SNIA PERSISTENT MEMORY + SUMMIT 2021 COMPUTATIONAL STORAGE

FROM DATACENTER TO EDGE : VIRTUAL EVENT
APRIL 21-22, 2021

The Challenges of Measuring Persistent Memory Performance

SNIA Persistent Memory Performance Test Specification (under development) - Solid State Storage Technical Working Group

Eduardo Berrocal, Senior Software Engineer, Intel Corp.

Keith Orsak, Master Storage Technologist, HPE

1. Introduction

SNIA Solid State Storage Technical Working Group

- The Persistent Memory (PM) Performance Test Specification (PTS) is an under development technical work of the SNIA Solid State Storage Technical Working Group.
- The PM PTS is intended to set forth a standardized methodology for the set-up, test and reporting of PM storage performance.
- The PM PTS also lists Reference Test Platforms (RTP) for the test of PM Storage based on commercially available third party servers that support the test methodology and operation of PM as described in the PM PTS.
- The PM PTS is intended to allow application and storage professionals to design, integrate and deploy architectures based on and including Persistent Memory.

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

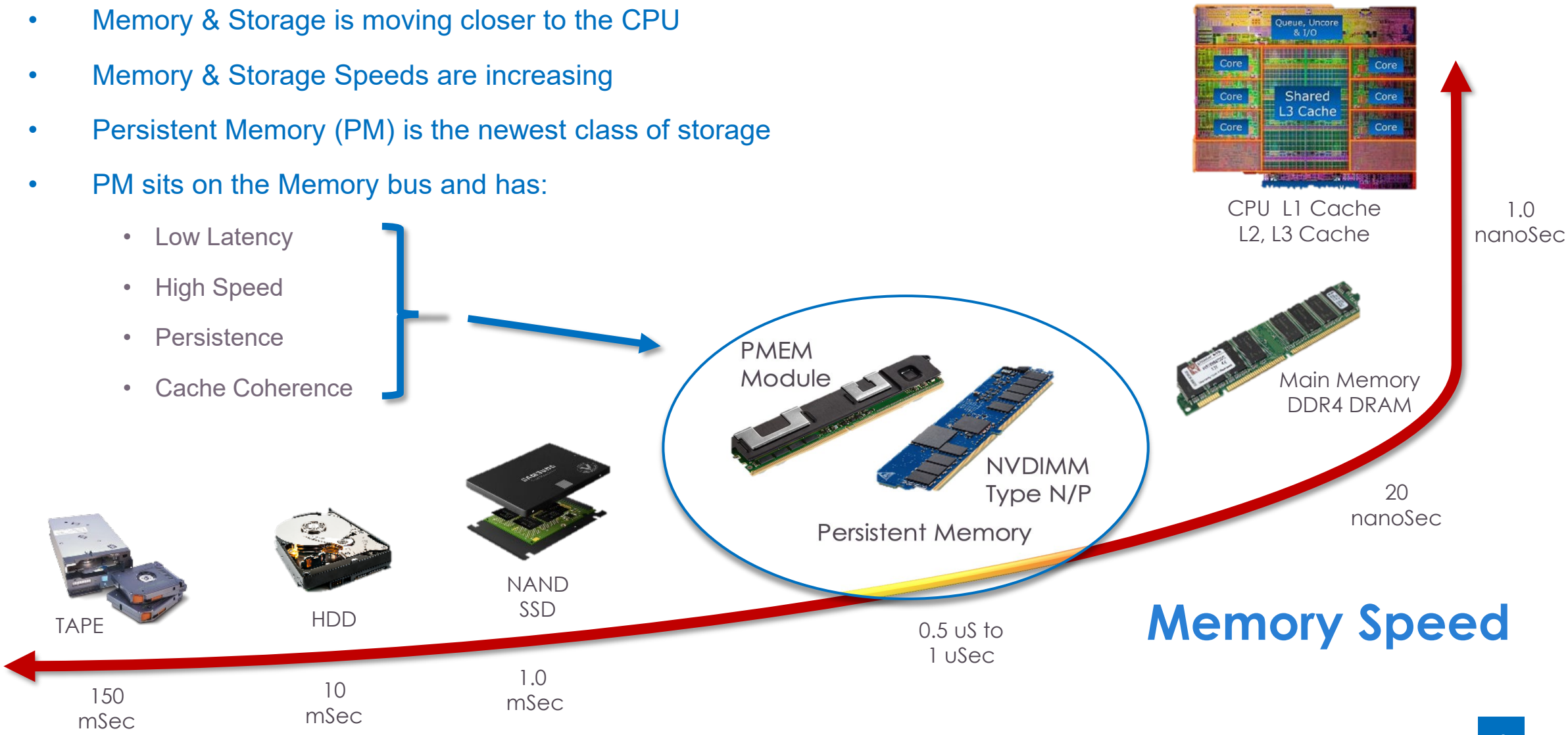
Part 1: Persistent Memory Configuration

Storage Access Modes: Data Path, Software Stack Configuration & PM Media

Eduardo Berrocal, Intel

2. Introduction to Persistent Memory

- Memory & Storage is moving closer to the CPU
- Memory & Storage Speeds are increasing
- Persistent Memory (PM) is the newest class of storage
- PM sits on the Memory bus and has:
  - Low Latency
  - High Speed
  - Persistence
  - Cache Coherence



© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

2. Storage Access Modes - Data Path Perspective

PM PTS addresses 3 types of Storage Access Modes (2, 3 & 4 below)

1 - Traditional Block IO

- Shown for reference Only
- Not addressed in PM PTS
- Shows traditional storage access

