SNIA Emerald™ Power Efficiency Measurement Specification for Data Center Storage Devices

Version 0.0.36

ABSTRACT*:* This document describes a standardized method to assess the power efficiency of storage devices used in data center applications in both active and idle states of operation. A taxonomy is defined that classifies storage devices in terms of operational profiles and supported features. Test definition and execution rules for measuring the power efficiency of each taxonomy category are described; these include test sequence, test configuration, instrumentation, benchmark driver, IO profiles, measurement interval, and metric stability assessment. Resulting power efficiency metrics are defined as ratios of idle capacity or active operations during a selected stable measurement interval to the average measured power.

Publication of this Working Draft for review and comment has been approved by the Green Storage Technical Work Group. This draft represents a "best effort" attempt by the Green Storage Technical Work Group 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

August 19, 2024

USAGE

DISCLAIMER

Revision History

Revision	Date	Changes
V0.0.36	August 19, 2024	Working Draft approved by Technical Council for release for public review and periodic call for IP.

Draft	Date	Changes – working list of changes, delete before publication
0	August 10, 2021	Create initial draft based on SNIA template, ISO requirements, and Emerald V4.0.0 drafts.
1	August 11, 2021	Update title to SNIA Emerald™ Power Efficiency Measurement Specification for Enterprise Storage Devices. Update some formatting.
2	February 23, 2022	Add Introduction text reviewed during Green TWG call on 2/23/2023.
3	March 25, 2023	Start over, using entire text of SNIA Emerald V4.0.0 and applying TC-provided document template.
4	April 4, 2023`	Merge Chuck Paridon's deletions of sections 5.2,5.6,5.7,5.8,6,7.6-7.8,8.5-8.7,B.2,I, J. Basically, delete major sections on COM, near-online, tape libraries, and virtual libraries as not needed.
5	April 5, 2023	Update taxonomy tables based on TWG walkthrough on Apr 5, 2023.
0.1.8	September 16, 2023	Add Eden's work done into Emerald System spec into this document. Add Keith's Taxonomy tables and definitions. Do some, but barely begun, editing of format and style (including to ISO requirements).
0.1.9	September 28, 2023	Edit during Sep 20 TWG call. Then: Globally change most "product" to "device", Globally change "pre[-]conditioning" to "conditioning". Use "SNIA Emerald Device Power Efficiency Measurement" globally to refer to a test or (with appended test result) a test result. Update terminology style to ISO requirements. Add diagrams as requested during Sep 27 TWG call. Move clauses 12 and 13 to immediately follow Clause 8.
0.0.10	October 6, 2023	Edit during Oct 4 TWG call. Then: Add/ember new Figures 1-3 and 6 as provided by Eden Kim. Replace Table 2 with one provided by Keith Orsak. Do additional changes requested by Keith. Miscellaneous editorial work.
0.0.11		
0.0.12	October 31, 2023	Continue work. Embed additional figures. Revise taxonomy tables. Globally, use device rather than drive, except when specifically address HDD. Globally, use DUT for item under test. Propose Test Process Concepts language.
0.0.13	November 4, 2023	Edit during Nov 1 TWG call. Then: Revise taxonomy classification tables. Insert new version of Conditioning Run… figure. Insert new version of Figures 7 and 8. Partially make figure number consistent. Make text box describing test process into normal text. Add Eden's requested text changes and clause headings.
0.0.14	November 8, 2023	Edit during Nov 8 TWG call. Then: accept changes per TWG discussion plus those Dave does not see as needing TWG discussion. Change text in Figure 2 to use "to" for range rather than "-". Restore text for HDD classifications that had been deleted. Rework classification text and tables to use names rather than digit-letter designations. Review formatting and make some corrections.
0.0.15	November 16, 2023	Edit during Nov 15 TWG call. Then: Make agreed changes to nomenclature throughout taxonomy clause. Change consumer to client globally. Change mid-line to midline globally. Accept changes to taxonomy tables as agreed during TWG call. Update SFF and LFF definitions to refer to Bibliography entries.
0.0.16	November 21, 2023	Work on test process (Clauses 6 and 7) in small group. DWT does work requested by small group.

Draft	Date	Changes – working list of changes, delete before publication
0.0.17	December 5, 2023	Make taxonomy nomenclature uniform. Clean up formatting of taxonomy tables. Revise/add material in sections 7.2.5 through 14 as provided by Eden and edit the Chuck and Eden. More editing working on format, etc.
0.0.18	December 12, 2023	Work on spec in 12/6/2023 TWG call. Add lots of material from Eden and do additional editing with Eden Kim, as requested by TWG, and on style/formatting.
0.0.19	December 19, 2023	Work on spec in 12/13/2023 TWG call. Dave and Eden did much removal of unneeded material from Emerald V4.0.0 and reorganized many sections. Additional editing by Dave.
0.0.20	January 2, 2024	Added material provided by Eden, Dave and Eden reorganized material and made many changes. Document considered ready for TWG internal review.
0.0.21	January 30. 2024	Merge comments from internal review. Update front matter and footers based on new template from TC.
0.0.22	February 1, 2024	TWG reviewed some of the internal review feedback during 1/31/2024 call. Dave performed editing requested by TWG, including moving pseudo-code form Annex to be 7.5 and moving 8. to be Annex B.
0.0.23	February 9, 2024	Review and editing During 2/7/2024 TWG call. Dave: Integrate Eden Kim input received 2/6/2024. Edit as requested by TWG, including accepting movement of large blocks of text listed above, globally change Power Board Server to Power Analyzer, Andy Norrie proposed replacement for Figures 4 and 5 - Dave offers integrated version, and text from Eden on CTS Lite and test scripts.
0.0.24	February 23, 2024	Changes made by Eden Kim. Revised/edited by TWG on 2/21/2024. Dave edited as requested by TWG, including changing Multi-stream Workload to Complex Workload and revising format of 7.4 and 7.5.2.
0.0.25	March 11, 2024	Revised/edited by TWG on 3/6/2024. Dave edited as requested by TWG, including accepting revised Figure 4 and adjacent text and revising indentation of pseudo-code to indicate binding of OR operator.
0.0.26	March 23, 2024	Eden Kim made changes to Clause 9. Edited during 3/20/2024 TWG call. Dave edited as requested by TWG, including moving the Workload, Parameters, and Run-Time Rules clause to be a new Annex and reordering existing Annexes.
0.0.27	April 3, 2024	Edited during 4/3/2024 TWG call – reviewed Clauses 10, 11, part of 12.
0.0.28	April 24, 2024	Edited during 4/17/2024 TWG call – reviewed Annex A Report Header and Clause on reporting requirements. Dave began work in Annex A on a revised report header as requested – this is inserted as accepted at the beginning of Annex A before the pre-existing material. Reviewed Definitions and Abbreviations, primarily to identify unused or definitions inconsistent with uses.
0.0.29	May 10, 2024	Review/edit definitions during 5/1/2024 TWG call. Add material provided by Eden as a table in 12.2. Copy Clause 12 to newly added Annex B.
0.0.30	May 15, 2024	Review/edit definitions 3.2.10 through 3.2.36 during 5/15/2024 TWG call.
0.0.31	May 29, 2024	Review/edit definition following 3.2.26, symbols and abbreviations, 12.2, and other aspects of the document during 5/29/2024 TWG call. Post call, Dave: 1) review all uses of Test Software and workload generator for consistency with definitions; 2) globally change test operator to test sponsor; 3) Move Environmental and Temperature Clauses; and sort definitions and symbols.
0.0.32	June 21, 2024	Edit Table 10 based on input from Eden Kim. Add Contributors list based on TWG discussion of June 19, 2024. Use “CTSlite” globally (except for the sections required reporting that are still under consideration for deletion) as the name of the test software.
0.0.33	June 26, 2024	Edit during June 26, 2024 TWG call, including changing title and removing test for client devices. After call, edit as requested by TWG, including working on the text and format of Annex A, now that this Annex is Informative.
0.0.34	July 17, 2024	At Eden’s request, make changes so that Pre-Fill is done for HDD the same as for SSD. Dave and Eden did some review and cleanup; globally changed Ready Idle test time from 10 minutes to 30 minutes. Edit during July 17, 2024 TWG call. After call, edit as requested by TWG.

Draft	Date	Changes – working list of changes, delete before publication
0.0.35	July 31, 2024	Update Clause 12 General Environmental… as proposed by Keith Orsak. Change “CTSlite” to “CTS Lite” globally as requested by Eden Kim. Review/edit during July 31, 2024 TWG call.
0.0.36	August 14, 2024	Review/edit during August 14, 2024 TWG call. After the call, Dave performed edits requested by TWG and minor editorial cleanup (there remain unaddressed issues with equation numbering, figuring/table numbering in Annexes, and probably other areas) to create draft to send to TC requesting Public Review and Call for IP.
0.0.36	August 20, 2024	Format as Working Draft for release for public review and for periodic call for IP; as approved by SNIA Technical Council on August 19, 2024.

Suggestions for revisions should be directed to https://www.snia.org/feedback/.

CONTACTING SNIA

SNIA Emerald™ Program Web Site

The SNIA Emerald™ Program web site is https://www.snia.org/emerald. SNIA Emerald™ Program-related downloads are available at https://www.snia.org/emerald/download, the *Download Materials for SNIA Emerald™ Testing* web page.

SNIA Address

Requests for interpretation, suggestions for improvement and addenda, or defect reports are welcome. They should be sent via the SNIA Feedback Portal at http://www.snia.org/feedback or by mail to SNIA, 5201 Great America Parkway, Suite 320, Santa Clara, CA, 95054 U.S.A.

Acknowledgements

The SNIA Green Storage Technical Working Group, which developed and reviewed this document, recognizes the significant contributions made by the following individuals and companies:

Contributors

Individuals

Wayne Adams
Bill Bailey
Don Capps
Jeffrey Doolittle
John Geldman
Donald Goddard
Michael Gregoire
Patrick Hery
Eden Kim
Tai Lee
Chris Lueth
Andrew Norrie
Keith Orsak
Chuck Paridon
Carlos Pratt
David Thiel

Companies and Organizations

Calypso Systems, Inc.
Dave Thiel, Consultant LLC
Dell Inc.
Ellisys Group SA (SANBlaze / SerialTek / Quarch)
Hewlett Packard Enterprise
Kioxia Corporation
NetApp
Seagate Technology
Solidigm
Toshiba America Electronic Components, Inc.
Ultrastratech

Contents

List of Tables, Figures and Equations

Introduction

The rapid growth in data storage and management has led to the development of more advanced storage devices, which are designed to meet the demands for increased capacity, performance, and reliability. However, as storage devices continue to evolve, the need for optimizing their power efficiency has become increasingly important.

To address this challenge, it is necessary to have a comprehensive and standardized method for measuring the power efficiency of storage devices at the individual device level. This document provides a detailed description of the procedures, protocols, and metrics that are used to measure the power efficiency of individual storage devices.

This document also outlines the requirements for conducting accurate and consistent power efficiency measurements, including the test environment, equipment, and data collection and reduction methods. This document serves as a reference for manufacturers, system integrators, regulatory bodies, test organizations and information technology professionals who are interested in evaluating the power efficiency of individual storage devices.

The methodologies of this document are based on those of

- International Standard ISO/IEC 24091:2019 *Information technology — Power efficiency measurement specification for data center storage*[7] and
- SNIA Standard *SNIA Emerald™ Power Efficiency Measurement Specification* Version 4.0.0[8]

with modifications that include a self-optimizing procedure and Principles of Demand Intensity and performance evolution based on

- SNIA Standard *Solid State Storage (SSS) Performance Test Specification (PTS)* Version 2.0.2[5] and
- SNIA Standard *Real World Storage Workload (RWSW) Performance Test Specification for Datacenter Storage* Version 1.0[6].

1 Scope

1.1 Introduction

This document defines a standardized method to assess the energy efficiency of individual storage devices in both active and idle states of operation. A taxonomy is defined that classifies storage devices in terms of operational profiles and supported features. Test definition and execution rules for measuring the power efficiency are described; these include test sequence, test configuration, instrumentation, benchmark driver, IO profiles, measurement interval, and metric steady state assessment. Resulting power efficiency metrics are defined as ratios of idle capacity or active operations during a selected steady state measurement interval to the average measured power.

An individual storage device is broadly defined as a storage product that is contained in a standard device form factor, connects to a standard computer bus, supports a recognized storage device protocol, and presents storage to a host server as a single block device.

A storage device is said to be in an “active” state when the storage device is processing externally initiated requests for data transfer. A storage device is said to be in “steady state” when the workload-dependent performance is relatively time-invariant. A storage state is said be in a “ready idle” state when it is not receiving host IO commands.

This document includes:

- A generalized taxonomy for storage devices (clause 5) that addresses solid state and magnetic disk storage;
- Measurement and data collection guidelines for assessing the power efficiency of storage devices in both active and idle states (clause 7);
- Metrics describing storage device power efficiency (clause 8); and
- Required disclosures for a test result published as a SNIA Emerald™ Device Power Efficiency Measurement[1] test result (clause 13).

This document is intended for use by individuals and companies engaged in assessing the power utilization of storage devices.

1.2 Purpose of a SNIA Emerald™ Device Power Efficiency Measurement Test Result

The purpose of a SNIA Emerald™ Device Power Efficiency Measurement test result is to provide a reproducible and standardized assessment of the energy efficiency of storage devices used in data center applications in both active and ready idle states. Disclaimer

A SNIA Emerald™ Device Power Efficiency Measurement test result provides a high-level assessment of the power efficiency of the tested storage device in specific idle and active states. Actual performance and power consumption is highly dependent upon workload, environmental and usage parameters.

[1] SNIA Emerald™ is a trademark of SNIA.

2 Normative References

The following documents are referred to in the text in a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

Table 1 lists these documents.

Table 1 – Normative References

Author/Owner	Title	Revision	URL
ISO/IEC	*ISO/IEC Directives Part II*	Ninth edition, 2021	https://www.iso.org/directives-and-policies.html

3 Definitions, Symbols, Abbreviations, and Conventions

3.1 Overview

For the purposes of this document, the terms and definitions given in *The SNIA Dictionary*[9] and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

• ISO Online browsing platform: available at https://www.iso.org/obp

• IEC Electropedia: available at http://www.electropedia.org/

The terms and definitions defined in this document are based on those found in *The SNIA Dictionary*[9]. They have been extended, as needed, for use in this document. In cases where the current definitions in the SNIA dictionary conflict with those presented in this document, the definitions in this document shall apply.

3.2 Definitions

3.2.1

ActiveRange

range of LBA's that may be accessed by the test code

Note: ActiveRange is specified as ActiveRange(start:end), where "start" and "end" are percentages, where the starting LBA# = start%*MaxUserLBA and the ending LBA# = end%*MaxUserLBA.

3.2.2

ART Ceiling

maximum allowable ART

3.2.3

cache

temporary data storage, not directly addressable by end-user applications, used to store data for expedited access to or from slower media

3.2.4

Complex Workload

a mix of different IO sizes, read/write mixes, and access patterns with a skewed access across a range of blocks

3.2.5

Conditioning

process of writing data to the device to prepare it for Steady State measurement

3.2.6

device under test

storage device that is undergoing test

3.2.7

DI Curve

Demand Intensity Curve in total OIO versus ART & IOP

3.2.8

EDSFF

industry standard Enterprise and Data Center Standard Form Factor device defined by SNIA Standard SFF-TA-1009[1]

3.2.9

Fresh Out-of-the-Box (FOB)

state of device prior to being put into service for the first time

3.2.10

latency

time interval between when a workload generator makes an IO request and when it receives notification of the request's completion

3.2.11

LFF

industry standard 3.5 inch Large Form Factor device defined by SNIA Standard SFF-8301[1].

3.2.12

Logical Block Address (LBA)

address of a logical block

3.2.13

Measurement Window

interval, measured in Rounds, during which test data is collected

Note: Measurement Window is bounded by the Round in which the device has been observed to have maintained Steady State for the specified number of Rounds (Round x), and five Rounds previous (Round x-4).

3.2.14

non-volatile storage

storage that retains data in the event of the unexpected loss of power

3.2.15

OIO/Thread

number of OIO per Thread

3.2.16

outstanding IO (OIO)

IO operations issued by a workload generator and awaiting completion

3.2.17

power efficiency

ratio of useful work to the power required to do the work

3.2.18

Pre-Fill

test step that writes data to a DUT to prepare that DUT for subsequent Conditioning and test steps

3.2.19

Purge

process of returning a storage 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

3.2.20

ready idle

operational state of no host IO commands in which a device is capable of satisfying an arbitrary IO request

3.2.21

Round

one complete pass through all the prescribed test intervals for any given steady state calculation

3.2.22

sequential read

IO load consisting of consecutively issued read requests to logically adjacent data

3.2.23

sequential write

IO load consisting of consecutively issued write requests to logically adjacent data

3.2.24

SFF

industry standard 2.5 inch Small Form Factor device defined by SNIA Standard SFF-8201[2]

3.2.25

SNIA Emerald™ Device Power Efficiency Measurement

test performed according to this document

3.2.26

SNIA Emerald™ Device Power Efficiency Measurement test result

result of a test performed according to this document

3.2.27

Solid State Drive (SSD)

storage capability built from non-volatile solid state storage

3.2.28

stable storage
storage that maintains data across power failures or other transient failures

3.2.29

Steady State

state where the workload-dependent performance becomes relatively time invariant

3.2.30

storage device

solid state drive or hard disk drive

3.2.31

Test Software

software used to execute the workloads, test steps, power aggregation, data reduction, and reporting

3.2.32

test sponsor

operator, company, or agent that performs a test according to this document

3.2.33

Thread Count (TC)

number of Threads specified by a workload generator

3.2.34

Total OIO (TOIO)

number of outstanding IO Operations

Note: Total OIO is defined as (Thread Count) * (Queue Depth)

3.2.35

U.2

industry standard 2.5 inch Small Form Factor device defined by SNIA Standard SFF-8639[4]

3.2.36

user capacity

LBA range directly accessible by a workload generator

3.2.37

workload generator

component of the Test Software that creates the command sequences applied to the DUT

3.3 Symbols and Abbreviated Terms

ART	average response time
CR	Conditioning Run
CTS Lite	Test Software to be used with this document
DI	Demand Intensity
DUT	Device Under Test
FOB	Fresh-Out-of-the-Box
HDD	Hard Disk Drive
IOPS	I/O Operations per Second
IOPS/W	IOPS per watt
LBA	logical block address
MaxTTFD	Maximum Time to First Data
MiB/s	mebibytes per second
MiB/s/W	mebibytes per second per watt
ms	millisecond
OIO	outstanding IO
QD	queue depth
R/W	read/write
SCSI	Small Computer System Interface
SSD	solid state drive
TC	Thread Count
TOIO	total outstanding IO
TP	throughput
USB	Universal Serial Bus

3.4 Expression of Provisions

This document uses the verbal forms for expressions of provisions as defined by ISO/IEC Directives Part 2 (Ninth edition, 2021). These verbal forms include shall, shall not, should, should not, may, may not, can, cannot, and must.

3.5 Storage Capacity Conventions

The associated units and abbreviations used in this document are shown in Table 2.

Table 2 – Storage Capacity Conventions

Storage Capacity Conventions					
Decimal			**Binary**		
Name	**Abbreviation**	**Value**	**Name**	**Abbreviation**	**Value**
kilobyte	kB	10^3 B	kibibyte	KiB	2^{10} B
megabyte	MB	10^6 B	mebibyte	MiB	2^{20} B
gigabyte	GB	10^9 B	gibibyte	GiB	2^{30} B
terabyte	TB	10^{12} B	tebibyte	TiB	2^{40} B

In this document

- Storage capacities are represented in base-10.
- IO transfer sizes and offsets are represented in base-2.
- IO throughput is reported in base-2.

4 Taxonomy

4.1 Introduction

This clause defines a market taxonomy that classifies block access individual storage devices in terms of operational profile and supported features. The purpose of the taxonomy is to define specific groups into which storage devices can be placed for test comparison.

Individual storage devices taxonomy classes are defined broadly for hard disk drive (HDD) and solid state storage device (SSD) products. Within each class, additional categories are provided for client and enterprise class devices across SATA, SAS, or NVMe protocols, that are housed in 2.5 inch, 3.5 inch, EDSFF, or M.2 standard storage device form factors.

While this taxonomy is broad and defines a framework for devices that range from consumer solutions to enterprise installations, it is not intended to address all storage device types. For example, SSD storage devices in this specification are limited to NAND Flash based media and specifically exclude Persistent Memory devices.

The taxonomy is structured as a 2-level hierarchy of Category and Classification. A Category is a broad grouping of devices such as HDD and SSD. A Classification is a level of device sophistication, size, and complexity that corresponds to market delineations. The Classifications distinguish performance, power utilization, and life-span.

The taxonomy addresses a broader range of storage devices than is covered by the power efficiency tests defined in this document.

4.2 Taxonomy Categories

4.2.1 General

Taxonomy categories define broad market segments that can be used to group devices that share common functionality or performance requirements, and within which meaningful device comparisons can be undertaken. This document defines two broad taxonomy categories (SSD and HDD) summarized in Figure 1.

Each of the categories contains additional classifications. It is intended that storage devices tested to this specification shall only be compared to storage devices that share the same category and classification.

While HDD and SSD categories have unique differentiators, devices satisfying either category can be used in similar applications. The relative performance of HDD and SSD storage devices are shown in the bubble chart in Figure 1.

Each class covers a range of sustained IOPS and average Latency performance. These are general estimates of relative class performance.

