix (SK Hynix)		x						
HyperIO	x	x						
IBM	x							
Intel	x	x						
Lecroy		x						
Lenovo	x							
Lotes								
LSI	x	x						
Lunastar		x						
Marvell	x	x						
Micron	x	x						
Molex	x	x						
Mushkin								
Objective Analysis	x							
OCZ	x							
Oracle	x	x						
Phison	x	x						
Renesas	x	x						
Samsung	x	x					Time: 4:00 – 4:05	

Attendance

SNIA
Advancing storage & information technology

Company	09APR12	23APR12	07MAY12	21MAY12	04JUN12	18JUN12	02JUL12	16JUL12
Seagate	x							
Smart Storage		x						
SNIA	x	x						
STEC	x	x						
Taejin	x	x						
Tektronix								
TMS	x							
Toshiba		x						
Tyco Electronics	x	x						
Unigen	x	x						
Viking								
Virident	x	x						
WDC	x	x						
Bitsprings		x						

Time: 4:00 – 4:05

Announcements - Eden Kim

SNIA
Advancing storage & information technology

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

Time: 4:05 – 4:10

Minutes Meeting No. 1
09APR12

SNIA
Advancing storage & information technology

1. **Attendance:**
   a. **(38) Companies present of (54)**
   b. **Presentations / Speakers - Marvel, Seagate, Calypso, STEC, Virident**

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
         Meetings No. 4 - 8 - tbd
      3. **Links:** Email Reflector: pciessd@snia.org PCIe Task Force Homepage: www.snia.org/forums/sssi/pcie

3. **SSSI PCIe Task Force Initial Survey - see slide deck for statistics**

4. **Overview of PCIe -**
   a. **Overview of PCIe Standards - Paul Wassenberg - Marvell - see slide deck**
   b. **PCIe 101 - Marty Czekalski - Seagate - see slide deck**
   c. **Introduction to Meeting Topics - Eden Kim, Calypso; Hany Eskander, STEC; Tony Roug, Virident**

5. **Next Actions -**
   a. **Feedback on Meeting Topics - post to reflector**
   b. **Agenda / Topics - Meeting No. 2**

Close 5:30PM

Time: 4:00 – 4:10

Announcements
23APR12

SNIA
Advancing storage & information technology

1. Meeting Schedule
    1. (8) Meetings Open to Public (Membership Required in Aug 2012)
    2. Every Other Monday
    3. Meetings No. 3 - 8: 07May; 21May; 04June; 18June; 02July; 16July

2. Topics
    1. Meeting No. 2 - Standards: A closer look; PCIe Hardware RTP Refresh
    2. Meeting No. 3 - PCIe Performance Issues; System Integration Issues
    3. Meetings No. 4 - 8 - TBD

3. Reminder - General Survey Period
    1. Charter is Discovery / Discussion of PCIe Issues
    2. Deliverables include Recommendations to SSSI and SSS TWG
    3. This is a GENERAL SURVEY PERIOD - actual work will be in Committee Aug - Dec 2012

4. FMS & SDC Round Tables & Panels
    1. Opportunity to Participate at FMS (Aug 2012) and SDC (Sep 2012)
    2. FMS - (5) Person Panel & (5) Individual speaker spots: deadline 01MAY12
    3. Open to PCIe Task Force Members that become SSSI Members
    4. Contact pciechair@snia.org

Time: 4:05 – 4:10

Charter
Standing slide

SNIA
Advancing storage & information technology

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

Time: 4:05 – 4:10

Structure
Standing Slide

SNIA
Advancing storage & information technology

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 16JUL12
    2. Post Meeting Survey's for feedback and agenda preparation
    3. Email reflector becomes SSSI member only starting 16JUL12

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

Time: 4:05 – 4:10

Links
Standing slide

SNIA
Advancing storage & information technology

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

Time: 4:05 – 4:10

PCIe Standards: A Closer Look
Paul Wassenberg, Marvell

1. SATA Express and NVMHe / AHCI

Questions:

1. ***Thunderbolt = 1 lane PCIe & 1 lane HDMI***
2. ***SATA 8b/10b encoding***
3. ***PCIe Gen 3 - 128 b / 130 b encoding***
4. ***Is SATA part of the PCIe payload? - No, SATA Express has no additional overhead, its pure PCIe, no SATA protocol. SATA is not being packetized.***
5. ***SATA lExpress is native PCIe. No SATA interfaces***

Time: 4:10 - 4:25

SATA Express & NVMe / AHCI
Paul Wassenberg, Marvell

SNIA
Advancing storage & information technology

- SATA Express standardizes PCIe as a client storage interface in an HDD-type form factor

- SATA-IO is defining host and device connectors for SATA Express
  - Slightly modified standard SATA connectors
  - Mechanically compatible with SATA connector, enabling SATA and SATA Express to co-exist
  - The new host connector supports two SATA ports or up to two PCIe lanes
  - Host connector pins for the SATA ports are multiplexed with PCIe pins
  - Separate drive-driven signal that tells the host if the device is SATA or SATA Express (PCIe) so the host knows what "language" to speak
  - Motherboard can have a single connector that supports up to two current SATA drives or a SATA Express drive

- Improved performance is the main motivation for SATA Express
- PCIe can be quickly scaled up in performance by adding lanes
- The new encoding for Gen3 further improves the actual data throughput
  - SATA = 6Gb/s = 0.6GB/s
  - PCIe Gen3 (one lane) = 8Gb/s = 1.0GB/s
  - PCIe Gen3 (two lanes) = 16Gb/s = 2.0GB/s
  - PCIe Gen 4 (3-4 years) = 16Gb/s per lane
- Finally, SATA Express is pure PCIe
  - No SATA link or transport layer, so no translation overhead – users will see the full performance of PCIe

Time: 4:10 – 4:25

SATA Express & NVMe / AHCI
Paul Wassenberg, Marvell

SNIA
Advancing storage & information technology

- The operating system driver interface is beyond the scope of the SATA Express specification, but developers must decide between two options – AHCI & NVMe

- AHCI (www.intel.com/content/www/us/en/io/serial-ata/ahci.html) is the driver interface used for SATA, and is built into most operating systems
  - A SATA Express device with AHCI would be compatible with SATA software environments
  - However, AHCI was designed around HDDs 10 years ago, and is not optimized for SSD performance