2 - PM Block IO - Sector Mode

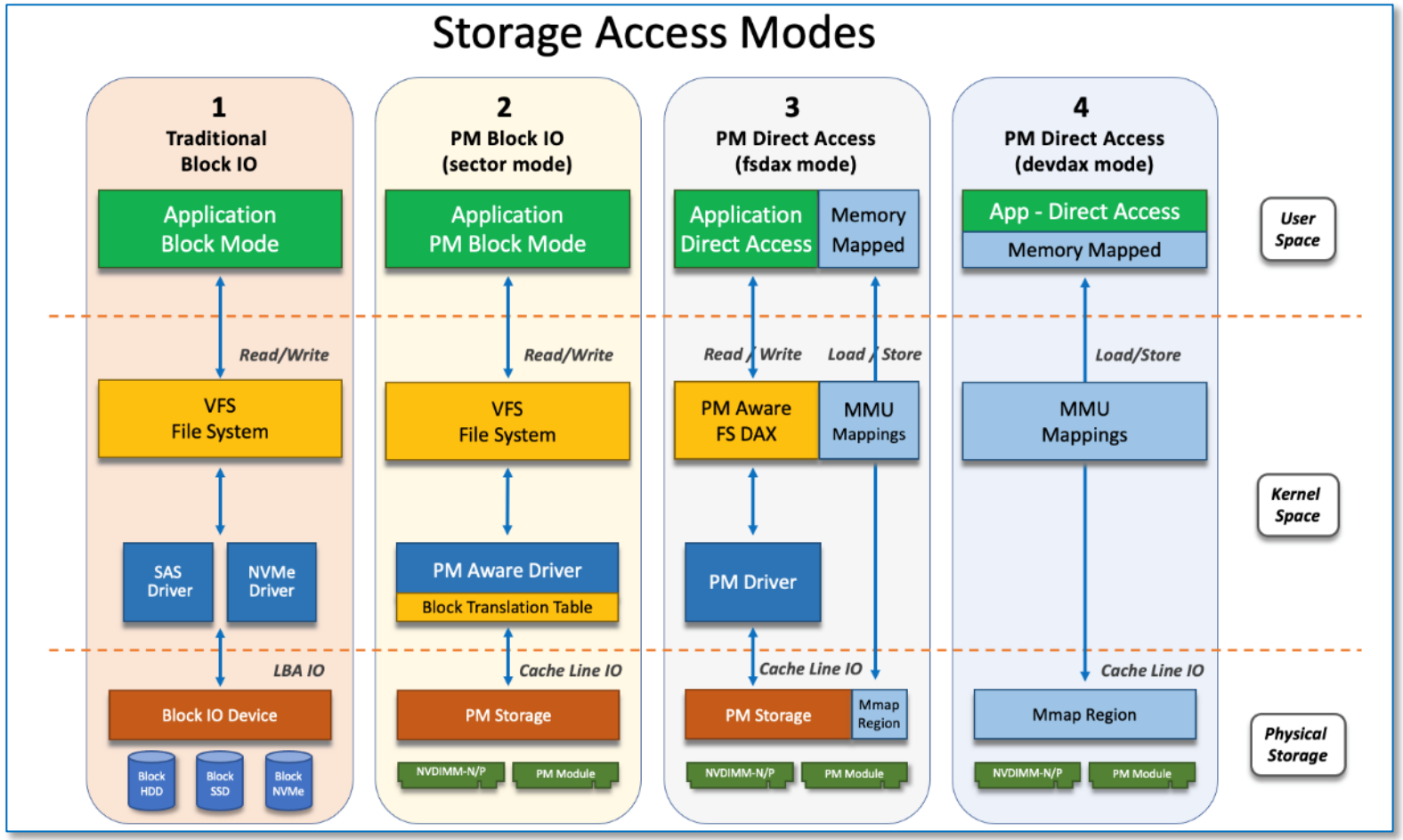
- PM Aware Driver
- Block Translation Table (BTT)
- Sector Atomicity

3 - PM Direct Access w/ File System

- Application Direct Access
- Memory Mapped
- Uses PM Aware File Systems

4 - PM Direct Access w/o File Systems

- Mmap entire Region
- Used mostly for legacy DAX



© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

3. Storage Access Modes - Software Stack Perspective

Linux DAX File System

1 - Configured Region

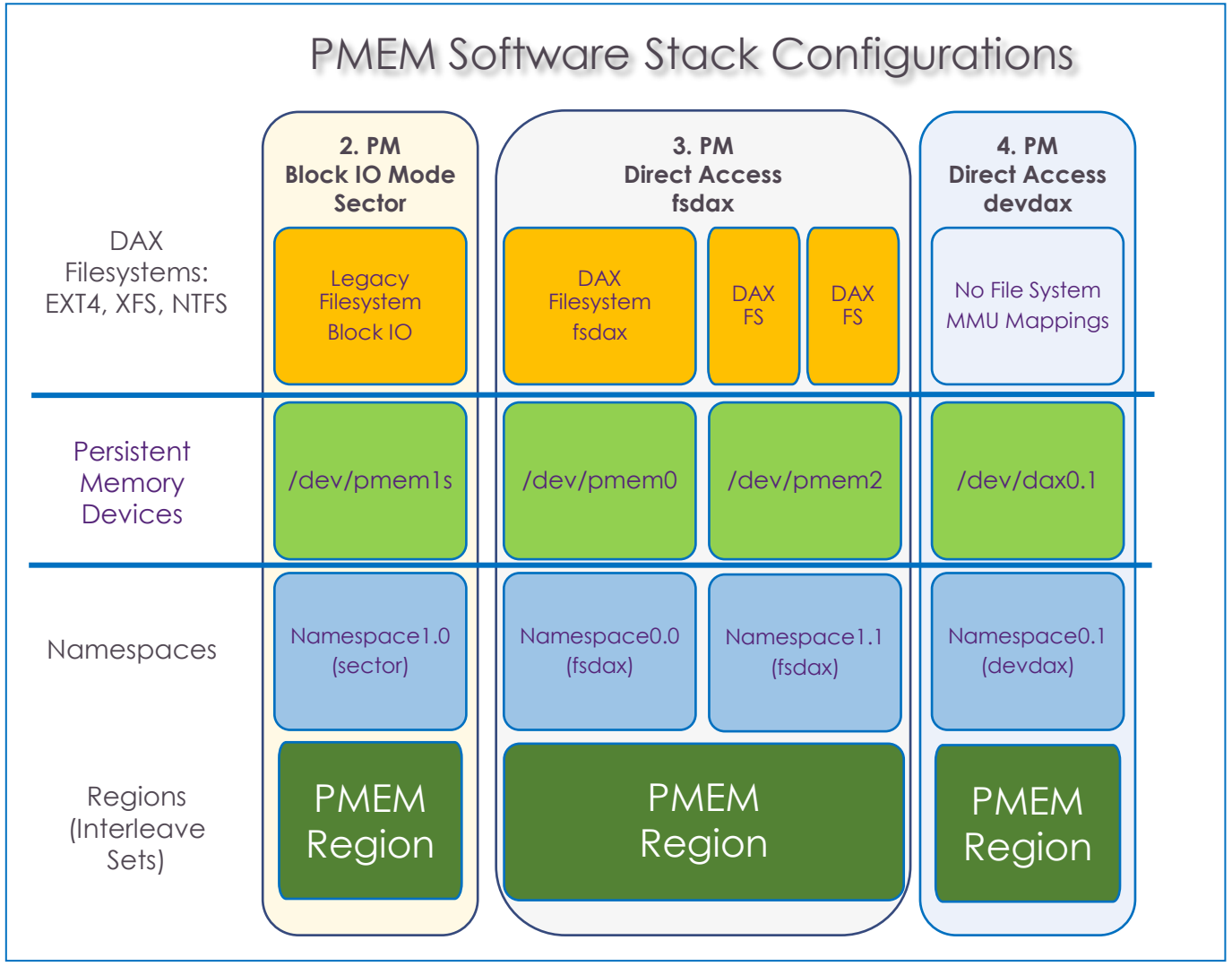
- Interleaved v Non-interleaved
- Regions: 0, 1, 2 …

