SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- 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.
Abstract

- A variety of parameters can influence the performance behavior of a solid state drive: current and previous workloads, fragmentation, block size, read/write mix, and queue depth to name a few.
- SNIA’s Performance Test Specification allows for performance benchmarking that result in repeatable and consistent test results.
- This presentation will provide an overview of the SNIA SSD Performance Test Specification for both client and enterprise SSDs.
Definition of SSS

SSS = Solid State Storage

Traditional hard disk drive

Solid state hard drive
### The Why and How of SSD Performance Benchmarking

**The Performance Landscape**

**Random or Sustained?**

<table>
<thead>
<tr>
<th>Block Size</th>
<th>IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>8k / 2k / 1k</td>
<td>Up to 80 MB/s</td>
</tr>
<tr>
<td>50k / 50k</td>
<td>Up to 100 MB/s</td>
</tr>
</tbody>
</table>

**Random Precondition Sustained Speed?**

<table>
<thead>
<tr>
<th>Block Size</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>8k / 2k / 1k</td>
<td>Up to 52,000 IOPS</td>
</tr>
<tr>
<td>50k / 50k</td>
<td>Up to 17,000 IOPS</td>
</tr>
</tbody>
</table>

**MB/s or Mb/s?**

- Read and Write IOPS Specifications:
  - Bandwidth Performance Specifications:
    - Random 4K Reads: Up to 35 K IOPS
    - Random 4K Writes: Up to 35 K IOPS
    - Sequential 4K Reads: Up to 250 MB/s
    - Sequential 4K Writes: Up to 160 MB/s

**160 MB/s**

- Read and Write IOPS Specifications:
  - Random 4K Reads: Up to 35 K IOPS
  - Random 4K Writes: Up to 35 K IOPS
  - Sequential 4K Reads: Up to 250 MB/s
  - Sequential 4K Writes: Up to 160 MB/s

**IOPS**

- Random 4K Reads: Up to 35 K IOPS
- Random 4K Writes: Up to 35 K IOPS
- Sequential 4K Reads: Up to 250 MB/s
- Sequential 4K Writes: Up to 160 MB/s

**MB/s**

- Read and Write IOPS Specifications:
  - Bandwidth Performance Specifications:
    - Random 4K Reads: Up to 35 K IOPS
    - Random 4K Writes: Up to 35 K IOPS
    - Sequential 4K Reads: Up to 250 MB/s
    - Sequential 4K Writes: Up to 160 MB/s

**MB/s**

- Read and Write IOPS Specifications:
  - Bandwidth Performance Specifications:
    - Random 4K Reads: Up to 35 K IOPS
    - Random 4K Writes: Up to 35 K IOPS
    - Sequential 4K Reads: Up to 250 MB/s
    - Sequential 4K Writes: Up to 160 MB/s

**SNIA**

© 2011 Storage Networking Industry Association. All Rights Reserved.
Variables influencing Performance

- **Platform**
  - Test Hardware (CPU, interface, chipset, etc)
  - Software (OS, drivers)
- **SSS Device Architecture**
  - Flash geometry, cache, flash management algorithm, etc
Variables influencing Performance

- **Platform**
  - Test Hardware (CPU, interface, chipset, etc)
  - Software (OS, drivers)
- **SSS Device Architecture**
  - Flash geometry, cache, flash management algorithm, etc
- **Workload**
  1. Write history & preconditioning: State of device before testing
The need for Preconditioning

Performance States for Various SSDs

- NM (MLC)
- NS (SLC)
- JS (SLC)
- PSM (MLC)
- JM (MLC)

Normalized IOPS (IOPS/Max(IOPS))

Time (Minutes)

Steady State (desirable test range)
Write History - I

4K Random to 128K Sequential Transition

F.O.B. (~1hr)

Random to Sequential Transition (~1.5hr)

4K Steady State

128K Steady State
Write History - 2

128K Sequential to 4K Random Transition

- F.O.B. (~10 hrs)
- 128K Steady State
- 4K Steady State
- Sequential to Random Transition
Variables influencing Performance

- **Platform**
  - Test Hardware (CPU, interface, chipset, etc)
  - Software (OS, drivers)

- **SSS Device Architecture**
  - Flash geometry, cache, flash management algorithm, etc