- NVM Express (www.nvmexpress.org) was designed specifically for PCIe SSDs, and will provide optimal performance
  - NVMe has several optimizations targeted at enhancing performance with SSDs, mainly by reducing latencies
    - Parallelism and multiple threads
    - Out of order data transfers
    - Increases queue depth from 32 to 64K queues of 64K each
  - NVM Express is not widely used yet, but there are drivers for Windows, Linux, and VMware

Time: 4:10 – 4:25

SATA Express & NVMe / AHCI
Paul Wassenberg, Marvell

SNIA
Advancing storage & information technology

- **Why SATA Express instead of 12Gb/s SATA?**
  1. **With the next speed increase, the SATA infrastructure would have to change in any case**
     - **New cables, connectors, backplanes, more complex protocol**
  2. **Some client SSDs will require more than 6Gb/s within 2012**
     - **Despite the work on 12Gb/s SAS, 12Gb/s SATA would not be available in 2012**
     - **PCIe Gen3 IP is already available**
  3. **The portion of client SSDs that will require greater than 6Gb/s is fairly small**
     - **12Gb/s SATA would be low volume & high cost for some time**
     - **PCIe Gen3 is already cost effective**
  4. **6Gb/s SATA will be more than adequate for HDDs for the foreseeable future**
     - **HDDs comprise the vast majority of client storage device shipments**
  5. **SATA & SATA Express must be low cost to serve the client storage market**
     - **In limited volumes, 12Gb/s SATA would be too expensive**

Time: 4:10 – 4:25

PCIe Standards: A Closer Look
Mike Fitzpatrick, Toshiba

SNIA
Advancing storage & information technology

1. SFF 8639 Connector

SFF spec to include form factor for multi port 8639 (2 SAS/SATA or 4 PCIe devices)

Rev expected within "weeks"

SFF 8639 slot defined as 25W slot (cf 2.5" SAS connector = 9W)

Main point: 8639 will be utilized either for SAS and/or PCIe

Questions:

?25W per slot?? - it's a capability, not a requirement. You can go up to 25W per 8639 slot

? How deal w/ temperature assoc w/ 25W high density SSDs ?

Time: 4:25 - 4:40

SFF 8639 Connector
Mike Fitzpatrick, Toshiba

- Item
- Item

Time: 4:25 – 4:40

PCIe Standards: A Closer Look
Questions & Discussion

SNIA
Advancing storage & information technology

Questions?

Time: 4:40 - 4:45

Overview: PTS & RTP
Eden Kim, Calypso

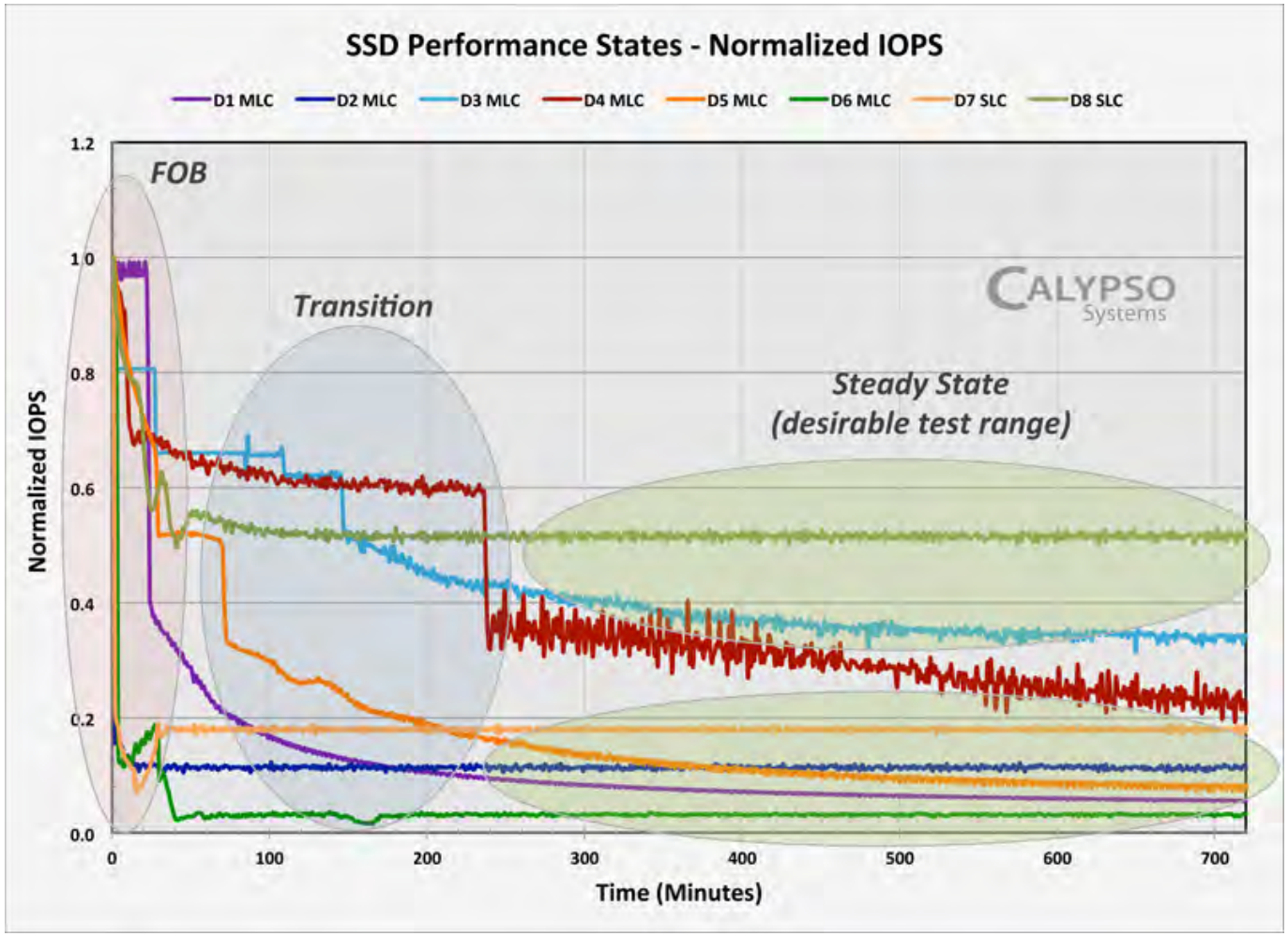
SNIA
Advancing storage & information technology