2 - Defined OS Namespace

- Sector
- Fsdax
- Devdax

3 - Format File System

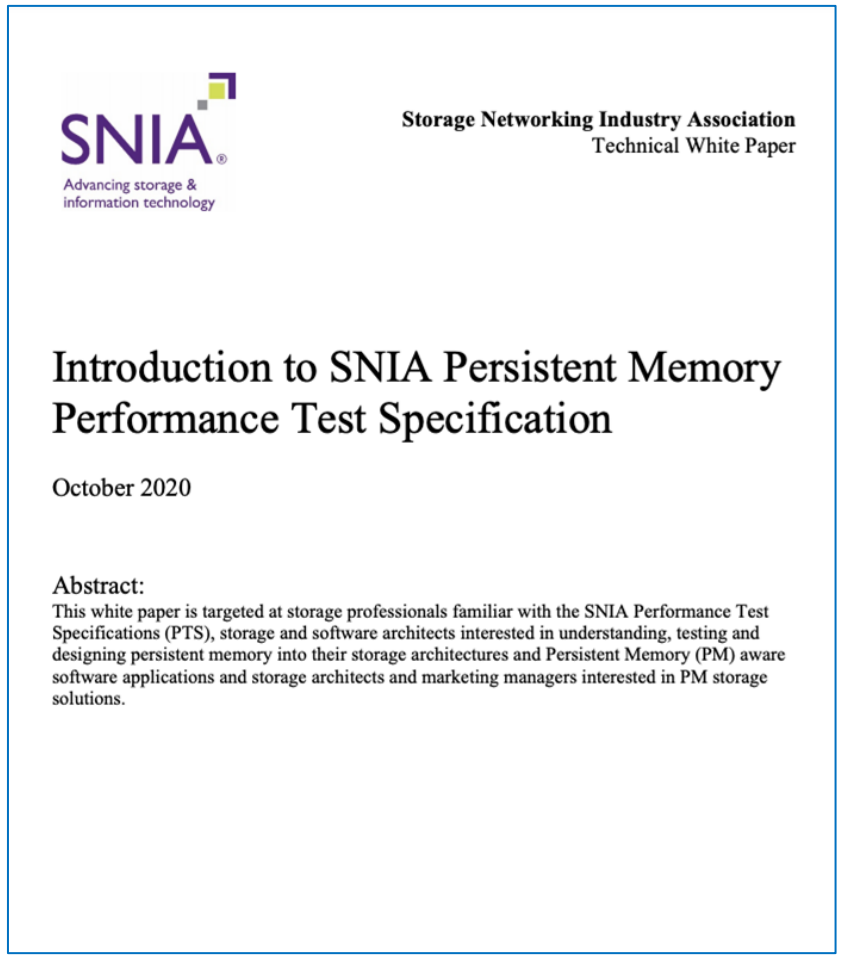
- Legacy v DAX FS



© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

4. SNIA PM Performance Test Specification - PM PTS

White Paper & Draft Specification



SNIA
Advancing storage & information technology

Storage Networking Industry Association
Technical White Paper

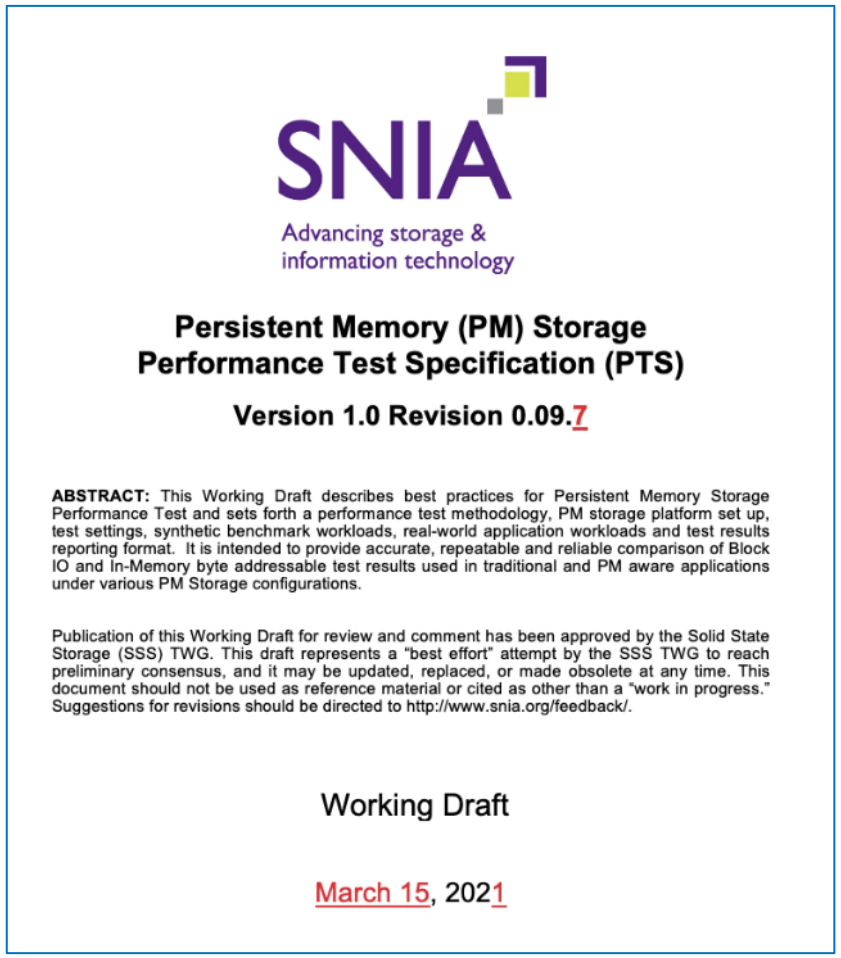
Introduction to SNIA Persistent Memory Performance Test Specification

October 2020

Abstract:
This white paper is targeted at storage professionals familiar with the SNIA Performance Test Specifications (PTS), storage and software architects interested in understanding, testing and designing persistent memory into their storage architectures and Persistent Memory (PM) aware software applications and storage architects and marketing managers interested in PM storage solutions.

Introduction to SNIA Persistent Memory Performance Test Specification White Paper

https://www.snia.org/white-paper/introduction-snia-pm-performance-test-specification



SNIA
Advancing storage & information technology

Persistent Memory (PM) Storage Performance Test Specification (PTS)

Version 1.0 Revision 0.09.7

ABSTRACT: This Working Draft describes best practices for Persistent Memory Storage Performance Test and sets forth a performance test methodology, PM storage platform set up, test settings, synthetic benchmark workloads, real-world application workloads and test results reporting format. It is intended to provide accurate, repeatable and reliable comparison of Block IO and In-Memory byte addressable test results used in traditional and PM aware applications under various PM Storage configurations.