- **Workload**
  1. Write history & preconditioning: State of device before testing
  2. Workload pattern: Read/write mix, transfer size, sequential/random
Workload Pattern

Performance depends on
- Read/Write Mix
- Block Size
- Queue Depth (not shown)
Variables influencing Performance

- Platform
  - Test Hardware (CPU, interface, chipset, etc)
  - Software (OS, drivers)
- SSS Device Architecture
  - Flash geometry, cache, flash management algorithm, etc
- Workload
  1. Write history & preconditioning: State of device before testing
  2. Workload pattern: Read/write mix, transfer size, sequential/random
  3. Data Pattern: The actual bits in the data payload written to the device
Dependency on data content - I

3D IOPS Surface Profile (IOMETER 2008)

- 30,000 - 35,000
- 25,000 - 30,000
- 20,000 - 25,000
- 15,000 - 20,000
- 10,000 - 15,000
- 5,000 - 10,000
- 0 - 5,000

Transfer Size (KB)

R/W Mix %
IOMeter 2008
Low Entropy Data Content

IOMeter 2006
High Entropy Data Content
The Need for Industry Standardization!

- **SNIA Technical Working Group (TWG)**
  - Created in early 2009
- **Specification for tests procedures to enable comparative testing of SSS performance**
  - **Agnostic** – Does not favor any one technology
  - **Relevant & Repeatable** – Meaningful to end users
  - **Practical** – Complete with reasonable time and effort
- **Performance Test Spec (PTS) 1.0 Client Released**
- **PTS 1.0 Enterprise Released**
  - PTS 1.1 in progress, target release 4Q11
Benchmark Types

- **Synthetic**
  - IOMeter, VDBench
  - Test specific scenario (QD, block size, transfer rate)
  - Good to determine corner case behavior

- **Application-based**
  - SysMark, PCMark
  - Test performance of specific application (ignores QD, transfer size, etc.)
  - Illustrates real world differences

- **Trace-based**
  - Storage Bench
  - Measures performance as drive is used (traces)
  - Most valid when similar applications are run (no two user workloads are the same)

SNIA PTS focuses on synthetic based benchmark tools
SSSI Reference Test Platform

Intel S5520HC

Single Intel W5580, 3.2GHz, Quad-core CPU

12GB, 1333MHz, ECC DDR3 RAM

LSI 9212-4e4i 6Gb/s SAS HBA

Intel ICH10R 3Gb/s SATA

8X Gen-II PCI-e

CentOS 5.5

Calypso RTP Backend V1.5

Calypso Test Suite (CTS) V6.5
The V1.0 Specification encompasses:

- A suite of basic SSS performance tests
- Preconditioning and Steady State requirements
- Standard test procedures
- Standard test reporting requirements
What Is NOT Covered In the Spec

- Application workload tests
- Matching to user workloads
- Energy efficiency
- Required test platform (HW/OS/Tools)
- Certification
- Device endurance, availability, data integrity

- Performance Test Specification v1.0 – Section 1.4
### Basic Test Flow

1. **Purge**
   - Security Erase, Sanitize, Format Unit, other proprietary methods where indicated

2. **Set Conditions**
   - Set user selectable test parameters, such as Active Range, Data Pattern, Demand Intensity

3. **Pre-Condition**
   - Workload independent (WIPScan)
   - Workload dependent (WDPC)

4. **Run Until SS**
   - Reiterate loops until Steady State is reached, or run to a prescribed maximum number of loops

5. **Collect Data**
   - Collect data from Steady State Measurement Window

6. **Generate Reports**
   - Use standard report formats and include required and optional elements
Key Concepts Used in the Spec.

A. Purge

B. Pre-Condition
   - Workload independent
   - Workload dependent

C. Active Range
   - Pre-conditioning
   - Test

D. Steady State
   - Measurement window
   - Data excursion condition
   - Slope excursion condition
A: Purge

As per the PTS V1.0 Specification, purge is defined as:

“The process of returning an SSS device to a state in which subsequent writes execute, as closely as possible, as if the device had never been used and does not contain any valid data”

Example implementation includes: ATA Security Erase, Sanitize, SCSI Format Unit
Pre-Conditioning is a key requirement in getting repeatable, representative results.