1. SNIA SSS Performance Test Specification
2. NAND Flash Performance Test Issues
3. PTS Test Methodology
4. PTS Reference Test Platform (RTP)
5. Issues for PCIe RTP Refresh

Time: 4:45 - 4:55

NAND Flash Based Performance States

SNIA
Advancing storage & information technology

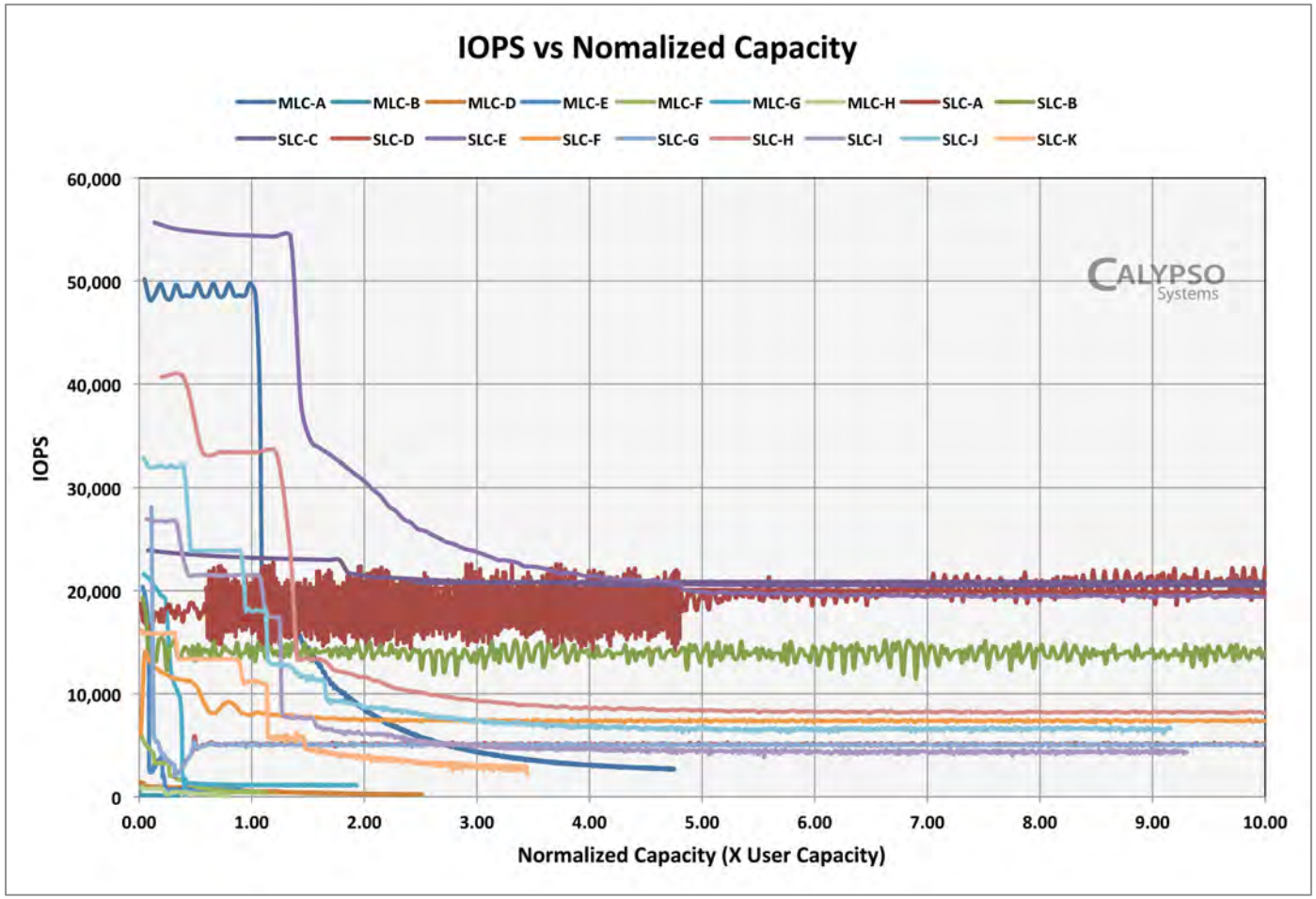


Performance States

1. **FOB**
2. **Transition**
3. **Steady State**

Time: 4:45 – 4:55

SSD Performance Dependencies

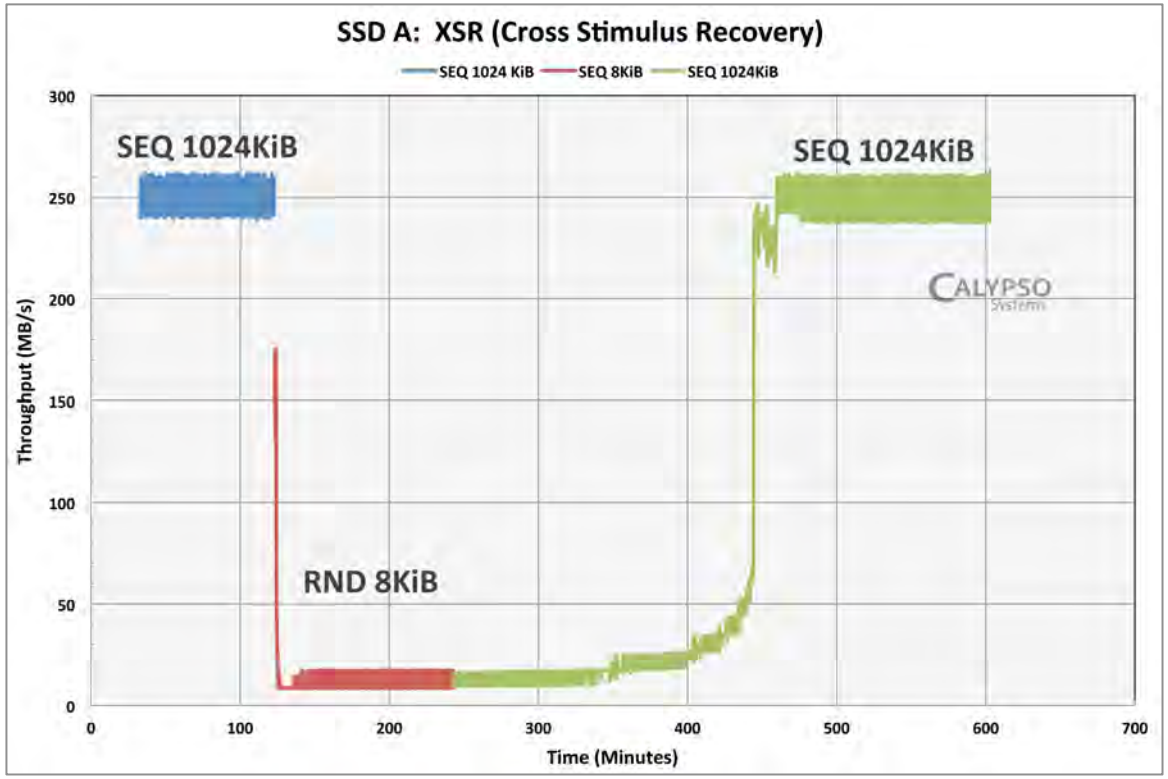


Performance HIGHLY Affected by 3 Factors:

- **Write History**
- **Measured Workload**
- **Test Environment**

Time: 4:45 – 4:55

Write History & Performance States



Write History

Previously written data may have more impact on performance than the measured IO command

Measured Workload

