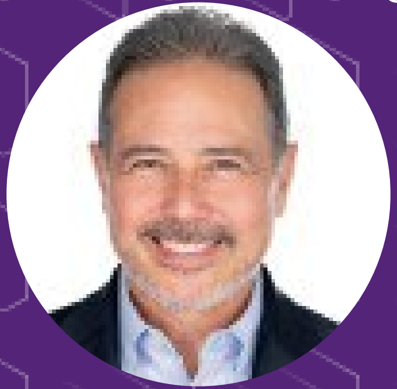
SNIA | Compute, Memory, and Storage

Unlocking Sustainable Data Centers: Optimizing SSD Power Efficiency and Liquid Cooling for AI Workloads

SNIA CMS Webinar
Tuesday, June 10, 2025
10:00 am PT

Nicole Ross, Samsung

Jonmichael Hands, Solidigm



Cameron T. Brett, KIOXIA America

Our Presenters

Jonmichael Hands

Co-Chair

SNIA SSD Special Interest Group

AI Technologist, Solidigm

Nicole Ross

SSD Architect,
Samsung Semiconductor

Hosted by:

Cameron Brett

Co-Chair

SNIA SSD Special Interest Group

Director of Marketing,
KIOXIA America, Inc.

The SNIA Community

200
Corporations, universities, startups, and individuals

2,500
Active contributing members

50,000
Worldwide IT end users and professionals

SNIA | CMS
A SNIA Community

SNIA | Compute, Memory, and Storage

What We Do

- Engage technology users
- Educate on compute, memory, and storage technologies
- Accelerate SNIA standards
- Propel technology adoption

How We Do It

- Demonstrate at industry events
- Host Programming Workshops and Hackathons
- Present educational webinars, podcasts, and blogs

Learn more at www.snia.org/groups/cms

SNIA Solid State Drive (SSD) Special Interest Group (SIG)

- The place to go for information on:
  - SSD Form Factors
    - snia.org/forums/cmsi/knowledge/formfactors
  - NVMe SSD Classification
    - snia.org/forums/cmsi/knowledge/nvme-ssd-classification
  - Total Cost of Ownership (TCO)
    - snia.org/forums/cmsi/programs/TCOcalc
- Why you should join
  - Engage and educate the industry
  - Expand reach of SSD technologies to end users
- Learn more
  - askcms@snia.org

SNIA | CMS A SNIA Community

SNIA Legal Notice

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be, construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

- Sustainability definition, PUE
- Data center challenges
- Shift to liquid cooling
- How to measure SSD power efficiency, perf/W, TB/W
- Total Cost of Ownership (TCO)
- Why Power Efficiency Matters
- NVMe® Power Management and Power States
- SSD Support and Innovation
- Power Scheduling and Optimization Examples
- Key Takeaways

SNIA | CMS A SNIA Community

Jonmichael Hands

SNIA / OCP Partner Update - Sustainability Project

Transparency, Reporting and Metrics

For data center operators: Reporting on energy and water usage and carbon (GHG) emissions - scope 1, 2, and 3

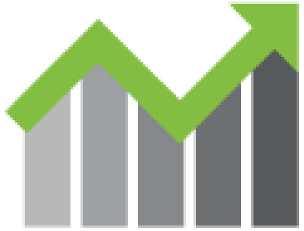
For suppliers: focus on Life Cycle Assessments (LCA) & upstream reporting accuracy

Circularity

Materials maintaining their highest value possible

Products are designed to extend the use period of a product and consider the next use

Extension of use (life), reuse, repair, refurbish, remanufacture, disassembly, and recycling



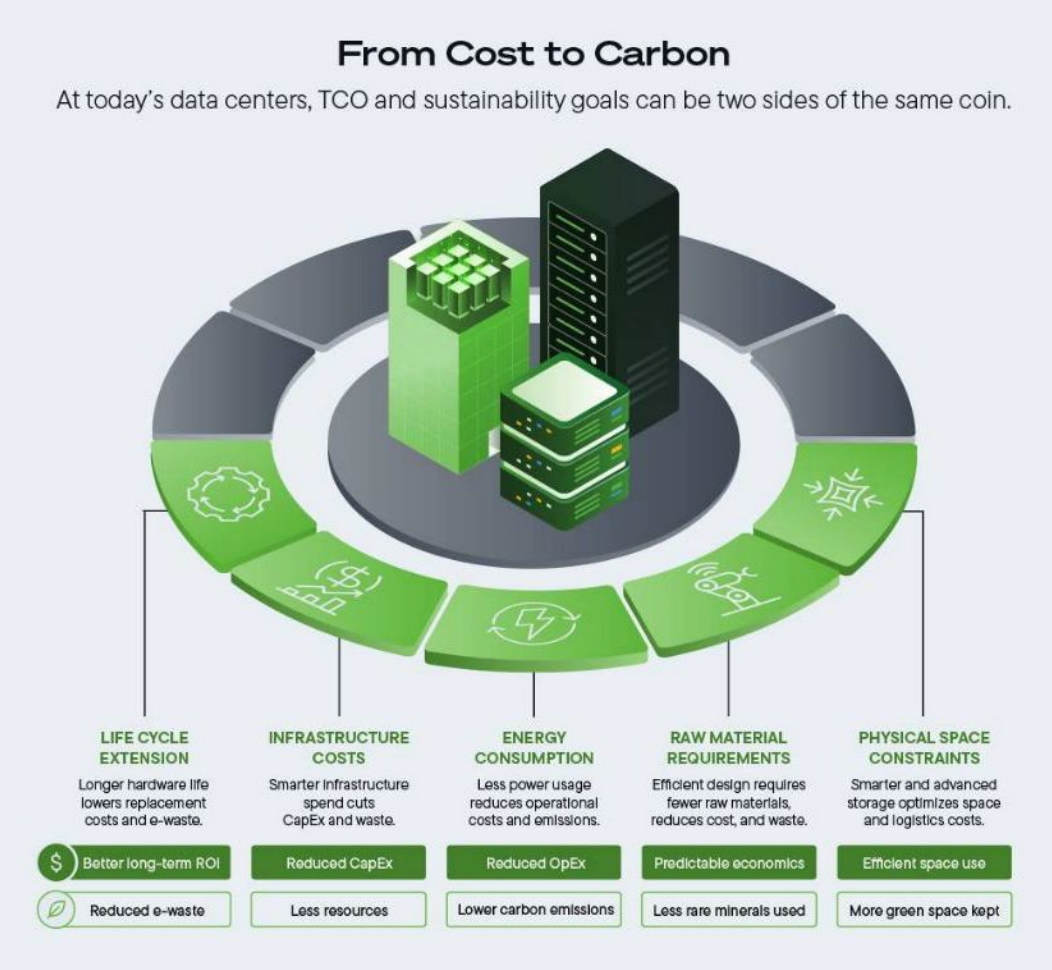
Efficiency & Interoperability

Efficiency metrics beyond PUE and focus on impact of reporting, and gen over gen improvements

