Introducing and Validating SNIA SSS Performance Test Suite

Esther Spanjer
SMART Modular
Abstract

- SSS Performance Benchmarking Learning Objectives
  - Get a good understanding of the various parameters that influence the performance characteristics of SSDs
  - Get a full understanding of the proposed SNIA Performance Measurement Specification
  - Provide step-by-step guidance on how to set up a test benchmark that enables comparison among the various SSS devices
Definition of SSS

SSS = Solid State Storage
The Performance Landscape

**IOPS?**

**Random Precondition**

**Sustained Speed?**

**Up to?**

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

**Random or**

**Sustained?**

**Block Size?**

**PEAK/Sustained IOPS - Sector 4KB aligned (random preconditioned Sustained speed)**

<table>
<thead>
<tr>
<th>Sector Size</th>
<th>PEAK IOPS 4KB READ</th>
<th>PEAK IOPS 4KB WRITE</th>
<th>Sustained 4KB READ</th>
<th>Sustained 4KB WRITE</th>
<th>Sustained 8KB READ</th>
<th>Sustained 8KB WRITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4KB</td>
<td>50K / 50K</td>
<td>50K / 50K</td>
<td>50K / 32K</td>
<td>50K / 11K</td>
<td>50K / 11K</td>
<td>28K / 28K</td>
</tr>
<tr>
<td>8KB</td>
<td>23K / 23K</td>
<td>23K / 23K</td>
<td>23K / 23K</td>
<td>23K / 23K</td>
<td>23K / 23K</td>
<td>23K / 23K</td>
</tr>
</tbody>
</table>

**Performance**

- Average Access Time: 20-120 microseconds
- Sustained Read Throughput: 240,000 Mb/s
- Sustained Write Throughput: 300 MB/s

- Random IOPS Read Operations: 45,000 / 10/sec, sustained
- Random IOPS Write Operations: 16,000 / 10/sec, sustained

**Prominent product specifications include:**

- Up to 52,000 Sustained Random Read IOPS
- Up to 17,000 Sustained Random Write IOPS
- Sequential read: Up to 250 MB/sec
- Sequential write: 170 MB/sec
Market Segmentation

**Client SSD**
- Low cost
- C-MLC
- 0-7% over provisioning
- No supercap
- No Enterprise features
- No customization
- Warranty 1-3 yrs

**Enterprise SSD**
- Higher cost
- E-MLC/SLC
- 28-50% over provisioning
- Supercap support
- Enterprise features
- Customization
- Warranty 5 yrs
Performance Comparison
Enterprise vs. Client SSD

Sequential Read

- Enterprise SSD (SATA)
- Enterprise SSD (SAS)
- Client SSD 1 (SATA)
- Client SSD 2 (SATA)

Sequential Write

- Enterprise SSD (SATA)
- Enterprise SSD (SAS)
- Client SSD 1 (SATA)
- Client SSD 2 (SATA)

Random Read

- Enterprise SSD (SATA)
- Enterprise SSD (SAS)
- Client SSD 1 (SATA)
- Client SSD 2 (SATA)

Random Write

- Enterprise SSD (SATA)
- Enterprise SSD (SAS)
- Client SSD 1 (SATA)
- Client SSD 2 (SATA)
Variables Influencing Performance

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

FOB
Transition
Steady State
(desirable test range)
Write History - 1

4K Random to 128K Sequential Transition

- F.O.B. (~1hr)
- Random to Sequential Transition (~1.5hr)
- 4K Steady State
- 128K Steady State

IOPS vs. Time (Minutes)
128K Sequential to 4K Random Transition

128K Steady State

F.O.B.

~10 hrs

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**
  - Write history & preconditioning: state of device before testing
  - Workload pattern: read/write mix, transfer size, sequential/random
Workload Pattern

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

Note: Shown 3D IOPS image courtesy of Calypso Systems
Dependency on data content

3D IOPS Surface Profile (IOMETER 2008)

3D IOPS Surface Profile (IOMETER 2006)
The need for industry standardization!