The IO Access Pattern (Block Size / Read/ Write Mix) can profoundly affect SSD performance - e.g. Large Block SEQ v Small Block RND

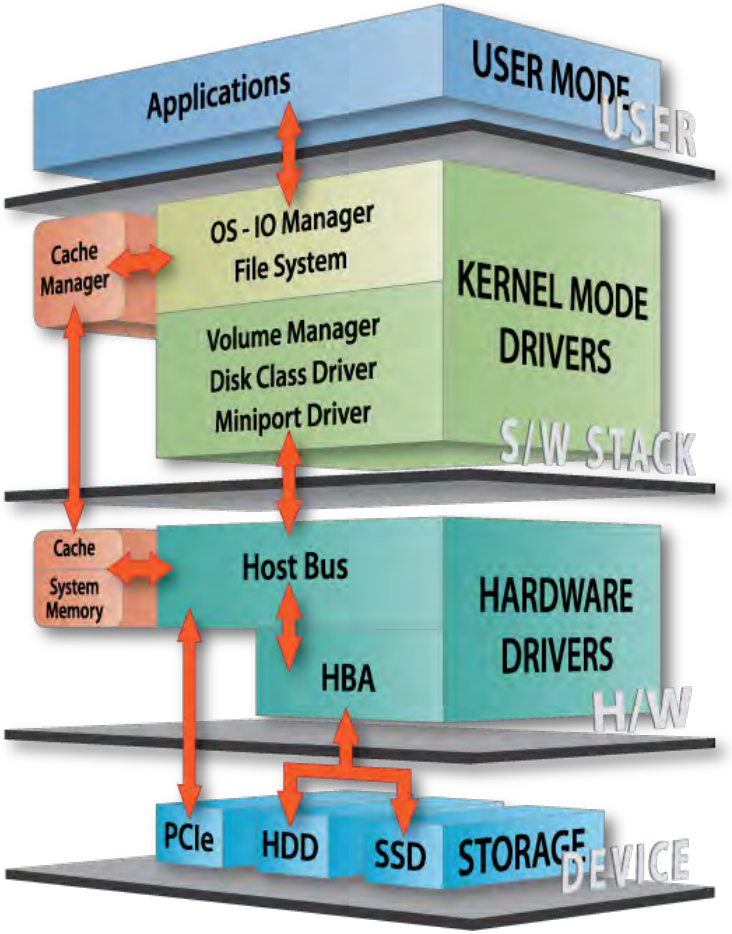
Hardware/Software Environment

H/W S/W should minimally affect measurements - is there sufficient bandwidth and host processing resources to generate the necessary IO loads? How much software overhead is there?

Time: 4:45 – 4:55

Hardware & Software IO Stack
Synthetic Device Level Test

SNIA
Advancing storage & information technology



File System Test

- Specific File IO operations issued in the File System
- IOs targeted at the Device traverse the SW/HW Stack
- IOs are subject to cache, OS task switching & timing, driver fragmentation & coalescing
- Original IO can be different at the Device level
- Can lose 1:1 correspondence original IO & Physical Device IO