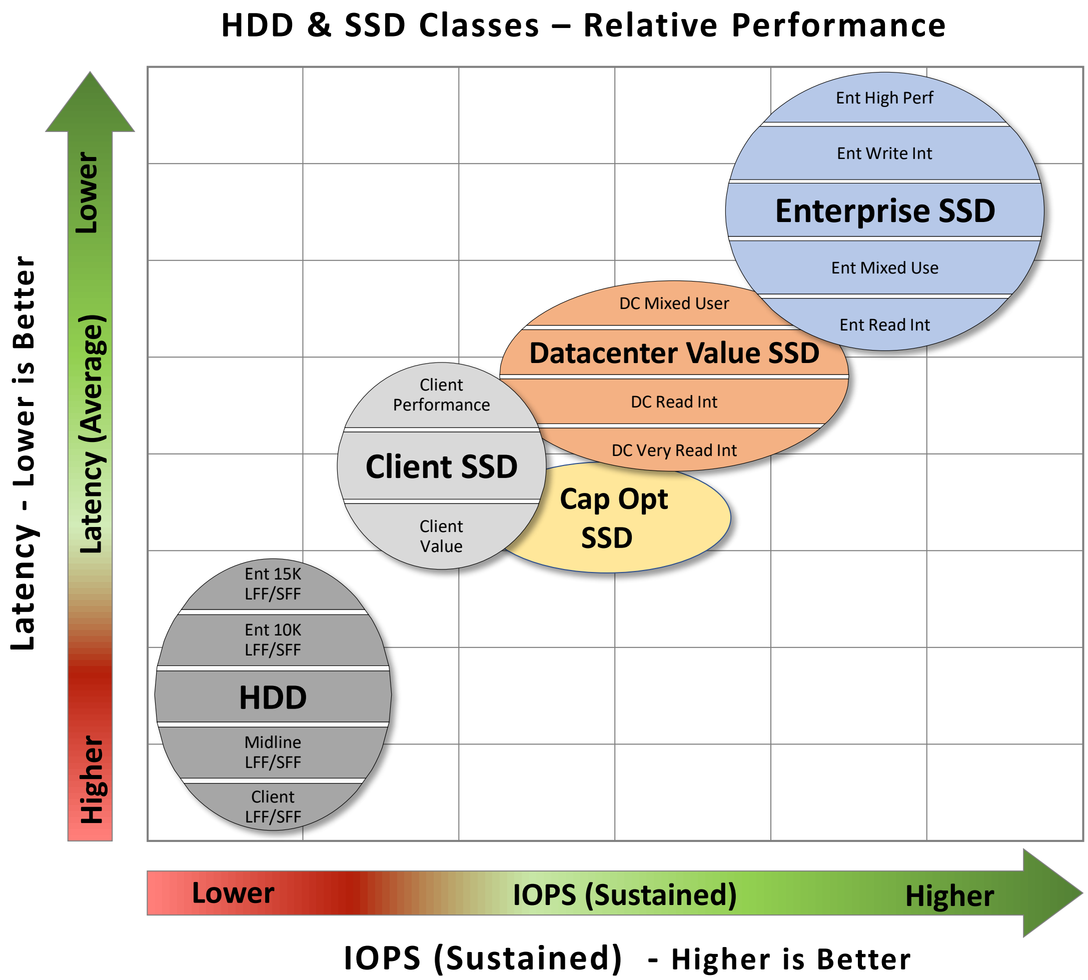

Figure 1 – Taxonomy Classes - Relative Performance

Taxonomy charts specific to HDD and SSD classes are listed in Figure 2 and Figure 3 and show relative power, life cycle, and performance.

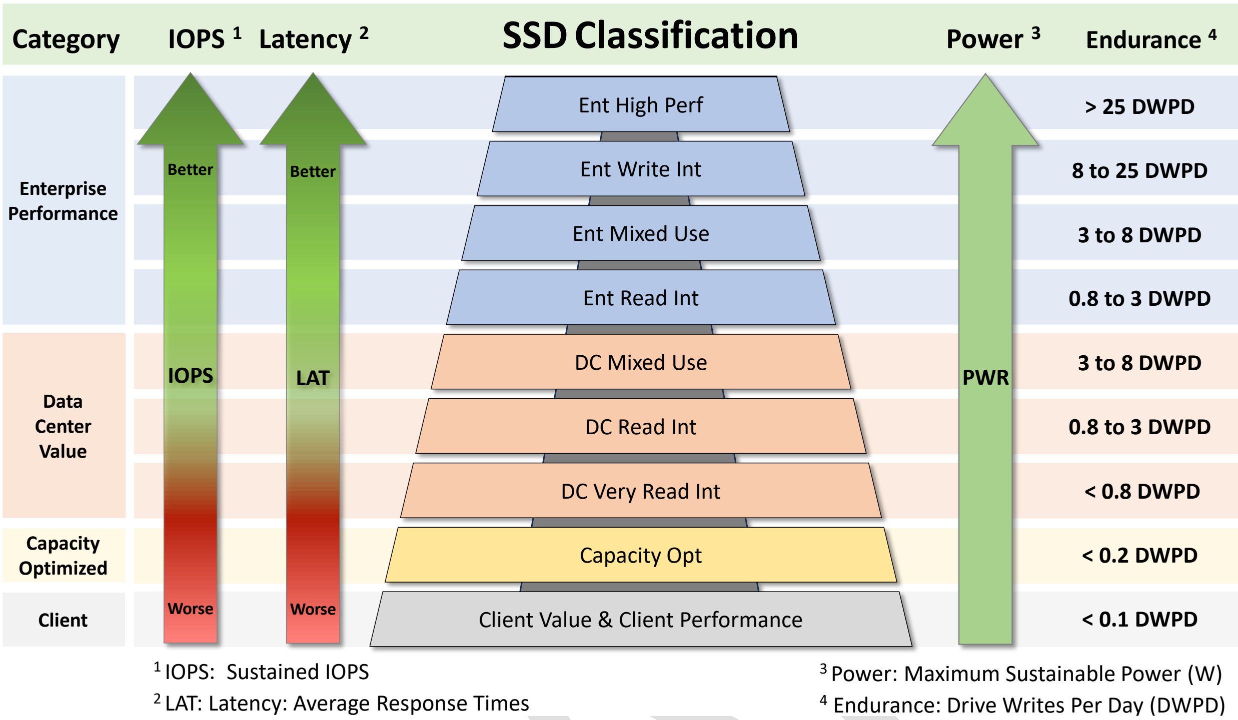

Figure 2 – SSD Taxonomy

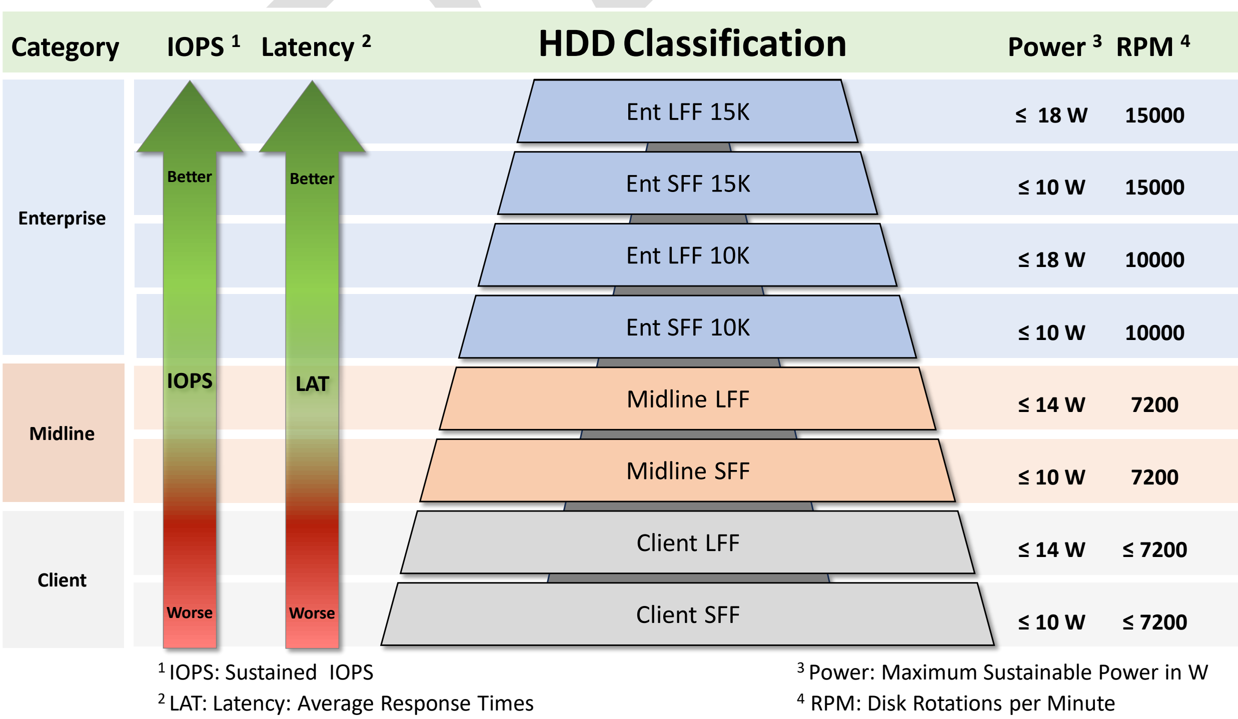

Figure 3 – HDD Taxonomy

4.3 Taxonomy Classifications

Classifications define combinations of settings or values for the attributes within a Category.

Taxonomy Classifications discriminate between different devices. The goal of the Classifications is to differentiate between devices of differing form factor, functionality, and target market.

4.3.1 SSD Category

4.3.1.1 Client SSD Classifications

Client devices are any of a wide variety of SSDs which are manufactured primarily for personal, family, household, or small-business purposes ranging from personal storage, personal computing, and boot to some Artificial Intelligence (AI) and Machine Learning (ML) applications. These devices typically have lower performance and endurance capabilities than those offered in Capacity Optimized, Data Center Value, or Enterprise Performance Classification SSDs.

a) **Client Value** Classification devices consist of lower capacity and lower performance client SSDs.

b) **Client Performance** Classification devices consist of higher capacity and higher performing client SSDs.

4.3.1.2 Capacity Optimized SSD Classification

Capacity Opt Classification devices are SSDs with lower performance and price than Enterprise Performance Classification devices. Capacity OPT devices are optimized for read intensive and large transfer applications where cost of capacity is critical. Common applications include AI/ML, Deep Learning, data streaming, and data analytics.

4.3.1.3 Data Center Value SSD Classifications

Data Center Value devices are any of a wide variety of cost optimized SSDs that are optimized for read intensive workloads. Common applications include Hyperscale, IoT, big data analytics, OLTP, and streaming media. These devices typically have lower performance and endurance capabilities than those offered by Enterprise Performance Classification devices.

a) **DC Very Read Int** Classification devices are Data Center Value SSDs that are optimized for capacity and very read intensive workloads at the expense of performance and endurance.

b) **DC Read Int** Classification devices are Data Center Value SSDs that are optimized for capacity and read intensive purposes with higher performance and endurance capabilities than DC Very Read Int Classification devices.

c) **DC Mixed Use** Classification devices are Data Center Value SSDs that are optimized for higher performance and endurance purposes than other Data Center Value SSD Classification devices.

4.3.1.4 Enterprise Performance SSD Classifications

Enterprise Performance devices are any of a wide variety of performance and endurance optimized SSDs which are manufactured primarily for enterprise applications such as data streaming, web servers, analytics, High Performance Computing, caching and logging. These devices typically have higher throughput, IOPS, and lower latency performance and endurance capabilities than those offered in the Data Center Value Classifications.

a) **Ent Read Int** Classification devices are Enterprise Performance SSDs that are optimized for capacity and read intensive workloads at the expense of performance and endurance.

b) **Ent Mixed Use** Classification devices are Enterprise Performance SSDs that are performance and endurance optimized over Read Intensive Classification devices.

c) **Ent Write Int** Classification devices are Enterprise Performance SSDs that are optimized for high performance and endurance optimized for applications such as Analytics, High Performance Computing, and server storage.

d) **Ent High Perf** Classification devices are Enterprise Performance SSDs that are manufactured for Enterprise purposes requiring the highest endurance and performance capabilities.

4.3.2 HDD Category

4.3.2.1 Client HDD Classifications

Client devices are any of a variety of HDDs which are manufactured primarily for personal, family, household, or small-business purposes such as PC, video, audio, and gaming.

a) **Client SFF** Classification devices are SFF client HDDs.

b) **Client LFF** Classification devices are LFF client HDDs.

4.3.2.2 Midline HDD Classifications

Midline devices are any of a variety of HDDs which are manufactured primarily for business-critical applications such as backup, archive, and video surveillance.

a) **Midline SFF** Classification devices are SFF midline HDDs.

b) **Midline LFF** Classification devices are LFF midline HDDs.

4.3.2.3 Enterprise HDD Classifications

Enterprise devices are any of a variety of HDDs which are manufactured primarily for enterprise applications such as transaction processing and high-performance computing.

a) **Ent SFF 10K** Classification devices are SFF 10000 RPM enterprise HDDs.

b) **Ent LFF 10K** Classification devices are LFF 10000 RPM enterprise HDDs.

c) **Ent SFF 15K** Classification devices are SFF 15000 RPM enterprise HDDs.

d) **Ent LFF 15K** Classification devices are LFF 15000 RPM enterprise HDDs.

4.4 Taxonomy Details

4.4.1 SSD Classifications

Table 3 and Table 4 define the SSD taxonomy classifications and their attributes.

Table 3 – SSD Category, Client Classifications

Attribute	Category: SSD	
	Classification	
	Client	
	Client Value[a]	Client Performance
Form Factor	M.2, U.2, EDSFF, 2.5 inch	M.2, U.2, EDSFF, 2.5 inch
Endurance (DWPD)	< 0.1	< 0.1
Max Sustainable Power (W)	9 to 25	9 to 25
Read Response Time[b] (µs)	< 130	< 130
Write Response Time[c] (µs)	< 50	< 50
Multiple Power States	Low power	Low power
Media Type	NAND Flash	NAND Flash
Access Paradigm	Block	Block
Protocol	USB, SATA, NVMe	SATA, SAS, NVMe
Power Loss Protection[d] Required	No	No
Persistence[e] Required	No	No
Hot Plug Required	No	No

[a] This document specifies a test procedure for Client Classification devices used in data center applications.

[b] Read Response Time is the average response time for a random 4 KiB read, with a queue depth of 1.

[c] Write Response Time is the average response time for a random 4 KiB write, with a queue depth of 1.

[d] Power Loss Protection is a device technology consisting of hardware and/or software to prevent data loss on data residing in a volatile location in the event of an unexpected power loss scenario.

[e] Persistence is the device property that writes are reported as complete only after data will be retained by the device in the event of an unexpected power loss.

Table 4 – SSD Category, Data Center Classifications

Attribute	Category: SSD							
	Classification							
	Capacity Optimized	Data Center Value			Enterprise Performance			
	Capacity Opt	DC Very Read Int	DC Read Int	DC Mixed Use	Ent Read Int	Ent Mixed Use	Ent Write Int	Ent High Perf
Form Factor	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF	U.2, EDSFF
Endurance (DWPD)	< 0.2	< 0.8	0.8 to 3	3 to 8	0.8 to 3	3 to 8	8 to 25	> 25
Max Sustainable Power (W)	9 to 25	9 to 25	9 to 25	9 to 25	25 to 40	25 to 40	25 to 40	25 to 40
Read Response Time[a] (µs)	< 130	< 110	< 110	< 110	< 90	< 90	< 90	< 90
Write Response Time[b] (µs)	< 50	< 30	< 30	< 30	< 20	< 20	< 20	< 20
Multiple Power States	Low power	Low power	Low power	Low power	Low to high power	Low to high power	Low to high power	Low to high power
Media Type	NAND Flash	NAND Flash	NAND Flash	NAND Flash	NAND Flash	NAND Flash	NAND Flash	NAND Flash
Access Paradigm	Block	Block	Block	Block	Block	Block	Block	Block
Protocol	SATA, SAS, NVMe	SATA, SAS, NVMe	SATA, SAS, NVMe	SATA, SAS, NVMe	SAS, NVMe	SAS, NVMe	SAS, NVMe	SAS, NVMe
Power Loss Protection[c] Required	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Persistence[d] Required	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hot Plug Required	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a] Read Response Time is the average response time for a random 4 KiB read, with a queue depth of 1.

[b] Write Response Time is the average response time for a random 4 KiB write, with a queue depth of 1.

[c] Power Loss Protection is a device technology consisting of hardware and/or software to prevent data loss on data residing in a volatile location in the event of an unexpected power loss scenario.

[d] Persistence is the device property that writes are reported as complete only after data will be retained by the device in the event of an unexpected power loss.

4.4.2 HDD Classifications

Table 5 defines the HDD taxonomy classifications.

Table 5 – HDD Classifications

Attribute	Category: HDD							
	Classification							
	Client[e]		Midline		Enterprise			
	Client SFF	Client LFF	Midline SFF	Midline LFF	Ent SFF 10K	Ent LFF 10K	Ent SFF 15K	Ent LFF 15K
Form Factor	SFF	LFF	SFF	LFF	SFF	LFF	SFF	LFF
Workload Rate[f] (TB / year)	< 500	< 500	≥ 500	≥ 500	> 500	> 500	> 500	> 500
Rotational Rate (RPM)	≤ 7200	≤ 7200	< 10000	< 10000	≥ 10000	≥ 10000	15000	15000
Max Sustainable Power (W)	≤ 10	≤ 14	≤ 10	≤ 14	≤ 10	≤ 18	≤ 10	≤ 18
Power Loss Protection[b] Required	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Hot Plug Required	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Persistence[c] Required	No	No	Yes	Yes	Yes	Yes	Yes	Yes

[e] This document specifies a test procedure for Client Classification devices used in data center applications.

[f] Workload rate (TB / year) is maximum workload that the device is designed to accommodate. One TB = 10^{12} bytes.

[g] Power Loss Protection is a device technology consisting of hardware and/or software to prevent data loss on data residing in a volatile location in the event of an unexpected power loss scenario.

[h] Persistence is the device property that writes are reported as complete only after data will be retained by the device in the event of an unexpected power loss.

4.5 Taxonomy Rules

To be considered to be in a particular category and classification, a device shall satisfy all the attributes for the designated category and designated classification, i.e., satisfy all of the attributes of a column in one of the taxonomy tables, i.e., Table 3, Table 4, or Table 5.

If a device satisfies the attributes of multiple classifications, the device may be considered to be in any or all of those classifications.

If a device does not satisfy all of the row attribute parameters listed in the relevant classification column listed in Table 3, Table 4, or Table 5, then the device shall be considered to be in the taxonomy category Other and taxonomy classification Other.

5 Test Definition and Execution Rules

5.1 Overview

A SNIA Emerald™ Device Power Efficiency Measurement shall be a good faith effort to accurately characterize the power requirements of the tested device. The precise configuration used in a SNIA Emerald™ Device Power Efficiency Measurement is left to the sponsor of a test. Any commercially released or soon to be released components may be used, and a focus on new or emerging components or technologies is encouraged.

This clause defines the test definition and execution rules that shall be used. Test specific parameter settings, reports, and other requirements are documented in other sections that follow.

5.2 Basic Test Flow

The Basic Test Flow defines test specific parameter settings, and reporting requirements.

1) **Establish Physical Configuration**
   a) Connect Power Analyzer to interposer bus or DUT power port.
2) **Map Device Under Test (DUT)**
   a) Identify and map the DUT and record storage server, software and DUT configuration information for later disclosure and reporting.
3) **Execute Test**
   a) Start the test.
   b) Wait for test completion.
4) **Report Specified Data**
   a) Report storage server and DUT information as specified in Clause 7 and Clause 9 .

5.3 General Requirements

5.3.1 Configuration Guidelines

Figure 4 shows a DUT connection via a direct connection or an interposer board to the device power port.

The sample configuration shown in Figure 4 is provided as a guideline. Test sponsors may modify the configuration to meet their particular needs and equipment, provided no other requirement of this document is violated.

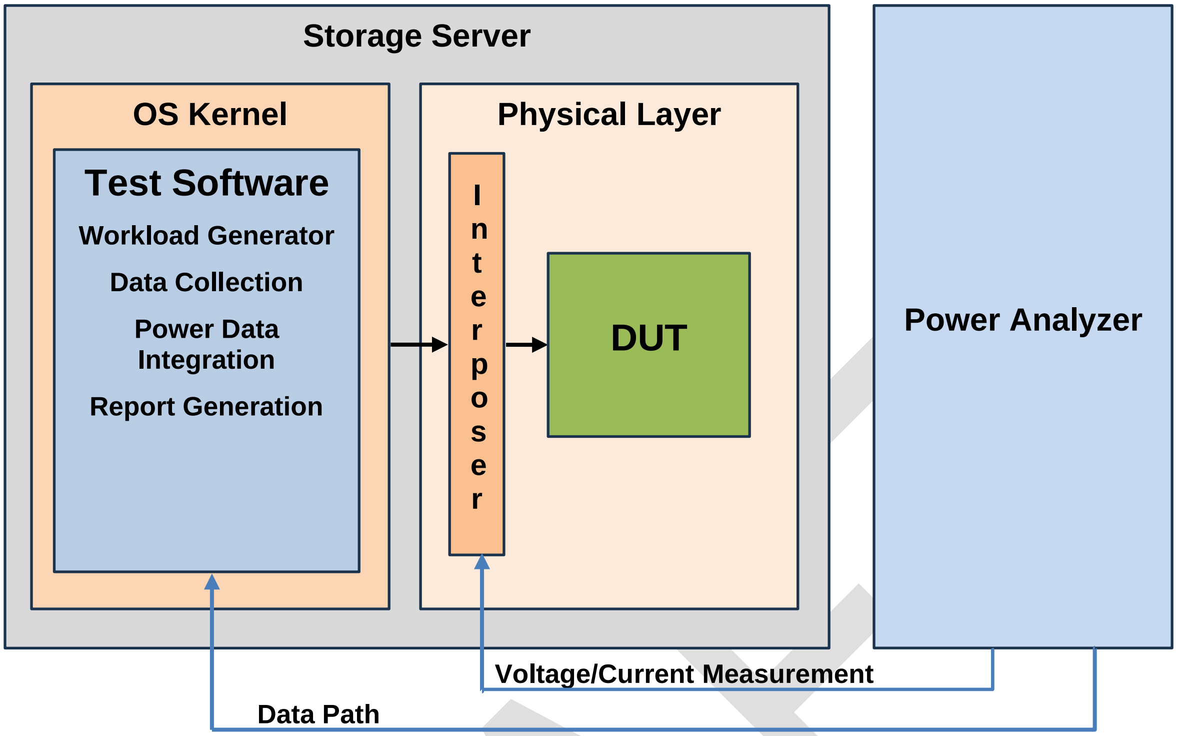

Figure 4 – Test Configuration

6 Test Process Concepts

6.1 Introduction

The performance of a storage device can be highly dependent on its prior usage, write history, caches, the pre-test state of the device and test parameters. This section defines key test process concepts. The next sections define Test Flow and Test Methodologies.