OCP standard firmware for multiple customers, open source tools. Hardware building blocks for servers and racks

SNIA | CMS A SNIA Community

Power Efficiency vs Embodied Carbon



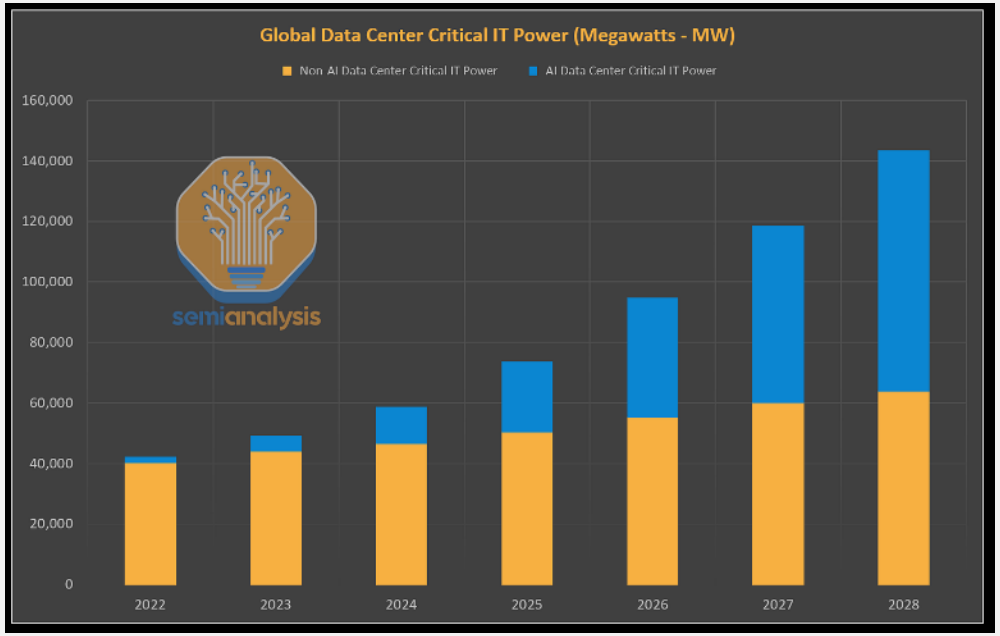
Storage Media	Embodied Carbon by Product (Kg CO_2)	Embodied Carbon per TB (CO_2/TB)	Embodied Carbon per TB per Year (CO_2/TB/year)
SSD[2]	4,915	160	32
Hard Drive[3]	29.7	<1	<0.2
LTO Tape[4]	48	2.66	<0.6

Source: https://www.seagate.com/resources/decarbonizing-data-report/

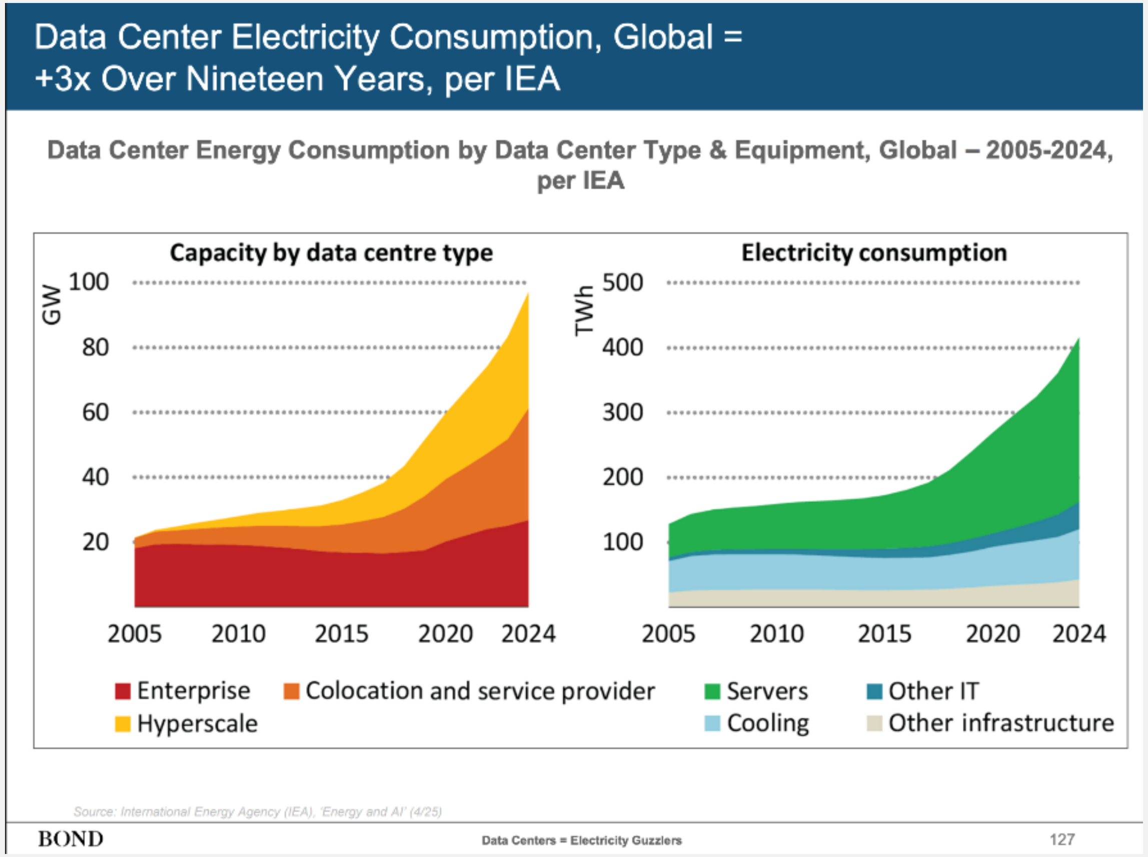
SNIA | CMS A SNIA Community

Finding AI Energy in an Already Challenged Grid

Projected **Non-AI** vs **AI** Data Center Energy Use (MW)[1]