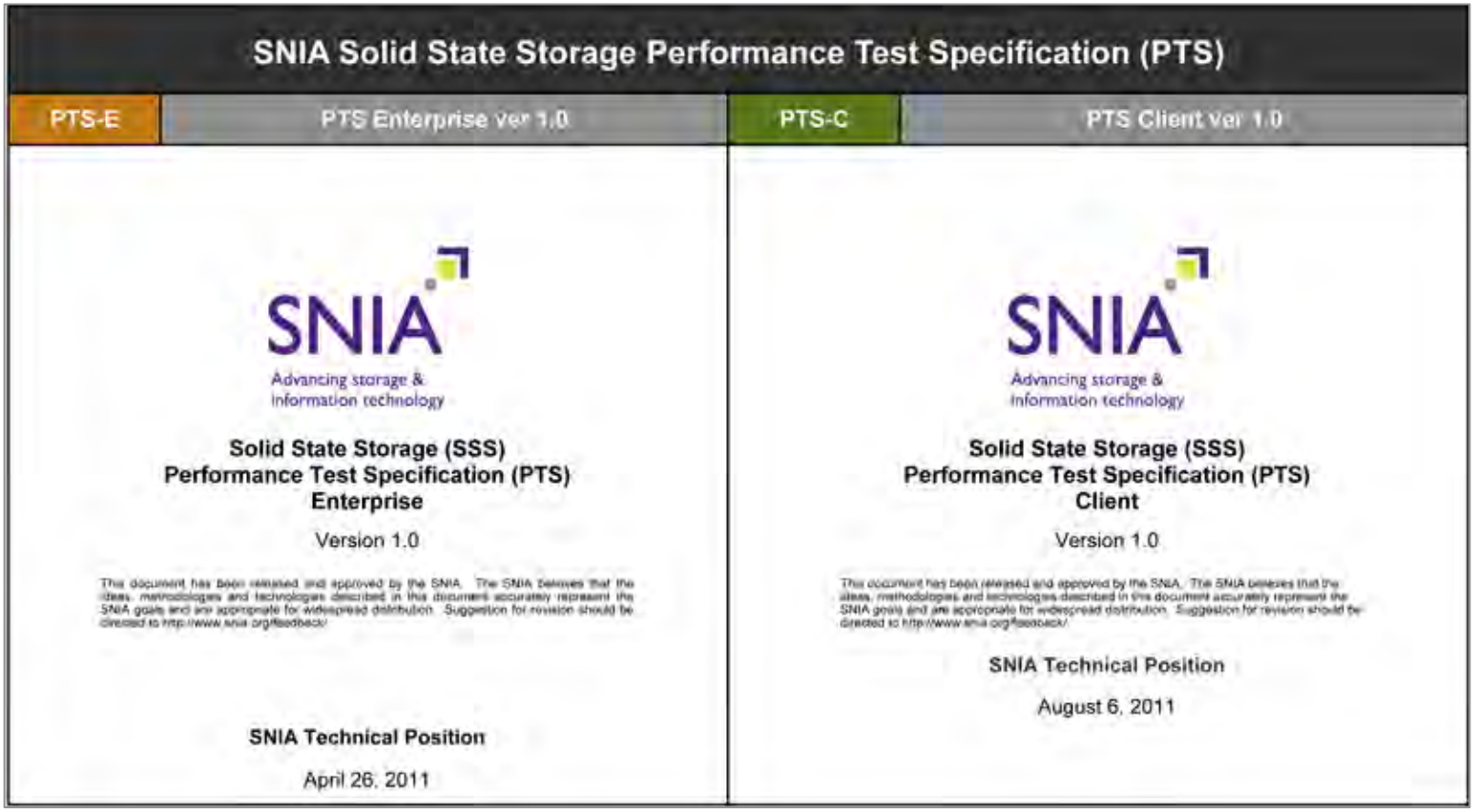
Synthetic Device Level Test

- Applies a known and repeatable test stimulus
- Targets Block IO Devices (not File System devices)
- Uses Specified Test Workloads (Access Patterns, Data Pattern)
- Specifies LBAs allowed to be used (ActiveRange, AR Amount)
- Prescribes the Test Methodology

Time: 4:45 – 4:55

SNIA SSS Performance Test Spec (PTS)

SNIA
Advancing storage & information technology

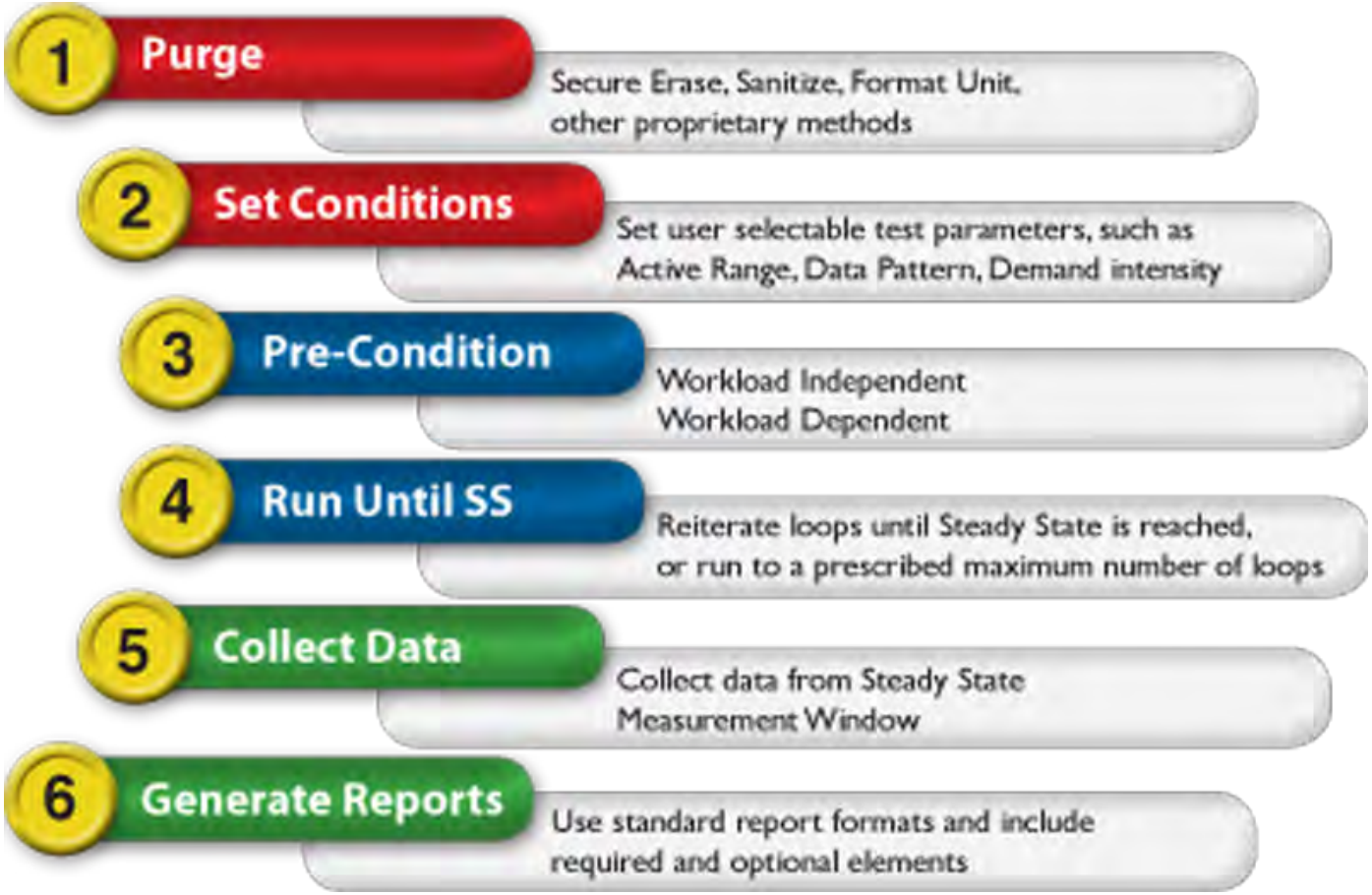


- **Based on Synthetic Device Level Test**
- **Standardized Pre Conditioning**
- **Specified Test Workloads**
- **Test Hardware Specific / Test Software Agnostic**