The test process defined by this document applies to storage devices used in data center applications.

6.2 Steady State

Storage devices that are Fresh Out-of-the-Box (FOB), or in an equivalent state, can exhibit an initial transient period of elevated performance, or peak performance.

For example, SSDs typically exhibit an initial peak performance, relative to the workload being applied, followed by a transition that evolves to Steady State.

The Steady State is a state where the workload-dependent performance becomes relatively time invariant.

It is important that the test data be gathered during a time window when the device is in Steady State (or during the Steady State Window), for two primary reasons:

1) To ensure that a device's initial performance (FOB or other peak) will not be reported as "typical", since this is transient behavior and not a meaningful indicator of the device's performance during the bulk of its operating life.
2) To enable test sponsors and reviewers to observe and understand trends. For example, oscillations around an average are "steady" in a sense, but might be a cause for concern.

Accordingly, the DUT shall achieve steady state wherein the linear least squares best-fit curve for five consecutive 30 minute Rounds does not exceed a ± 10% deviation from average IOPS of the five Rounds and the slope of linear least squares best-fit curve over the 5 Rounds does not exceed 10 %. This steady state determination procedure is referred to as the **5 Rounds Steady State Determination**.

The steady state procedure is contained within the Conditioning Run (CR) step of the test. The Conditioning Run Steady State report is automatically generated in the Test Audit File and reported as Plot P4.

Figure 5 shows Plot P4 with seven 30 minute Conditioning Run Rounds for an SSD DUT. The 5 Round Steady State Window occurs within the Conditioning Run and is shown occurring during Rounds 3 through 7. Round 1 shows IOPS at FOB peak performance. Rounds 1 and 2 show IOPS settling to a stable value. Conditioning Run Rounds 3 through 7 show the Steady State (SS) Window. Here, the SS Window occurs during SS Window Rounds 1 through 5.

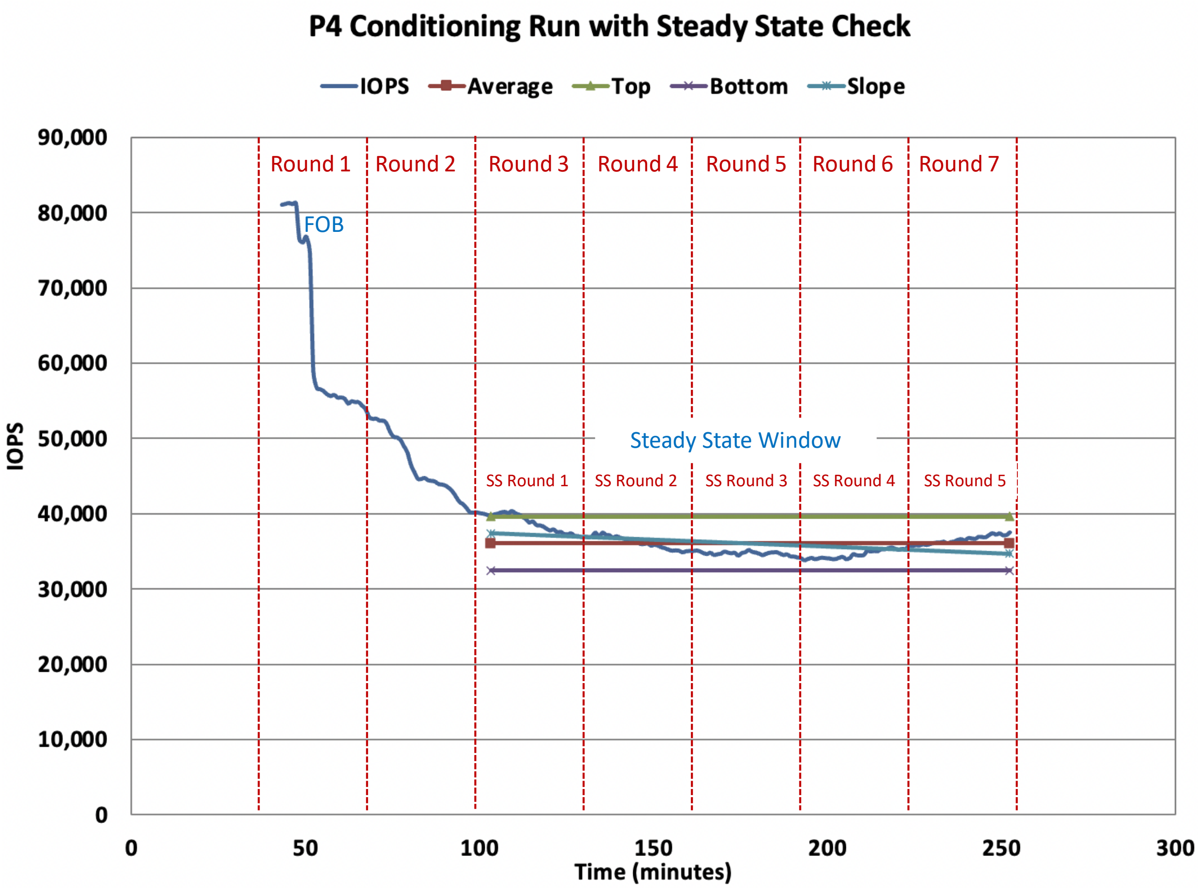

Figure 5 – Conditioning Run and Steady State Check

6.3 Purge

The purpose of Purge is to put the device in a consistent 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. The Purge step, if used, shall be executed prior to the Pre-Fill step to facilitate a clear demonstration of Steady State convergence behavior.

If a) the DUT does not support any kind of Purge method and b) the test sponsor chooses to continue the test, then the fact that Purge was not supported/run shall be documented in the test report.

The test sponsor may select any valid method of implementing the Purge step, including, but not limited to, the following:

a) ATA: SECURITY ERASE, SANITIZE DEVICE (BLOCK ERASE EXT)
b) SCSI: FORMAT UNIT
c) Vendor specific methods

The Test Software detects and applies Security Erase, Format Unit, or no Purge command, depending on the DUT self-reporting and this is noted in the Test Audit file test report headers. The test sponsor shall note and disclose the use of any other valid PURGE method employed.

6.4 Pre-Fill and Conditioning

6.4.1 Pre-Fill

The goal of Pre-Fill is to facilitate convergence to Steady State prior to running the measurement intervals of the Test. The Pre-Fill step is intended to prepare the DUT for subsequent steady state and test measurement steps.

6.4.2 Conditioning Run

The Conditioning Run is intended to apply the Complex Workload followed by a Steady State procedure where the DUT is run for a sequence of consecutive 30 minute Rounds during which five (5) consecutive Rounds show relatively time invariant performance.

6.4.3 Steady State

Steady State criteria are intended to put the DUT in a condition where performance is relatively time invariant so that subsequent measurements are consistent and repeatable.

6.5 Active Range

Active Range refers to the span of logical block addresses (LBA) that are accessed by the test code.

It is important to test the performance characteristics of devices using workloads that issue IO across 100% of the LBA space of the device.

6.6 Data Patterns

The workload generator uses data patterns to generate test code workloads. It is important to use data with known and specified data compression ratio because the compressibility of the data impacts performance.

6.7 Ready Idle

Ready Idle measurements shall be taken during a period of specified length when no host commands are issued to the DUT.

6.8 Caching

Memory caches in a data path significantly impact DUT performance. Therefore, cache settings are specified and reported.

7 Test Steps

7.1 Overview

7.1.1 DUT Configuration

The test sponsor shall identify one taxonomy set, category and classification for the DUT.

The DUT shall be configured to satisfy the requirements of the selected taxonomy set, category and classification.

The DUT shall represent a customer orderable configuration whose use within the selected one taxonomy set, category and classification is supported by the device vendor.

7.1.2 DUT Consistency

The physical and logical configuration of the DUT shall not be changed during a test unless explicitly allowed in the definition of a test step.

7.1.3 No Non-Test Activity

Other than booting/starting the device under test and any test equipment employed during the test, no substantive work shall be performed on the DUT between the test steps defined in this document, unless explicitly allowed in the definition of a test step.

7.1.4 Test Step Sequence

All test steps shall be executed as an uninterrupted sequence.

7.2 Single Continuous Test

The test steps shall be run as a single, continuous test with no time gaps or delays between the test steps. See Table 6.

Table 6 – Test Steps

Test Steps	
Test Step	**Description**
1. Setup Test Environment	Connect the DUT. Setup and start the Power Analyzer.
2. Purge	Apply the Purge, if supported, to the DUT.
3. Pre-Fill	Apply the SEQ 256 KiB Write Workload to the DUT for twice the user capacity. For HDDs, this test step is optional.
4. Conditioning Run with Steady State	Apply the Complex Workload for a minimum of 5 Rounds of 30 minutes each. Execute Steady State procedure at Round 1 during the Steady State Conditioning Run. If the DUT is not in steady state, perform additional Rounds (up to a total of 25) until Steady State is achieved.
5. Complex	Apply the Complex Workload for a warm-up interval of 10 minutes that includes a self-optimizing sweep of Total OIO followed by a measurement interval of 30 minutes.
6. RND 8 KiB Write	Apply the RND 8 KiB Write Workload for a warm-up interval of 10 minutes that includes a self-optimizing sweep of Total OIO followed by a measurement interval of 30 minutes.
7. RND 8 KiB Read	Apply the RND 8 KiB Read Workload for a warm-up interval of 10 minutes that includes a self-optimizing sweep of Total OIO followed by a measurement interval of 30 minutes.
8. SEQ 256 KiB Write	Apply the SEQ 256 KiB Write Workload for a warm-up interval of 10 minutes that includes a self-optimizing sweep of Total OIO followed by a measurement interval of 30 minutes.
9. SEQ 256 KiB Read	Apply the SEQ 256 KiB Read Workload for a warm-up interval of 10 minutes that includes a self-optimizing sweep of Total OIO followed by a measurement interval of 30 minutes.
10. Ready Idle	Apply Ready Idle for an interval of 30 minutes and report test measurements.
11. Post Process and Report	Import the power data, post process the power data, and generate the report.

7.3 Self-Optimizing TOIO Sweep

For the Complex, RND 8 KiB Write, RND 8 KiB Read, SEQ 256 KiB Write, and SEQ 256 KiB Read Test Steps, a self-optimizing TOIO sweep is executed during each test step warm-up interval. The self-optimizing TOIO sweep determines the TOIO setting to be used during the corresponding test step measurement interval. The TOIO setting shall be the maximum IOPS point where the average response time is less than 20 ms.

The Test Software selects the TOIO setting to be used in the immediately following measurement interval.

Note: the self-optimizing TOIO sweep determines the optimal TOIO setting during a single test run instead of having to run multiple tests at different TOIO settings to achieve the same result.

7.4 Test Steps Definition

7.4.1 Setup Environment

Connect the DUT to the test server. Connect the Power Analyzer to the DUT. Start the Power Analyzer.

See the pseudo-code in Subclause 7.5.2.1.

7.4.2 Purge

If supported, the Purge command or equivalent may be applied to the DUT.

This step places the DUT in a Fresh-Out-of-Box (FOB) state as if no writes have occurred. If no Purge command is applied, "No Purge" shall be reported in the test reporting.

See the pseudo-code in Subclause 7.5.2.2.

7.4.3 Pre-Fill

This step prepares the DUT for convergence to steady state.

Apply the SEQ 256 KiB Write Workload to the DUT for twice the user capacity.

For HDDs, this test step may be omitted. If an HDD test has not performed the pre-fill test step and subsequently fails the steady state conditional run test step, then it might be beneficial to include the pre-fill test step when the test is rerun.

See the pseudo-code in Subclause 7.5.2.3.

7.4.4 Steady State Conditioning Run

The **5 Rounds Steady State Determination** shall show relatively time invariant (stable) performance for five consecutive rounds.

Apply the Complex Workload for a minimum of 5 Rounds of 30 minutes each. Execute the **5 Rounds Steady State Determination** (see Subclause 6.2) procedure at Round 1. If the DUT has not achieved steady state for 5 consecutive Rounds, perform additional Rounds (up to a total of 25) until steady state is achieved for 5 consecutive Rounds. If steady state is not achieved after 25 Rounds, the DUT shall be deemed to have failed the test.

See the pseudo-code in Subclause 7.5.2.4.

7.4.5 Complex

Apply the Complex Workload that is defined in Clause 8 . The IO steam composition of the Complex Workload is intended to emphasize the impact of read caches by skewing IO Block Size distributions and adjusting the LBA Range accesses of the IO load.

The Complex step starts with a 10 minute warm-up interval that includes a self-optimizing TOIO sweep to determine the TOIO setting to be used during the measurement interval.

See the pseudo-code in Subclause 7.5.2.5.

7.4.6 RND 8 KiB Write

Apply random 8 KiB writes for a warm-up interval of 10 minutes followed by a measurement interval of 30 minutes.

The RND 8 KiB Write step starts with a 10 minute warm-up interval that includes a self-optimizing TOIO sweep to determine the TOIO setting to be used during the measurement interval.

See the pseudo-code in Subclause 7.5.2.6.

7.4.7 RND 8 KiB Read

Apply small block Random Reads for a warm-up interval of 10 minutes followed by a measurement interval of 30 minutes.

The RND 8 KiB Read step starts with a 10 minute warm-up interval that includes a self-optimizing TOIO sweep to determine the TOIO setting to be used during the measurement interval.

See the pseudo-code in Subclause 7.5.2.7.

7.4.8 SEQ 256 KiB Write

Apply large block Sequential Writes for a warm-up interval of 10 minutes followed by a measurement interval of 30 minutes.

The SEQ 256 KiB Write step starts with a 10 minute warm-up interval that includes a self-optimizing TOIO sweep to determine the TOIO setting to be used during the measurement interval.

See the pseudo-code in Subclause 7.5.2.8.

7.4.9 SEQ 256 KiB Read

Apply large block Sequential Reads for a warm-up interval of 10 minutes followed by a measurement interval of 30 minutes.

The SEQ 256 KiB Read step starts with a 10 minute warm-up interval that includes a self-optimizing TOIO sweep to determine the TOIO setting to be used during the measurement interval.

See the pseudo-code in Subclause 7.5.2.9.

7.4.10 Ready Idle

The Ready Idle test step is an interval of no host IO commands of 30 minutes duration.

Changes in Ready Idle power can indicate the impact of garbage collection, write amplification, and power throttling activities on power consumption.

See the pseudo-code in Subclause 7.5.2.10.

7.4.11 Post Process and Plot

Import the power data and perform post-processing.

Plot the reports and steady state rounds data as required by Clause 13 .

See the pseudo-code in Subclause 7.5.2.11.

7.5 Test Software

7.5.1 Purpose

The purpose of Subclause 7.5 is to define the requirements for the Test Software to be used. The Test Software required by this document creates and applies the required workloads to the DUT. The Test Software will be used to accurately capture and analyze the dynamic power characteristics of DUTs tested pursuant to this specification. See Annex C, Annex D, and Annex E for more information.

7.5.2 Test Pseudo-Code

The Test Software shall execute the test according to the following pseudo-code. The pseudo-code language is described in Annex E.

Test Steps

7.5.2.1 Setup Environment

```
# See the Test Step Definition in Subclause 7.4.1.
TBD
```

7.5.2.2 Purge

```
# See the Test Step Definition in Subclause 7.4.2.
TBD
```

7.5.2.3 Pre-Fill

```
# See the Test Step Definition in Subclause 7.4.3.
Apply SEQ 256 KiB Writes for 2x drive user capacity.
For HDDs, this test step may be omitted.
```

7.5.2.4 Steady-State Conditioning Run

```
# See the Test Step Definition in Subclause 7.4.4.
Do
   Apply Complex Workload at T4Q32 for 30 min; store IOPS measurement
   Perform Steady State analysis (IOPS shall be within ±10% excursion around
   average and slope shall be less than 10%)
While quantity of rounds is less than 5 or Steady State is not reached

Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
   Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
         Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
         ART measurements
      OR
         Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
         2, 1]  QD [32, 16, 8, 4, 2, 1]
      Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
      measurements
```

7.5.2.5 Complex

```
# See the Test Step Definition in Subclause 7.4.5.
Perform DI analysis with determination of the optimum TC/QD parameters
Apply Complex Workload with optimum TC/QD parameters

Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
   Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
         Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
         ART measurements
      OR
         Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
         2, 1]  QD [32, 16, 8, 4, 2, 1]
      Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
      measurements
```

7.5.2.6 RND 8 KiB Write

```
# See the Test Step Definition in Subclause 7.4.6.
Perform DI analysis with determination of the optimum TC/QD parameters
```

```
Apply Random 8 KiB Write workload with optimum TC/QD parameters

Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
   Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
         Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
         ART measurements
      OR
         Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
         2, 1]  QD [32, 16, 8, 4, 2, 1]

      Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
      measurements
```

7.5.2.7 RND 8 KiB Read

```
# See the Test Step Definition in Subclause 7.4.7.
Perform DI analysis with determination of the optimum TC/QD parameters

Apply Random 8 KiB Read workload with optimum TC/QD parameters

Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
   Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
         Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
         ART measurements
      OR
         Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
         2, 1]  QD [32, 16, 8, 4, 2, 1]
      Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
      measurements
```

7.5.2.8 SEQ 256 KiB Write

```
# See the Test Step Definition in Subclause 7.4.8.
Perform DI analysis with determination of the optimum TC/QD parameters

Apply Sequential 256 KiB Write workload with optimum TC/QD parameters

Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
   Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
         Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
         ART measurements
      OR
         Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
         2, 1]  QD [32, 16, 8, 4, 2, 1]
      Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
      measurements
```

7.5.2.9 SEQ 256 KiB Read

```
# See the Test Step Definition in Subclause 7.4.9.
Perform DI analysis with determination of the optimum TC/QD parameters

Apply Sequential 256 KiB Read workload with optimum TC/QD parameters
```

```
Default setting Loop for TC [64, 48, 40, 32, 16, 8, 4, 2, 1] do
    Loop for QD in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1] do
            Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and
            ART measurements
        OR
            Optional setting Loop & TC test sponsor choice (e.g., TC [32, 16, 8, 4,
            2, 1]  QD [32, 16, 8, 4, 2, 1]
        Apply Complex Workload with selected TC/QD for 5 seconds; store IOPS and ART
        measurements
```

7.5.2.10 Ready Idle

```
# See the Test Step Definition in Subclause 7.4.10.
Stay idle for 30 minutes
```

7.5.2.11 Post Process and Plot

```
# See the Test Step Definition in Subclause 7.4.11.
Import & aggregate power data

./merge_power <report_file> <power_file> <power_field_name>

Where <report_file> - is the path to the HTML file generated by CTS Lite,
<power_file> - is a path to the CSV file with power measurements.

#The CSV file should have a header line describing fields/columns. The first column
is considered a timestamp. The tool will use values from the column specified with
<power_field_name> argument.
```

8 Test Workloads

8.1 Overview

The particular IO stimuli used to drive the DUT during a test step are specified in terms of an IO profile (a.k.a. workload) made up of multiple attributes:

- Name: the name of the IO pattern for this stimulus. The identifier for the associated test phase is included parenthetically, when appropriate;
- IO Size: the number of bytes requested by a given read or write operation;
- Read/Write Percentage: the mixture of read/write IO requests within an IO profile;
- Transfer Alignment: Minimum granularity of IO transfer addresses. All transfer addresses within an IO stream shall be a multiple of this value;
- Access Pattern: either one or the other of the following two alternatives:
  - Random: Randomly distributed throughout the address space of the device under test;
  - Sequential, as defined in subclause 3.2.22 sequential read and subclause 3.2.23 sequential write;
- Data Pattern: the data patterns written shall be random.

8.2 Random Read and Random Write

Random Reads and Random Writes refer to two different types of operations that can be performed on data stored in a storage medium.

Random Reads involve retrieving or reading data from non-contiguous or randomly chosen logical block addresses (LBAs) within a storage device.

Random Writes refer to the process of writing data to non-contiguous or randomly chosen logical block addresses (LBAs) within a storage device.

8.3 Sequential Read and Sequential Write

Sequential Reads and Sequential Writes refer to two different types of operations that can be performed on data stored in a storage device.

Sequential Reads refers to the process of reading data from a storage device in consecutive logical block address (LBA) order.

Sequential Writes refers to the process of writing data to a storage d in consecutive logical block address (LBA) order.

8.4 Complex Workload

The Complex Workload consists of a mix of different IO sizes, read/write mixes, and access patterns with a skewed access across a range of blocks.