Publication of this Working Draft for review and comment has been approved by the Solid State Storage (SSS) TWG. This draft represents a "best effort" attempt by the SSS TWG to reach preliminary consensus, and it may be updated, replaced, or made obsolete at any time. This document should not be used as reference material or cited as other than a "work in progress." Suggestions for revisions should be directed to http://www.snia.org/feedback/.

Working Draft

March 15, 2021

SNIA Solid State Storage Technical Working Group
https://www.snia.org/tech_activities/work/twgs#s3

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

Part 2: PM PTS Test Methodology

Reference Test Platforms, Test Set-up, Test Methodology, Tests, Results & Reporting

Keith Orsak, Master Storage Technologist, HPE

1. Some Challenges of Measuring PM Performance

- Software stack complexity, including BIOS, Drivers & set-up, can greatly influence the measurement of PM Performance
- Test platform, test software and test set-up affects accuracy, repeatability and consistency
- Selection of IO Engine and workloads are also key to performance
- Note that drive/storage preparation, such as pre writes and steady state, while useful do not have as large an impact on PM performance as occurs with NAND Flash SSD performance.

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

2. PM Performance Test Specification

Standardized Performance Test of Persistent Memory

Draft PM PTS v1.0:
Preliminary & Subject to Change

Storage Server

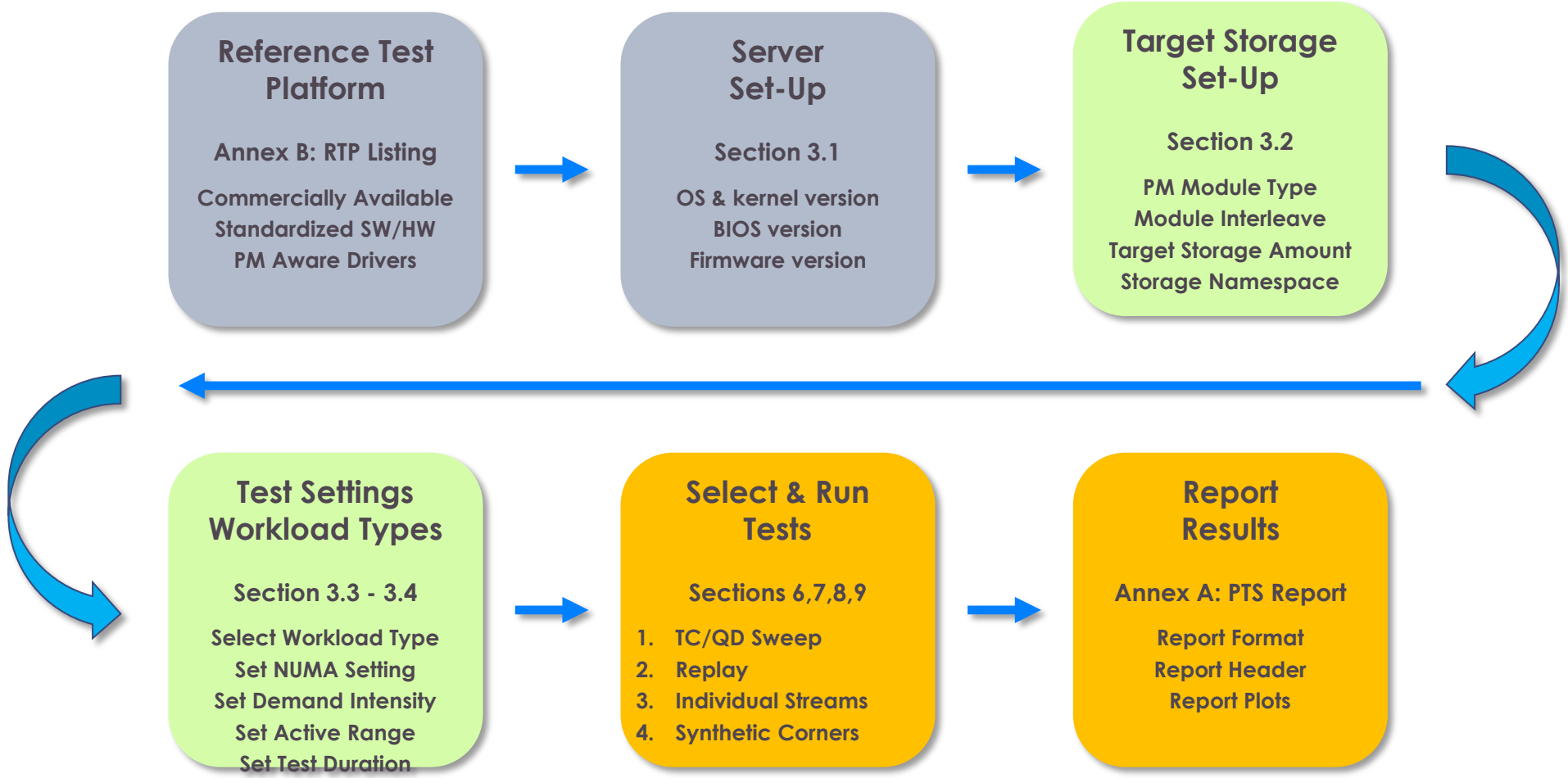
- RTP
- Server set-up

Test Storage

- PM Storage
- Test Settings
- Workloads

Tests

- Test Types
- Results Reporting



© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

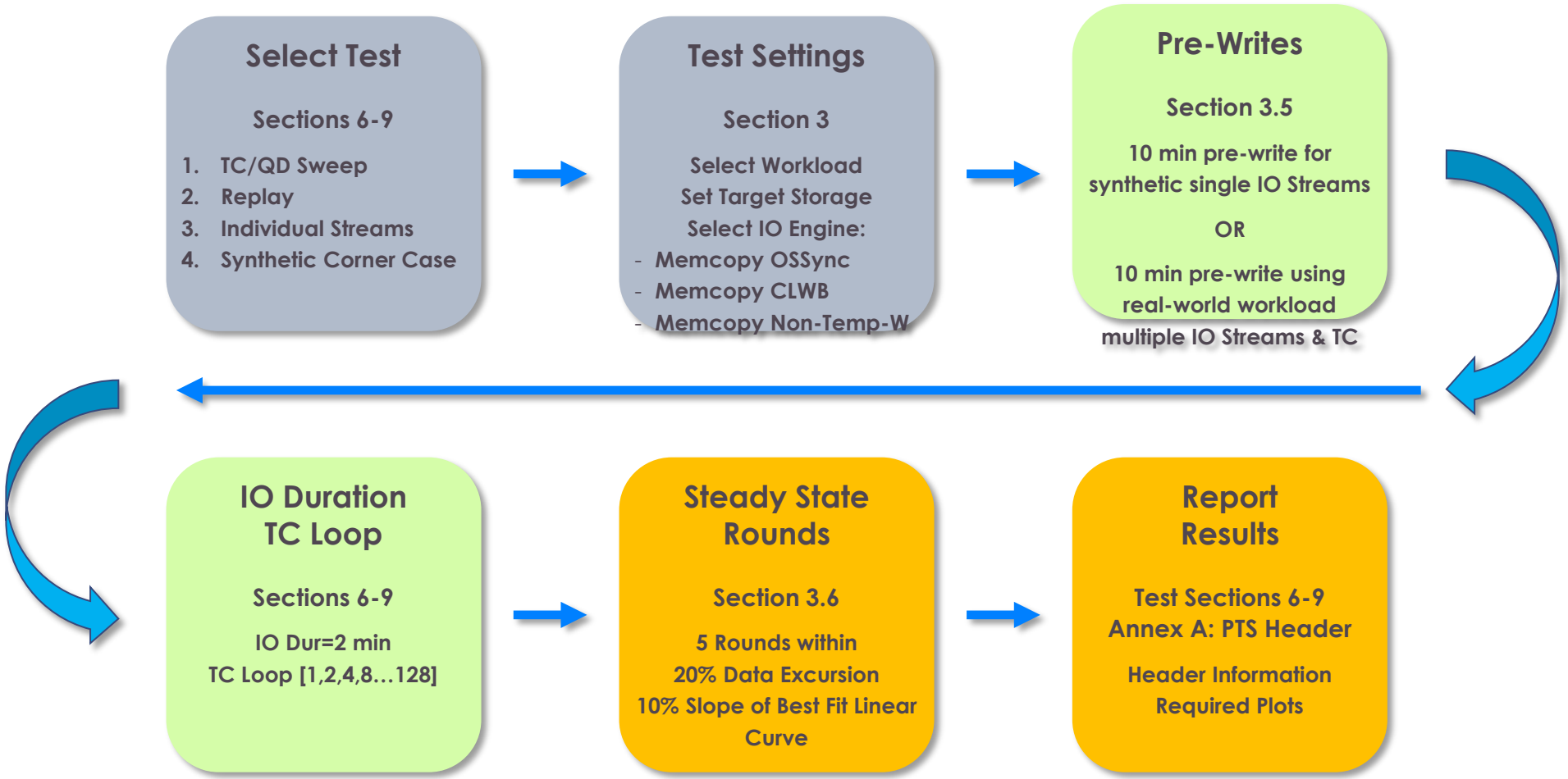
3. PM PTS - Test Methodology
Test Flow

Draft PM PTS v1.0:
Preliminary & Subject to Change

Tests
- Select Test Type
- Synthetic or Real World Workload
- Select IO Engine

Test Steps
- Pre-writes
- IO Duration
- TC Loop

Results
- Rounds (repeatability)
- Reporting (SNIA Format)



© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

7. PM PTS - Tests

Test Type, Description, Purpose

Draft PM PTS v1.0:
Preliminary & Subject to Change

Comment

- Item
- Item

Comment

- Item
- Item
- Item

Comment

- Item
- Item

Test Type	PTS Section No.	Test Description	Purpose of Test
Replay	Section 6.0	Apply Sequence & Combination of IO Stream Subset observed	Observe Performance Relative to Actual Real-World Workloads