- SNIA Technical Working Group (TWG)
  - Created in early 2009
- Spec
  - Agnostic
    - Does not favor any one SSS technology
  - Repeatable & Practical
    - Complete with reasonable time and effort
- Spec 0.9 open for review now
- Spec 1.0 focuses on further detailed test metrics
## Benchmark Suites

<table>
<thead>
<tr>
<th>Test Suite</th>
<th>Client SSD</th>
<th>Enterprise SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCMark</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HDD Score, OS and application loading timing, user simulation (surfing web, windows media player, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SysMark</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>System-level test. Measures performance based on average response time, gives score (0-250)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOMeter</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Sequential/Random performance, workload simulation (file server, web server workload, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDTach/H2benchw</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Performance stability, Sequential/Burst performance, Access Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD Tune</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance stability, Sequential/Burst performance, Access Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Everest</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Random Access Time (Read/Write)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VDBench</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Workload generator, performance on DAS and NAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calypso CTS</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Device (RAW) level, direct IO synthetic stimulus generator for both client and enterprise</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is NOT covered in spec

- Application Tests
- Matching to user workloads
- Energy efficiency
- Required Test platform (HW/SW tools)
- Certification
- Device endurance, data integrity, availability
**Test Flow**

1. **Prepare the Device**
   - Purge/Erase/Format → put SSD back into “original” state

2. **Set Conditions**
   - Set Active Range, Data Pattern, Demand Intensity

3. **Preconditioning**
   - Bring device to known state

4. **Steady State Testing**
   - Run Test Loop up until steady state is achieved (± 10% margin)
   - Testing for throughput, IOPS & latency

5. **Standardized Reporting**
   - Steady state convergence & verification
   - Performance measurement (2D/3D)
Preconditioning

- Preconditioning is key to get repeatable results
- Preconditioning needed to get drive in Steady State, after which performance can be measured
Preconditioning

- Preconditioning is key to get repeatable results
- Preconditioning needed to get drive in Steady State, after which performance can be measured

- Two types of preconditioning
  - Workload independent – write 2x capacity with 128KB sequential writes
  - Workload dependent – run workload itself until steady state is achieved
Steady State

- Measurement window is interval for last 5 measured rounds (i.e. test loops) that show steady state results
- Steady State is achieved if BOTH conditions are met:
  - Variation of $y$ in measurement window is within 20% of average
  - Trending of $y$ within measurement window is within 10% of average
Performance Workloads & Tests

Client Test

- Random IOPS
  - 100/0, 95/5, 65/35, 50/50, 35/65, 5/95, 0/100
  - 1024K, 128K, 64K, 32K, 16K, 8K, 4K, 0.5K
- Sequential MB/s
  - 100/0, 0/100
  - 1024K
- Latency (random access)
  - 100/0, 65/35, 0/100
  - 8K, 4K, 0.5K

Enterprise Test

- Random IOPS
  - 100/0, 95/5, 65/35, 50/50, 35/65, 5/95, 0/100
  - 1024K, 128K, 64K, 32K, 16K, 8K, 4K, 0.5K
- Sequential MB/s
  - 100/0, 0/100
  - 1024K
  - 64K, 8K, 4K, 0.5K
- Latency (random access)
  - 100/0, 65/35, 0/100
  - 8K, 4K, 0.5K

Version 1.0 will focus further on differentiating Client vs. Enterprise Testing
Steady State Convergence Plot

Steady State Verification
Performance Measurements 3D

Average Latency vs. Block Size and R/W Mix %
ActiveRange = (x,y); OIO/Thread = x; Thread Count = x; Data Pattern = x

Ave Latency (ms)