The goal of the Complex Workload is to provide continuity of workloads as defined in the *ISO/IEC 24091:2019*[7] and *SNIA Emerald™ Power Efficiency Measurement Specification Version 4.0.0*[8] standards and those defined in this document. Accordingly, the Complex Workload considers, among other things, the contribution of device mechanisms that exploit access locality, e.g., read caching.

The Complex Workload is adopted from the Hot Band IO Profile of the *ISO/IEC 24091:2019*[7] and *SNIA Emerald™ Power Efficiency Measurement Specification Version 4.0.0*[8] standards with terminology changes to match this document. Table 15, Table 16, and Table 17 of those documents correspond to Table 7. Table 8, and Table 9, respectively, of this document.

Table 7 shows information concerning IO streams of the Complex Workload.

Table 7 – IO Streams Within the Complex Workload

IO Streams	% of workload (skew)	Read/Write Percentage	IO Size (KiB)[a]	Access Pattern	Usable Address Range
Write Stream 1	5	0/100	See Table 8 & 9	Sequential	0 % to 100 %
Write Stream 2	5	0/100	See Table 8 & 9	Sequential	0 % to 100 %
Write Stream 3	5	0/100	See Table 8 & 9	Sequential	0 % to 100 %
Read Stream 1	5	100/0	See Table 8 & 9	Sequential	0 % to 100 %
Read Stream 2	5	100/0	See Table 8 & 9	Sequential	0 % to 100 %
Read Stream 3	5	100/0	See Table 8 & 9	Sequential	0 % to 100 %
Read Stream 4	5	100/0	See Table 8 & 9	Sequential	0 % to 100 %
Read Stream 5	5	100/0	See Table 8 & 9	Sequential	0 % to 100 %
Uniform Random	6	50/50	See Table 8 & 9	Random	0 % to 100 %
Hot Band 1	28	70/30	See Table 8 & 9	Random	10 % to 18 %
Hot Band 2	14	70/30	See Table 8 & 9	Random	32 % to 40 %
Hot Band 3	7	70/30	See Table 8 & 9	Random	55 % to 63 %
Hot Band 4	5	70/30	See Table 8 & 9	Random	80 % to 88 %

[a] For a DUT using native or emulated 512 B sectors, see Table 8.
For a DUT using native 4 KiB sectors, see Table 9.

The IO transfer size (Xfer) used within the Complex Workload is listed in Table 8 and Table 9.

Table 8 – IO Transfer Size Within the Complex Workload For 512 Byte Native Devices

Xfer	Streaming Write	Streaming Read	Uniform	Hot Band
512 B			2 %	2 %
1 KiB			2 %	2 %
4 KiB	29 %	29 %	27 %	27 %
8 KiB	33 %	33 %	31 %	31 %
16 KiB	6 %	6 %	5 %	5 %
32 KiB	5 %	5 %	5 %	5 %
48 KiB			1 %	1 %
56 KiB			1 %	1 %
60 KiB			2 %	2 %
64 KiB	22 %	22 %	20 %	20 %
128 KiB	3 %	3 %	2 %	2 %
256 KiB	2 %	2 %	2 %	2 %

Table 9 – IO Transfer Size Within the Complex Workload For 4 KiB Native Devices

Xfer	Streaming Write	Streaming Read	Uniform	Hot Band
4 KiB	29 %	29 %	31 %	31 %
8 KiB	33 %	33 %	31 %	31 %
16 KiB	6 %	6 %	5 %	5 %
32 KiB	5 %	5 %	5 %	5 %
48 KiB			1 %	1 %
56 KiB			1 %	1 %
60 KiB			2 %	2 %
64 KiB	22 %	22 %	20 %	20 %
128 KiB	3 %	3 %	2 %	2 %
256 KiB	2 %	2 %	2 %	2 %

8.5 Ready Idle

Ready Idle is the operational state in which a) a storage device is capable of satisfying an arbitrary host-initiated IO request within the response time and MaxTTFD constraints of its selected taxonomy category and b) no host-initiated IO requests are being submitted to the storage device.

9 Metrics

9.1 Taxonomy Considerations

This document defines metrics for the HDD and SSD taxonomy categories.

9.2 Primary Metrics

Power efficiency for each test phase for HDD and SSD (see subclause ??):

- EP_{CP} for Complex (IOPS/W);
- EP_{RR} for Random Read (IOPS/W);
- EP_{RW} for Random Write (IOPS/W);
- EP_{SR} for Sequential Read (MiB/s/W);
- EP_{SW} for Sequential Write (MiB/s/W); and
- EP_{RI} for Ready Idle (GB/W).

9.3 Power Efficiency Metrics

9.3.1 Ready Idle Metric

The Ready Idle metric represents the amount of user capacity supported per watt of power required by the DUT. It is calculated as shown in Equation 8-1, as the ratio of:

- The user capacity of the DUT, measured in GB;
- The average power, from the Ready Idle test step, measured in watts.

Equation 8-1: Power Efficiency, Ready Idle

$$EP_{RI} = C_{R} / PA_{RI}\ (1800)$$

Where:

- EP_{RI} is the power efficiency metric for the Ready Idle test step;
- C_{R} is the user capacity of the DUT (see subclause ???);
- PA_{RI} (1800) is the average power over the 30 minute measurement interval for the Ready Idle test step.

9.3.2 Active Test Steps

The active tests steps are

1) Complex,
2) RND 8 KiB Write,
3) RND 8 KiB Read,
4) SEQ 356 KiB Write, and
5) SEQ 256 KiB Read.

For each active test step, the power efficiency metric represents the rate of data transfer supported per watt required by the DUT during a selected measurement interval. It is calculated, as shown in Equation 8-2, as the ratio of:

- The operations or throughput rate, during the measurement interval of the active test, expressed in IOPS or MiB/s;

- The average power, during the measurement interval of the active test, measured in watts.

Equation 8-1: Power Efficiency, Active

$$EP_i = \frac{O_i(1800)}{PA_i(1800)}$$

Where:

- EP_i is the power metric for an active test step;
- PA_i (1800) is the average power over the 30 minute measurement interval for the active test step $_i$;
- O_i (1800) is the average operations rate over the 30 minute measurement interval for the active test step $_i$.

9.4 Average Response Time

The average response time for a test step, ART, is calculated over a specified time interval T in seconds.

9.5 Average Power

The average power for a test step *i*, PA_i(T), is the arithmetic average of sampled power measurements taken over a specified time interval T in seconds, as illustrated in Equation 7-2.

Equation 7-2: Average Power

$$PA_i(T) = \frac{\sum W_s}{n}$$

Where:

- PA_i(T) is the average power during test step *i*, taken over a time interval of T seconds;
- W_s is power in watts measured at each sampling interval *s* taken during the time interval T;
- *n* is the number of samples gathered by the power meter during the time interval T.

9.6 Operations Rate

The operations rate for a test step *i*, O_i(T), is a measure of the average rate of completed work over a specified time interval T. It is different for random workloads and sequential workloads.

For random workloads, i.e., Complex, RND 8 KiB Write, and RND 8 KiB Read, the operations rate is the average rate of IO operation completions during time interval T.

For sequential workloads, i.e., SEQ 256 KiB Write and SEQ 256 KiB Read, the operations rate is the average rate of data transfer in mebibytes per second (MiB/s) within time interval T.

To provide a uniform basis for the metrics of a SNIA Emerald™ Device Power Efficiency Measurement, these two different measures of operations rate are both represented by O_i(T).

9.7 Measurement Interval

Each test step specifies a duration for the measurement interval. A measurement interval is a subset of a test step during which the data underlying a specific metric or calculation is gathered.

9.8 Reporting Granularity

The power efficiency metrics shall be reported to three significant digits.

10 Workload Generator Requirements

10.1 Purpose

The Workload Generator shall be capable of generating the IO test data necessary to carry out the measurement procedures specified by this document.

Annex D specifies the Workload Generator that shall be used.

10.2 Procedure

The workload generator shall operate within the Test Software. The workload generator shall produce IO data patterns and fill memory buffers that the Test Software uses to define and create test IO streams that are applied to a DUT.

11 Power Analyzer

11.1 Purpose

The Power Analyzer is used to accurately capture and analyze the dynamic power characteristics of DUTs tested pursuant to this specification.

The Power Analyzer incorporates voltage and current measurement apparatus, the power measurement software, the DUT interposer, and the interposer interconnect cable.

Annex F contains Power Analyzer requirements and a list of recommended Power Analyzers.

11.2 Procedure

The Power Analyzer shall be started prior to a test and continue to be active throughout the test and shall record:

- Input voltage to the DUT, to an accuracy of ± 2 % at the 95 % confidence level;
- Power consumption by the DUT, to the resolution specified in Annex F. The voltage, current, and power measurements shall be reported using a period of not more than 1 s. The power data timestamp shall be synchronized with the test performance data timestamp to a resolution of at least 1 s. All samples within each reporting period, taken at the specified sample rate, shall be averaged.

12 General Environmental Measurement Recommendations

All measurements should be conducted in a climate-controlled environment which complies with the environmental specification for the hardware under test.

13 Reporting Requirements

13.1 General Reporting Requirements

This clause lists the information that shall be disclosed in every published test result identified as an SNIA Emerald™ Device Power Efficiency Measurement test result. Test results identified as an SNIA Emerald™ Device Power Efficiency Measurement test result shall be generated according to this document.

When units are specified after an item, that item shall be reported in terms of the specified unit.

The following items, common to all tests, shall be included in the test report. These items only need to be reported once in the test report. Test-specific report items are defined in the relevant test sections themselves. Administrative

1) Test date
2) Report date
3) Test ID number
4) Test sponsor name
5) Auditor name, if applicable
6) Test specification version

13.1.1 Storage Server Hardware

1) Manufacturer and model number
2) System board manufacturer and model number
3) CPU model, number of cores, number of sockets
4) DRAM quantity in GiB and DRAM speed
5) Physical Interface to DUT
6) Primary storage model number
7) Peripherals

13.1.2 Power Analyzer Hardware

1) Manufacturer and model Number
2) Power Analyzer serial number
3) DUT connection interposer board
4) DUT connection cable

13.1.3 Storage Server Software

1) Operating system name and version
2) File system name and version
3) CTS Lite version

13.1.4 Device Under Test

1) Manufacturer
2) Model number
3) Serial number
4) Firmware revision
5) User capacity in GB or TB
6) Interface (e.g. U2)
7) Protocol (e.g., SAS, NVMe)
8) Form factor (e.g., 2.5”)
9) Media type (e.g., QLC NAND Flash)

13.1.5 Primary Test Metrics

1) Complex Workload Power Efficiency (IOPS/W and MiB/s/W)
2) Random 8K Write Power Efficiency (IOPS/W and MiB/s/W)
3) Random 8K Read Power Efficiency (IOPS/W and MiB/s/W)
4) Sequential 256K Write Power Efficiency (IOPS/W and MiB/s/W)
5) Sequential 256K Read Power Efficiency (IOPS/W and MiB/s/W)
6) Idle Average Power (GB/W)

13.2 Test Specific Reporting

A SNIA Emerald™ Device Power Efficiency Measurement test result shall include the information defined by Table 10.

Table 10 – Test Results

SNIA Emerald™ Device Power Efficiency Measurement Required Test Results							
Primary Test Results							
		Workload					
Metric	**Units**	**Complex**	**RND 8 KiB Write**	**RND 8 KiB Read**	**SEQ 256 KiB Write**	**SEQ 256 KiB Read**	**Ready Idle**
Power Efficiency	**IOPS/W**	<value>	<value>	<value>	<value>	<value>	N/A
	MiB/s/W	<value>	<value>	<value>	<value>	<value>	N/A
	GB/W	N/A	N/A	N/A	N/A	N/A	<value>
Supporting Data							
Average Power	**mW**	<value>	<value>	<value>	<value>	<value>	<value>
Performance	**IOPS**	<value>	<value>	<value>	<value>	<value>	N/A
	MiB/s	<value>	<value>	<value>	<value>	<value>	N/A
Thread Count	**Number**	<value>	<value>	<value>	<value>	<value>	
Queue Depth	**Number**	<value>	<value>	<value>	<value>	<value>	
Test Parameters							
Parameter	**Units**	**Data**		**Parameter**	**Units**	**Data**	
Steady State	**Yes/No**	<value>		**Purge**	**Yes/No**	<value>	
Pre-Fill	**Yes/No**	<value>		**Active LBA Range**	**Percent**	<value>	
Data Pattern	**Random, 2:1 Compressible**						

Annex A (Informative) CTS Lite Report Example

A.1 Introduction

CTS Lite generates a series of reports, examples of which are shown in this Clause.

CTS Lite automatically generates a Read-only Audit file that contains report plots labelled P1 – P17. Figure 6 shows the list of report plots generated by the Test Audit file. Corresponding reports are also contained in editable excel report files.

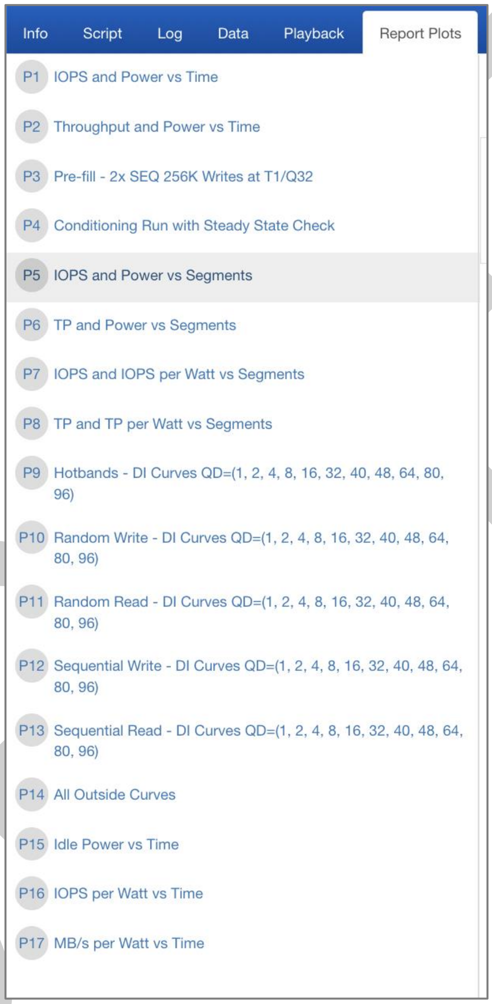

Figure 6 – Report Plots Tab

CTS Lite generates 17 reports plots and are grouped as follows:

1. Test Validation   Test Flow: P1, P2, P3, P4, P5, P6
2. Data Reporting   Test Results: P7, P8, P15, P16, P17
3. Data Analysis   DI Curve: P9, P10, P11, P12, P13, P14

The following Subclauses present an example of each of the 17 reports plots.

A.2 P1: IOPS and Power vs. Time

An example of P1 IOPS and Power vs. Time in shown in Figure 7.

P1 shows the continuous and uninterrupted test flow in IOPS from the start of the test to end of the test. The Ready Idle segment is power only and indicated by a dotted line.

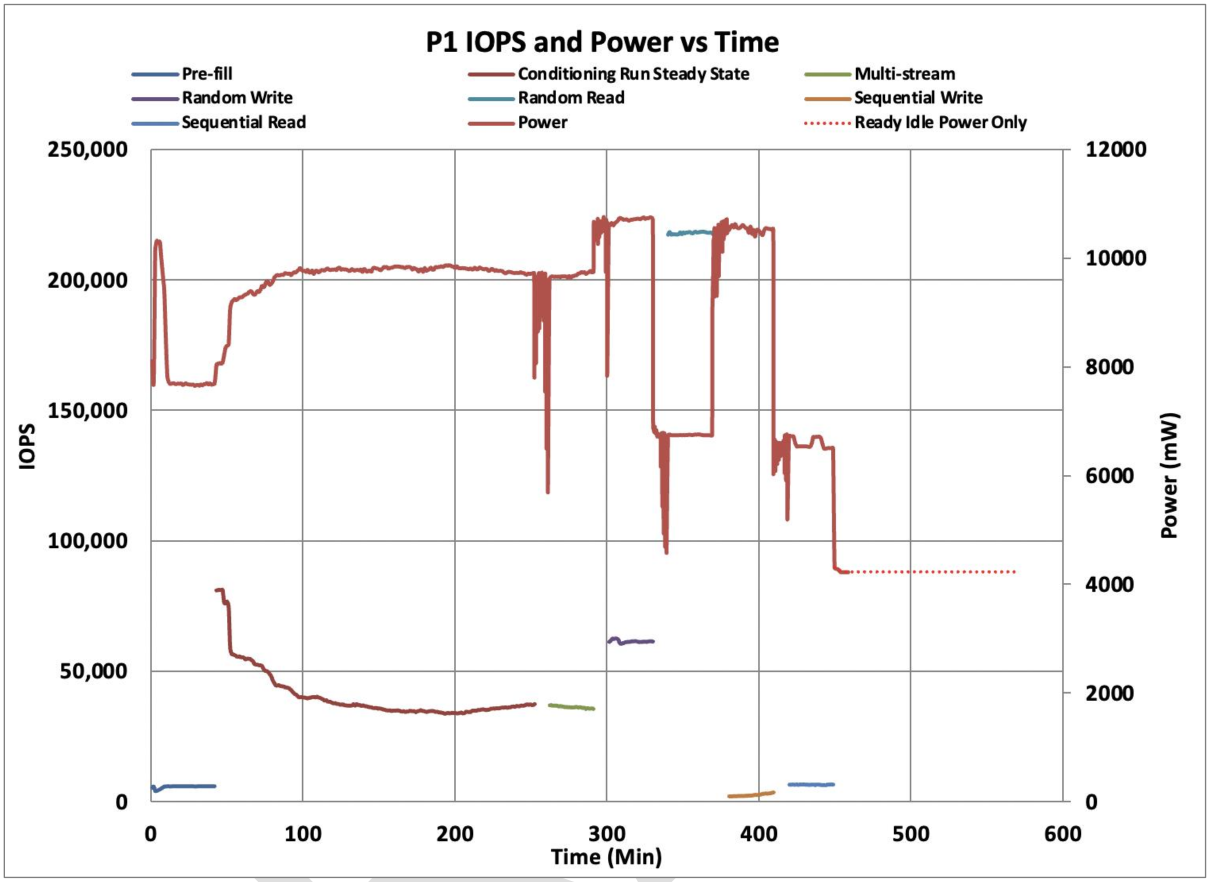

Figure 7 – IOPS and Power vs. Time

A.3 P2: Throughput and Power vs. Time

An example of P2 Throughput and Power vs Time in shown in Figure 8.

P2 shows the continuous and uninterrupted test flow in MiB/s from the start of the test to end of the test. The Ready Idle segment is power only and indicated by a dotted line.

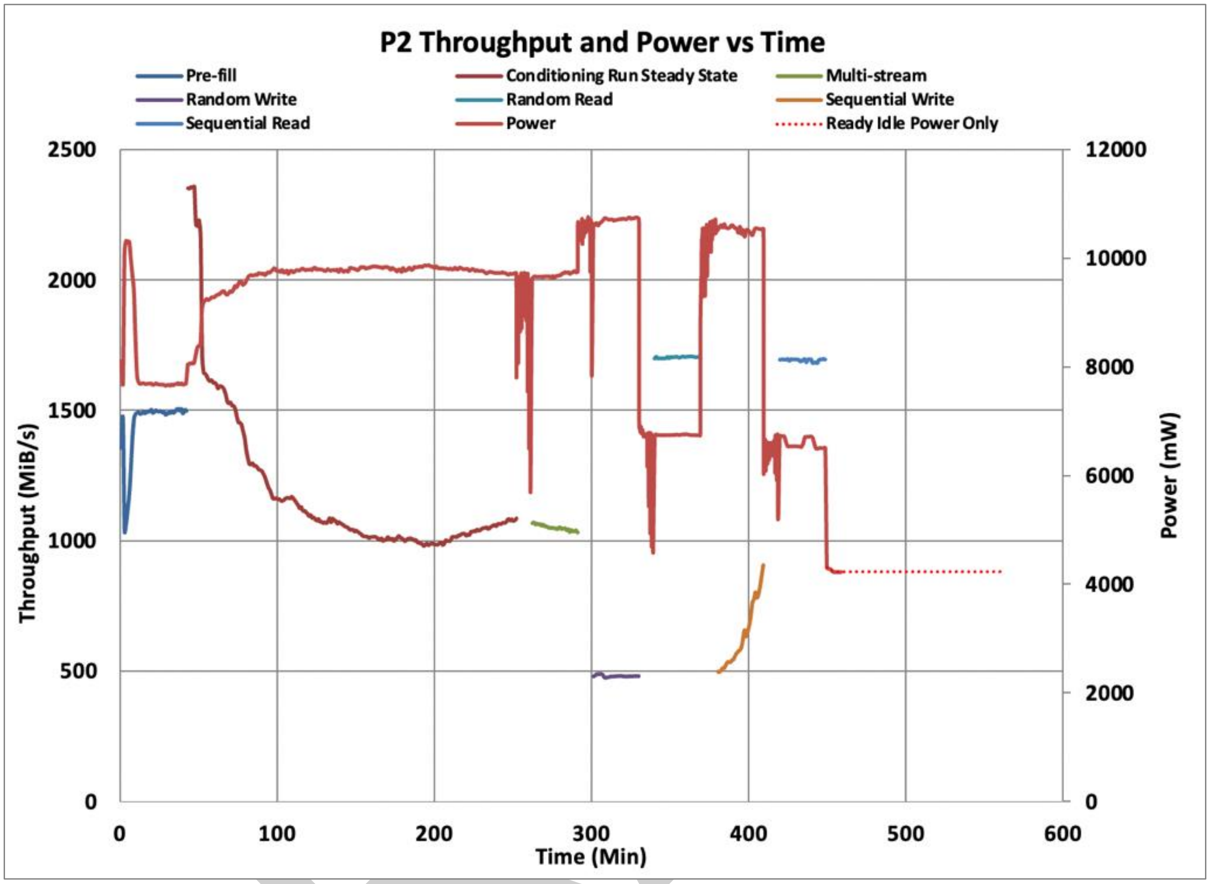

Figure 8 – Throughput and Power vs. Time

A.4 P3: Pre-Fill Test Step

An example of P3 Pre-fill – 2x SEQ 256K Writes at T1/Q32 is shown in Figure 9.