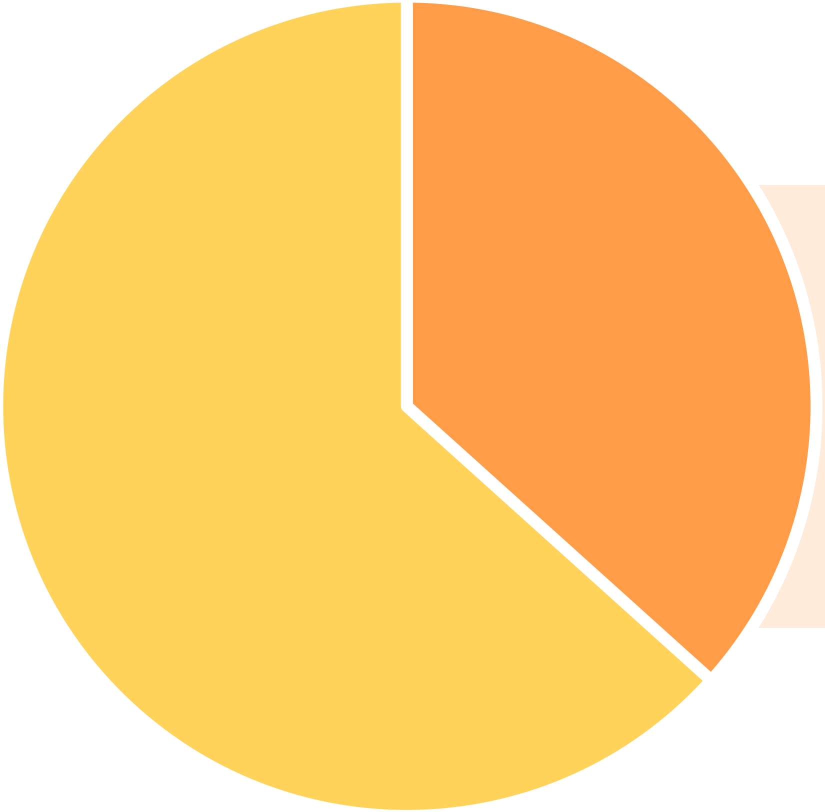
>50% of data center critical IT power[2] will be AI-driven by 2028



[1] Source – SemiAnalysis, Mar 2024
[2] Source – SemiAnalysis defines critical IT power as the power capacity required to support the essential IT infrastructure within datacenters, particularly focusing on the demand driven by AI accelerators
[3] Source – Bond Trends – Artificial Intelligence

 |

Storage is a **significant DC power component**



Microsoft Azure + Carnegie Mellon Study[1]

Storage-related emissions – including storage racks and local storage devices – make up **33% of operational emissions**

"We would probably build out bigger clusters than we currently can if we could get the energy to do it."
Mark Zuckerberg, Meta[2]

[1] Source - Microsoft and Carnegie Mellon University
[2] Source – Data Center Dynamics

SNIA | CMS A SNIA Community

AI Datacenters are Distinctly Different from Traditional Datacenters

Feature	Traditional Data Center	AI-Focused Data Center
Compute Hardware	x86 CPUs (Intel, AMD)	GPUs (NVIDIA), TPUs (Google), Accelerators (ASIC, NPU)
Key Workloads	Web hosting, databases, ERP/CRM, cloud VMs and containers	AI/ML model training, inference
Memory Architecture	DRAM	High Bandwidth Memory (HBM) on accelerators + DDR system DRAM
Backend Networking	Standard Ethernet (10-100 Gbps)	High-Speed Fabrics (InfiniBand, 400/800G+ Ethernet w/ RoCE, Spectrum-X)
Rack Power Density	Low-Moderate (5-20 kW/rack)	High-Extreme (40-120+ kW/rack)
Primary Cooling	Air Cooling (CRAC/CRAH, Economizers)	Liquid Cooling (Direct-to-Chip, Immersion) essential
Key Software Stack	OS, Databases, Enterprise Apps, Virtualization (VMware)	ML Frameworks (PyTorch, TF), CUDA, Kubernetes
Security Focus	Infrastructure Protection, Data Confidentiality/Integrity	ML Lifecycle Security (Data Poisoning, Model Theft, Adversarial)
Economics (Drivers)	Balanced CapEx (Servers/Storage/Network), OpEx (Power)	CapEx dominated by Accelerators/Networking, OpEx by Power

SNIA CMS A SNIA Community

OCP Ready™ Data Center Requirements for AI/ML

- Higher-Power Density Racks (65kW-136 kW)
- Liquid Cooling Solutions (Cold Plate, Door Heat Exchange, Immersion)
- Optimized Power and Networking
- Enhanced Data Center Physical Infrastructure
- Sustainability Metrics and Reporting

OCP AI/ML Physical Infra Workshop 1

600kW to 1MW racks are coming

SNIA | CMS A SNIA Community

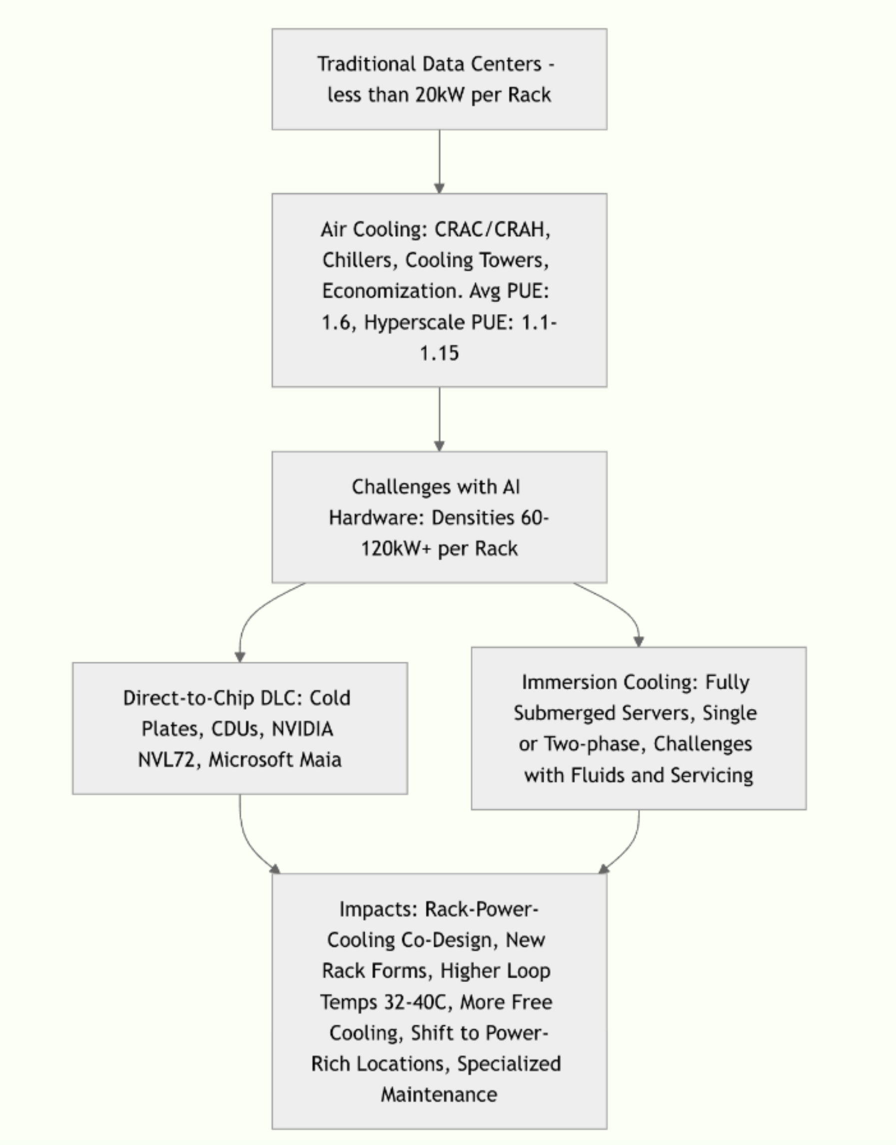
Meta Prineville Data Center

LEED **Gold** certification

Our operational data centers, on average, in 2023 exhibited a Power Usage Effectiveness (**PUE**) of **1.08** and Water Usage Effectiveness (**WUE**) of **0.18**.