Goal is to put drive into “Steady State”, using:

- **Workload independent** – PTS v1.0 Section 3.3
  - Use a prescribed workload unrelated to the test loop
  - Write 2X user capacity using SEQ/128KiB blocks

- **Workload dependent** – PTS v1.0 Section 3.3
  - Run test workload itself as pre-conditioning (self pre-conditioning)
C: Active Range

As per the PTS V1.0 Specification, Active Range is defined as:

“… ActiveRange is the range of LBA’s that may be accessed by the preconditioning and/or test code…”

They are normally defined as % of the maximum LBA available to the user

Note Pre-conditioning and Test can have different Active Ranges
D: Steady State Definition

- Premise is that reported data should be take only AFTER the test loop results shows the drive has reached and maintained “Steady State”
- The Measurement Window is the interval, measured in Rounds, when the test results have entered and maintained Steady State for 5 Rounds
Steady State is reached only if BOTH of the following conditions are satisfied (assuming “y” is the variable being tracked):

1. Variation of y within the Measurement Windows is within 20% of the Average
   " Max(y) - Min(y) within the Measurement Window is no more than 20% of the Ave(y) within the Measurement Window; and "

2. Trending of y within the Measurement Windows is within 10% of the Average
   " [Max(y) as defined by the linear curve fit of the data within the Measurement Window] – [Min(y) as defined by the best linear curve fit of the data within the Measurement Window] is within 10% of Ave(y) within the Measurement Window. “
Steady State Measurement Windows (Rounds 12-16)

- **IOPS**
- **Average**
- **110% Average**
- **90% Average**
- **Slope**

**Data Excursion**
**Slope Excursion**

10% of Average
20% of Average

Round

The Why and How of SSD Performance Benchmarking
© 2011 Storage Networking Industry Association. All Rights Reserved.
D: Steady State Definition

Compare

- [Data Excursion] with [20% of Average]
- [Slope Excursion] with [10% of Average]

Note

- This method is slightly more tolerant than +10% and – 10% data excursion method and +5% and – 5% slope excursion method
D: How Good is the Steady State

200G-Class MLC: 72 Rounds Pre-conditioning Report: 100% Writes

"Steady State"
Workload Schematics

**Write Saturation**
- Random Access
- R/W: 100% Writes
- BS: 4K

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

**Enterprise TP**
- Sequential Access
- R/W:
  - 100/0, 0/100
- BS:
  - 1024K, 64K, 8K, 4K, 0.5K

**Enterprise Latency**
- Random Access
- R/W:
  - 100/0, 65/35, 0/100
- BS:
  - 8K, 4K, 0.5K
IOPS RW/BS Sequence

Enterprise IOPS Block Sequencing

The Why and How of SSD Performance Benchmarking
© 2011 Storage Networking Industry Association. All Rights Reserved.
TP RW/BS Sequence

Enterprise TP Block Size Sequencing
Example: Enterprise IOPS

**DUT:**
- 100GB-Class Enterprise SLC drive

**Test Parameters:**
- Active Range = [0, 100%]
- Thread Count = 2
- Queue Depth (Outstanding IO/Thread) = 16
- DP = RND
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle Recovery</td>
<td>See how the drive responds to host idle time amidst continuous access</td>
</tr>
<tr>
<td>Cross Stimulus Recovery</td>
<td>See how drive handles switching between sustained access patterns</td>
</tr>
<tr>
<td>Demand Intensity</td>
<td>See how drive responds to increasing host demands</td>
</tr>
<tr>
<td>Response Time Histogram</td>
<td>Get detailed response time statistics during specific stimulus</td>
</tr>
<tr>
<td>IOPS/W</td>
<td>Measures power efficiency of the device</td>
</tr>
<tr>
<td>Trace-Based Workloads</td>
<td>Captures or uses captured workloads traces and provide a consistent way to playback such traces</td>
</tr>
<tr>
<td>Enterprise Composite Synthetic Workload</td>
<td>Synthetic composite workload for Enterprise environments similar to JEDEC workload for endurance testing</td>
</tr>
</tbody>
</table>
Q&A / Feedback

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

Many thanks to the following individuals for their contributions to this tutorial.
- SNIA Education Committee

Eden Kim
Easen Ho
Esther Spanjer