0/100  65/35  100/0
0.5KiB  4KiB  8KiB

Block Size

Performance Measurements 2D

Enterprise IOPS Test - Ave IOPS vs. Block Size & R/W Mix %
ActiveRange = (x,y); OIO/Thread = x; Thread Count = x; Data Pattern = x

IOPS

100,000  10,000  1,000  100  10

1  2  4  8  16  32  64  128  256  512  1,024

Block Size (KiB)
Sample Test Report included

<table>
<thead>
<tr>
<th>Device Under Test (DUT)</th>
<th>ABC Co.</th>
<th>SSS TWG PTS Summary Report</th>
<th>SNIA SSS TWG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model No.:</td>
<td>ABC123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form Factor:</td>
<td>2.5&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAND Capacity:</td>
<td>256 GB MLC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUT Interface:</td>
<td>SATAI, SAS HBA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing Summary: Tests Run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Preparation</td>
<td>Test Loop Parameters</td>
<td></td>
</tr>
<tr>
<td>8.1 Client IOPS</td>
<td>Secure Erase</td>
<td>2X 128K SEQ</td>
<td>100%</td>
</tr>
<tr>
<td>8.2 Client IOPS OPT - AR</td>
<td>Secure Erase</td>
<td>2X 128K SEQ</td>
<td>10%</td>
</tr>
<tr>
<td>8.3 Client IOPS OPT - Data</td>
<td>Secure Erase</td>
<td>2X 128K SEQ</td>
<td>100%</td>
</tr>
</tbody>
</table>

General Device Description

<table>
<thead>
<tr>
<th>Device Under Test (DUT)</th>
<th>System Hardware Configuration</th>
<th>System Software Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>ABC Co.</td>
<td>System Mfg:</td>
</tr>
<tr>
<td>Model No.</td>
<td>ABC123</td>
<td>Calypso Systems, Inc.</td>
</tr>
<tr>
<td>Sata No.</td>
<td>123</td>
<td>Operating Sys:</td>
</tr>
<tr>
<td>Firmware Rev No.</td>
<td>1.12</td>
<td>Linux CentOS 5.4</td>
</tr>
<tr>
<td>User Capacity</td>
<td>256 GB</td>
<td></td>
</tr>
<tr>
<td>Interface/Speed</td>
<td>6Gbs SATAI</td>
<td></td>
</tr>
<tr>
<td>Form Factor</td>
<td>2.5&quot;</td>
<td></td>
</tr>
<tr>
<td>Media Type</td>
<td>MLIC</td>
<td></td>
</tr>
<tr>
<td>Major Features</td>
<td>DUTFP</td>
<td></td>
</tr>
</tbody>
</table>

| Pre-conditioning Convergence Report - All Rounds |

8.1.1 Steady State Convergence Plot - All Block Sizes

8.1.2 Steady State Convergence Plot - 4K Block Sizes
Focus of Next Revision

- Random 4K Write Saturation
  - IOPS performance over time and with total amount of data written
- IOPS/W
  - Measure power efficiency
- Client Active Range restriction
  - Simulation of client usage, limit LBA range
- Cross Stimulus Recovery
  - Switching between random/sequential and large/small block
- Demand Intensity
  - Trade off between max. IOPS and max. response time
- SSD Figure of Merit
  - Direct comparison between SSD’s (similar to HDDs 10K, 15K RPM)
Other Standardization Initiatives

- SSSI Group of SNIA
  - Technical Work Group (TWG) → Performance Benchmark Spec
  - Tech Dev Group → Performance Test Platform
- JEDEC 64.8
  - Specification for SSD endurance measurement
- SSDA
  - Testing of reliability (power cycling, data retention, endurance, etc) and OS compatibility (Windows 7)
Summary

- SSS Performance is dependent on many variables
- Comparing vendors is not trivial → industry standard required
- SNIA performance spec allows apple to apples comparison
  - Spec for review at http://www.snia.org/tech_activities/publicreview
  - Send your feedback to http://www.snia.org/tech_activities/publicreview