P3 shows the throughput of the DUT while twice the user capacity is filled using SEQ 256 KiB Writes at a TOIO of 32 (Thread Count 1 / Queue Depth 32).

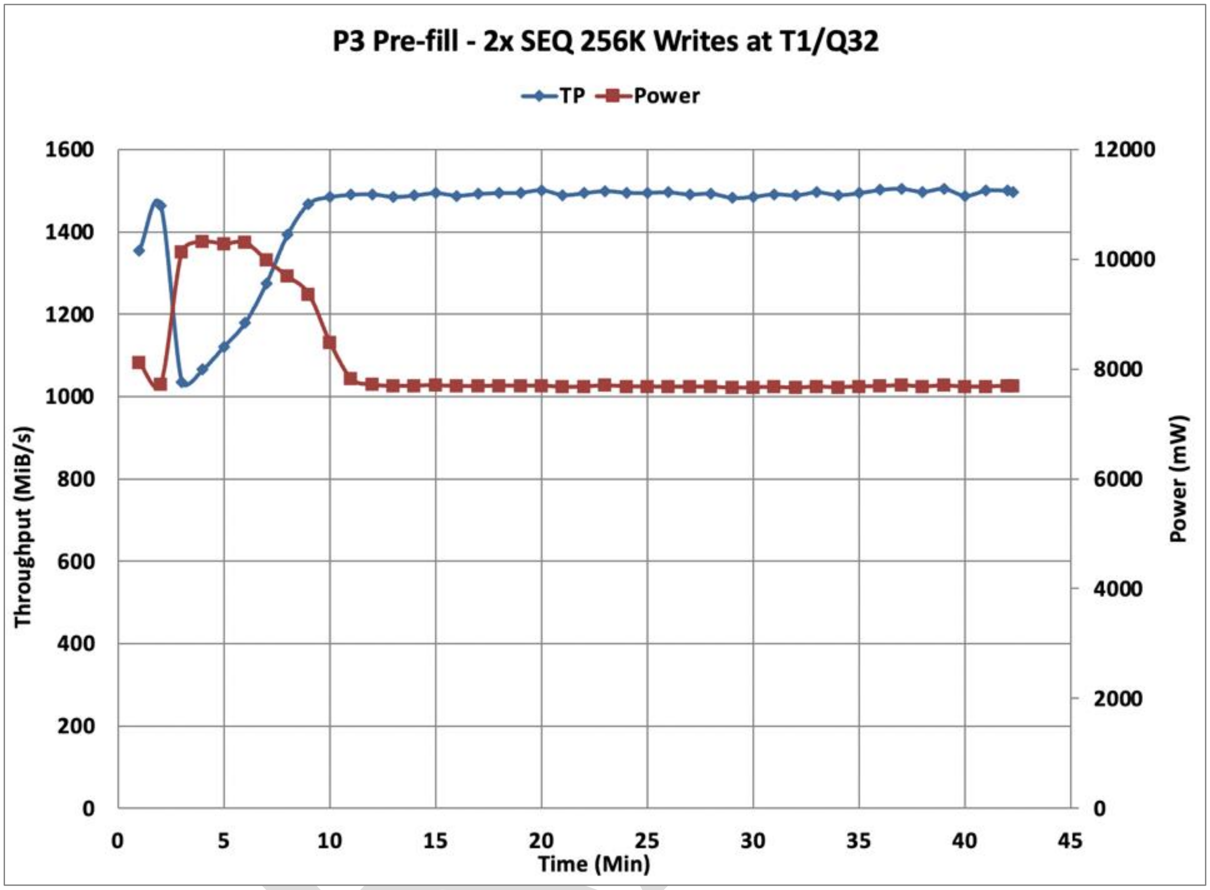

Figure 9 – Pre-Fill

A.5 P4: Steady State Verification

An example of P4 Condition Run with Steady State Check is shown in Figure 10.

P4 documents the convergence of the complex workload to steady state.

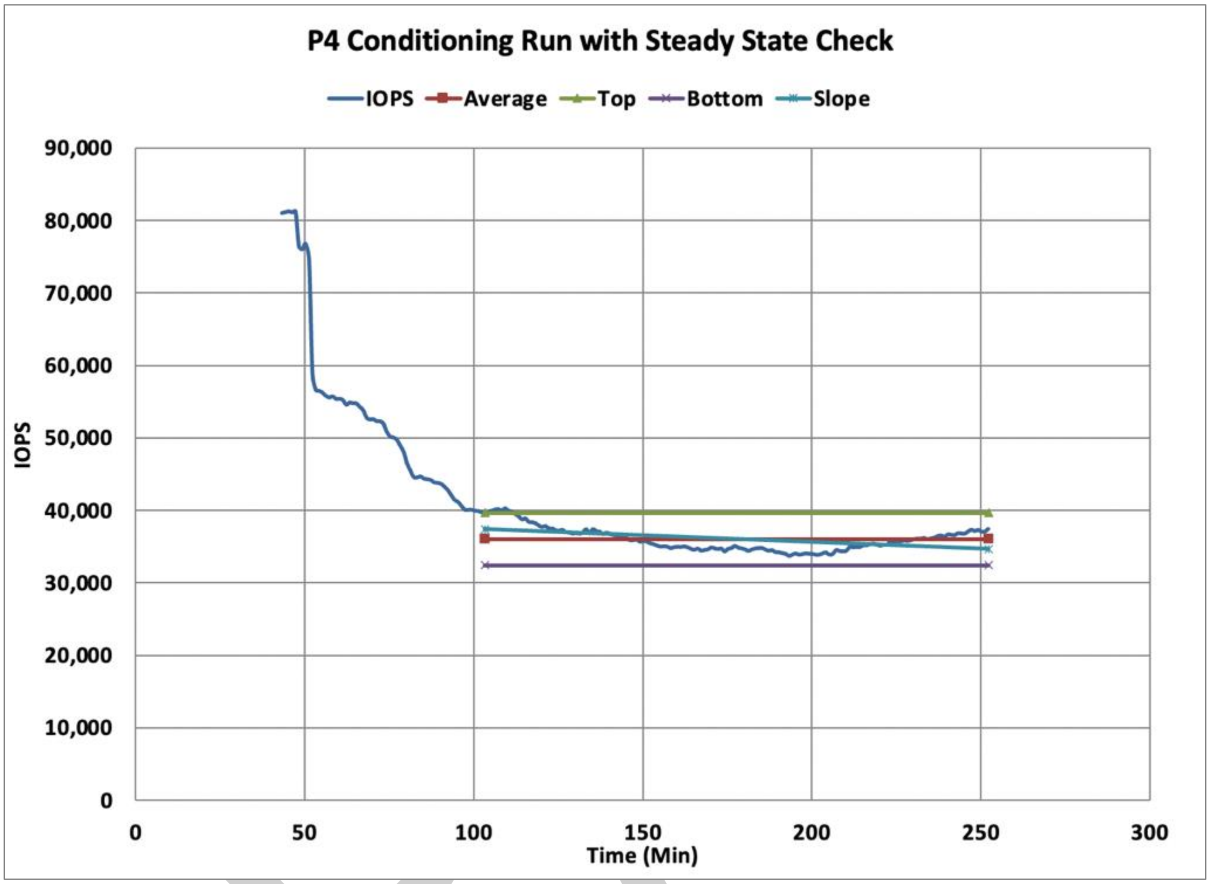

Figure 10 – Conditioning Run with Steady State

A.6 P5: IOPS and Power vs. Test Step

An example of P5 IOPS and Power vs Test Steps (Segments) in shown in Figure 11.

P5 shows the average IOPS, average power, and TOIO for each test step workload measurement interval and average power only for Ready Idle.

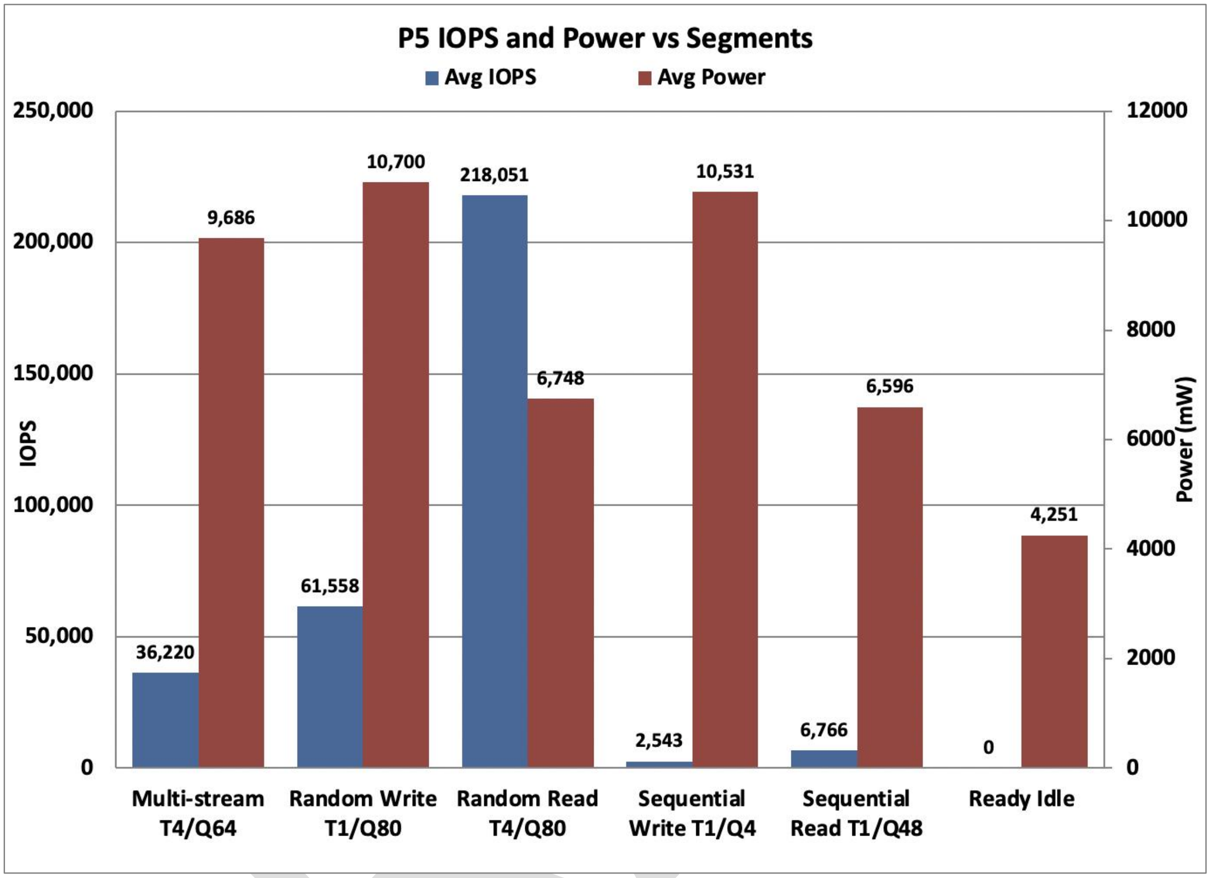

Figure 11 – IOPS and Power vs. Test Steps (Segments)

A.7 P6: Throughput and Power vs Segment

An example of P6 TP and Power vs Segments is shown in Figure 12.

P6 shows the average throughput, average power, and TOIO for each test step measurement interval and average power only for Ready Idle.

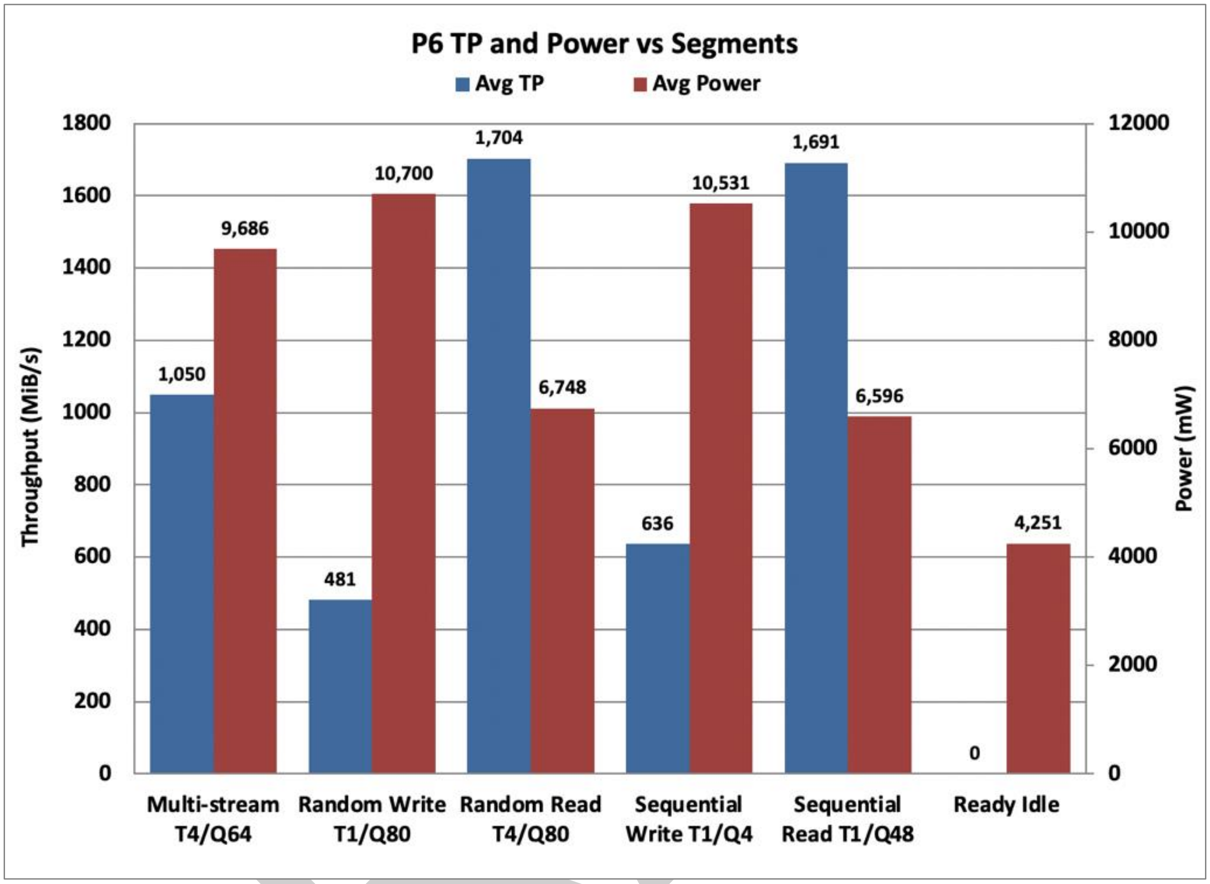

Figure 12 – Throughput and Power vs. Segments

A.8 P7: IOPS And IOPS per Watt vs. Test Steps (Segments)

An example of P7 IOPS and IOPS per Watt vs Segments is shown in Figure 13**.**

P7 shows the average IOPS and average IOPS/watt for each test step workload.

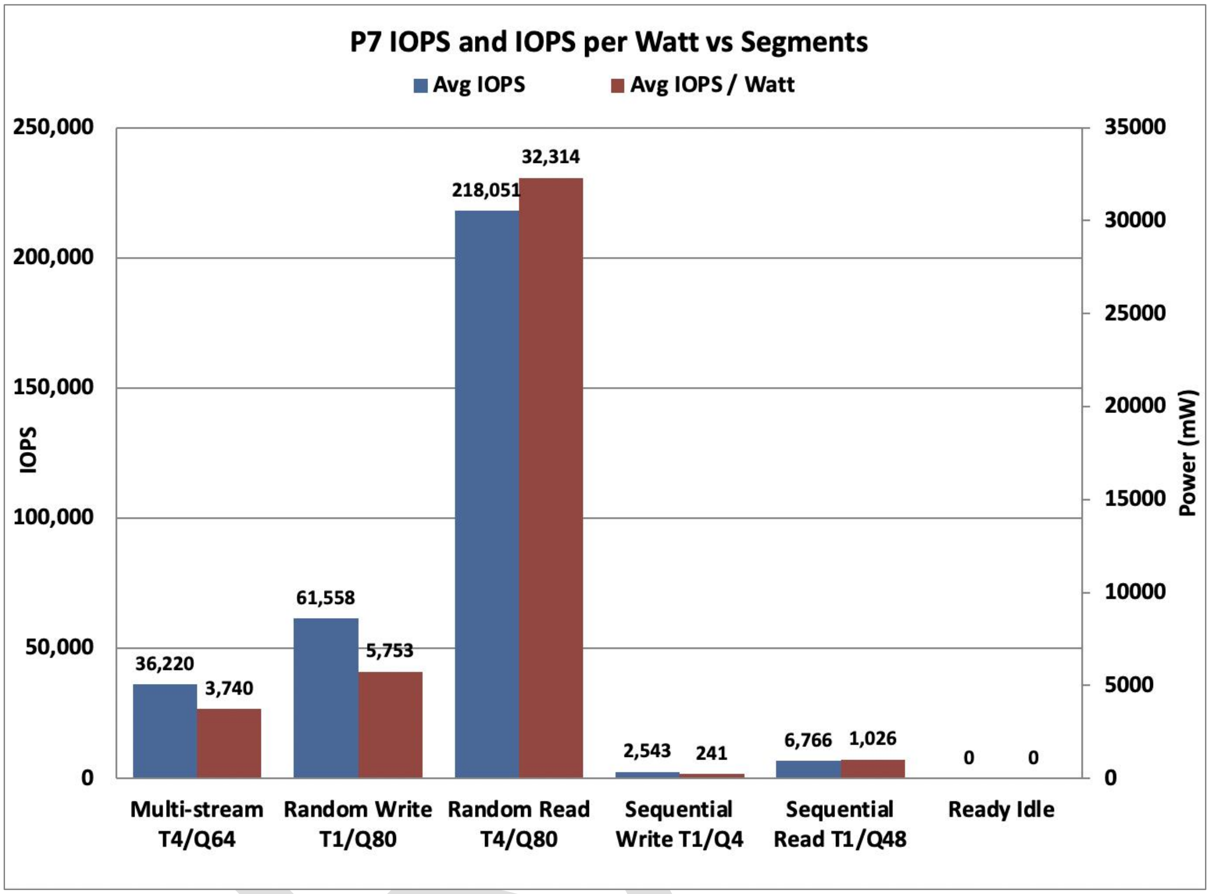

Figure 13 – IOPS and IOPS/Watt vs. Test Steps (Segments)

A.9 P8: Throughput and Throughput/Watt Vs. Test Steps (Segments)

An example of P8 TP and TP per Watt vs Test Steps (Segments) is shown in Figure 14**.**

P8 shows average throughput and average throughput/watt for each test step workload.

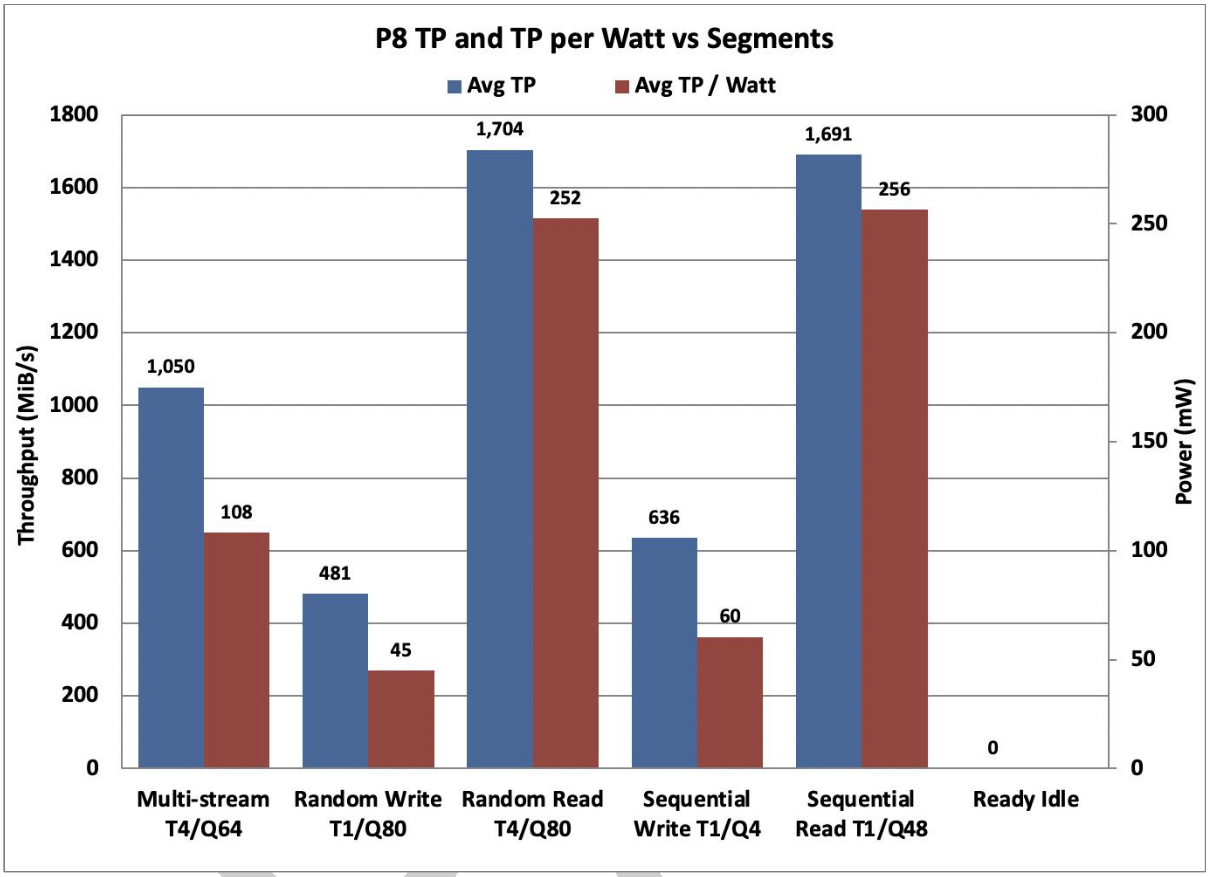

Figure 14 – Throughput and Throughput/Watt vs. Test Steps (Segments)

A.10   P9: Complex – DI Curves

An example of P9 Multi-stream – DI Cures is shown in Figure 15**.**

In this document, the Complex Workload is the name used for what is called Multi-stream in P9.