Time: 4:45 – 4:55

PTS Test Methodology

SNIA
Advancing storage & information technology

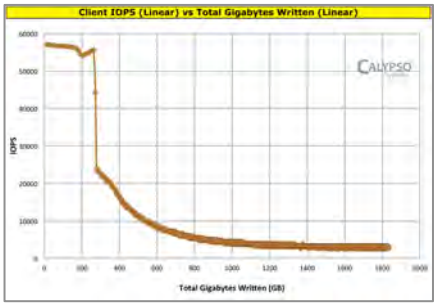


Time: 4:45 – 4:55

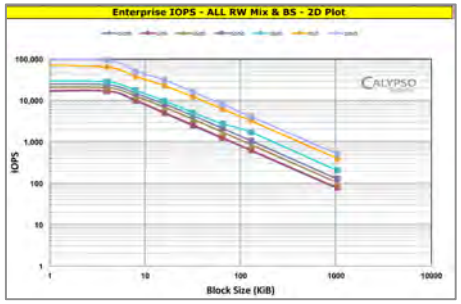
PTS - Performance Tests

SNIA
Advancing storage & information technology

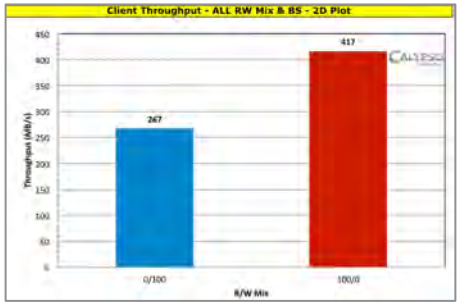
PTS rev 1.0 Performance Tests			
Test	**Test Description**	**Purpose**	**Metric**
WSAT	Continuous RND 4KiB W from FOB, No PC	FOB Performance Evolution over Time	IOPS
IOPS	Large & Small Block RND IOs at Steady State	Steady State IO Transfer Rate per second	IOPS
Throughput	Large Block SEQ R/W Data Transfer at Steady State	Steady State Bandwidth Speed	MB/Sec
Latency	AVE & MAX Response Times measured at a single OIO	Steady State IO Response Time Latency	mSec



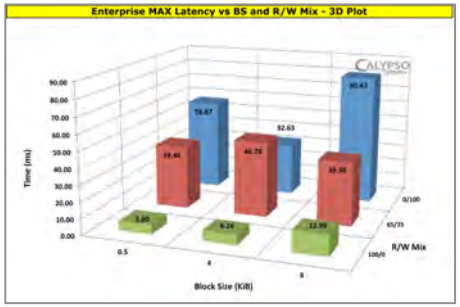
WSAT



IOPS



TP



LAT

Time: 4:45 – 4:55

PTS - Reference Test Platform SAS/SATA & PCIe Gen2

SNIA
Advancing storage & information technology

Reference Test Platform (RTP 2.0)			
Hardware		**Software**	
Processor	Single Intel Xeon 5580W 3.2 Ghz 4 core	**Operating System - Back End**	CentOS 5.6
Motherboard	Intel 5520 HC	**Test Software - Back End**	CTS 6.5
RAM	12 GB ECC DDR3	**Front End - GUI**	Chrome Browser
HBA	6 Gb/s LSI 9212-4e-4i	**Front End: OS, Database**	Windows 7 / MySQL

Time: 4:45 – 4:55

PCIe - Reference Test Platform Refresh

SNIA Advancing storage & information technology

PTS RTP -

- **Hardware defined for SAS/SATA and PCIe Gen 2 Motherboard**
- **Software Tools Requirements set forth**
- **SSD Performance Test Methodology Defined**

PCIe RTP Refresh -

- **Hardware Refresh for Gen 3 Motherboards**
- **Interposer Board for PCIe Power Tap?**
- **Other?**

Performance Test Methodology - PCIe Issues? See Meeting No. 3 Agenda

Time: 4:45 – 4:55

PCIe Hardware Reference Test Platform
Tony Roug, Virident

SNIA
Advancing storage & information technology

PCIe Hardware RTP Refresh Focus

- Gen 3 PCIe (i.e. 8 Ghz)
- Maximum PCIe connections
  - x8 G2 (e.g. 16 slots) half height, half width ATX FF
  - x4 G3 (e.g. 16 slots) half height, half width ATX FF
  - x1/x2 G3 (e.g. 32 slots) 2.5” FF
- Non-blocking PCIe expander (e.g. PLX)

Time: 4:55 - 5:05

PCIe Hardware Reference Test Platform
Tony Roug, Virident

SNIA
Advancing storage & information technology

	Version 1.0 RTP		PCIe RTP proposal	
Component	**Part**	**Units**	**Part**	**Units**
Chassis	Intel SC5660DP	1	TBD	1
Motherboard	Intel S5520HC	1	TBD	1
Processor	Intel 3.2Ghz W5580	1 or 2	Intel E5-2600	1 or 2
Main Memory	1333M DDR3, ECC	12-96GB	1600M DDR3, ECC	16-128GB
Boot Drive	160GB, 7.2K	1	Intel 320 SSD 40GB	1
HBA	LSI 9212-4i4e	1	n/a or PLX switch?	
DUT	SATA/SAS SSD	1	PCIe SSD	1-32
OS	Linux	CentOS 5.4	Linux	CentOS 5.6

Time: 4:55 - 5:05

PCIe Hardware Reference Test Platform

Rob Vizena / Min-jei Chong, Agilent

SNIA
Advancing storage & information technology

1. Item
2. Item

Time: 5:05 - 5:15

SNIA SSS Performance Test Spec RTP
Rob Vizena / Min-jei Chong, Agilent

SNIA
Advancing storage & information technology

- **PCIe Interposer Boards -** The need for them in a reference test platform & Selection Criteria