Cooling

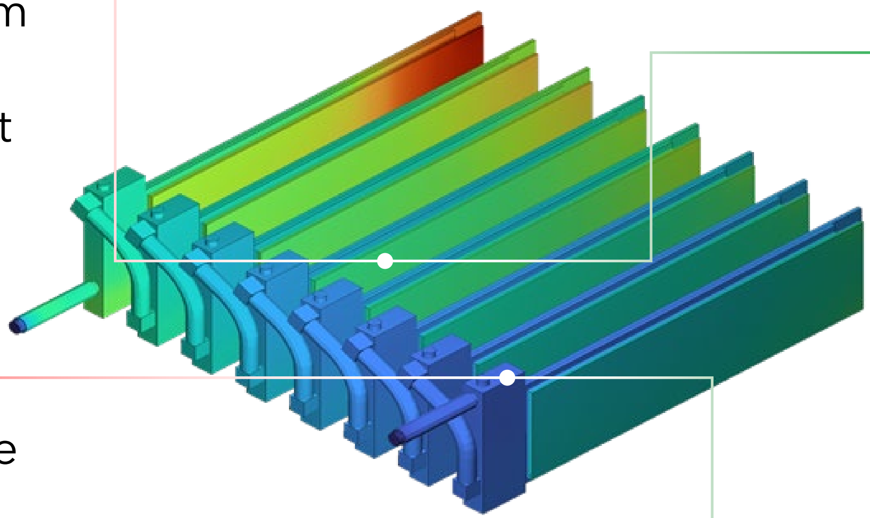
- Direct-to-Chip (DLC)
  - Circulates coolant through cold-plates mounted directly on GPUs/CPUs. Heat is removed to a facility loop via CDUs. DLC already ships with NVIDIA's 120 kW NVL72 racks
- Immersion Cooling
  - Servers are submerged in dielectric fluid – single-phase (liquid only) or two-phase (boiling + condensation). Efficiency is high but servicing, material compatibility and cost remain barriers
- Solve rack power density and efficiency (PUE, IUE)



Cooler Than the Other Side of the Pillow

Problem

SSDs using liquid cooling dissipate heat primarily from the side of the drive touching the cold plate, but both sides get hot.



Solution

The world's first SSD enclosure with a thermal solution that actively cools both sides of an E1.S drive using a single cold plate.

Problem

SSDs using liquid cold plate cooling are not easily hot-swappable, increasing maintenance cost.

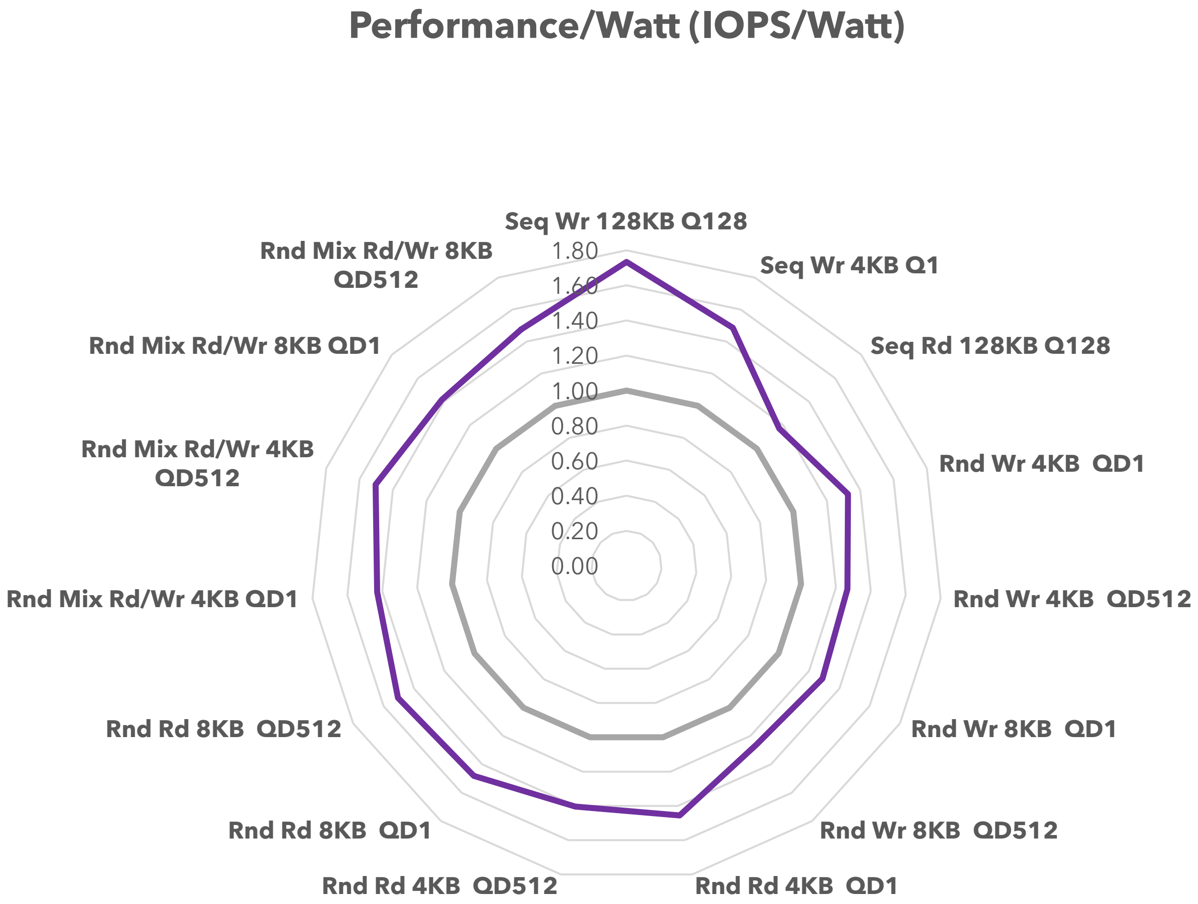
Solution

A unique, spring-loaded mechanism that allows for quick release and swapping.