Thread Count (TC) Sweep	Section 7.0	Apply Fixed Composite IO Streams v TC Sweep to Steady State	IO, Bandwidth, Response Time & Thread Count Saturation
Individual Streams	Section 8.0	Run Observed IO Streams to Steady State followed by TC Sweep	Real-World IO Streams vs Corner Case/Mfgr Specs
Synthetic Corner Case	Section 9.0	Run Selected IO Streams to Steady State followed by TC Sweep	Synthetic IO Streams vs Corner Case/Mfgr Specs

Note: Individual IO Stream is a single IO access pattern consisting of a RND or SEQ access of a R or W IO of a data transfer size - See PTS Specification Definitions. Real-World Workload capture, analysis and workload creation are discussed in the PM PTS, RWSW PTS for Datacenter Storage and the PM PTS White Paper.

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

8. Reporting Requirements
SNIA PM PTS Report Header

Draft PM PTS v1.0:
Preliminary & Subject to Change

Comment
- Item
- Item

Comment
- Item
- Item
- Item

Comment
- Item
- Item

Test Run Date:			11/15/2020 12:08:00 PM		Report Run Date:		12/02/20 11:14 AM	
Demand Intensity Response Time Histogram - Thread Count/Queue Depth Sweep (REQUIRED) - Report Page								
SNIA SSS TWG	Persistent Memory Performance Test Spec (PM PTS)		DIRTH TC/QD Sweep Test: Retail Web Portal 9 IO Stream				Rev.	PM PTS 1.0
							Page	1 of 12
Vendor:	ABC Co.	PM Storage Model:	PMEM ABC123		TEST SPONSOR		XYZ Test Co.	
Test Platform		**Device Under Test**		**Set Up Parameters**		**Test Parameters**		
Ref Test Platform	PM RTP Server 1.0	Mfgr	ABC Co.	Workload	Rtl Web Portal	Active Range	100%	
Motherboard	Intel Cascade Lake	Model No.	PMEM 12345678	IO Streams	9 IO Streams	Max IOPS	623,539 IOPS	
CPU	Dual Socket Intel 8176 2.1GHz	S/N	PA1B2C3-D4E5F6	Max OIO	TC=128/QD=1	OIO	T8Q1	
Memory	128 GB DDR4 2166	Firmware ver	1.00.11.00.11	Min OIO	TC=1/QD=1	5 9s QoS	120 mS	
Operating System	RHEL 7.5	Capacity	256 GB	Pre-Conditioning	SEQ 128K W - (10) min	Mid IOPS	604,498 IOPS	
Page Size Memory	2 MB	Total Modules	6	Steady State	SSS PTS 2.0.1	OIO	T4Q1	
NUMA	Enabled	Interleaved	Interleaved	Data Excursion	120%	5 9s QoS	100 mS	
Block / Byte	Block IO	DAS/Remote	DAS	Slope	120%	Min IOPS	158,274 IOPS	
IO Access Type	Mmap	LUN	1536 GB	Rounds	Five (1) min Rounds	OIO	T1Q1	
File System	DAX FS	RAID	Striped	Data Pattern	RND Once	5 9s QoS	50 mS	
Pre Conditioning IOPS Plot								

Note: SNIA format report headers are required for PM PTS reporting. In addition to administrative information, PM PTS headers shall disclose key information on Test Platform, Device Under Test, Set-up Parameters & Test Parameters.