- 1. PCIe card performance test
  2. Need for interposer cards
  3. Interposer card functionality
  4. Analyzer / Exerciser capabilities

Time: 5:05 – 5:15

PCIe Hardware Reference Test Platform
John Weidermeier, LeCroy

LeCroy PCIe SSD Protocol Analysis and Test

Need for PCIe Interposer Cards

Time: 5:15 - 5:25

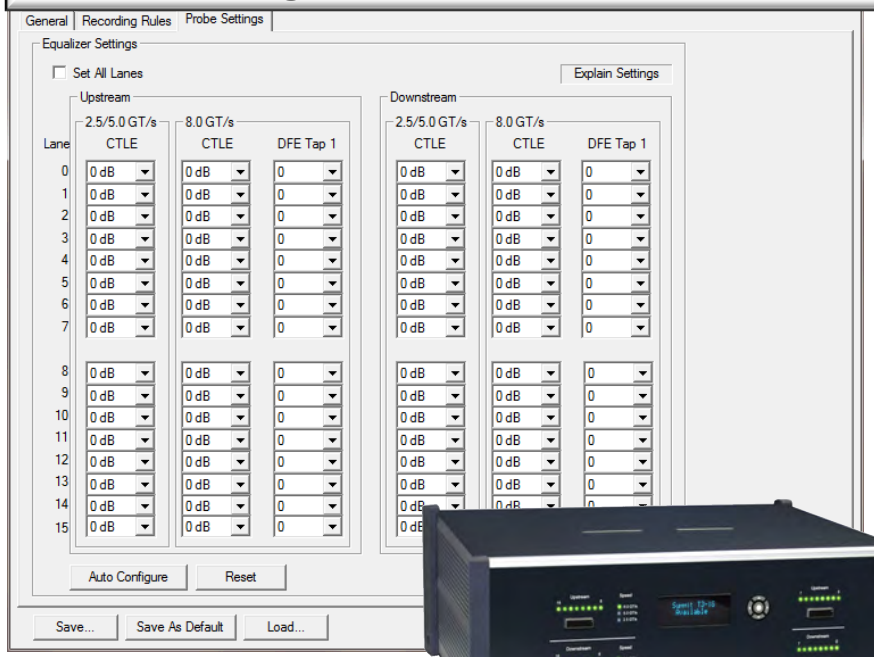
Need for Interposer Cards

- There are 3 approaches to tapping a PCI Express link.
  - Interposer
  - Midbus probe
  - Multi-lead probe
- The interposer is usually the best option for tapping an SSD card/drive with a PCIe connector on it.
- Important considerations when choosing an interposer.
  - Should be electrically transparent as much as possible in the system
  - Should be able to capture traffic at highest transmission rate of system
  - Should support appropriate lane width of bus of device/host
  - Connector on interposer should be compatible with card/drive connector
  - Fast interposer auto configuration for PCIe 3.0
- LeCroy PCI Express protocol analyzer/interposers can support the following:
  - PCIe 3.0 x1/x4/x8/x16 SSD devices
  - PCIe 2.0 x1/x4/x8/x16 SSD devices

Interposer Card Functionality

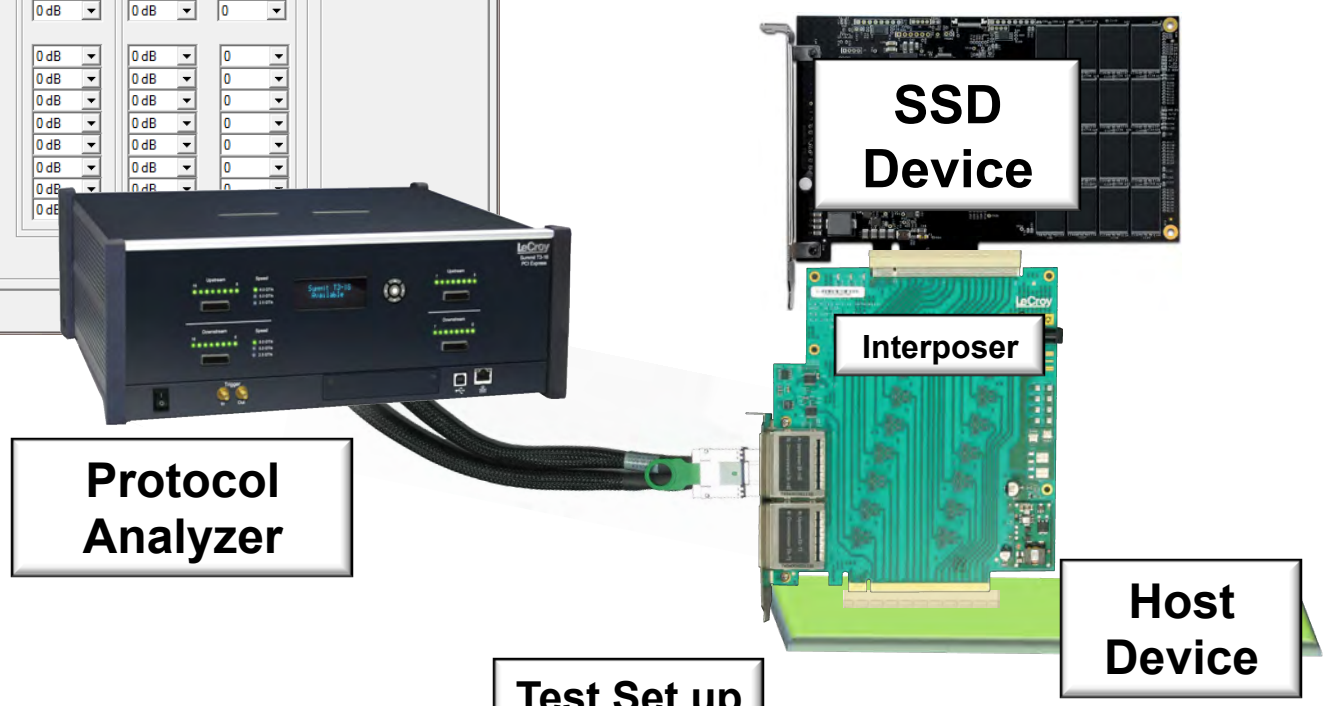
SNIA
Advancing storage & information technology