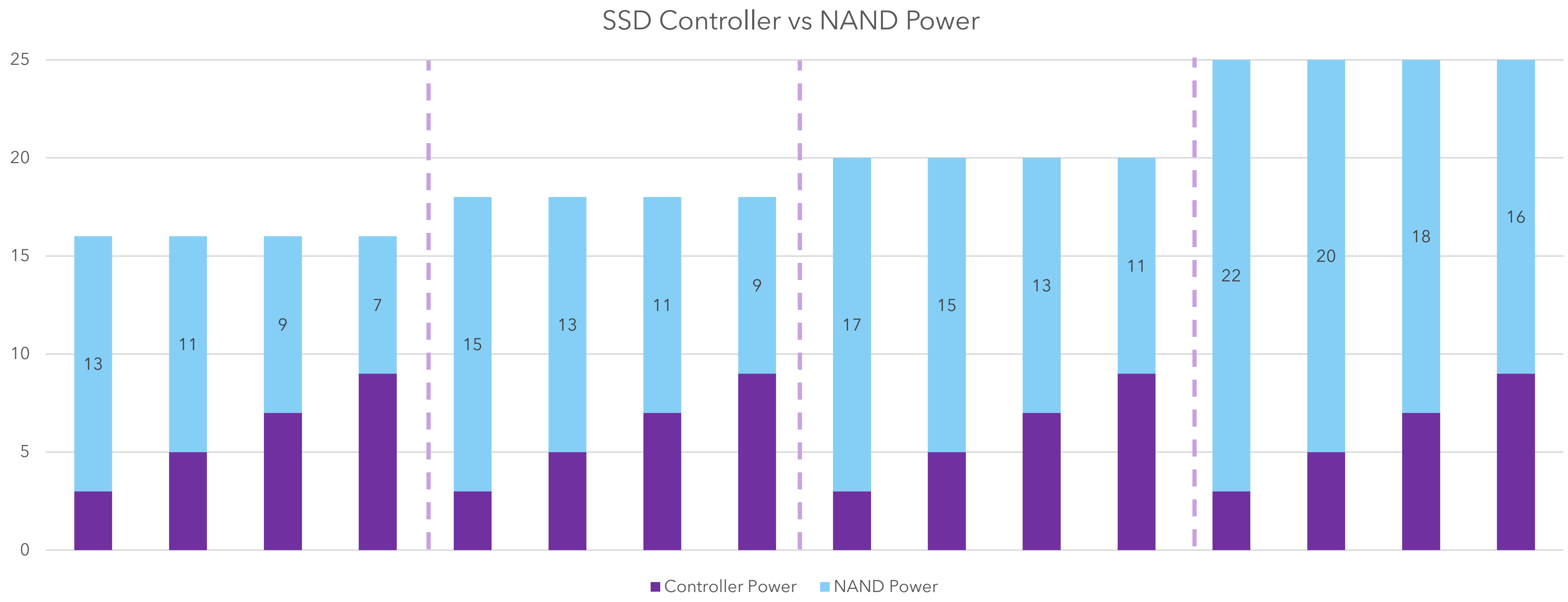
Fanless designs enable **more compact servers and reduced facility-level ambient air conditioning**

Efficiency for SSDs

- Compute SSDs – IOPS/W, Bandwidth/W
- Storage SSDs – TB/W, TCO
- Consumer SSDs – idle power / battery life



Controller vs NAND Power



New OCP 2.6 DSSD NVMe SSD Power States

Inverse Numbering vs. NVMe:
- Higher-numbered **DSSD states = higher power**
- Higher-numbered **NVMe states = lower power**

Simple Watt-Based Mapping:
- Set Features → DSSD PS N ⇒ controller drops to **highest NVMe PS whose Max Power ≤ N watts**
- Gives host granular watt-level control without expanding NVMe PS table

More DSSD States than NVMe States:
- Fills gaps between NVMe power ceilings (e.g., DSSD 18 W maps to NVMe PS 6 @ ≤ 16 W)
- If DSSD value isn't an exact NVMe MP, device auto-selects closest lower NVMe state

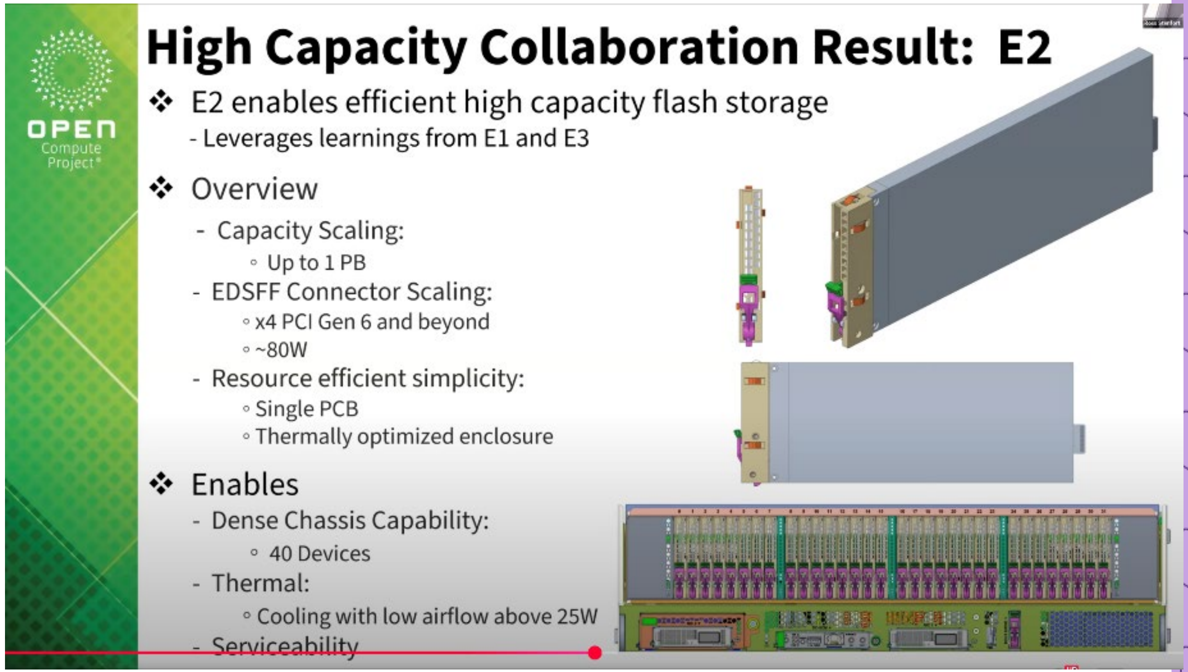
Special Values & Validity:
- **DSSD 0** ≠ allowed (0 W = power-off)
- Lowest valid DSSD state defined in **DCLP-8**
- States below the lowest NVMe PS power are invalid

Descriptor Format (per state):
- Bit 7 = Valid flag (1 = ≥ Minimum Valid DSSD)
- Bits 4-0 = Mapped NVMe Power State index

Visibility & Reporting:
- Device reports both its current **DSSD PS** (watts) and **NVMe PS** (index) to the host