© 2021 SNIA Persistent Memory + Computation Storage Summit. All Rights Reserved.

9. Compare IO Engines

Memcopy OSSync v CLWB v Non-Temp-W

Draft PM PTS v1.0:
Preliminary & Subject to Change

Memcopy_Non-Temp Writes

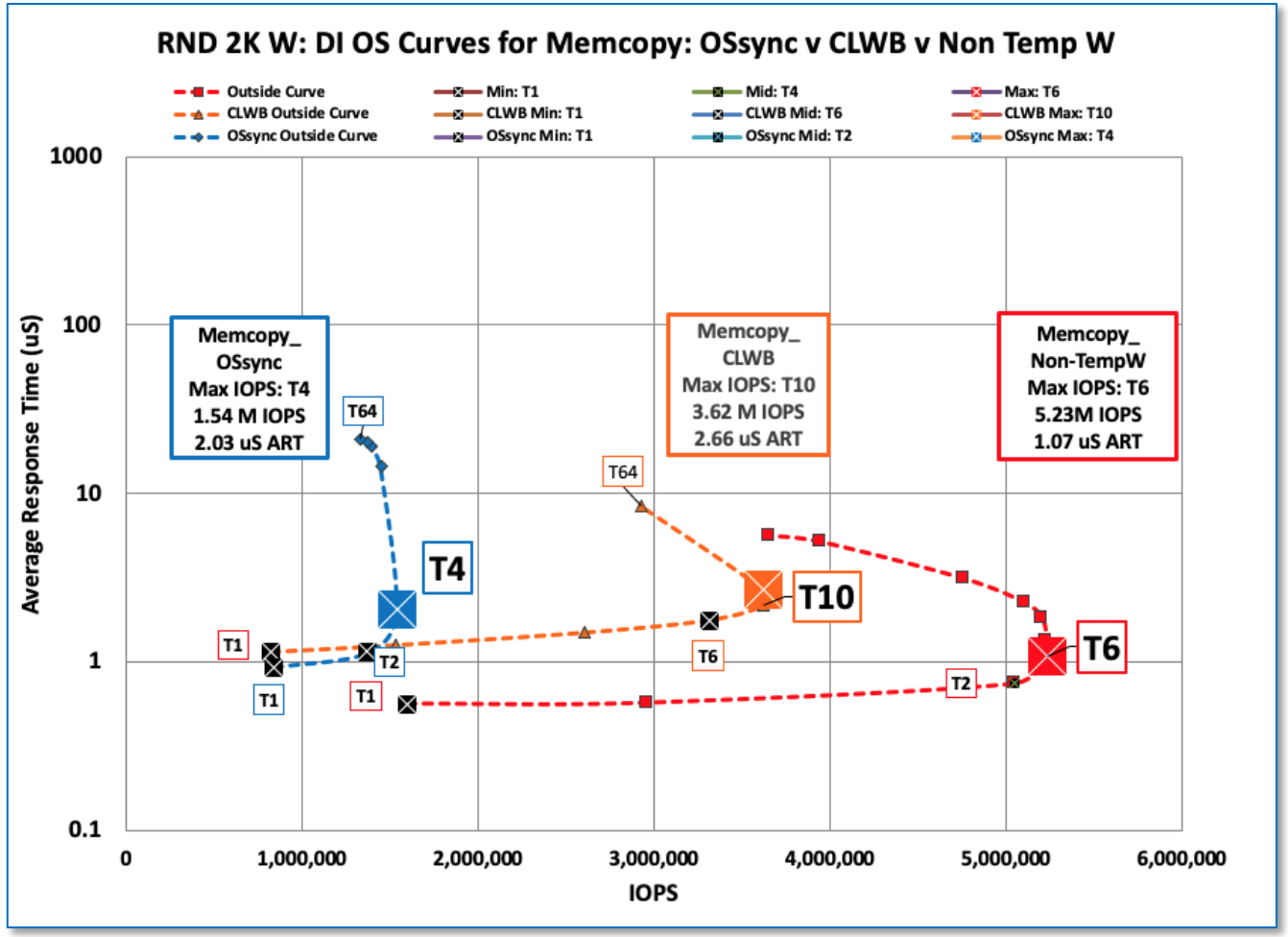
- Read data goes to cache
- Write data bypasses cache
- Non Temp W assumes data is accessed w/o temporal locality
- Synchronization done in User space by sfence
- Used when you wish to remove the effect of caches on the transfer of IOs

Memcopy_OSSync

- Synchronization is done by calling system calls
- Performance varies between msync, fsync and fdatasync
- Used when you evaluate performance in legacy applications - no control of caching

Memcopy_CLWB

- Cache line write back
- Synchronization done via periodic flushes
- Used when you desire persistent cache write back



Note: Demand Intensity (DI) curves show IO and Bandwidth saturation as Thread Count (TC) is increased. The optimal DI point is where IOPS or Bandwidth is highest and before Response Times begin to dramatically increase