Auto Configuration screen for PCIe 3.0



Future Interposer

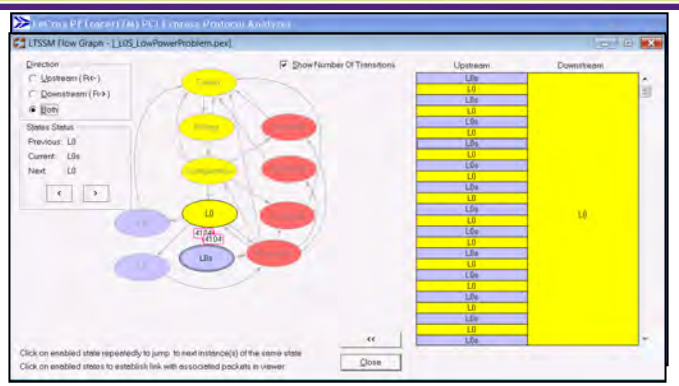
- SFF 8639 Multiprotocol Connector



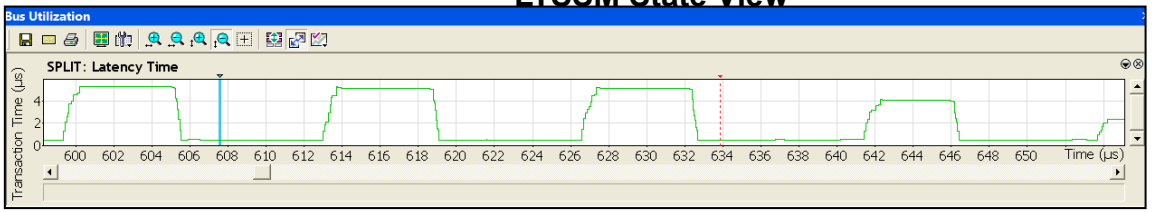
PCIe Card Performance Testing

SNIA
Advancing storage & information technology

- Packet Metrics
- Timing Calculator
- Link Transaction Performance Tool
- Post Capture Bus Utilization Graph Tool
  - Split View
  - Latency View
  - Throughput View
- Real Time Statistics Graph Tool
- Flow Control View
- LTSSM State View



LTSSM State View



Bus Utilization Graph Tool

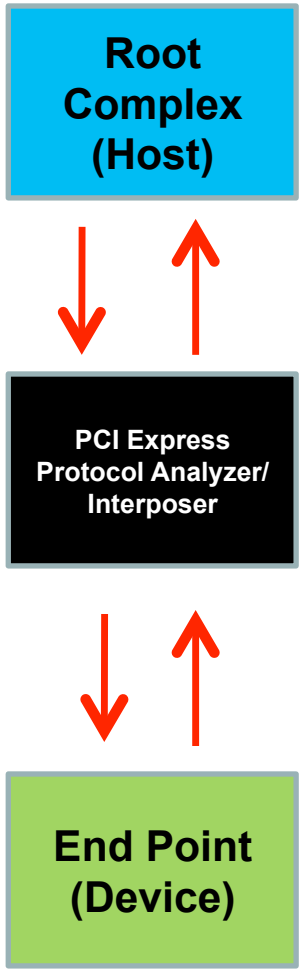
Bus Utilization

	Upstream	Downstream
Link Utilization	45.127 %	44.546 %
Time Coverage	45.072 %	44.493 %
Bandwidth	9025.43 Mb/s	8909.10 Mb/s
Data Throughput	592.29 MB/s	598.72 MB/s
Packets/second	19696168.45	24833612.81

Bandwidth, Link Utilization and Data Throughput View

LeCroy Protocol Analysis/Exerciser Capabilities

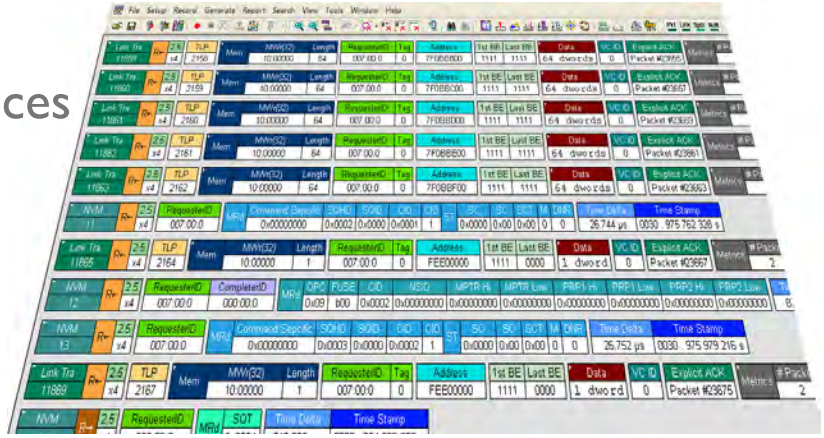
- **LeCroy Protocol Analyzers are used for**:
  - **Performance and Latency** timing characteristics/ measurement
  - Decoding a bit stream into **intelligible/readable structures**
  - **Understanding** training sequence requests and handling
  - Determining adherence to **Flow Control**
  - **Identifying** packet structures and timing issues
  - Retry/Replay sequences and error correction handling
  - General Fault Finding / **Trouble shooting**
- **LeCroy Exercisers are used for:**
  - Canned tests and scripting language provide **corner case** testing and **Validation testing**



LeCroy's Solution for PCIe SSD Decoding

SNIA
Advancing storage & information technology

- LeCroy's Summit T28/T3-8/T3-16 protocol analyzers support decoding for the following technologies:
  - SCSI Express(PQI) devices
  - NVMExpress 1.0B devices
  - SATA Express(SATA ACHI) devices



Decoded NVMExpress Trace

Open Discussion
Next Actions

SNIA
Advancing storage & information technology

Open Discussion:

Next Actions:

- **Feedback on Meeting No. 1 & 2**
- **Agenda & Topics for Meeting No. 3**

Time: 5:25 - 5:30

Supplemental Slides