P9 shows a DI Curve for each Thread Count / Queue Depth value of the Complex Workload and the DI OS Curve.

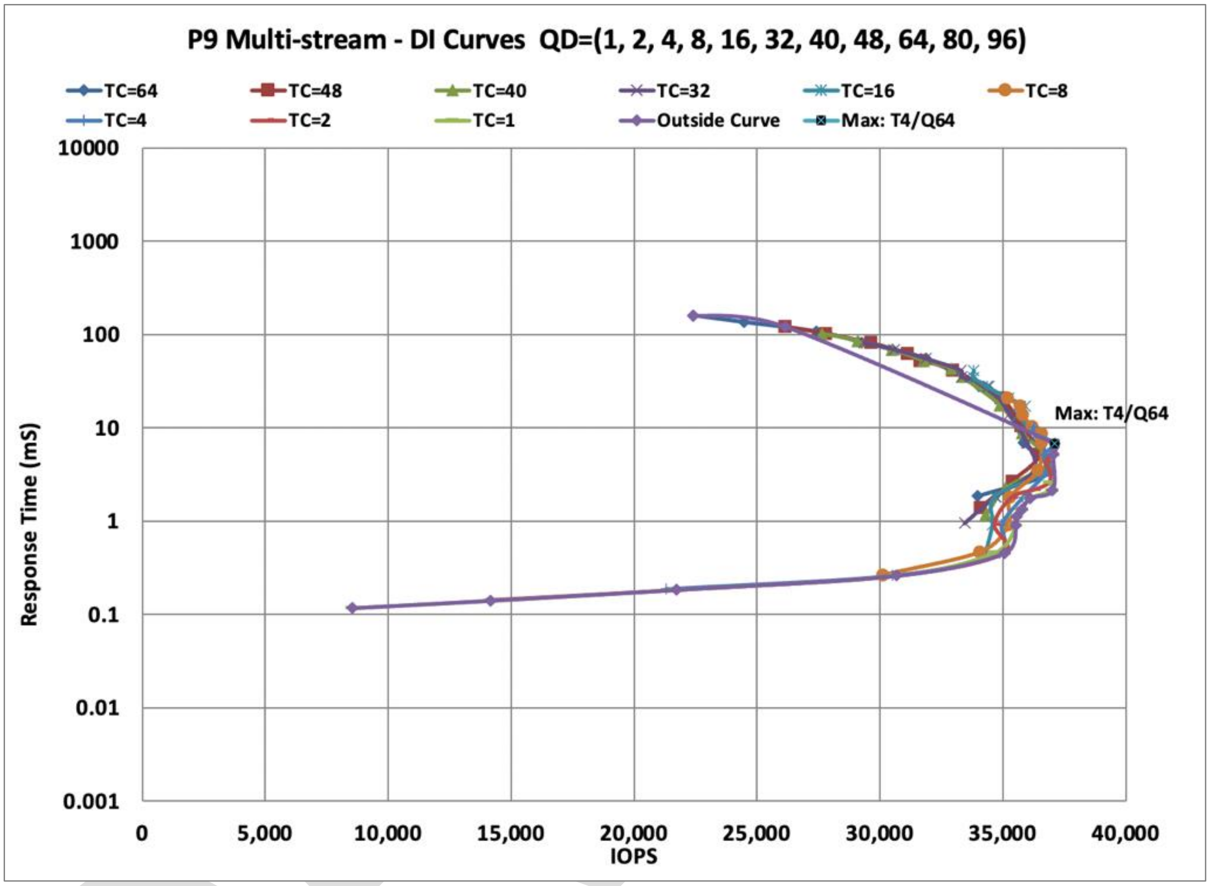

Figure 15 – Complex Workload – DI Curves

A.11 P10: Random Write – DI Curves

An example of P10 Random Write – DI Curves is shown in Figure 16**.**

P10 shows a DI Curve for each TOIO value of the RND 8 KiB Write workload and the DI OS Curve.

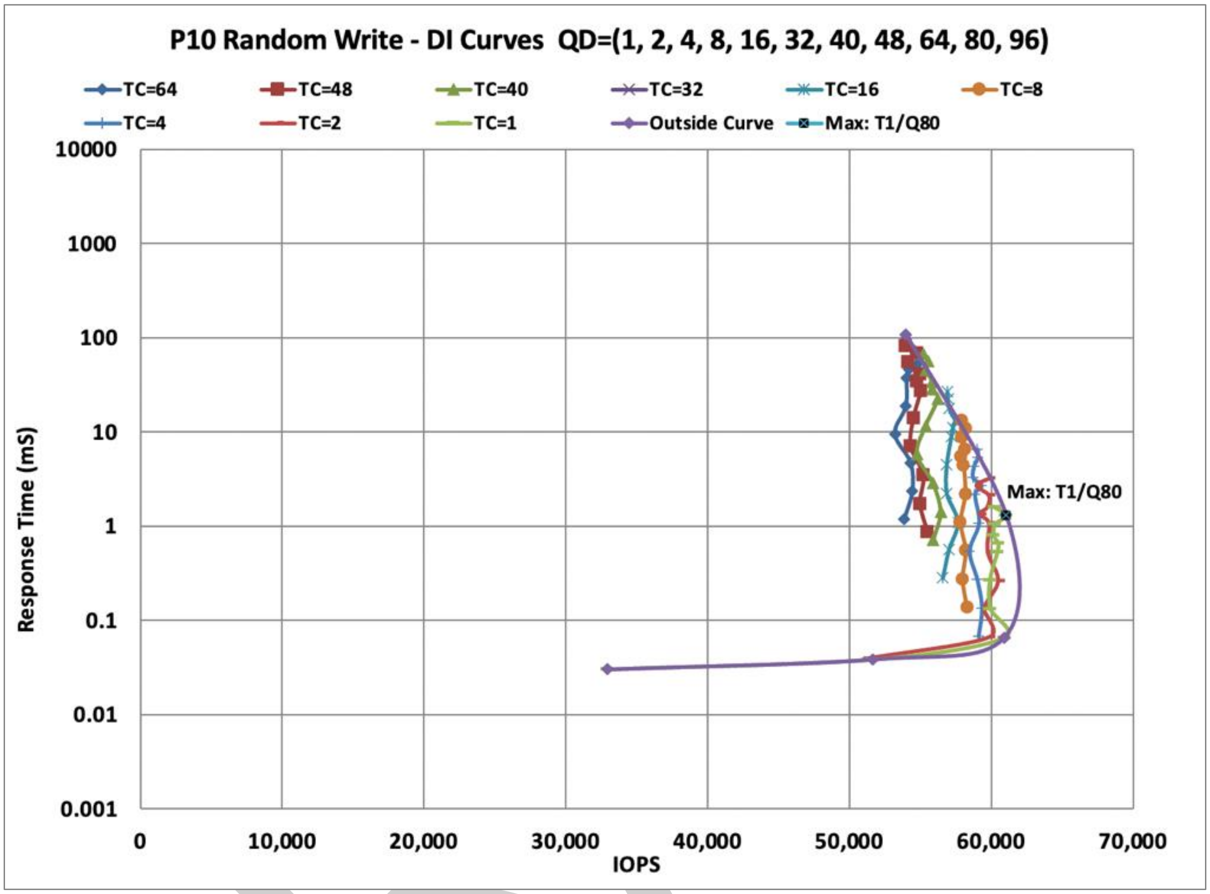

Figure 16 – Random Write – DI Curves

A.12 P11: Random Read – DI Curves

An example of P11 Random Read – DI Curves is shown in Figure 17**.**

P11 shows DI Curves for each TOIO value of the RND 8 KiB Read workload and the DI OS Curve.

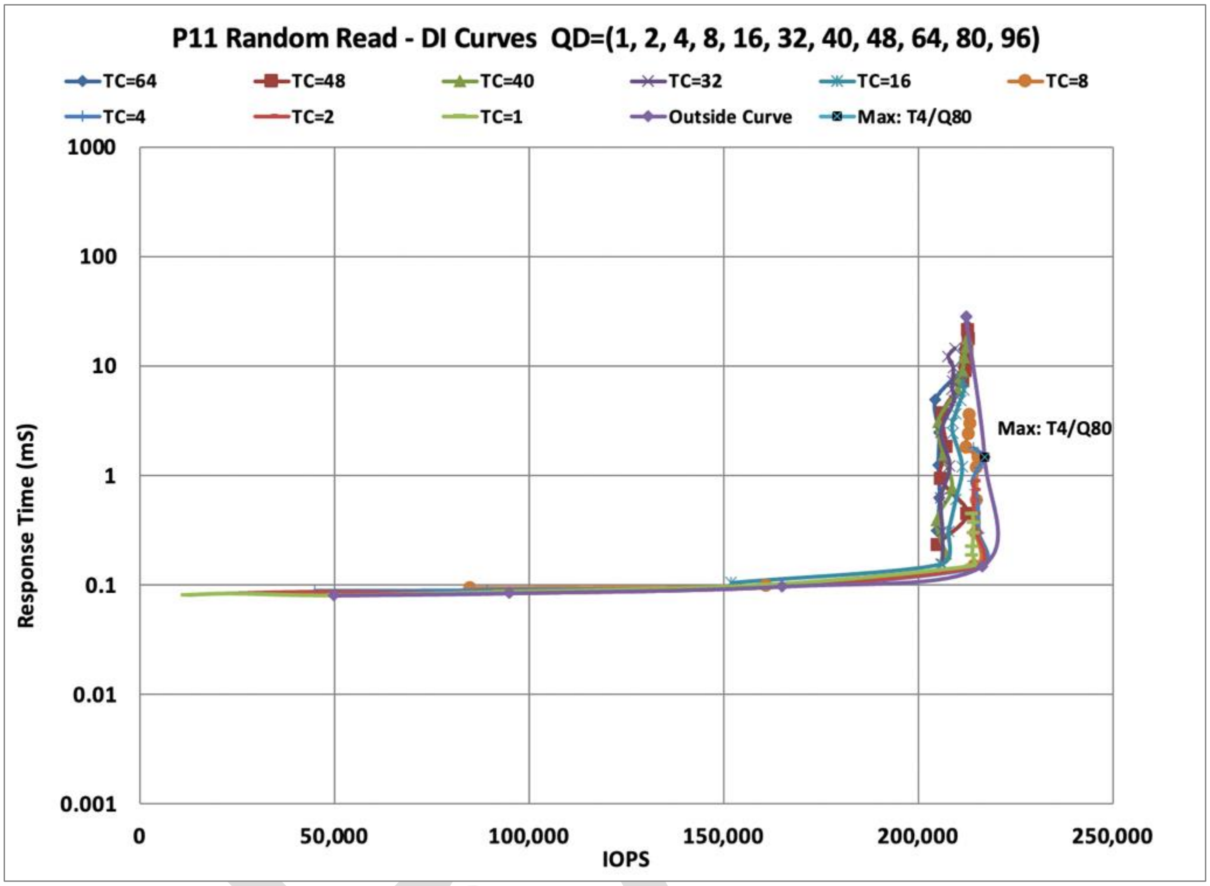

Figure 17 – Random Read – DI Curves

A.13 P12: Sequential Write – DI Curves

An example of P12 Sequential Write – DI Curves is shown in Figure 18**.**

P12 shows DI Curves for each TOIO value of the SEQ 256 KiB Write workload and the DI OS Curve

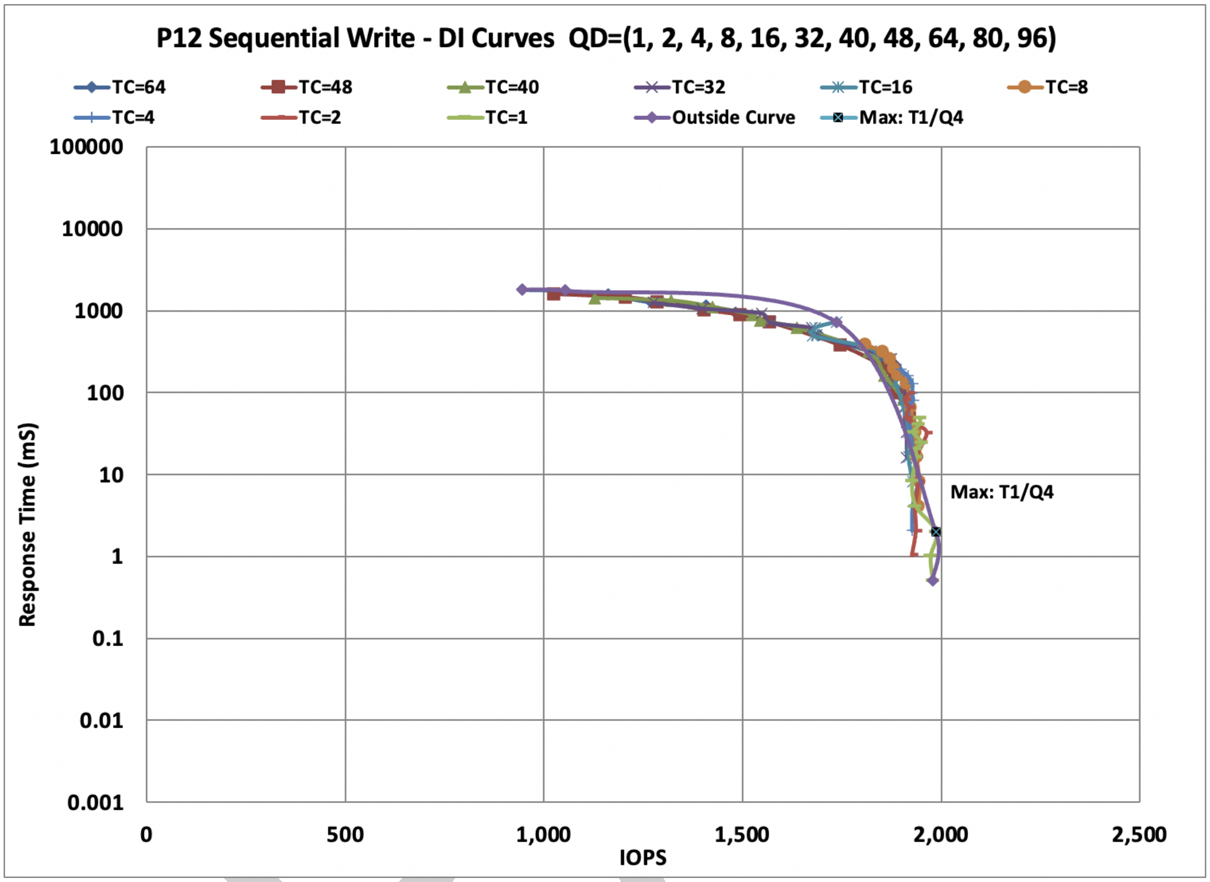

Figure 18 – Sequential Write – DI Curves

A.14 P13: Sequential Read – DI Curves

An example of P13 Sequential Read – DI Curves is shown in Figure 19**.**

P13 shows DI Curves for each TOIO value of the SEQ 256 KiB Read workload and the DI OS Curve.

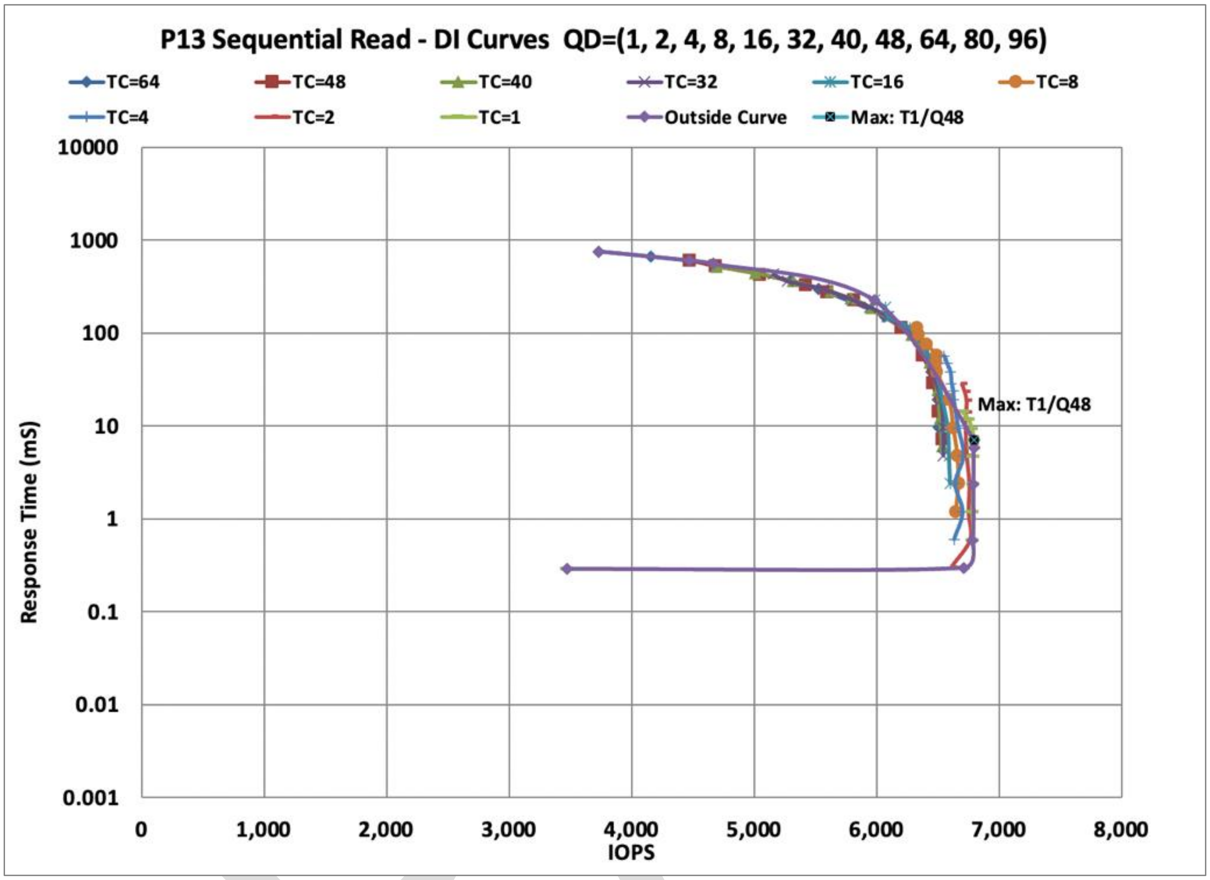

Figure 19 – Sequential Read – DI Curves

A.15 P14: All Outside DI Curves

An example of P14 All Outside Curves is shown in Figure 20**.**

P14 shows an Outside DI Curve for each workload.