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

10. Sample Data: Replay Test

Memcopy OSSync: Retail Web Portal - 9 IO Stream 24 hr

Draft PM PTS v1.0:
Preliminary & Subject to Change

IO Streams v Time

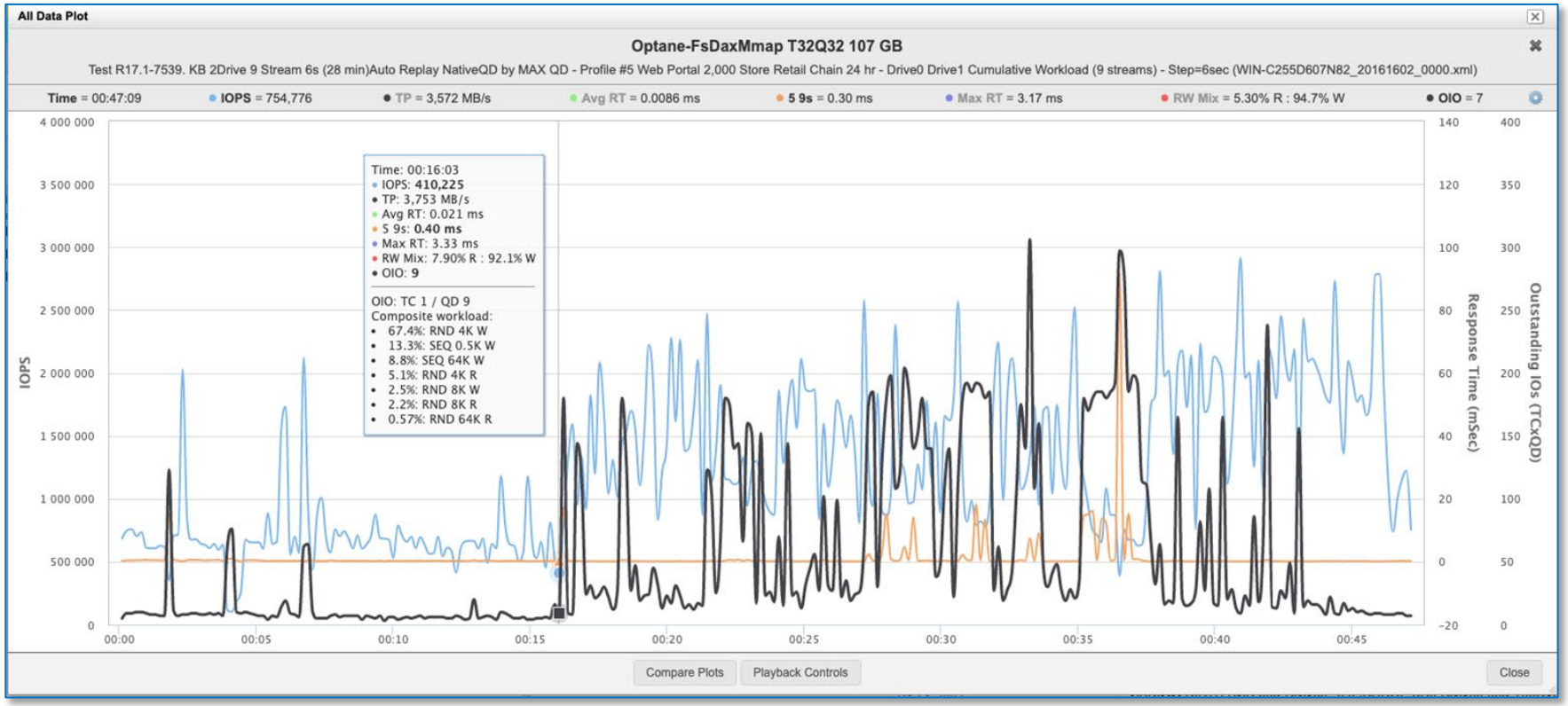
- Shows Changing combinations of IO Streams and TC
- Each data point is the average for the 9 IO Streams observed in the IO Capture period

Purpose

- See IO Stream combinations that occur in real world
- See changes over time
- Associate IO metrics with usage

Purpose

- Understand workload content
- Optimize PMEM for IO Stream content
- E.g. direct Reads to PMEM and Writes to DRAM



Note: Replay test shows the application of IOs and Demand Intensity over time. Above shows IOPS, Outstanding IO and 5 9s Response Time Quality of Service over time for a composite 9 IO Stream Retail Web Portal workload.

© 2021 SNIA Persistent Memory + Computation Storage Summit. All Rights Reserved.

11. Sample Data: Thread Count Sweep Test

Memcopy OSSync: RND 64 byte Write

Draft PM PTS v1.0:
Preliminary & Subject to Change

Test Flow

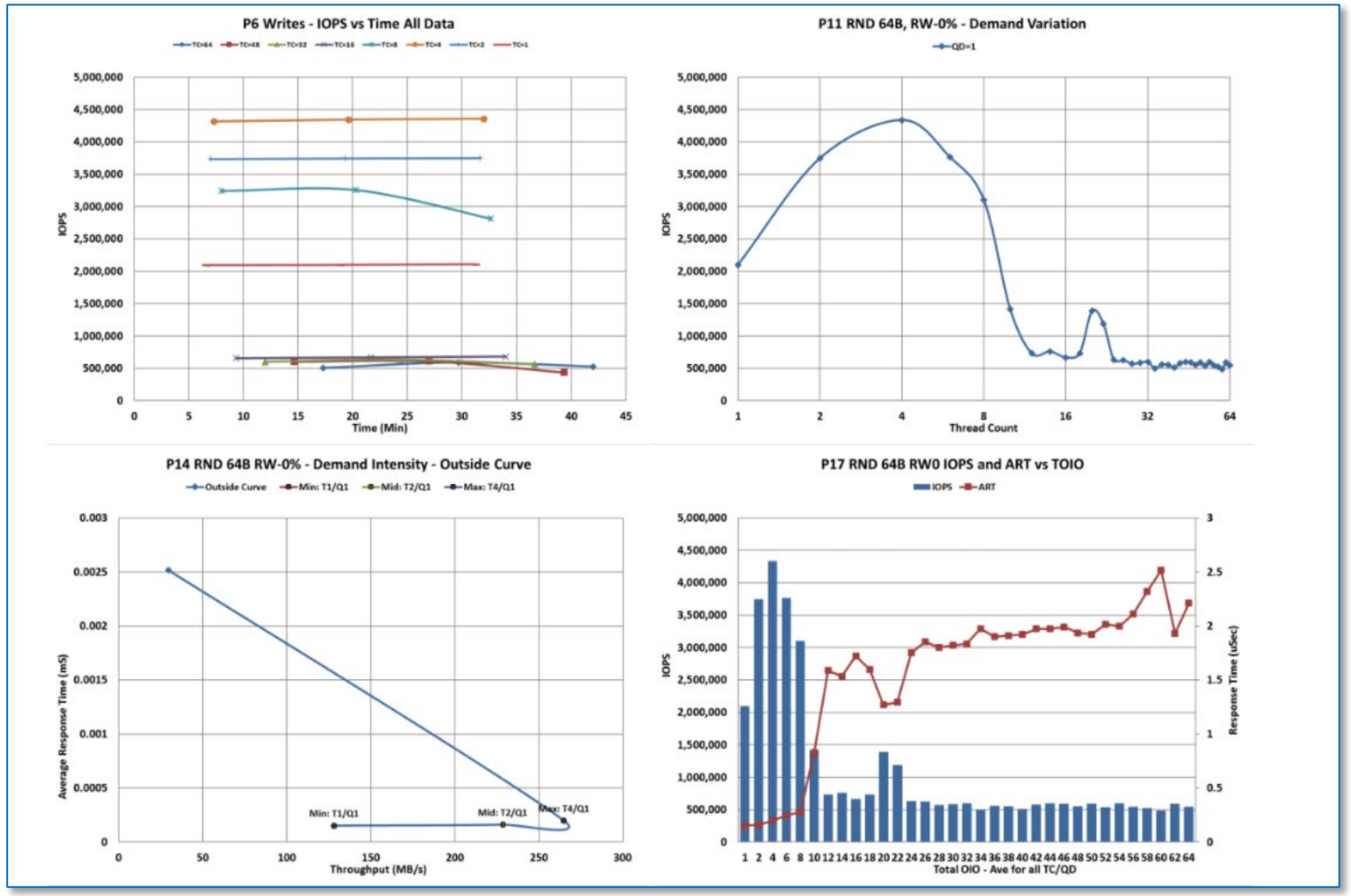
- Pre Writes 10 min
- TC loop [1,2,4…128]
- Run 5 loops

Results

- IOPS per TC Round
- IOPS per TC average
- Demand Intensity Curve
- IOPS & ART v Total OIO

Purpose

- Observer TC, IO & RT saturation
- Select optimal TC performance
- Observe TC Repeatability



Note: For RND 64-byte Writes, P6 shows IOPS over time for each Thread Count. P11 shows total IOPS for each TC. P14 shows Demand Intensity Outside Curve. P17 shows IOPS and Average Response Time vs Total Thread Count.