SNIA | CMS A SNIA Community

Hyperscale TCO – HDD vs SSD

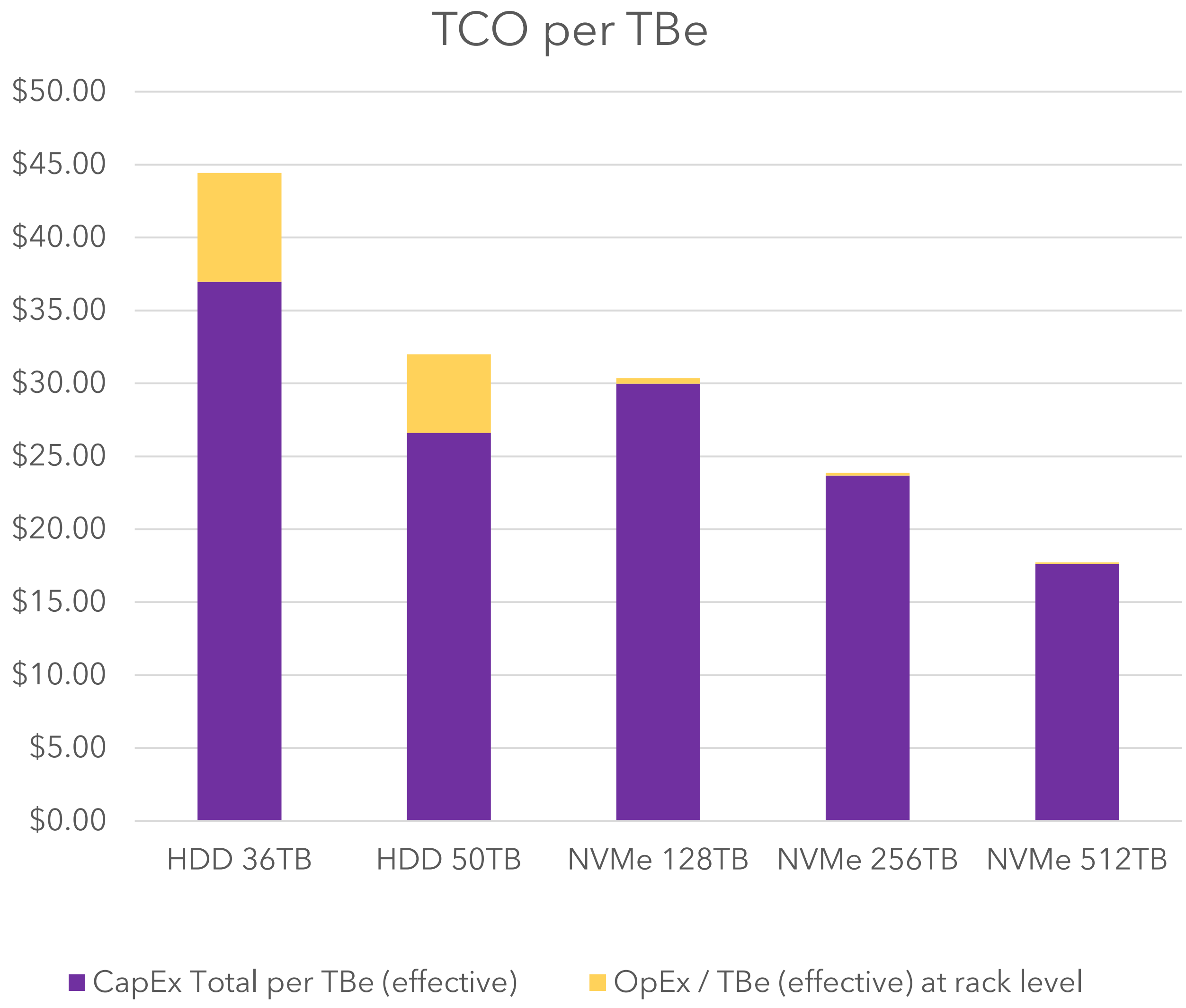


Large Capacity SSD Trends

- 122TB SSDs shipping
- E2 form factor designed for 2U max density
- QLC adoption growing (hyperscalers, enterprise, AI)
- Features to minimize write amplification (e.g. NVMe ZNS, FDP)
- Cost reduction (lower DRAM) with larger IU size
- Large capacity SSDs improve TB/W
- System + SSD improves PB/rack
- SSD BW/TB currently >>> HDD, TBD targets for >512TB SSDs

Cloud Storage TCO Model for 2027

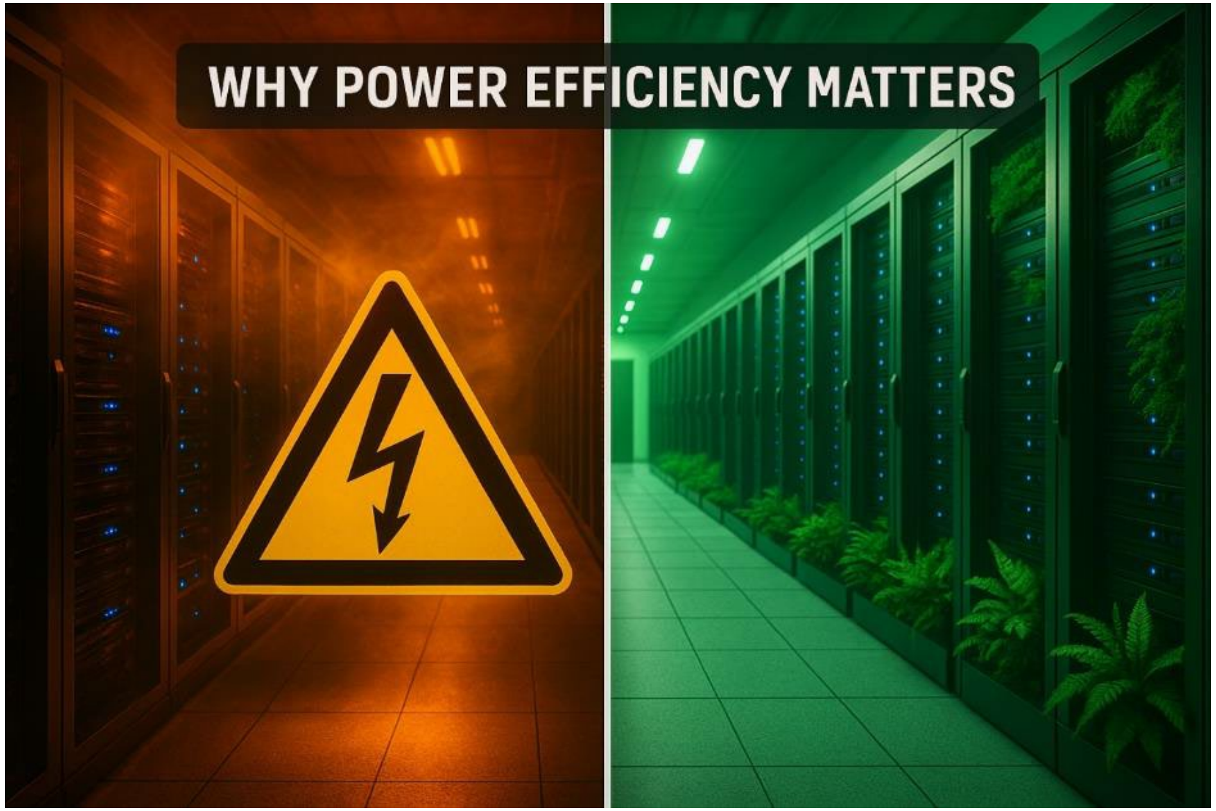
- Showing path for warm storage
- Compression, data reduction improved on SSD
- Reduced OpEx % of TCO
- EB raw per MW: 15x between 36TB HDD and 512TB SSD