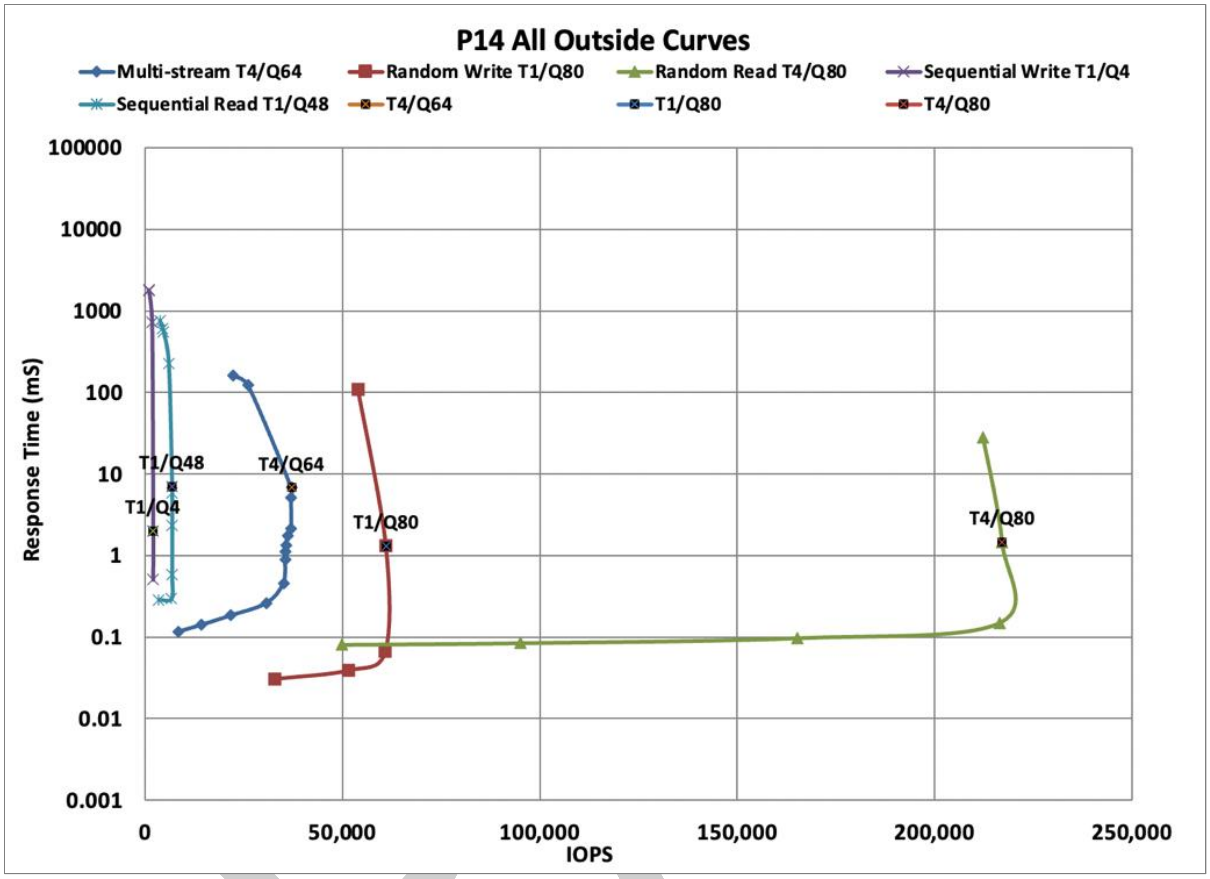

Figure 20 – All Outside DI Curves

A.16 P15: Ready Idle Power vs. Time

An example of P15 Ready Idle Ave Power v Time is shown in Figure 21**.**

P15 shows Ready Idle Power vs. Time.

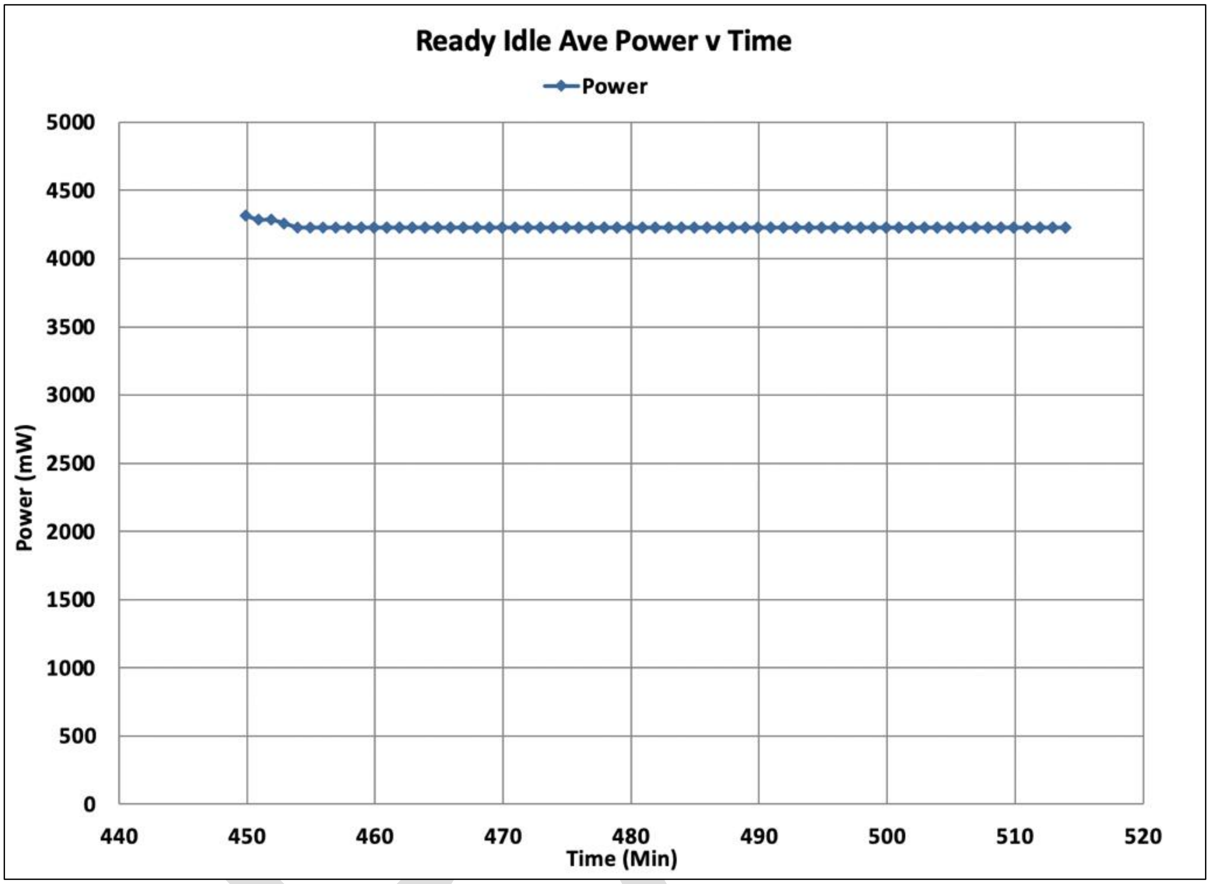

Figure 21 – Ready Idle Power vs. Time

A.17 P16: IOPS/Watt vs. Time

An example of P16 IOPS per Watt vs Time is shown in Figure 22**.**

P16 shows continuous IOPS/watt vs. time for each test step.

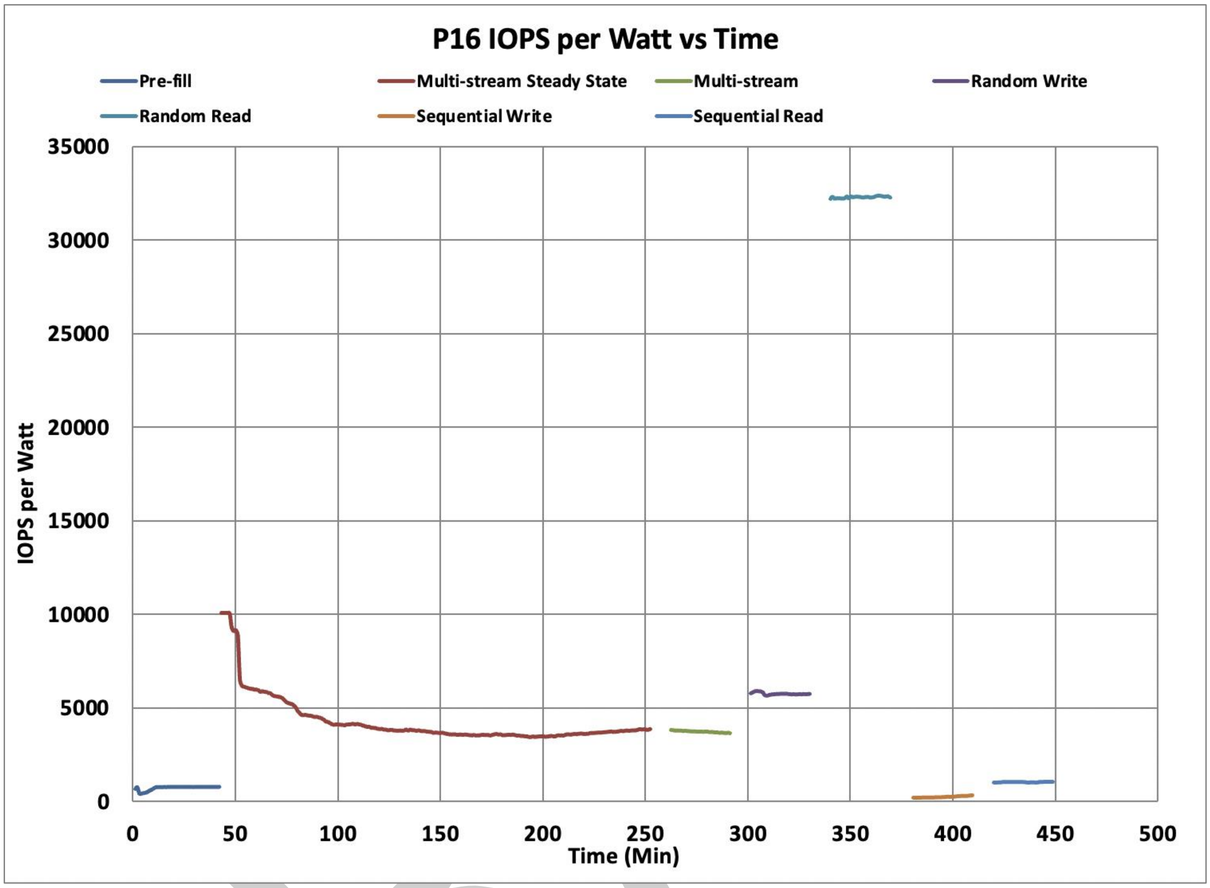

Figure 22 – IOPS/Watt vs. Time

A.18 P17: MiB/s/Watt vs. Time

An example of P17 MiB/s per Watt vs Time is shown in Figure 23**.**

P17 shows continuous MiB/s/watt vs. time for each test step.

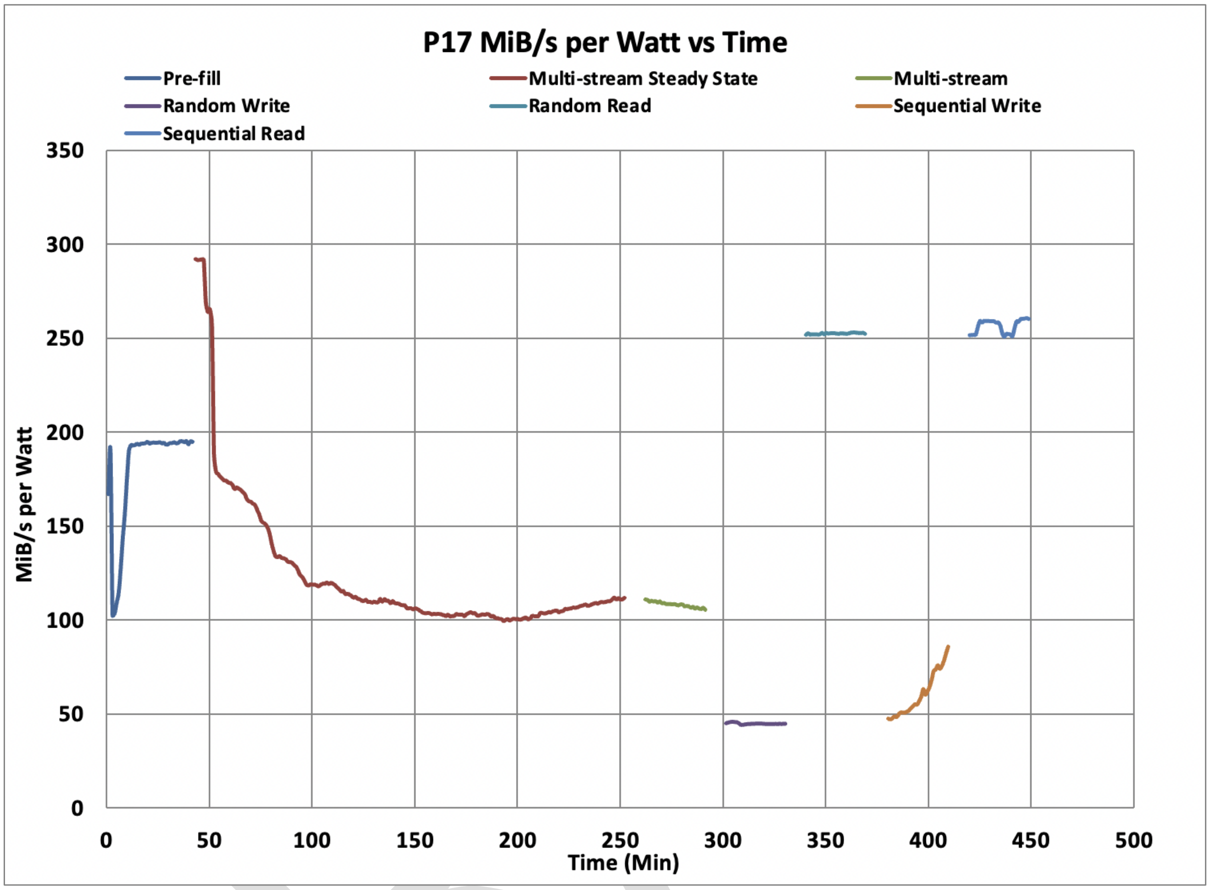

Figure 23 – MiB/s/Watt vs. Time

Annex B (Informative) Test Methodologies

B.1 Introduction

Principles guiding test processes derived from existing standards[7][8][5][6] include Steady State algorithms, block size sequencing, NAND flash-based FOB and write history, impact of thread count and queue depth and the impact of workloads on HDD and SSD device performance.

B.2 Test Flow and Methodology

The basic test flow and workloads are based on the ISO/IEC 24091:2019[7] and *SNIA Emerald™ Power Efficiency Measurement Specification Version 4.0.0*[8] for storage systems. Table 11 shows the correspondence between the names used for the workloads in this document and those used in ISO/IEC 24091:2019 and in *SNIA Emerald™ Power Efficiency Measurement Specification*.

Table 11 – Workload Name Mapping

ISO/IEC 24091:2019 and SNIA Emerald Specification	This Document
Hot Band IO Profile	Complex Workload
Random Read	RND 8 KiB Read Workload
Random Write	RND 8 KiB Write Workload
Sequential Read	SEQ 256 KiB Read Workload
Sequential Write	SEQ 256 KiB Write Workload
ready idle	Ready Idle Workload

The 12 hour Conditioning Run of standards[7][8] is modified to a minimum of 2.5 hours (five 30 minute rounds) up to a maximum duration of 12.5 hours (twenty five 30 minute rounds) depending on the Steady State determination criteria.

The 20 ms average response time ceiling of standards[7][8] is adopted for the Conditioning Run IOPS performance.

A self-optimizing TOIO sweep eliminates the need (implicit in [7] and [8]) to run multiple tests to determine the TOIO setting with the highest IOPS at less than 20 ms average response time.

B.3 Substitutable Test Workloads for Future Revisions

This basic test flow applies five specific workloads (Complex, Random 8 KiB Write, Random 8 KiB Read, Sequential 256KiB Write and Sequential 256 KiB Read) and is intended to provide easy comparison to power efficiency measurements defined by existing standards[7][8] for measuring storage system power efficiency.

The test flow, methodologies and reporting requirements of this document are structured to enable future revisions where one or more of the test workloads are substituted with new workloads deemed relevant and significant by industry and governing bodies. The Test Software allows any given workload to be substituted with a different test workload and provides the ability to set test parameters and settings for future tests released under this test methodology.

B.4 Steady State Algorithm

The **5 Rounds Steady State Determination** algorithm in the Conditioning run ensures the DUT achieves relatively time invariant performance of the Complex Workload.

B.5 Demand Intensity Curve

The Self-optimizing TOIO Sweep is based on a Demand Intensity Curve for IOPS and Average Response Times plotted as a function of varying TOIO.

Demand Intensity (DI), a.k.a., TOIO, is the product of Thread Count and Queue Depth for a given applied workload. The plotting of IOPS at different TOIO values against IOPS and average response time generates a Demand Intensity Curve. As DI is increased, the IO rate in IOPS and the average response times (ART) increases until the storage becomes “saturated”. At this saturation point, IOPS no longer increases (and may regress or “fold back”) while ART begins to dramatically rise.

By setting a given ART ceiling, the Test Software determines the Thread Count and Queue Depth settings that generate the highest IOPS for a given workload while staying below the designated ART ceiling.

Figure 24 shows a Demand Intensity Curve for an 1800 GB HDD. IOPS are on the x-axis with ART in ms is on the y-axis. The DI Threat Count points are 2, 4, 8, 16, 32, 48, 64, 80 and 96 with the ART ceiling shown at 20 ms. Queue Depth is set to 1 for each Thread Count in this example.

In this DI Curve, the optimal TOIO setting is Thread Count of 48 which yields 2559 IOPS at an ART of 18.75 ms. Accordingly, the measurement interval for this Complex Workload as determined by the self-optimizing TOIO sweep would be set at an TOIO of 48.

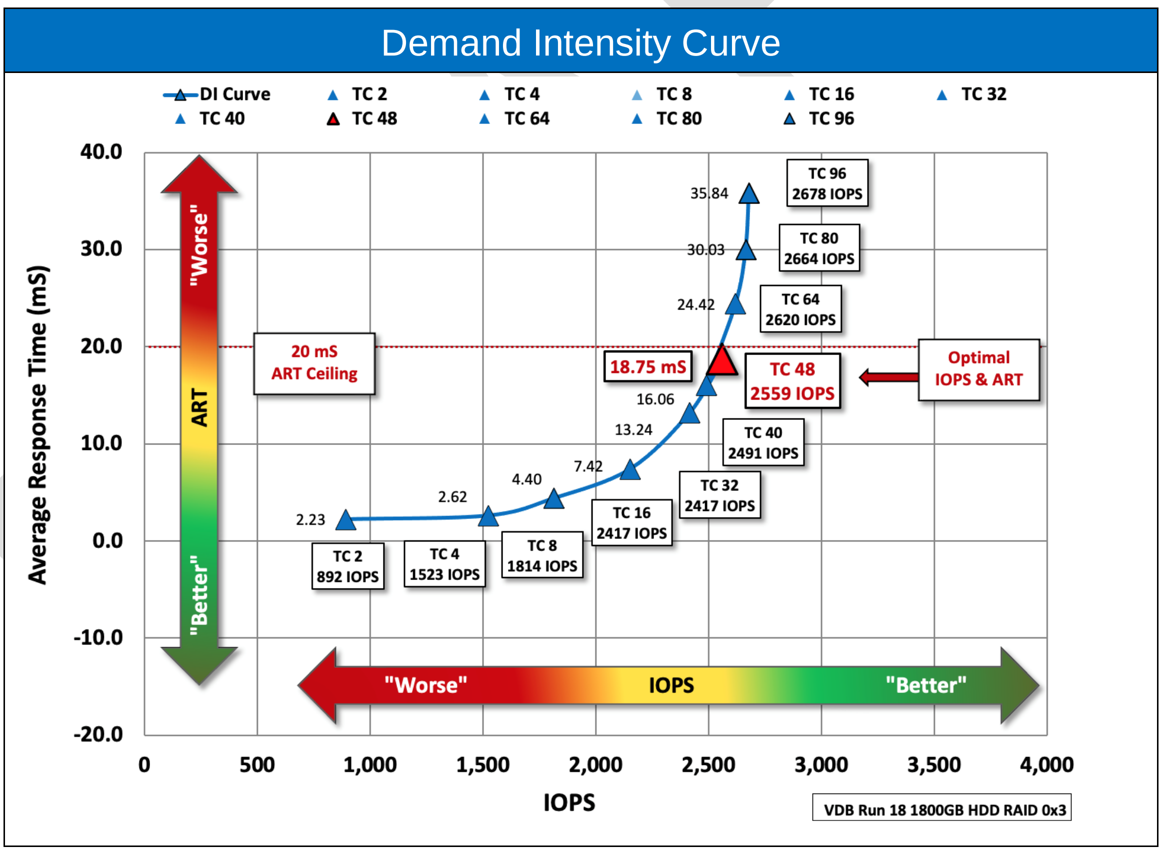

Figure 24 – Demand Intensity Curve

B.6 Demand Intensity Outside Curve

A Demand Intensity Curve is generated for each Thread Count and Queue Depth combination. Therefore, multiple DI Curves are generated for a multi-Thread Count, multi-Queue Depth self-optimizing TOIO sweep.

A Demand Intensity Outside Curve (DI OS Curve) is the best fit curve for all of the individual Thread Count/Queue Depth DI Curves. A DI OS Curve shows the increasing IOPS and ART relative to a given ART ceiling and is useful for observing the overall performance saturation for a given workload. Figure 25 is an example plot of an SSD device with a Complex DI Curve for each TOIO value.