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

12. Sample Data: Individual Streams Test

Memcopy OSSync: Retail Web Portal - 9 IO Stream

Draft PM PTS v1.0:
Preliminary & Subject to Change

Test Flow

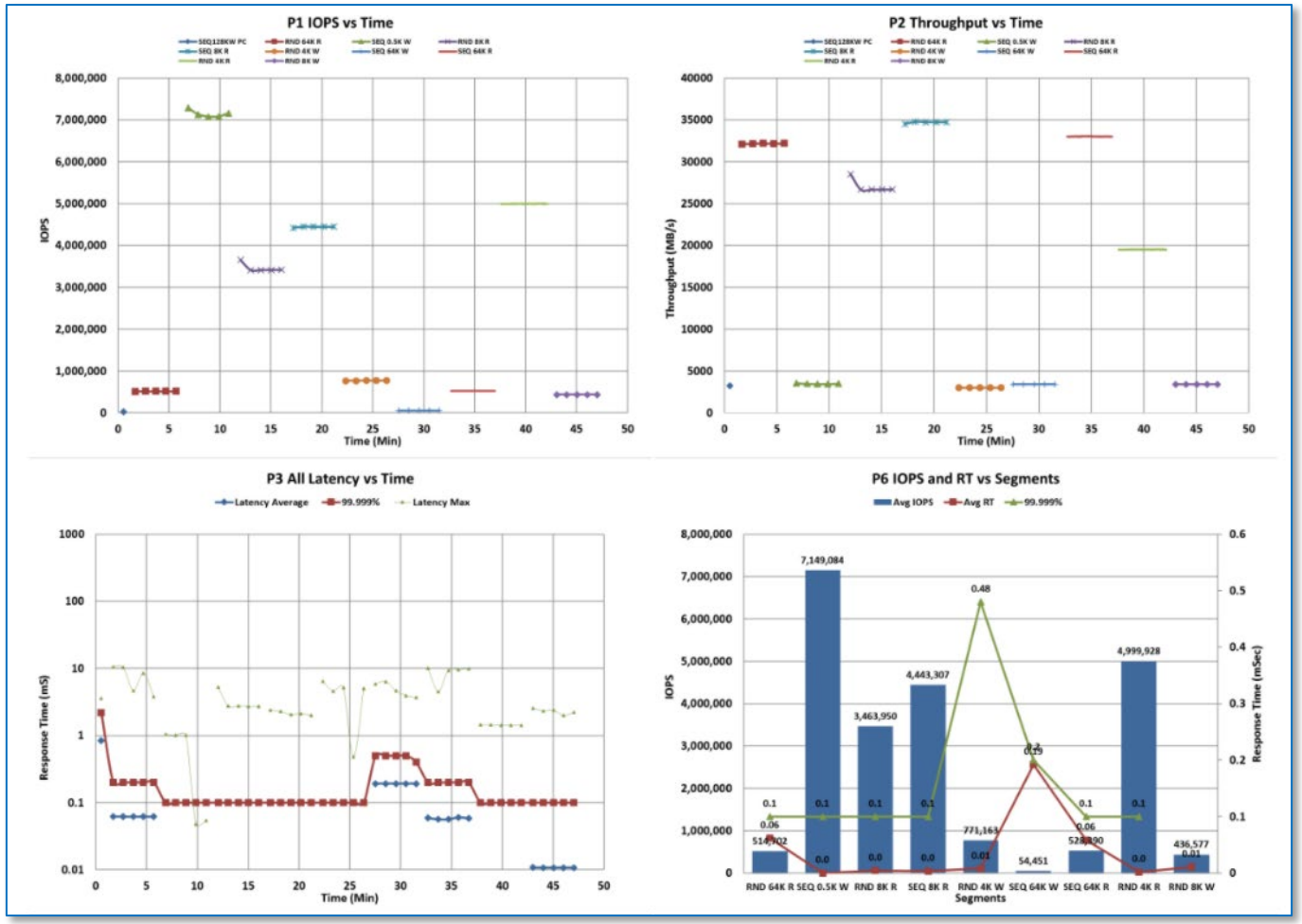
- Each IO Stream to SS
- Run IO Streams Sequentially

Results

- IOPS, TP & LAT for each IO Stream
- IOPS, ART & 5 9s for each IO Stream

Purpose

- Observe SS performance of each IO Stream in workload
- Compare IO Stream performance to mfgr specs



Σ	Cumulative Workload		
☑	RND 64K R	18.5%	842,361
☑	SEQ 0.5K W	17.0%	775,127
☑	RND 8K R	10.0%	456,175
☑	SEQ 8K R	8.4%	382,972
☑	RND 4K W	4.0%	182,251
☑	SEQ 64K W	3.7%	169,571
☑	SEQ 64K R	3.4%	155,798
☑	RND 4K R	2.92%	132,777
☑	RND 8K W	2.78%	126,550
☐	SEQ 4K R	2.20%	100,309
☐	RND 16K R	1.85%	84,245
☐	RND 32K R	1.55%	70,613

Total IOs of 5,086 streams: 4,551,062
RW mix: 66% R : 34% W
Selected 9 streams: 3,223,582 (71%) E

Note: The 9 IO Streams for the Retail Web Portal workload are shown in the Cumulative Workload Description box. P1 shows IOPS v Time. P2 shows Throughput (Bandwidth) v Time. P3 shows Latency v Time. P6 shows Average IOPS and Response Time for each IO Stream.

© 2021 SNIA Persistent Memory + Computation Storage Summit. All Rights Reserved.

Conclusion & Call to Action

- This PTS is an extension of previous PTS for NAND Flash SSDs and Real-World workloads
- The PM PTS is designed to allow repeatable, consistent performance comparisons
- The PM PTS is intended to allow application & storage engineers to understand & implement PM
- Reference Test Platforms, OS, BIOS & Drivers are commercially available servers
- Additional information is available in the PM PTS White Paper of Oct. 2020

Call to Action:

1. Join SNIA CMSI and/or the SSS TWG
2. Submit your Real-World Workload captures to SNIA SSS TWG
3. Comment on PM PTS when published for public review at snia.org/feedback

© 2021 SNIA Persistent Memory + Computational Storage Summit. All Rights Reserved.

Thank you

Please visit www.snia.org/pmsummit for presentations