Nicole Ross

Why Power Efficiency Matters[1]

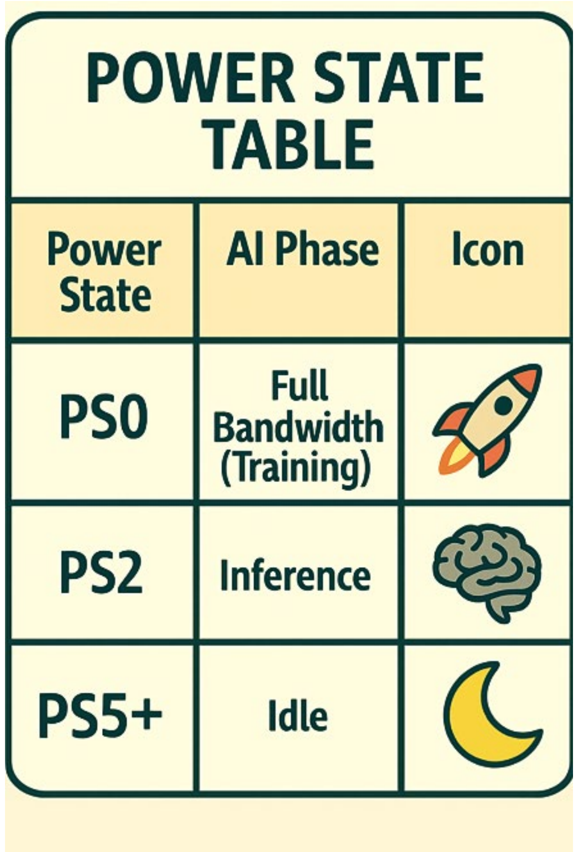
- Explosive growth of AI requires >GB/s sustained throughput
- GPU dense servers exhaust the cooling capacity
- Storage forced to operate with leftover thermal envelope
- Data center spends 40% of operating budget on power[1]
- Align with sustainability goals and regulations



1. U.S. Chamber of Commerce
https://www.uschamber.com/assets/archived/images/ctec_datacenterrpt_lowres.pdf

NVMe® Power Management and Power States

- Historically use:
  - Rack level power budgeting
  - Idle state efficiency
- 32 Power States defined by SSD vendor
- Each state caps maximum drive power and defines entry/exit latencies
- Host can switch states explicitly via Set Features
- Proposed Addition: Workload Aware Power States/Profiles



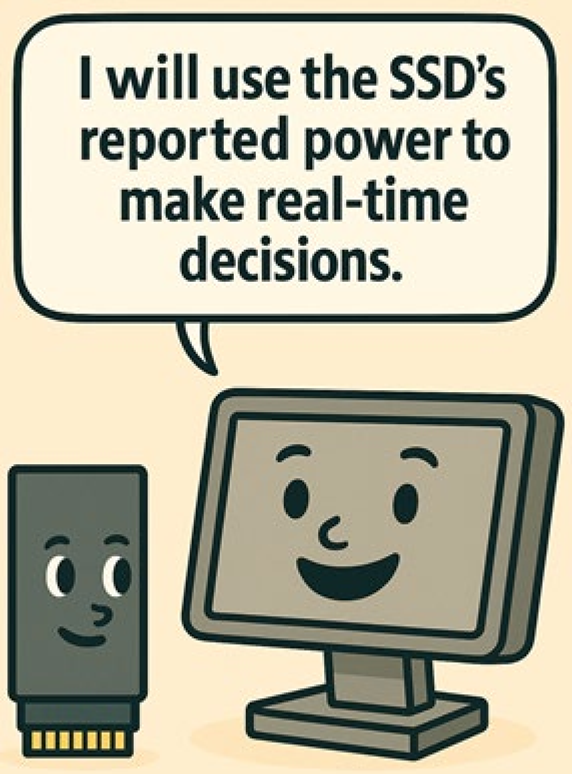
POWER STATE TABLE

Power State	AI Phase	Icon
PS0	Full Bandwidth (Training)	
PS2	Inference	
PS5+	Idle	

SNIA | CMS A SNIA Community

Upcoming NVMe® TP4199: Self Reported Drive Power

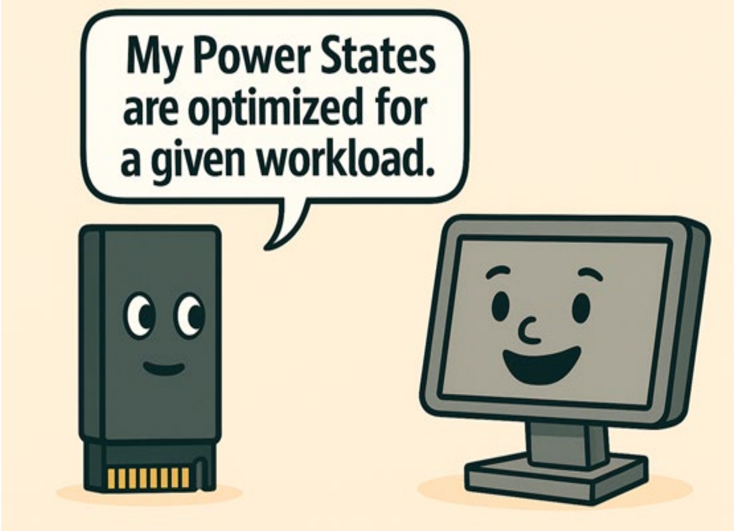
- Today: No standardized scalable way to report power
- TP4199: Adds new functionality and reporting capabilities to NVMe®:
  - One second power averages reported via SMART log page
  - Lifetime Energy Consumed via SMART log page (carbon costs)
  - Power threshold exceedance interrupts via AEN
  - Power log page
    - Average power
    - Maximum power
    - Histogram
- Standardized and Scalable!



- NOTE: Technical Proposal not ratified and can change. See 2025 OCP Storage Tech Talks for more information