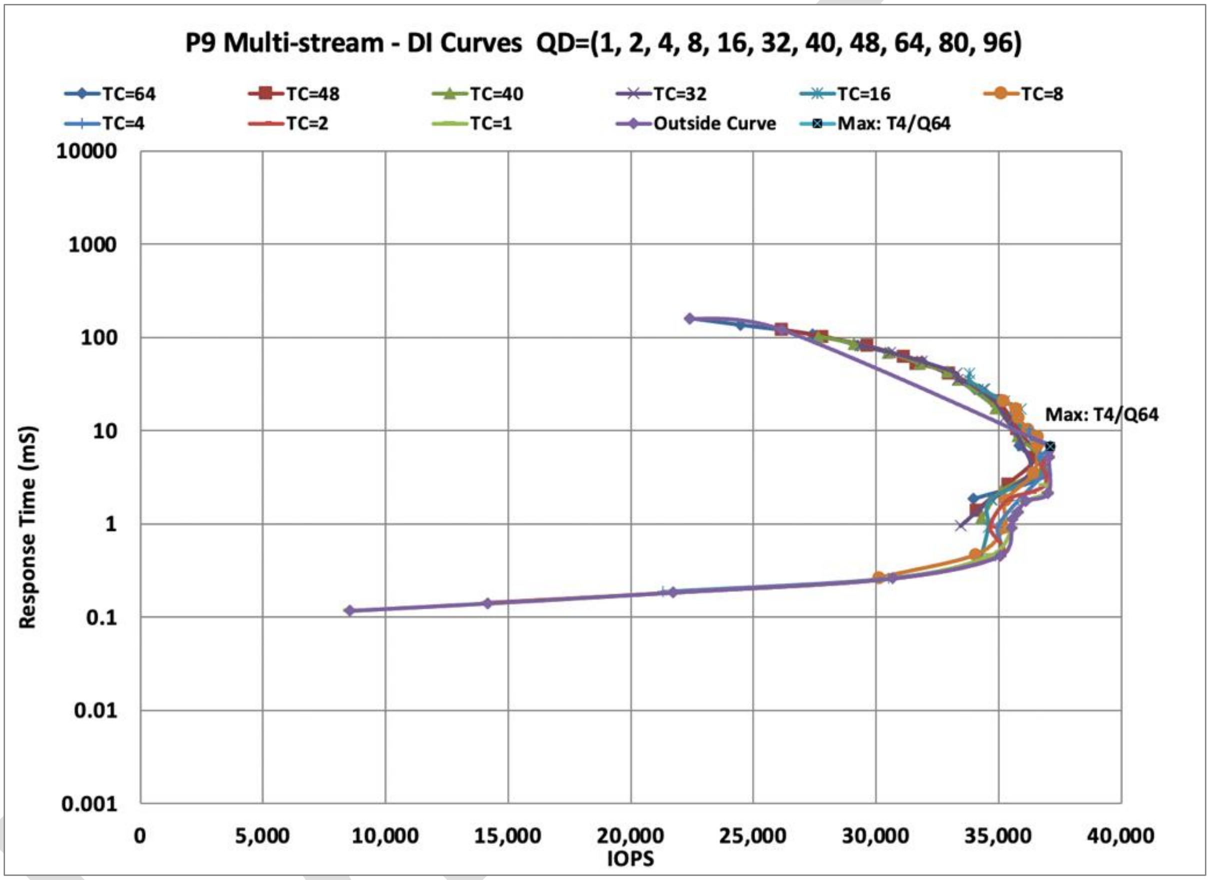

Figure 25 – DI Curve Complex

Figure 26 is an example DI OS Curve for a Complex Workload. The individual OS Curves are de-selected to only show the Complex DI OS Curve. The optimal TOIO value is 128 at T4/Q64.

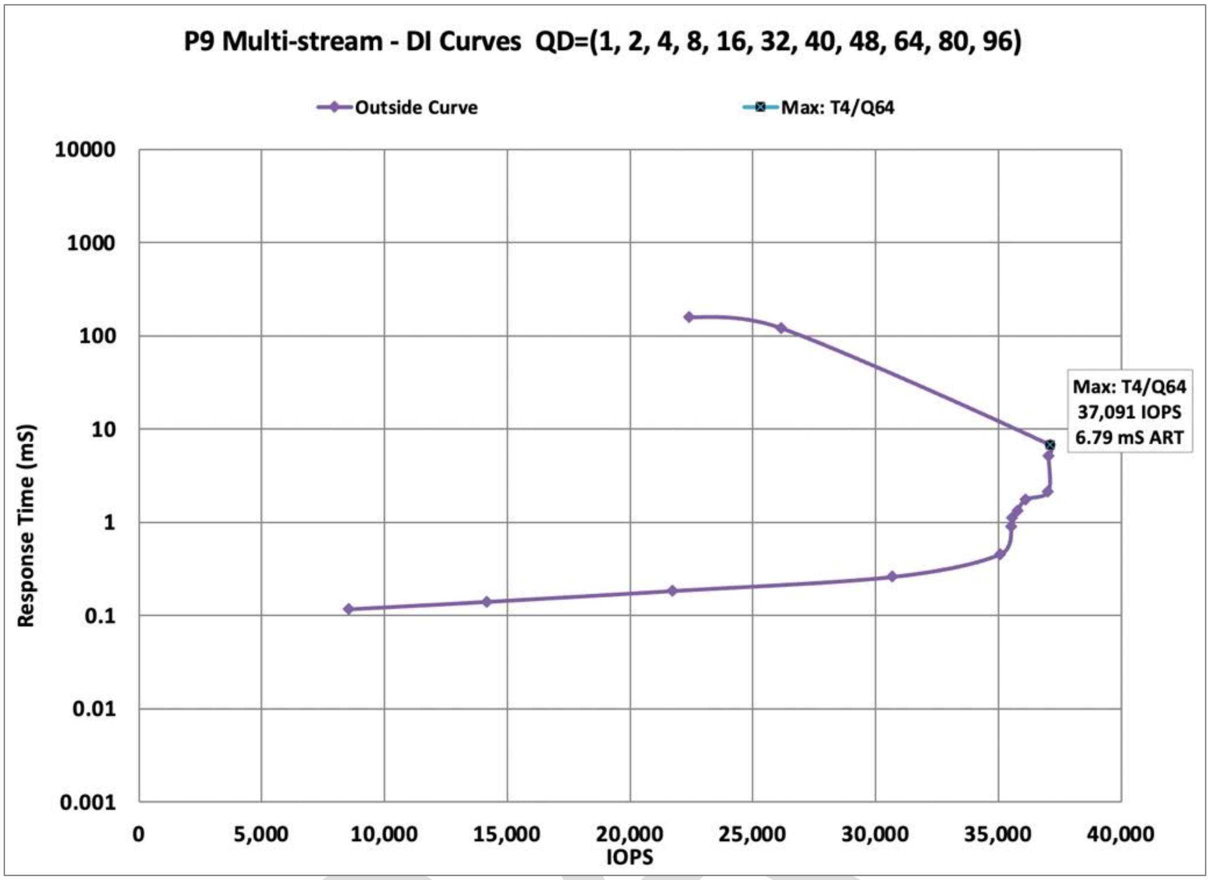

Figure 26 – DI OS Curve Complex

Figure 27 is an example All DI OS Curves (Complex, Random 8 KiB Write, Random 8 KiB Read, Sequential 256 KiB Write and Sequential 256 KiB Read).

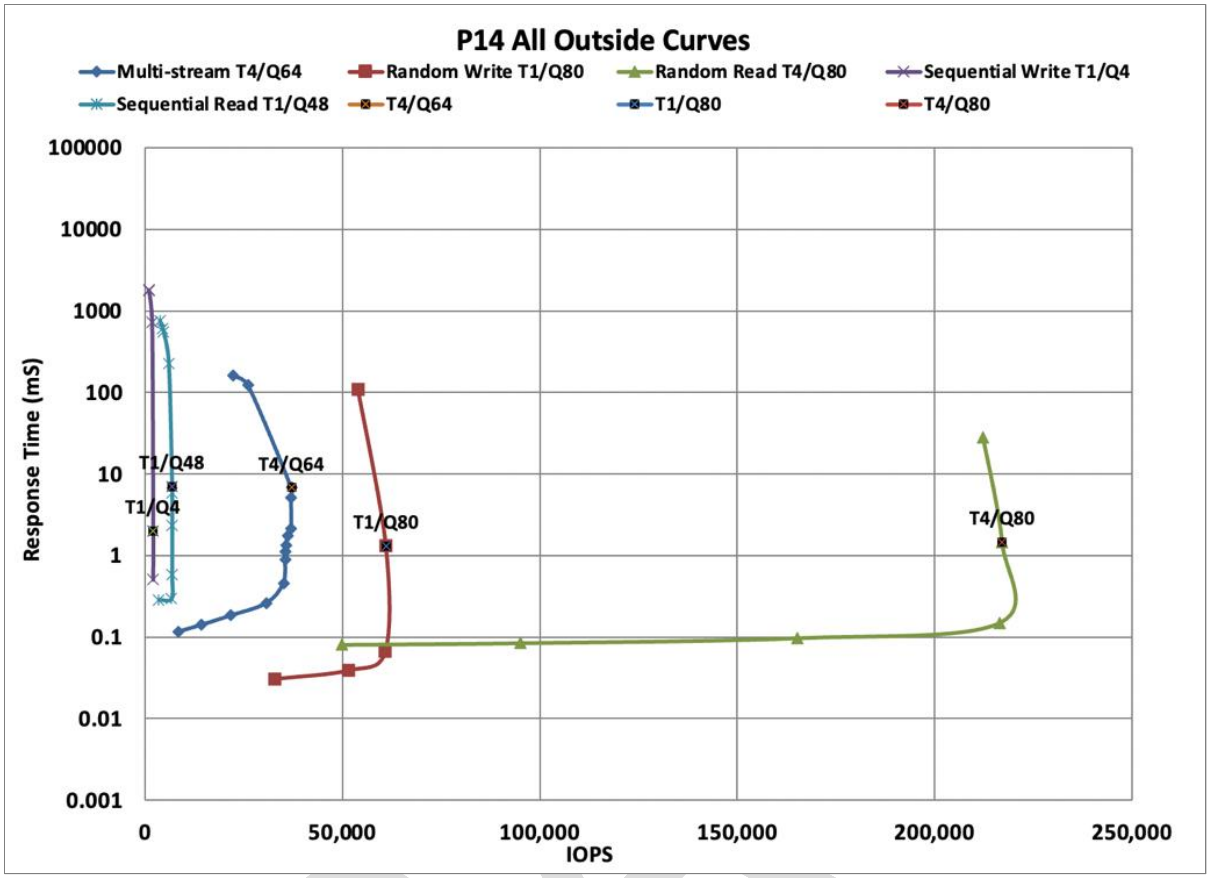

Figure 27 – All DI OS Curves

B.7 Report Generation

The Test Software automatically generates a read-only html audit file report and an editable excel report file. Both report formats provide required information for test review and process validation. The audit file report also provides a dynamic playback file.

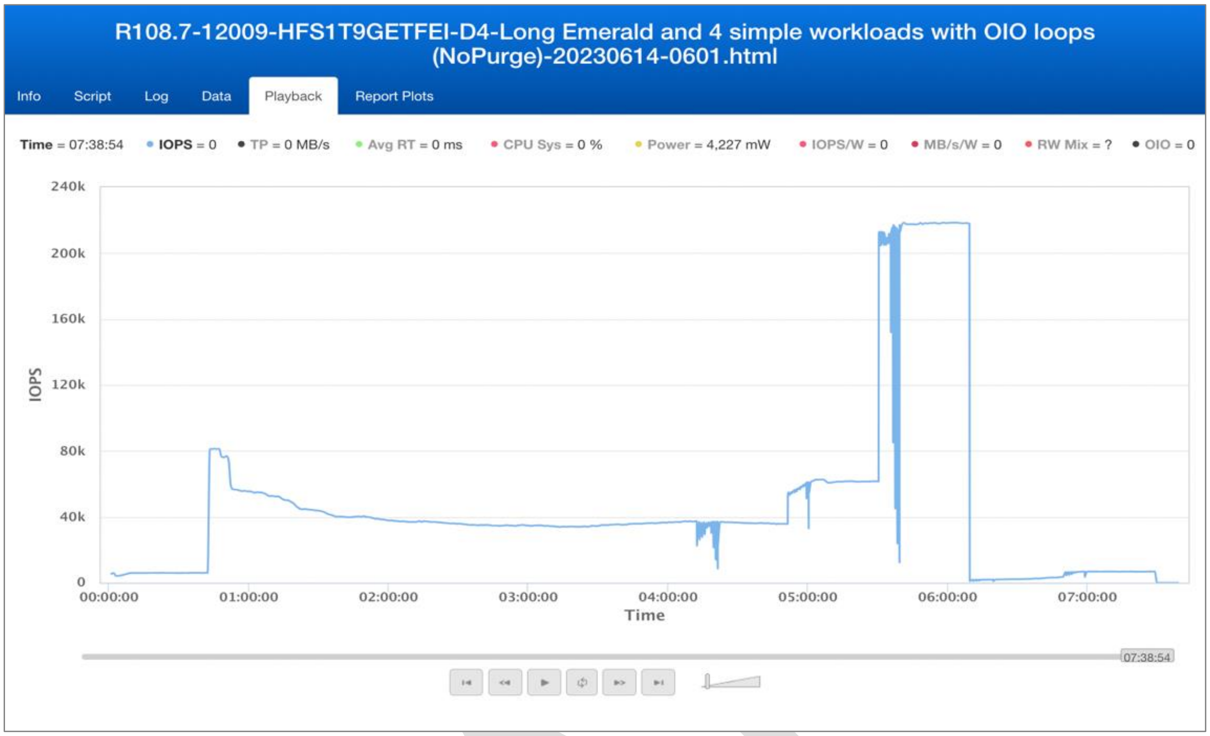

Figure 28 – Audit File Playback

Annex C (Normative) Workload Generator

C.1 Required Workload Generator

The Test Software (see Annex E) contains the Workload Generator that shall be used.

C.2 Workload Generator Requirements

The Workload Generator shall be able to:

1) test storage devices pursuant to the requirements of this document;
2) act as workload stimulus generator as well as data recorder;
3) issue Random and Sequential block level I/O;
4) restrict LBA accesses to a particular range of available user LBA space;
5) limit “total unique LBAs used” to a specific value;
6) randomly distribute a number of equally sized LBA segments across the test active range;
7) set R/W percentage mix %;
8) set Random/Sequential IO mix %
9) set IO Transfer Size;
10) set data pattern compression ratio;
11) generate and maintain multiple outstanding IO requests and ensure that all steps in the test sequence can be executed immediately one after the other, to ensure that devices are not recovering between processing steps, unless recovery is the explicit goal of the test;
12) provide output, or output that can be used to derive, IOPS, MiB/s, maximum latency and average response time (latency if TOIO=1) within some measurement period;
13) run a TOIO sweep (in Thread Count and Queue Depth), save results in an array, generate a Demand Intensity Curve (DI Curve) based on a 20 ms ART Ceiling, and set the workload segment TOIO to the values ascertained in the aforementioned DI Curve; and
14) integrate and aggregate power measurements (time stamp, power (W) and current (mA); and
15) All IOs issued shall complete successfully.

The random function for generating random LBA #’s during random IO tests shall be:

1) seedable;
2) have an output >= 48-bit; and
3) deliver a uniform random distribution independent of capacity.

Annex D (Normative) Required Test Software

CTS Lite is an IO performance test software product developed by Calypso Systems, Inc. This document describes the use of CTS Lite as the Workload Generator. CTS Lite consists of block access test software and associated workload generator. CTS Lite executes the test specified by this document.

The CTS Lite software specified on the *Download Material for SNIA Emerald™Testing*[10] web page for use with this document shall be used as the Workload Generator. The web page also provides access to the CTS Lite device user guide and other related testing materials.

CTS Lite single seat and enterprise licenses are available for purchase from Calypso Systems, Inc.[14] The CTS Lite web page[14] shows license terms and fees and provides access to the CTS Lite software and related materials.

Annex E (Normative) Pseudo-Code Language

E.1 Purpose

The purpose of this clause defines the pseudo-code language used to define the tests executed by the Test Software.

E.2 Pseudo Code Language

The specification uses an informal pseudo code to express the test loops. It is important to follow the precedence and ordering information implied by the syntax. In addition to nesting/indentation, the main syntactic construct used is the “For” statement.

A “For” statement typically uses the syntax: For (variable = x, y, z). The interpretation of this construct is that the test sponsor sets the variable to x, then performs all actions specified in the indented section under the “For” statement, then sets the variable to y, and again performs the actions specified, and so on. Sometimes a “For” statement will have an explicit “End For” clause, but not always; in these cases, the end of the For statement’s scope is contextual.

Take the following loop as an example:

```
For (TC in [64, 48, 40, 32, 16, 8, 4, 2, 1])
    For (Queue Depth in [96, 80, 64, 48, 40, 32, 16, 8, 4, 2, 1])
        1) Execute random IO, per (TC, Queue Depth), for 1 minute
        2) Record Ave IOPS (TC, Queue Depth)
```

This loop is executed as follows:

1) Set Thread Count to 64 >>>>> Beginning of outer loop, iteration 1
2) Set Queue Depth to 96 >>>>> Beginning of inner loop, iteration 1
3) Execute random IO…
4) Record Ave IOPS…
5) Set Queue Depth to 80
6) Execute…
7) Record…
8) …
9) Set Queue Depth to 1
10) Execute…
11) Record… >>>>> End of inner loop, iteration 1
12) Set Thread Count to 48 >>>>> Beginning of outer loop, iteration 2
13) Set Queue Depth to 96
14) Execute…
15) Record…
16) …

Annex F (Normative) Power Analyzer

F.1 Purpose

The purpose of this clause is to outline the requirements for device level Power Analyzer for measuring power consumption in storage devices. The Power Analyzer will be used to accurately capture and analyze the dynamic power characteristics of storage devices tested pursuant to this specification.

F.2 General Description

The Power Analyzer incorporates an independent power supply solely for the DUT, the power measurement software, and the DUT interconnect cable or interposer board.

The Power Analyzer that shall be started prior to a test and continue to be active throughout the test and shall record:

- Input voltage to the DUT and
- Power consumption by the DUT.

The power and voltage measurements shall be recorded to durable media using a period of not more than 1 s and shall use a timestamp that is synchronized the test performance data with the Power Analyzer power data to a resolution of at least 1 s.

F.3 General Specifications

Table 12 defines the specifications that the Power Analyzer shall meet.

Table 12 – Power Analyzer Requirements

Power Analyzer Specifications		
Item	**Requirement**	**Description**
DUT Power Supply Voltage Range	3.3 V, 5 V, 12 V	Should cover the voltage levels of SSD & HDD storage devices during workload measurement steps and ready idle steps (see clause xxx).
DUT Power Supply Current	An appropriate external supply, or the host power may be used to provide power to the DUT	Should cover the current levels of the DUT during read, write, and idle states.
Meter Voltage Range	0.1 V to 14.4 V	Should cover the voltage levels of SSD & HDD storage devices during workload measurement steps and ready idle steps (see clause xxx) and also cover the full operating range of the drive
Meter Current Range	100 μA to xx A	Should cover the current levels of the DUT during read, write, and idle states.
Voltage Accuracy	2 mV ± 1% of reading	Accuracy should be maintained across the specified voltage and current operating ranges.
Current Accuracy	2 mA ± 1% (25 uA + 1% below 1 mA)	Accuracy should be maintained across the operating voltage range

Meter Sampling Rate	Minimum 250 KHz	Adequate to capture rapid changes in power consumption during DUT operations.
Meter Bandwidth	Minimum 250 KHz	Sufficient to cover the frequency components of signals produced by DUTs.
Meter Resolution	1 mV, 1 µA	High resolution to capture small changes in power consumption during DUT operations.
Number of Channels	Minimum 1 channel	Sufficient channels to capture power consumption across multiple voltage and current rails in DUT.
Integration with Data Acquisition Systems	Integration with the Test Software	Compatibility with software for efficient data analysis and storage.
Communication Interfaces	USB, Ethernet, or other interfaces for data transfer and control	
Data Output	csv or other file format compatible with the Test Software	
Environmental Conditions - Temperature	operate within an ambient environment with temperature in the range 0 °C to 40 °C	
Environmental Conditions – Relative Humidity	operate within an ambient environment with relative humidity in the range 20 % to 80 % relative humidity, non-condensing	

F.4 Recommended Power Analyzers

A Power Analyzer from the list in Table 13 shall be used.

Table 13 – Recommended Power Analyzers

Recommended Power Analyzers		
Manufacturer	**Model**	**Description**
Quarch Technology	QTL1999	Single port Programmable power module
Quarch Technology	QTL1995	Multi-power Programmable Power Module
Quarch Technology	QTL2312	Single port Power Analysis Module

Bibliography

[1] SNIA (2018). SFF-8301 *Specification for 3.5" Form Factor Drive Dimensions.* Available from https://www.snia.org/technology-communities/sff/specifications.

[2] SNIA (2018). SFF-8201 *Specification for 2.5" Form Factor Drive Dimensions.* Available from https://www.snia.org/technology-communities/sff/specifications.

[3] SNIA (2024). SFF-TA-1009 *Enterprise and Datacenter Standard Form Factor Pin and Signal Specification.* Available from https://www.snia.org/technology-communities/sff/specifications.

[4] SNIA (2017). SFF-8639 *Specification for Multifunction 6X Unshielded Connector.* Available from https://www.snia.org/technology-communities/sff/specifications.

[5] SNIA (2020). SNIA Standard *Solid State Storage (SSS) Performance Test Specification (PTS) Version 2.0.2.* Available from https://www.snia.org/tech_activities/standards/curr_standards/pts.

[6] SNIA (2018). SNIA Standard *Real World Storage Workload (RWSW) Performance Test Specification for Datacenter Storage Version 1.0.7.* Available from https://www.snia.org/rwsw.

[7] ISO (2019). International Standard ISO/IEC 24091:2019 *Information technology — Power efficiency measurement specification for data center storage.* Available from https://www.iso.org/.

[8] SNIA (2020). SNIA Standard *SNIA Emerald™ Power Efficiency Measurement Specification Version 4.0.0.* Available from https://www.snia.org/tech_activities/standards/curr_standards/emerald.

[9] SNIA. *The SNIA Dictionary*. Available from https://www.snia.org/dictionary.

[10] SNIA. *Download Material for SNIA Emerald™Testing*. Available from https://www.snia.org/emerald/download.

[11] SNIA. *User Guide for the SNIA Emerald™ Power Efficiency Measurement Specification.* Available from https://www.snia.org/emerald/download.

[12] *The gzip home page*. Available from https://www.gzip.org/.

[13] SNIA. *Test Data Report Template.* Available from https://www.snia.org/emerald/download.

[14] Calypso Systems, Inc. *CTS Lite.* Available from https://www.ctslite.com.