SSD Support and Innovation

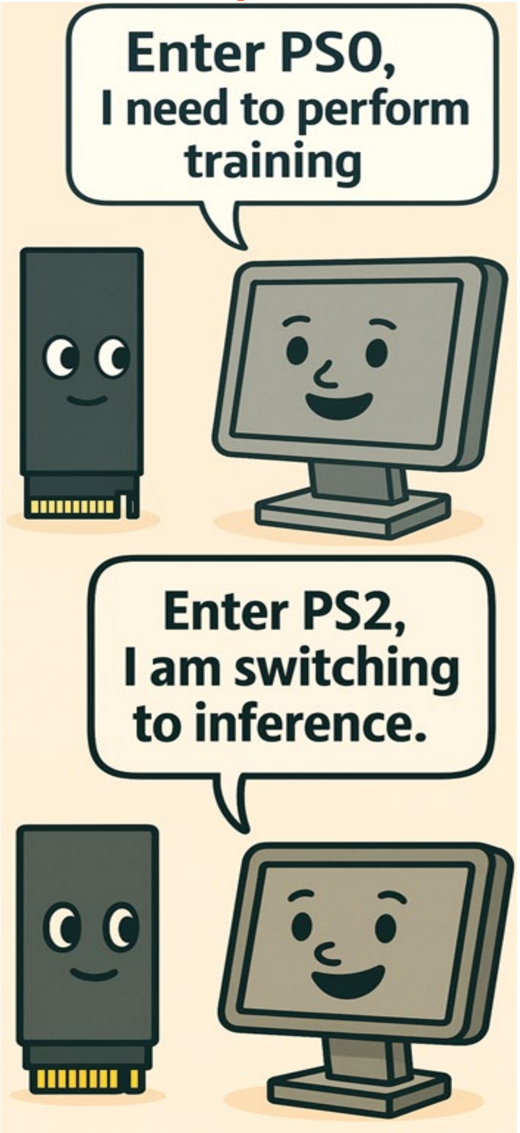
- Onboard power rail sensors
- NVMe® TP4199 support
- Workload specific NVMe® Power States
  - Vendor defined descriptors with clear caps, latencies and perf hints
  - States should map cleanly to AI phases
  - Drive optimization for specific workloads allows host to “dial in” exactly what the drive needs.



SNIA | CMS A SNIA Community

Workload Aware Power Scheduling Example

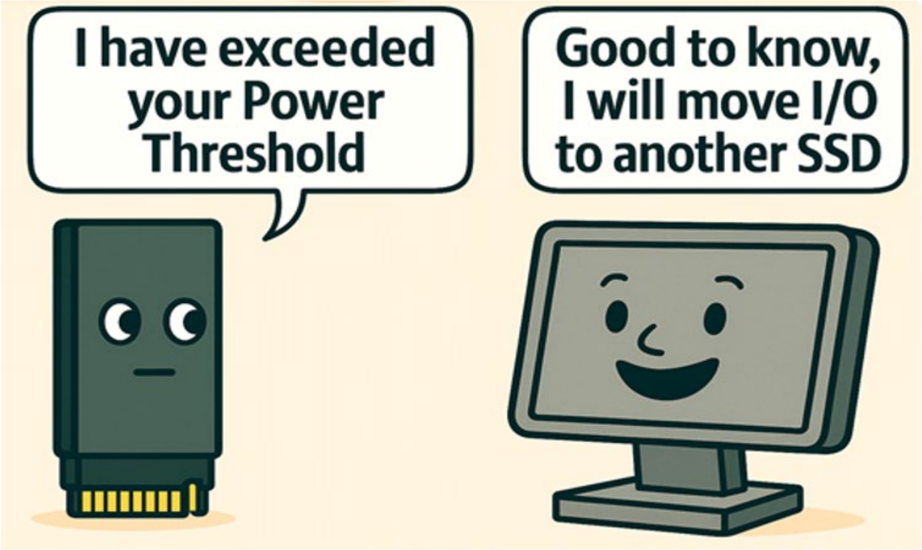
- Map known workloads to the optimal NVMe® PS
- Proposed AI Mapping
  - Training Checkpoint Flush (PS0):
    - Large sequential writes
    - Most power and cooling resources required
  - Inference (PS2):
    - Reads
    - Requires low latency and high throughput processing
    - Less cooling resources require
  - Idle (PS5+):
    - SSD not being utilized
    - Drive enters sleep states to free up power and cooling resources
- Polls power telemetry to validate expected power or react accordingly



SNIA | CMS A SNIA Community

Reactive Power Optimization Example

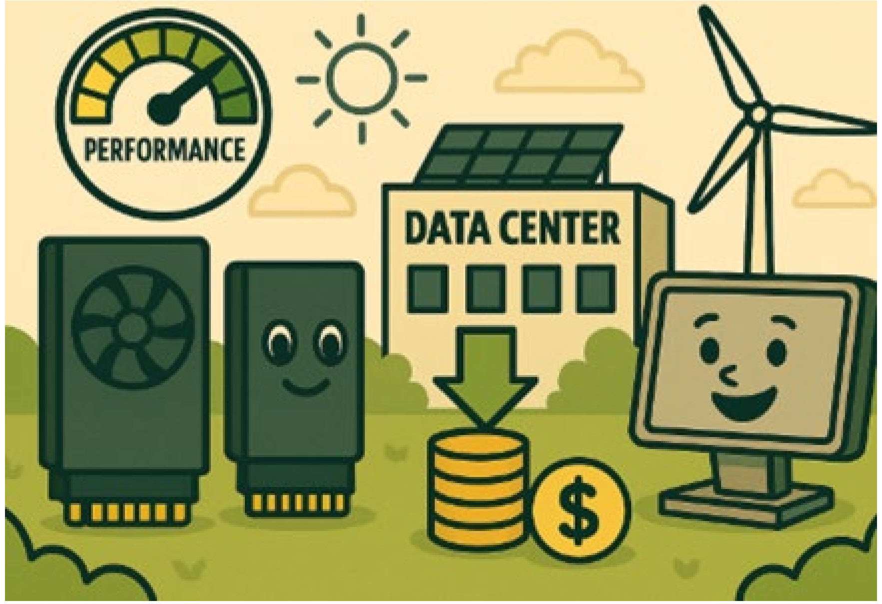
- Even the best planned PS assignments face runtime surprises
- Live power telemetry fuels real-time decision making
- Instantly respond to power budget breach → power balancing
- Pinpoint potential SSD anomalies
- Spot host to drive workload mismatches early
- Together, these reduce power, lowers TCO, and supports sustainability goals



SNIA | CMS A SNIA Community

Key Takeaways

- AI power demands → power is now a first class design metric
- TP4199 unlocks precise, standardized, scalable power telemetry
- I am starting the conversation of using dynamic PS management with workload aware power states
- This will increase power efficient and will cut energy consumption across the cluster
- Lower TCO + greener data centers without sacrificing IOPS



SNIA | CMS A SNIA Community

Questions?

Next Steps. . .

- ✓ Please rate this webinar and provide us with feedback
- ✓ Look for a Blog answering questions from this webinar at snia.org/blog
- ✓ Get more SNIA education!
  - **Live**
    - FMS: Future of Memory and Storage, August 4-7, 2025 Santa Clara CA futurememorystorage.com
    - SNIA Developer Conference, September 15-17, 2025 Santa Clara CA sniadeveloper.org
    - SC25, November 16-21, St. Louis MO sc25.supercomputing.org
  - **Online**
    - This webinar and many other videos and presentations on today's topics are in the **SNIA Educational Library** https://snia.org/educational-library
    - **SNIA Video You Tube Channel** https://www.youtube.com/user/sniavideo
- ✓ Join SNIA and the Solid State Drive (SSD) Special Interest Group
  - www.snia.org/join
  - www.snia.org/groups/cms
  - www.snia.org/forums/cmsi/solid-state-drive-special-interest-group

SNIA | CMS A SNIA Community

Thank You